

Ensemble Learning-based Algorithms for Traffic Flow Prediction in Smart Traffic Systems

Anas Saleh Alkarim

Information Systems Department, Faculty of Computing and Information Technology, King Abdulaziz University, Saudi Arabia | Department of Information Systems and Technology, College of Computer Science and Engineering, University of Jeddah, Saudi Arabia
aaalkarim@stu.kau.edu.sa (corresponding author)

Abdullah S. Al-Malaise Al-Ghamdi

Information Systems Department, Faculty of Computing and Information Technology, King Abdulaziz University, Saudi Arabia | Information Systems Department, School of Engineering, Computing and Design, Dar Al-Hekma University, Saudi Arabia
aalmalaise@kau.edu.sa

Mahmoud Ragab

Information Technology Department, Faculty of Computing and Information Technology, King Abdulaziz University, Saudi Arabia | Department of Mathematics, Faculty of Science, Al-Azhar University, Egypt
mragab@kau.edu.sa

Received: 18 December 2023 | Revised: 8 January 2024 | Accepted: 16 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6767>

ABSTRACT

Due to the tremendous growth of road traffic accidents, Intelligent Transportation Systems (ITSs) are becoming even more important. To prevent road traffic accidents in the long term, it is necessary to find new vehicle flow management techniques in order to optimize traffic flow. With the high growth of deep learning and machine learning, these methods are increasingly being used in ITSs. This research provides a novel conceptual ITS model that aims to predict vehicle movement through the collective learning usage to anticipate intersections. The proposed approach consists of three main stages: data collection through cameras and sensors, implementation of machine learning and deep learning algorithms, and result evaluation, utilizing the coefficient of determination (R-squared), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). To accomplish this, various machine learning and deep learning algorithms, such as Random Forest, LSTM, Linear Regression, and ensemble methods (bagging), were incorporated into the model. The findings revealed the enhancement due to the proposed method, which was observed through a significant performance improvement of 93.52%.

Keywords-intelligent transportation systems; smart traffic systems; traffic flow; prediction models; smart cities; bagging ensemble learning

I. INTRODUCTION

Road traffic accidents are dramatically augmented each year due to the massive increasing number of vehicles on the roads. This problem is considered a serious risk, a major source of trouble for individuals worldwide, and a significant global concern. [1]. Collecting and analyzing comprehensive data is essential for any initiative aiming to improve traffic safety [2]. With the rising number of vehicles on the roads and the resulting congestion issues, optimizing traffic flow has become

a pressing challenge in modern cities. Intelligent Transportation Systems (ITSs) have emerged as a promising solution to alleviate traffic congestion and enhance overall transportation efficiency [3-4]. The Vehicle Ad-Hoc Network (VANET) serves as a fundamental infrastructure for ITSs, enabling wireless connectivity among vehicles [5-6]. Additionally, intelligent transport systems are increasingly focused on addressing traffic congestion. Researchers have employed machine learning algorithms to predict traffic flow and reduce congestion at intersections. These models were evaluated using

the national road traffic dataset for the UK. An adaptive traffic light system was implemented, which adjusts green and red lights based on road width, traffic density, and vehicle categories. Simulations demonstrated a 30.8% decrease in traffic congestion [7].

Accurate traffic prediction is crucial for ameliorating the effectiveness of traffic systems and reducing energy consumption. Machine learning-based methods have become commonplace, but they often rely on historical data [8-9]. Furthermore, ML-based models are gaining popularity due to their ability to accurately forecast traffic conditions, thereby improving safety and infotainment applications. However, the efficacy of these models in predicting real-time traffic remains a subject of investigation [10].

Several research studies have focused on developing methods and models for traffic flow prediction and management. In [11], a framework is presented that utilizes Vector Auto Regression (VAR) and a CNN-LSTM hybrid neural network to predict short-term traffic flow. The CNN-LSTM model outperforms other models in forecasting short-term traffic flow and demonstrates predictive accuracy associated with spatial correlation in traffic flow. In [12], three proposed solutions are discussed to address the issue of missing data in traffic management. These solutions include a live-traffic simulation, a neural network traffic prediction and rerouting system based on pheromone principles, as well as a Weighted Missing Data Imputation (WEMDI) approach. The integration of WEMDI into the systems yields notable improvements in various traffic factors and demonstrates efficient routing to alternative destinations.

ML and neural networks play a significant role in solving traffic congestion issues. In this context, authors in [13] propose ML and DL algorithms for predicting intersection traffic flow. The models were trained, validated, and tested using public datasets, and the Multilayer Perceptron Neural Network (MLP-NN) produced the best results. Gradient Boosting, Recurrent Neural Networks, RF, LR, and Stochastic Gradient also showed promising performance.

ITSs require traffic flow monitoring for effective management and optimization. Conventional methods of data collection and analysis are being augmented with AI techniques, such as ensemble learning [14]. The IAROEL-TFMS methodology utilizes feature subset selection and optimal ensemble learning to predict traffic flow, outperforming other approaches with its low RMSE. Authors in [14] used Hybrid-LSSVM, AST2FP-OHDBN, and IAROEL-TFMS models for evaluation purposes, considering their respective performance indicators. Among the several models evaluated, it was observed that IAROEL-TFMS had the most superior predictive performance. In close succession, the AST2FP-OHDBN model exhibited robust performance, whereas, in contrast, the Hybrid-LSSVM model demonstrated a somewhat reduced level of prediction accuracy. Regarding predictive performance, the IAROEL-TFMS model had the best precision and accuracy in forecasting the target variable. The AST2FP-OHDBN model closely followed it. On the other hand, the Hybrid-LSSVM model exhibited slightly inferior predictive skills.

This paper utilizes four Machine Learning (ML) and Deep Learning (DL) models: Random Forest (RF), Linear Regression (LR), Long Short-Term Memory (LSTM), and ensemble bagging (RF). The objective is to utilize these predictions to enhance the efficiency of traffic light controllers in the context of traffic flow prediction at intersections. Experimental results demonstrate that all models exhibit a strong predictive capacity for estimating vehicular flow, highlighting their potential utility in smart traffic systems.

II. THE PROPOSED MODEL

This study has developed a model for monitoring traffic flow. The primary objective of this model is to predict traffic movement. To achieve this objective, the model operates in three distinct stages. Firstly, data collection can be accomplished using cameras or sensors. Secondly, ML and DL technologies are applied. Thirdly, the outcomes are evaluated using MAE, RMSE, and the coefficient of determination (R-squared). The workflow of the suggested approach is illustrated in Figure 1. Overall, the proposed model aims to monitor traffic flow by predicting its movement. It involves data collection through cameras or sensors, the application of machine learning and deep learning technologies, and the evaluation of outcomes using specific error metrics. Figure 1 provides a visual representation of the workflow.

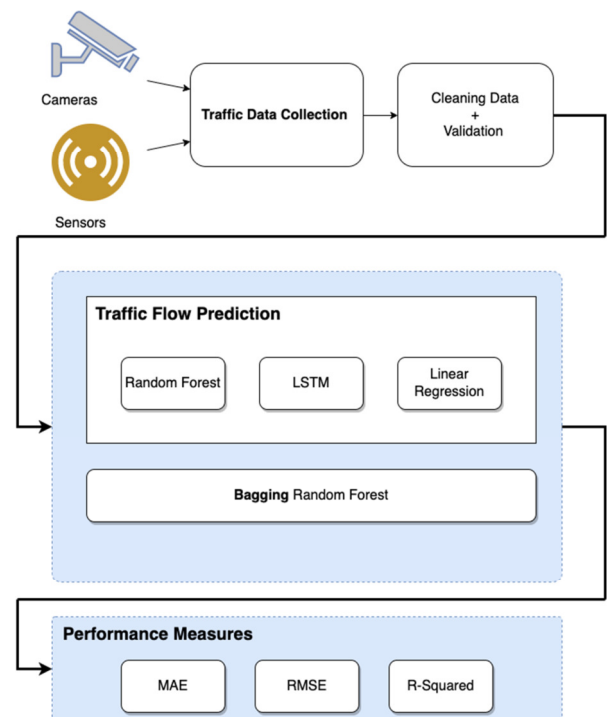


Fig. 1. The workflow of the proposed model.

A. Data Collection

The dataset used for traffic prediction was obtained from various traffic sensors provided by the Huawei Munich Research Center. The dataset plays a crucial role in predicting traffic patterns and making necessary adjustments to stop-light control settings, including cycle length, offset, and split

timings. The dataset consists of recorded data from six intersections located within an urban area, collected over a period of 56 days (Table I). The data are presented as a flow time series, which indicates the number of vehicles passing through each intersection every 5 minutes, spanning 24 hours. This results in 12 readings per hour, 288 readings per day, and a total of 16,128 readings over the course of the 56 days. For this study, 4 out of the 6 intersections were selected to replicate a 4-lane intersection scenario [15].

TABLE I. THE 6 INTERSECTIONS OF THE DATASET USING COLAB

	Cross 1	Cross 2	Cross 3	Cross 4	Cross 5	Cross 6
0	105.0	48.0	30	62.0	31	110.0
1	97.0	41.0	32	55.0	42	103.0
2	76.0	47.0	44	58.0	40	100.0
3	98.0	40.0	39	59.0	43	104.0
4	87.0	41.0	47	49.0	35	112.0

B. Data Preparation

Data cleaning is a critical step in the preprocessing phase, where incorrect, incomplete, duplicate, or erroneous data within a dataset are rectified. Fortunately, the collected data for this study do not contain any missing values. The dataset has been divided into two parts: 70% for training the model and the remaining portion for testing. To ensure consistency and optimal performance during training, the data were scaled using the MinMaxScaler from the scikit-learn library. This scaler transforms the data, making them range between zero and one [16].

C. Proposed Techniques

In this study, four regression models from the scikit-learn module in the Python programming language are employed. The scikit-learn module is a comprehensive Python library that offers a wide range of state-of-the-art machine learning algorithms designed to tackle various supervised and unsupervised challenges [16]. The authors applied four ML/DL techniques to the dataset: RF, LSTM, LR, and ensemble method (bagging). The following section provides an overview of the traditional ML and ensemble methods utilized in the experiment.

III. OVERVIEW OF TRADITIONAL MACHINE LEARNING AND ENSEMBLE METHODS

A. Random Forest

RF is a learning method that combines multiple tree predictors. Each tree in the forest is constructed based on the values of a random vector, sampled independently from the same distribution for all trees. Tree-based models form the core components of the random forest algorithm. A tree-based model involves iteratively dividing a given dataset into two distinct groups, guided by a specific criterion, until a predetermined stopping condition is met. The terminal nodes of decision trees are commonly known as leaf nodes or leaves [17].

B. Long Short-Term Memory (LSTM)

LSTM networks have found extensive applications in various domains, including image processing, speech recognition, manufacturing, autonomous systems, communication, and energy consumption, for dynamic system modelling purposes. LSTM has gained significant attention in recent years due to its effectiveness in modeling and predicting the dynamics of nonlinear time-variant systems. It incorporates the characteristics of short-term and long-term memory, the ability to make predictions several steps ahead, and the propagation of errors. Sequence-to-sequence networks with partial conditioning have been shown to outperform other techniques such as bidirectional or associative networks, making them well-suited for achieving the specified objectives [18].

C. Linear Regression

LR is a widely used and straightforward ML algorithm. The technique is a mathematical methodology employed to do predictive analysis. Moreover, LR is a statistical technique that enables the prediction of continuous or numerical variables. LR is a statistical technique employed to assess and quantify the association between variables under consideration [19].

D. Ensemble Method (Bagging)

Bagging, short for bootstrap aggregating, is a technique that involves creating multiple iterations of a predictor and combining them to form an aggregated predictor. In the aggregation process, the mean is calculated across the iterations when predicting a numerical outcome, while a majority vote is used when predicting a class. To generate multiple versions, bootstrap copies of the original learning set are created, and these replicates are then used as new learning sets [20].

IV. EVALUATION MEASURES

In model evaluation, the coefficient of determination (R^2), RMSE, and MAE are standard metrics [21].

$$R^2 = 1 - \frac{\sum_{i=1}^m (x_i - y_i)^2}{\sum_{i=1}^m (\bar{y} - y_i)^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3)$$

V. EXPERIMENTAL RESULTS AND DISCUSSION

Table II presents the results of the model using various ML and DL algorithms. It can be observed that RF achieved an MAE of 13.76, while LSTM and LR achieved 14.74 and 17.80, respectively. When Bagging (RF) was applied, the minimum MAE obtained was 13.69. In terms of RMSE, the models achieved values of 22.39, 23.50, and 27.04, while the Bagging model achieved a lower RMSE of 22.21. In terms of R^2 , the experimental results for the models were 0.9341, 0.9275, and 0.9040, respectively. The best R^2 value was obtained by the Bagging model, which achieved a value of 0.9352. The results show that the RF model and the Bagging model (using RF as the base model) outperformed the LSTM and LR models in terms of both MAE and RMSE. Additionally, the Bagging

model showed the highest R^2 value, suggesting a better fit to the data. Overall, these findings demonstrate the effectiveness of the RF algorithm and the potential benefits of using ensemble methods like Bagging for traffic flow prediction.

TABLE II. RESULTS

ML/DL Model	MAE	RMSE	R^2
RF	13.76	22.39	0.9341
LSTM	14.74	23.50	0.9275
LR	17.80	27.04	0.9040
Bagging (RF)	13.69	22.21	0.9352

Figure 2 illustrates the numerical values of MAE measurements of the considered models. It can be observed that the LSTM model has a slightly higher MAE (14.74) compared to the RF (13.76) and Bagging (RF) (13.69) models. This suggests that the LSTM model may not perform optimally in this particular scenario. On the other hand, the LR model has the highest MAE score (17.80). This indicates that it may not excel at accurately predicting the target variable. These results suggest that the RF and Bagging models (using RF as base) perform better than the LSTM and LR models in terms of MAE. It is important to consider these findings when selecting the most suitable model for traffic prediction in this context.

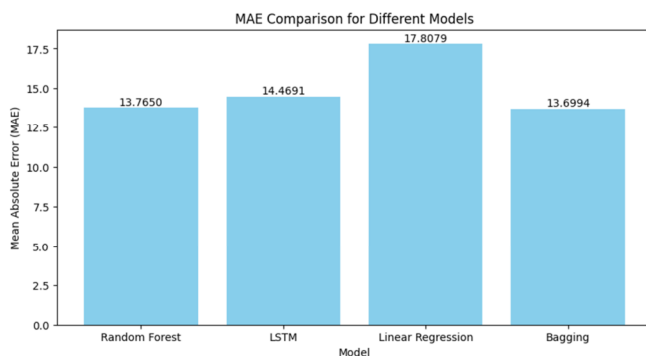


Fig. 2. MAE of the different models.

The figures presented in Figure 3 illustrate the RMSE values of the considered models. It can be observed that the Bagging (RF) model has the lowest RMSE score (22.21), indicating that, on average, its predicted values deviate the least from the actual values. This suggests that the model exhibits strong predictive accuracy. The RF model also performs well, although it has a slightly higher RMSE (22.39) compared to the Bagging model. On the other hand, the LSTM model shows a larger RMSE (23.50), indicating potentially inferior performance in terms of predictive accuracy. The LR model has the largest RMSE value (27.04), suggesting a potentially lower level of accuracy in predicting the target variable. These findings again suggest that both the Bagging (RF) and RF models perform well in terms of RMSE, indicating their ability to provide accurate predictions. However, the LSTM and LR models may have limitations in accurately predicting the target variable based on their higher RMSE values.

Figure 4 shows the R^2 values of the considered models. R^2 ranges from 0 to 1, with a value of 1 indicating a perfect fit.

Among the models presented, it is evident that the Bagging (RF) model shows the highest R^2 value (0.9352), designating its superior ability to fit the data accurately. The RF (0.9341) and LSTM (0.9275) models also demonstrate high R^2 values, suggesting their effectiveness in explaining a significant proportion of the observed variability in the dependent variable, while LR performs satisfactorily (0.9040), its R^2 value is slightly lower compared to the alternative models. Overall, it can be noticed that all of the models exhibit strong performance in elucidating the variability in the dependent variable. However, it is noteworthy that Bagging (RF) emerges as the most prominent performer among them.

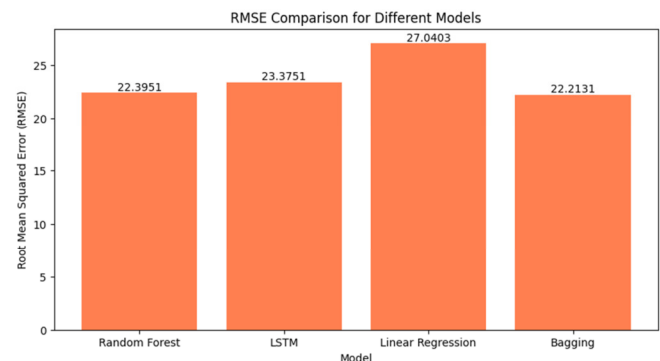
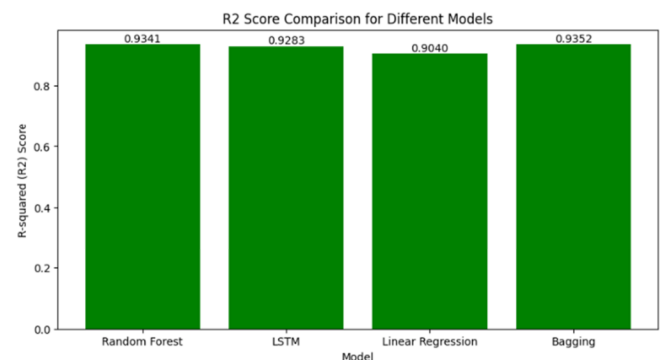


Fig. 3. RMSE of the different models.

Fig. 4. R^2 of the different models.

Compared to [13], this research has improved the results by more than 0.5%. In [13], the researchers used the same dataset and applied 5 ML methods. Gradient boosting was the most successful, with 93.05%. The proposed model gets 93.41% by utilizing an RF model. Bagging (RF) has the highest result of 93.52%. In the future, researchers in this field could use a combination of ML and DL models to improve model performance [22].

VI. CONCLUSION

In this article, we mainly presented a new model to enhance intelligent traffic systems. The main purpose of this method is to recognize the traffic flow prediction or vehicle movement at intersections applying an ensemble learning technique. The proposed framework consisted of three primary phases: data collection through cameras and sensors, implementation of

machine learning and deep learning algorithms, and evaluation of the outcomes using metrics such as R-squared, RMSE, and MAE. The model utilized Random Forest, LSTM, Linear Regression, and Bagging (Random Forest), to achieve its objectives. To safeguard the better performance of the proposed procedure, a series of tests was involved. The comprehensive results highlighted the improved performance of the proposed method by achieving the significant accuracy of 93.52%. Regarding future work, combining multiple machine learning and deep learning algorithms could be explored to further enhance the performance of this model for more effective intelligent traffic systems.

REFERENCES

- [1] M. Angin and S. I. A. Ali, "Analysis of Factors Affecting Road Traffic Accidents in North Cyprus," *Engineering, Technology & Applied Science Research*, vol. 11, no. 6, pp. 7938–7943, Dec. 2021, <https://doi.org/10.48084/etasr.4547>.
- [2] N. K. Al-Shammari and S. M. H. Darwish, "In-depth Sampling Study of Characteristics of Vehicle Crashes in Saudi Arabia," *Engineering, Technology & Applied Science Research*, vol. 9, no. 5, pp. 4724–4728, Oct. 2019, <https://doi.org/10.48084/etasr.2939>.
- [3] P. Sen, "Optimization of Traffic Flow Using Intelligent Transportation Systems," *Mathematical Statistician and Engineering Applications*, vol. 70, no. 1, pp. 720–727, Jan. 2021, <https://doi.org/10.17762/msea.v70i1.2530>.
- [4] Y. A. Hanafy, M. Mashaly, and M. A. Abd El Ghany, "An Efficient Hardware Design for a Low-Latency Traffic Flow Prediction System Using an Online Neural Network," *Electronics*, vol. 10, no. 16, Jan. 2021, Art. no. 1875, <https://doi.org/10.3390/electronics10161875>.
- [5] A. K. Kazi and S. M. Khan, "DyTE: An Effective Routing Protocol for VANET in Urban Scenarios," *Engineering, Technology & Applied Science Research*, vol. 11, no. 2, pp. 6979–6985, Apr. 2021, <https://doi.org/10.48084/etasr.4076>.
- [6] S. S. Sepasgozar and S. Pierre, "Network Traffic Prediction Model Considering Road Traffic Parameters Using Artificial Intelligence Methods in VANET," *IEEE Access*, vol. 10, pp. 8227–8242, 2022, <https://doi.org/10.1109/access.2022.3144112>.
- [7] I. Moumen, J. Abouchabaka, and N. Rafalia, "Adaptive traffic lights based on traffic flow prediction using machine learning models," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 5, Oct. 2023, Art. no. 5813, <https://doi.org/10.11591/ijece.v13i5.pp5813-5823>.
- [8] Z. Wang, P. Sun, Y. Hu, and A. Boukerche, "A Novel Mixed Method of Machine Learning Based Models in Vehicular Traffic Flow Prediction," in *Proceedings of the 25th International ACM Conference on Modeling Analysis and Simulation of Wireless and Mobile Systems*, New York, NY, USA, Jul. 2022, pp. 95–101, <https://doi.org/10.1145/3551659.3559047>.
- [9] F. Sheriff, "ELMOPP: an application of graph theory and machine learning to traffic light coordination," *Applied Computing and Informatics*, Mar. 2021, <https://doi.org/10.1108/ACI-07-2020-0035>.
- [10] P. Sun, N. Aljeri, and A. Boukerche, "Machine Learning-Based Models for Real-time Traffic Flow Prediction in Vehicular Networks," *IEEE Network*, vol. 34, no. 3, pp. 178–185, Feb. 2020, <https://doi.org/10.1109/MNET.011.1900338>.
- [11] Z. Cheng, J. Lu, H. Zhou, Y. Zhang, and L. Zhang, "Short-Term Traffic Flow Prediction: An Integrated Method of Econometrics and Hybrid Deep Learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 6, pp. 5231–5244, Jun. 2022, <https://doi.org/10.1109/tits.2021.3052796>.
- [12] R. K. C. Chan, J. M.-Y. Lim, and R. Parthiban, "A neural network approach for traffic prediction and routing with missing data imputation for intelligent transportation system," *Expert Systems with Applications*, vol. 171, Jun. 2021, Art. no. 114573, <https://doi.org/10.1016/j.eswa.2021.114573>.
- [13] A. Navarro-Espinoza et al., "Traffic Flow Prediction for Smart Traffic Lights Using Machine Learning Algorithms," *Technologies*, vol. 10, no. 1, Jan. 2022, Art. no. 5, <https://doi.org/10.3390/technologies10010005>.
- [14] M. Ragab, H. A. Abdushkour, L. Maghrabi, D. Alsaman, A. G. Fayoumi, and A. A.-M. Al-Ghamdi, "Improved Artificial Rabbits Optimization with Ensemble Learning-Based Traffic Flow Monitoring on Intelligent Transportation System," *Sustainability*, vol. 15, no. 16, Aug. 2023, Art. no. 12601, <https://doi.org/10.3390/su151612601>.
- [15] C. Axenie and S. Bortoli, "Road traffic prediction dataset," Zenodo, Oct. 07, 2020, <https://doi.org/10.5281/zenodo.3653880>.
- [16] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [17] M. Schonlau and R. Y. Zou, "The random forest algorithm for statistical learning," *The Stata Journal*, vol. 20, no. 1, pp. 3–29, Mar. 2020, <https://doi.org/10.1177/1536867X20909688>.
- [18] B. Lindemann, T. Müller, H. Vietz, N. Jazdi, and M. Weyrich, "A survey on long short-term memory networks for time series prediction," *Procedia CIRP*, vol. 99, pp. 650–655, Jan. 2021, <https://doi.org/10.1016/j.procir.2021.03.088>.
- [19] D. Maulud and A. M. Abdulazeez, "A Review on Linear Regression Comprehensive in Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 1, no. 4, pp. 140–147, Dec. 2020, <https://doi.org/10.38094/jastt1457>.
- [20] E. Hillebrand, M. Lukas, and W. Wei, "Bagging weak predictors," *International Journal of Forecasting*, vol. 37, no. 1, pp. 237–254, Jan. 2021, <https://doi.org/10.1016/j.ijforecast.2020.05.002>.
- [21] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Computer Science*, vol. 7, Jul. 2021, Art. no. e623, <https://doi.org/10.7717/peerj-cs.623>.
- [22] J. A. Fadhil and Q. I. Sarhan, "Internet of Vehicles (IoV): A Survey of Challenges and Solutions," in *2020 21st International Arab Conference on Information Technology (ACIT)*, Giza, Egypt, Aug. 2020, pp. 1–10, <https://doi.org/10.1109/ACIT50332.2020.9300095>.

Utilization of Multi-Mission Satellite Altimetry for Wave Energy with Site Suitability Analysis using the Analytic Hierarchy Process

Mat Nizam Ut

Geospatial Imaging and Information Research Group, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, Malaysia
mnizam65@graduate.utm.my

Ami Hassan Md Din

Geospatial Imaging and Information Research Group, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, Malaysia
amihassan@utm.my (corresponding author)

Norhakim Yusof

Geoinformation Department, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, Malaysia
norhakim@utm.my

Syarif Abdul Asaad Jairin

Geospatial Imaging and Information Research Group, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, Malaysia
saasaad2@graduate.utm.my

Received: 21 December 2023 | Revised: 7 January 2024 and 11 January 2024 | Accepted: 12 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6791>

ABSTRACT

Space technology advancements have enabled the acquisition of marine data that support the research on wave energy as an alternative to reduce fossil fuel dependency and mitigate climate change. Malaysia's ocean renewable energy potential lacks attention from local authorities due to insufficient in-situ data, posing challenges in investigating ocean characteristics, such as wave heights. This study investigated Malaysia's wave energy potential using extensive significant wave height data from multiple altimetry missions. The former assessed site suitability using the Analytical Hierarchy Process (AHP) multicriteria analysis, incorporating marine constraints, namely socioeconomic, physical, and environmental factors. The multicriteria findings were integrated into a Geographical Information System (GIS) to improve the site suitability analysis and generate a localized suitability index for wave energy. Validation of satellite altimeter data with in-situ measurements showed a strong correlation and low RMSE. AHP analysis indicated good consistency in the criteria analysis, with a consistency ratio of 0.045, which falls below the limit of 0.1. The coastal and offshore regions of the Malaysian seas are suitable for harnessing wave energy with energy ranges up to 4.21 kW/m. Therefore, this study provides valuable information to stakeholders and the government to increase their interest in wave energy.

Keywords-*analytic hierarchy process; satellite altimeter; wave energy; multi-criteria*

I. INTRODUCTION

Changes in the global socioeconomic environment, the environment, natural sources, and social life have increased the demand for renewable energy, particularly for electric generation. Natural sources of electricity generation in Malaysia are believed to experience a shortage in the future due

to environmental impacts and population growth [1-2]. There are various types of ocean energy, including tidal barrage, tidal stream, wave energy, salinity gradient, and ocean thermal energy [3]. The feasibility study of renewable energy requires ocean information and large spatial and temporal resolution to characterize ocean conditions. Coarse spatial information is not ideal for describing spatial variations in sea conditions [4]. To

date, the available information on ocean characteristics in Malaysian waters is based on observations from oil platforms, ships, and buoys [5]. According to [4], satellite measurements have offered large spatial and temporal data to characterize the ocean surface. This study used satellite altimeter data from various altimetry missions to retrieve Significant Wave Height (SWH) data. Although renewable ocean energy could help minimize carbon emissions, there is still concern about the impact of its development on marine space. For example, noise generated during the installation can affect marine life, as its habitat is disturbed, and turbines can block the movement of marine life or even reduce its population [6-7]. This study performs a multicriteria analysis considering environmental, physical, and socioeconomic factors to avoid conflicts, e.g., human activities, marine environment, conservation areas, ship routes, etc. so that the proposed sites for wave energy evade the conflicting areas.

II. DATA AND METHODS

Figure 1 shows the study area, which is circumscribed by latitude from 0° to 14° and longitude from 95° to 120°. The analysis of available resources and site suitability is performed inside the Malaysian Exclusive Economic Zone (EEZ). According to UNGA, the EEZ extends to 200 nautical miles from the territorial baseline and stretches to 12 nautical miles of territorial sea [8]. Within this area, the coastal nation has exclusive control and exploitation rights over natural resources. This study used various altimetry missions to retrieve SWH data over the Malaysian seas. The coverage period of the multi-mission satellite altimeter data employed was from January 1993 to December 2020 (28 years). The altimetry mission included ERS-1, TOPEX, Poseidon, ERS-2, Jason-1, Envisat1, Jason-2, Cyrosat-2, SARAL, Jason-3 and Sentinel-3A. Buoy measurements are crucial to assess the accuracy and reliability of other techniques through the validation process. Table I summarizes the buoys engaged in this study.

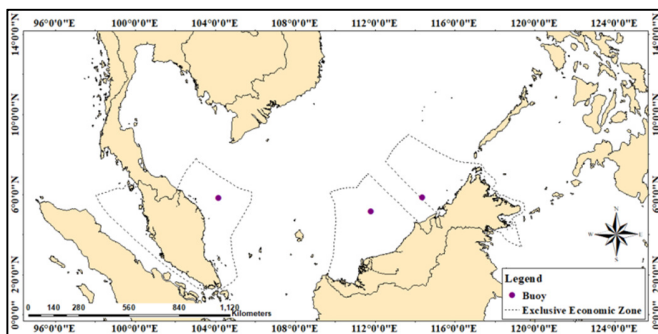


Fig. 1. The study area.

TABLE I. SELECTED BUOY DATA USED IN THIS STUDY

Buoy	Location	Period
Sabah	5.83 N 114.39 E	Nov 04-Dec 07
Sarawak	5.15 N 111.82 E	Jan 09-Jan 12

Validation plays a crucial role in proving the accuracy of the altimetric data. The relationship between the altimeter-

derived and the in-situ data translated into correlation coefficient r utilizing Pearson's coefficient [9]:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x}) - (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

where r is the correlation coefficient, x_i refer to in-situ data, \bar{x} is the mean value for measured data (in-situ), y_i is satellite altimeter data, and \bar{y} is the mean value for the observed data (satellite altimeter). RMSE determines how to fit the data around the line of the best fit. Some studies have shown that satellite altimeters are well correlated, have good RMSE values with in-situ data, and are applicable for energy studies [5,10-11].

The energy density derivation is part of the theoretical resource assessment. The energy density, called the potential energy of the wave energy, is calculated applying an existing equation. Wave energy flow is generally described as the average energy per meter of crest length of wave in kW/m. The exploitation of wave energy resources using Wave Energy Converters (WECs) requires the significant wave height H_s and wave period T_z as inputs. The wave energy density can be obtained utilizing (2). P_{wave} is the wave energy density (kW/m) parameterized in terms of H_s and T_z , expressed in meters and seconds, respectively.

$$P_{wave} = 0.49 H_s^2 T_z \quad (2)$$

According to [12], a satellite altimeter is unable to provide a direct measurement for the wave period T_z . This study applies the indirect estimation of wave period using the Fara-U algorithm (3). This algorithm is developed in the form of significant wave height and wind speed data from a satellite altimeter, employing the symbolic regression method to develop wave period data, and is suitable for the Malaysian sea [12].

$$T_z = 1.299a - 1.127 \quad (3)$$

$$a = 2.247 + \frac{4.947 + 25.399 H_s}{7.562 + H_s + U}$$

where the wave period is denoted as T_z (s), H_s is the significant wave height (m), and U is the wind speed (m/s).

Practical wave energy resource assessment considered the marine constraints as defined in [13-14]. The importance of this analysis is to identify a suitable location by avoiding existing conflicts. There are various applications related to site suitability analysis, such as flood shelter mitigation [15], agricultural management [16], landfill management [17], and coastal management and planning [18]. The selection of a suitable site for ocean renewable energy requires an understanding of the spatial distribution of the marine environment, human activities, and physical criteria in the marine space. However, marine data in Malaysia are difficult to access because there are no centralized and strategically developed marine data centers for marine applications due to a lack of cooperation within and among organizations [19]. This study involves the physical, environmental, and socioeconomic factors, which are represented by several types of data, as presented in Table II.

TABLE II. FACTORS AND SUB-FACTORS USED IN THE SITE SUITABILITY ANALYSIS

Factors	Sub-factors
Environmental	Marine protected areas
	National marine parks
	Coral reefs
	Sea turtle nesting
Physical	Underwater pipelines
	Submarine cables
	Shipping lanes
	Fish main landing point
	Oil and gas areas
	Ports
	Shoreline

The AHP technique comprises certain steps to create a model by breaking down complex problems into smaller, more manageable parts. These steps are the following: (1) Structure the problem hierarchy and create the decision hierarchy, (2) define criteria and create pairwise comparison matrices, (3) create matrices for alternatives based on objectives, (4) verify judgment consistency by calculating the ratio, and (5) calculate weighted average ratings for decision alternatives by selecting the highest-scoring one [20, 23]. Comparisons or adjustments between two elements in AHP can be made using the relative scale between two alternatives [21]. The pairwise comparison is performed using a square matrix, as in (4) [22].

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} = \begin{pmatrix} \frac{w_1}{w_1} & \frac{w_1}{w_2} & \dots & \frac{w_1}{w_n} \\ \frac{w_2}{w_1} & \frac{w_2}{w_2} & \dots & \frac{w_2}{w_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{w_n}{w_1} & \frac{w_n}{w_2} & \dots & \frac{w_n}{w_n} \end{pmatrix} = \begin{pmatrix} 1 & \frac{w_1}{w_2} & \dots & \frac{w_1}{w_n} \\ \frac{w_2}{w_1} & 1 & \dots & \frac{w_2}{w_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{w_n}{w_1} & \frac{w_n}{w_2} & \dots & 1 \end{pmatrix} \quad (4)$$

where $a_{ij} = w_i/w_j$ represents the weight of criterion i over criterion j . During the aggregating, this eigenvector is always an argument in the AHP process [22]. This argument is associated with a consistency issue, in which the weight values are dependent on the closeness of the eigenvalue towards the perfect consistency. Consistency Ratio (CR) and Consistency Index (CI) were introduced to counter this controversy:

$$CR = \frac{CI}{RCI} \quad (5)$$

$$CI = \frac{\lambda_{max} - n}{n - 1} \quad (6)$$

where RCI is the Random Consistency Index derived from a sample size of 500 randomly produced reciprocal matrices, λ_{max} is the maximum eigenvalue of the judgment matrix, and n is the number of criteria in the decision matrix.

AHP improves the understanding of a problem by building a hierarchy, which involves the connection of factors and sub-criteria to achieve the decision goal [24]. The selection of a suitable location is based on the study analysis using GIS. This part produces a conceptual suitability index. The selected suitable locations are analyzed to provide a regional suitability

index in the form of a statement generated from the assessment of site suitability using ArcGIS spatial analysis tools. This local indexing can improve the government's policy on ocean renewable energy and increase the effort to produce technology that befits Malaysia's seas. All generated maps and site suitability analyses were performed using ArcMap software version 10.3.

III. RESULTS AND DISCUSSION

This section provides the validation of satellite altimeter data with in-situ observation to evaluate their accuracy. The analysis of factors that hindered the ORE development is explained briefly by providing results based on the AHP assessment that is integrated into the GIS environment. The local site suitability index is generated to indicate suitable locations that may have the potential to harness energy.

A. Validation of Satellite Altimeter Data

Generally, in-situ measurements such as buoys have always been a benchmark to confirm measurements from space observations and a proven technique to assess satellite altimeters [11]. Figure 2 shows a comparison graph of significant wave height measurements from the satellite altimeters with buoy observation.

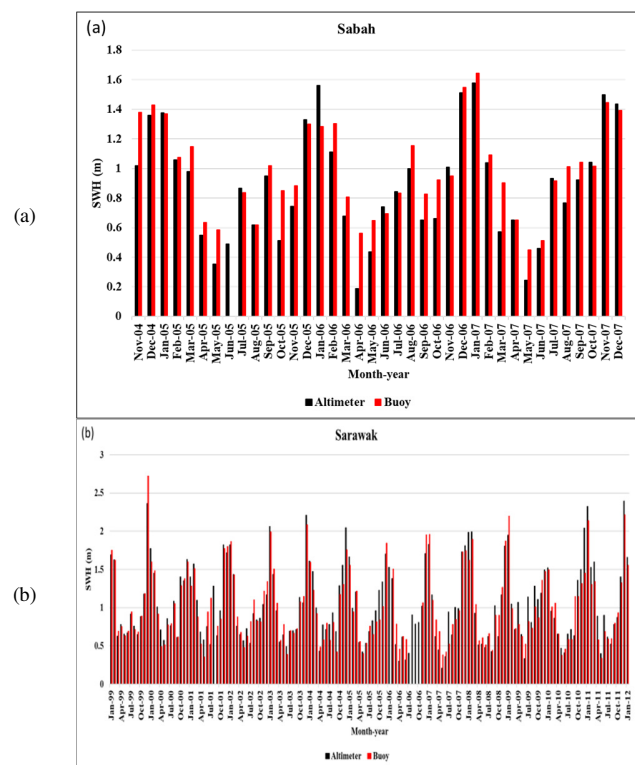


Fig. 2. Trend graphs for wind speed from satellite altimeter compared to buoys for (a) Sabah and (b) Sarawak.

The graphs indicate that the observations of the altimeter H_s are approximately equal to the buoy measurements. However, the monthly average values of the significant wave height from the altimeter of the Sabah buoy station underestimate the buoy measurements from December 2006 to March 2007. There are

missing observation data for the Sabah and Sarawak buoys from June 2005 and July 2006 to October 2006, respectively. These missing data were excluded in the correlation analysis to avoid degradation of the altimeter performance analysis. Figure 3 presents the scatter plot of the altimeter significant wave height against the buoy data. The correlation value agrees well with previous studies, such as [25-26], stating that satellite altimeter measurements for H_s are robust and widely accepted to be in close agreement with buoy measurements.

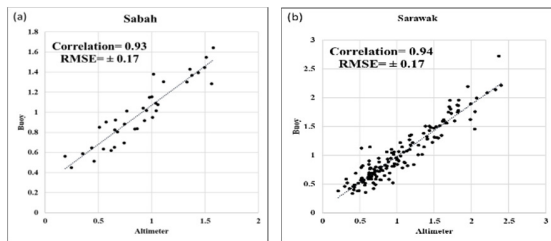


Fig. 3. Correlation graph for significant wave height altimeter and buoy measurements.

B. Theoretical Resource of Wave Energy

Figure 4 demonstrates that available wave energy is located at the offshore region of Sabah and Sarawak waters, ranging from 2.84 to 4.21 kW/m. This outcome is similar to the findings of [5]. In Peninsular Malaysia, most of the available wave energy ranges between 1.28 to 1.91 kW/m and 1.93 to 2.77 kW/m, located on the east coast facing the South China Sea.

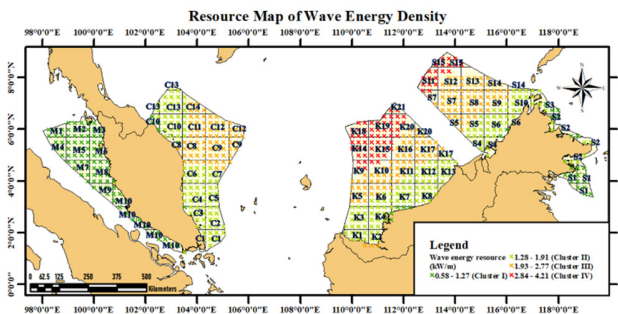


Fig. 4. Theoretical resource map for wave energy of Malaysian seas.

For Peninsular Malaysia, only Terengganu waters are available for wave energy density ranges from 1.93 to 2.77 kW/m. Kapas Island, Terengganu (5.94 km from the Marang shoreline) has an average available wave energy of 1.91 kW/m. Besides that, Tenggol Island, Terengganu (28.72 kilometers from Kuala Dungun) is also indicated with an average wave energy of 1.94 kW/m.

C. Restriction Map

In developing countries, marine space exploration is vital for economic growth, fostering activities, such as aquaculture, fisheries, ecotourism, and deep-sea resource exploration. Simultaneously, it offers valuable renewable energy resources for future electricity generation. As marine space use increases, ocean renewable energy development must avoid conflicts for sustainable and low-risk environmental impacts [27]. Based on

Figure 5, most of the activities operate on the east coast of Peninsular Malaysia and the north coast of Sabah and Sarawak, facing the South China Sea. Malacca Strait is important for Malaysia as it provides a route for global trade, which shortens the sea route between the Indian and the Pacific Oceans. Each criterion was buffered using the geoprocessing buffer tool of ArcGIS 10.3. Model Builder tools were applied by constructing workflows that sequentially arrange the geoprocessing tools, namely the Buffer, Feature to Raster, Is Null, and Raster Calculator tools, to generate the restriction map. The Raster Calculator was employed to create and execute the criteria layers to generate the restriction map. Figure 6 shows the criteria buffer map within the Malaysian EEZ.

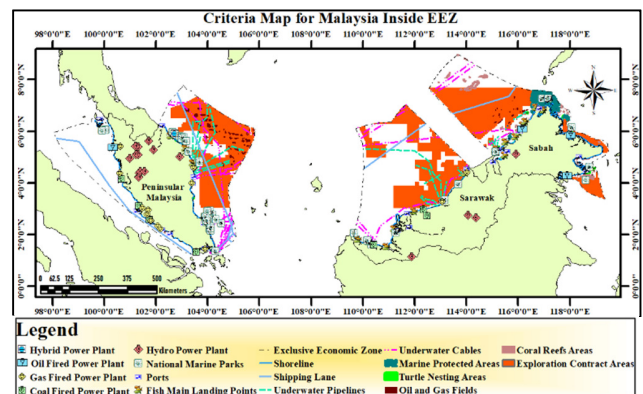


Fig. 5. Mapping of the considered criteria inside the EEZ.

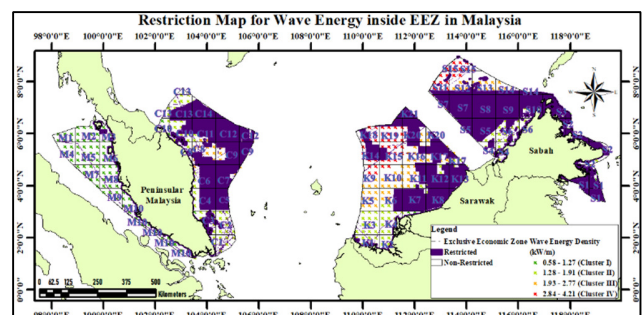


Fig. 6. Restricted and non-restricted zones from the buffer analysis.

The highest potential for wave energy within the range of 2.84 to 4.21 kW/m is offshore Sarawak and Sabah. Meanwhile, in Terengganu waters, most of the location has available wave energy within the range of 1.93 to 2.77 kW/m.

D. Integration of AHP with GIS for Site Suitability Analysis

The construction of a hierarchical tree, which involves criteria, sub-criteria, and alternatives, is the primary step of the AHP process [28]. The weightage value for the criteria is generated by averaging the comparison value of each criterion. To get the average value, the comparison matrix is then normalized by dividing the corresponding matrix value in column order with the sum of cell values in the same column. CR (5) was used to evaluate the weight consistency. According to [29], if the CR is greater than 10% or 0.10, the judgments of the criteria based on the pairwise comparison indicate inconsistency, and therefore a reanalysis of the judgment must

be performed. Table III demonstrates that CR is 0.045 or 4.5%, which is below the limit value of 0.1. Thus, the comparison of the criteria utilized for this study was validated and accepted. Table IV displays the weight values of the considered criteria. These weight values were then applied in the Weighted Sum Tool to generate the indexed suitability map.

TABLE III. CONSISTENCY RATIO FOR THE COMPARISON MATRIX

CI	RCI (n=11)	CR	Percentage
0.068	1.510	0.045	4.5%

TABLE IV. CRITERIA WEIGHT VALUES GENERATED FROM AHP

Criteria	Weightage
Port	0.024
Shoreline	0.043
Oil and gas field	0.138
Fish main landing point	0.021
Underwater cables	0.151
Underwater pipelines	0.151
National marine parks	0.065
Marine protected areas	0.065
Shipping lane	0.212
Coral reefs	0.065
Sea turtle nesting	0.065
Sum of columns	1

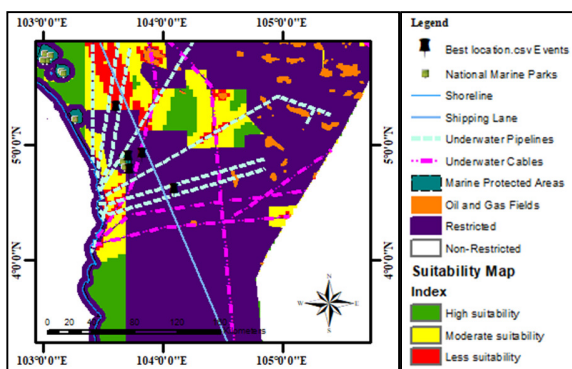


Fig. 7. The area proposed in [31] (indicated by the black pinpoint) and embedded into the suitability map generated by this study. The best location falls within the restricted area.

The deployment of Wave Energy Converters (WECs) depends not only on available resources, wave characteristics, and depth, but also on a detailed site assessment [30]. In [31], it was determined that Terengganu and Sarawak waters were the most suitable areas to deploy WECs. The selection was based on available wave resources, depth analysis, and hot-spot area analysis. Figure 7 portrays the five best locations in Terengganu waters that were integrated into the suitability map generated in this study. The analysis reveals that the locations proposed in [31] are in restricted areas for energy development, making them less suitable for WEC deployment. This study provides a more thorough analysis to identify suitable sites for wave renewable energy development compared to previous research. Based on the site suitability analysis for wave energy resources, Figure 8 shows the most suitable locations with high wave energy resources identified in Sarawak offshore areas (zones K9, K10, K15, K18, and K19). Among them, the K18

zone has the highest wave resource with an average of 3.99 kW/m and wave height in the range of 3.82 to 4.21 m.

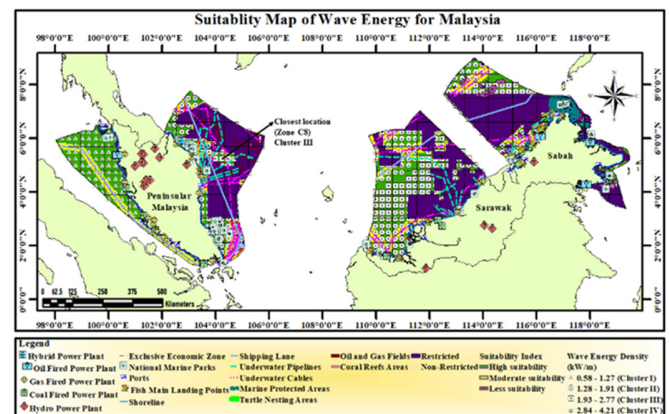


Fig. 8. Suitability map of wave energy resource for Malaysia

TABLE V. HIGH SUITABILITY AREA FOR WAVE ENERGY IN MALAYSIA'S EEZ

Location	Zone	Nearby conflicts
Terengganu	C8 (offshore)	Exploration contract areas Underwater cables Underwater pipelines Shipping lane
Sarawak	K9, K10, K11, K16, K18, K19 (offshore)	Exploration contract areas Underwater cables Underwater pipelines Shipping lane
Sabah	S11, S12, S13, S14, S15 (offshore)	Exploration contract areas Underwater cables Coral reefs

IV. CONCLUSION

This study significantly contributes to the exploration of Malaysia's wave energy potential, utilizing multimission satellite altimetry and the AHP for site suitability analysis. A comprehensive approach was adopted, leveraging extensive significant wave height data from various altimetry missions to address the challenges posed by the lack of in-situ data in Malaysia's ocean renewable energy development. The study showed that satellite altimeter data contribute to the accurate measurement of significant wave heights, having a good correlation coefficient and RMSE value. Additional analysis, using multi-criteria decision-making from the AHP approach to generate the criteria weight for further suitability analysis, revealed a good consistency ratio of 0.045, which is less than 0.1. This result indicates that the criteria judgment is good and acceptable to be used in the Weighted Sum tool of ArcGIS to generate the suitability map. Seven zones (K9, K10, K15, K18, K19, S11, and S15) were identified as highly suitable locations with high wave resources (2.84-4.21 kW/m). The outcomes of these suitability analyses generated a local conceptual suitability index, which is a useful evaluation indicator in site selection analysis. The combination of satellite altimetry, AHP, and GIS presents a comprehensive and tailored approach to overcome the specific challenges, particularly in the feasibility study of wave energy and the selection of suitable locations. This study also offers valuable information for stakeholders

and the government in advancing the development of wave energy as an alternative source of electricity.

ACKNOWLEDGMENT

The authors thank TU Delft Altimetrics LLC for providing altimetry data through the Radar Altimeter Database System (RADS). This project is funded by the Ministry of Higher Education (MOHE) Malaysia under the Universiti Teknologi Malaysia Fundamental Research Grant (UTMFR) (UTM Vote Number: Q.J130000.3852.22H75).

REFERENCES

- [1] C. S. Khor and G. Lalchand, "A review on sustainable power generation in Malaysia to 2030: Historical perspective, current assessment, and future strategies," *Renewable and Sustainable Energy Reviews*, vol. 29, pp. 952–960, Jan. 2014, <https://doi.org/10.1016/j.rser.2013.08.010>.
- [2] N. A. Basri, A. T. Ramli, and A. S. Aliyu, "Malaysia energy strategy towards sustainability: A panoramic overview of the benefits and challenges," *Renewable and Sustainable Energy Reviews*, vol. 42, pp. 1094–1105, Feb. 2015, <https://doi.org/10.1016/j.rser.2014.10.056>.
- [3] K. A. Samo, I. A. Samo, Z. A. Siyal, and A. R. H. Rigit, "Determination of Potential Tidal Power Sites at East Malaysia," *Engineering, Technology & Applied Science Research*, vol. 10, no. 4, pp. 6047–6051, Aug. 2020, <https://doi.org/10.48084/etasr.3674>.
- [4] S. Zen, E. Hart, and E. Medina-Lopez, "The use of satellite products to assess spatial uncertainty and reduce life-time costs of offshore wind farms," *Cleaner Environmental Systems*, vol. 2, Jun. 2021, Art. no. 100008, <https://doi.org/10.1016/j.cesys.2020.100008>.
- [5] O. Yaakob, F. E. Hashim, K. Mohd Omar, A. H. Md Din, and K. K. Koh, "Satellite-based wave data and wave energy resource assessment for South China Sea," *Renewable Energy*, vol. 88, pp. 359–371, Apr. 2016, <https://doi.org/10.1016/j.renene.2015.11.039>.
- [6] S. Horwath, J. Hassrick, R. Grismala, and E. Diller, "Comparison of Environmental Effects from Different Offshore Wind Turbine Foundations," Bureau of Ocean Energy Management, Fairfax, VA, USA, OCS Study BOEM 2020-041.
- [7] L. Y. Kai *et al.*, "Current Status and Possible Future Applications of Marine Current Energy Devices in Malaysia: A Review," *IEEE Access*, vol. 9, pp. 86869–86888, 2021, <https://doi.org/10.1109/ACCESS.2021.3088761>.
- [8] "United Nations Convention on the Law of the Sea." United Nations, Dec. 1982, [Online]. Available: https://www.un.org/Depts/los/convention_agreements/convention_overview_convention.htm.
- [9] J. J. Berman, *Principles and Practice of Big Data: Preparing, Sharing, and Analyzing Complex Information*. Academic Press, 2018.
- [10] P. R. Shanas, V. Sanil Kumar, and N. K. Hithin, "Comparison of gridded multi-mission and along-track mono-mission satellite altimetry wave heights with in situ near-shore buoy data," *Ocean Engineering*, vol. 83, pp. 24–35, Jun. 2014, <https://doi.org/10.1016/j.oceaneng.2014.03.014>.
- [11] J. Kang, R. Mao, Y. Chang, and H. Fu, "Comparative analysis of significant wave height between a new Southern Ocean buoy and satellite altimeter," *Atmospheric and Oceanic Science Letters*, vol. 14, no. 5, Sep. 2021, Art. no. 100044, <https://doi.org/10.1016/j.aosl.2021.100044>.
- [12] F. E. Hashim, "Assessment of wave energy resources in Malaysia using satellite altimetry," Universiti Teknologi Malaysia, Johor, Malaysia, 2016.
- [13] "IEC TS 62600-101:2015 Marine energy - Wave, tidal and other water current converters - Part 101: Wave energy resource assessment and characterization," International Electrotechnical Commission, 2015.
- [14] L. Kilcher, M. Fogarty, and M. Lawson, "Marine Energy in the United States: An Overview of Opportunities," National Renewable Energy Laboratory, NREL/TP-5700-78773, Feb. 2021, <https://doi.org/10.2172/1766861>.
- [15] N. A. B. Mabahwi, Y. Bhattacharya, and H. Nakamura, "GIS-based multi-criteria analysis to identify site suitability of flood shelters in Kuantan, Malaysia," *IOP Conference Series: Earth and Environmental Science*, vol. 799, no. 1, Mar. 2021, Art. no. 012027, <https://doi.org/10.1088/1755-1315/799/1/012027>.
- [16] A. O. Olaniyi, A. J. Ajiboye, A. M. Abdullah, M. F. Ramli, and A. M. Sood, "Agricultural land use suitability assessment in Malaysia," *Bulgarian Journal of Agricultural Science*, vol. 21, no. 3, pp. 560–572, 2015.
- [17] H. I. Mohammed, Z. Majid, Y. B. Yamusa, M. F. M. Ariff, K. M. Idris, and N. Darwin, "Sanitary Landfill Siting Using GIS and AHP: A Case Study in Johor Bahru, Malaysia," *Engineering, Technology & Applied Science Research*, vol. 9, no. 3, pp. 4100–4104, Jun. 2019, <https://doi.org/10.48084/etasr.2633>.
- [18] M. Bagheri, W. N. A. Sulaiman, and N. Vaghefi, "Land use suitability analysis using multi criteria decision analysis method for coastal management and planning: a case study of Malaysia," *Journal of Environmental Science and Technology*, vol. 5, no. 5, pp. 364–372, 2012.
- [19] Z. M. Tarmidi, A. R. Mohd Shariff, A. R. Mahmud, Z. Zaiton Ibrahim, and A. H. Hamzah, "Spatial data sharing implementation in Malaysia's marine organisations: a case study," *Journal of Spatial Science*, vol. 61, no. 1, pp. 209–216, Jan. 2016, <https://doi.org/10.1080/14498596.2015.1084248>.
- [20] F. Zahedi, "The Analytic Hierarchy Process—A Survey of the Method and its Applications," *Interfaces*, vol. 16, no. 4, pp. 96–108, Aug. 1986, <https://doi.org/10.1287/inte.16.4.96>.
- [21] T. L. Saaty and L. G. Vargas, "The Analytic Network Process," in *Decision Making with the Analytic Network Process: Economic, Political, Social and Technological Applications with Benefits, Opportunities, Costs and Risks*, T. L. Saaty and L. G. Vargas, Eds. Boston, MA: Springer US, 2013, pp. 1–40.
- [22] B. M. Thibadeau, "Prioritizing Project Risks Using AHP," presented at the PMI Global Congress North America 2007, Atlanta, GA, USA, 2007.
- [23] S. Alshehri, "Multicriteria Decision Making (MCDM) Methods for Ranking Estimation Techniques in Extreme Programming," *Engineering, Technology & Applied Science Research*, vol. 8, no. 3, pp. 3073–3078, Jun. 2018, <https://doi.org/10.48084/etasr.2104>.
- [24] R. Md. Sum, "Risk prioritisation using the analytic hierarchy process," *AIP Conference Proceedings*, vol. 1691, no. 1, Dec. 2015, Art. no. 030028, <https://doi.org/10.1063/1.4937047>.
- [25] S. A. Janssen Peter, "Monitoring Waves and Surface Winds by Satellite Altimetry: Applications," in *Satellite Altimetry Over Oceans and Land Surfaces*, CRC Press, 2017.
- [26] K. H. B. Shaeb, A. Anand, A. K. Joshi, and S. M. Bhandari, "Comparison of Near Coastal Significant Wave Height Measurements from SARAL/AltiKa with Wave Rider Buoys in the Indian Region," *Marine Geodesy*, vol. 38, no. sup1, pp. 422–436, Sep. 2015, <https://doi.org/10.1080/01490419.2015.1029655>.
- [27] H. H. Tsai, H. S. Tseng, C. K. Huang, and S. C. Yu, "Review on the Conflicts between Offshore Wind Power and Fishery Rights: Marine Spatial Planning in Taiwan," *Energies*, vol. 15, no. 22, Jan. 2022, Art. no. 8768, <https://doi.org/10.3390/en15228768>.
- [28] T. Chen, "A diversified AHP-tree approach for multiple-criteria supplier selection," *Computational Management Science*, vol. 18, no. 4, pp. 431–453, Oct. 2021, <https://doi.org/10.1007/s10287-021-00397-6>.
- [29] L. I. Tegou, H. Polatidis, and D. A. Haralambopoulos, "Environmental management framework for wind farm siting: Methodology and case study," *Journal of Environmental Management*, vol. 91, no. 11, pp. 2134–2147, Nov. 2010, <https://doi.org/10.1016/j.jenvman.2010.05.010>.
- [30] N. Guillou, G. Lavidas, and G. Chapalain, "Wave Energy Resource Assessment for Exploitation—A Review," *Journal of Marine Science and Engineering*, vol. 8, no. 9, Sep. 2020, Art. no. 705, <https://doi.org/10.3390/jmse8090705>.
- [31] N. A. M. Nasir and K. N. A. Maulud, "Wave power potential in Malaysian territorial waters," *IOP Conference Series: Earth and Environmental Science*, vol. 37, no. 1, Art. no. 012018, Mar. 2016, <https://doi.org/10.1088/1755-1315/37/1/012018>.

Investigating the Response Variability of Statically Determined Sandwich Beams considering two Random Fields of Elastic Modulus

Dao Ngoc Tien

Hanoi Architectural University, Vietnam
tiendn@hau.edu.vn

Tran The Hiep

University of Transport and Communications, Vietnam
tranthehiep@utc.edu.vn

Hoang Van Thanh

Vietnam Maritime University, Vietnam
hoangvanthanh@vimaru.edu.vn

Nguyen Van Thuan

Nha Trang University, Vietnam
thuannv@ntu.edu.vn (corresponding author)

Received: 21 November 2023 | Revised: 24 December 2023, 7 January 2024, and 15 January 2024 | Accepted: 16 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6652>

ABSTRACT

In this paper, the displacement variation in sandwich beams is determined by employing a semi-analytical approach. The classical displacement is calculated by integration using Mohr's equation, although the integration is complicated due to the inclusion of random fields in the inertial moment term. Using the trapezoidal rule to compute these integrals, the random fields are discretized into random variables at the nodal point of the beam segments. Thus, the expected displacement, standard deviation, and coefficient of variation can be computed. To validate the results, the random fields are simulated using a previously described spectral method. The results of numerical examples were compared with the semi-analytical method and the Monte Carlo simulation demonstrating the high accuracy of the proposed method. The results also illustrate the influence of the parameters of the random fields of elastic modulus on the variability of displacement.

Keywords-sandwich beam; random field; semi-analytical approach; MSC

I. INTRODUCTION

The main advantage of sandwich beams is their high stiffness-to-bending weight ratio, enabling numerous applications in automobile industry [1, 2] and structural engineering [3, 4]. Researchers who study the analysis of structural elements, such as beams, frames, piles, and plates, have the option to employ various methods, including analytical [5-10], semi-analytical [11, 12], and numerical techniques such as the Finite Element Method (FEM) [13-15]. For example, authors in [16] analyzed sandwich beams with functionally graded materials using the scaled boundary FEM. In reality, many input parameters in structural calculations are

often assumed to be random variables. The random oscillation of a moving load on a rough road surface has been investigated according to international standards using Matlab in [17]. In [18], the variation of deformation of non-uniform columns has been studied using the stochastic FEM with a weighted technique to discretize a random field. Authors in [19] developed stochastic finite elements using the point method to discretize a random field and determine the critical loads of non-uniform columns. Authors in [20] applied Karhunen-Loève expansion to develop the stochastic FEM for prestressed beams and frames [20]. In [21], the variation in the natural frequency of a beam was computed using stochastic FEM with a random field approach for the elastic modulus.

For sandwich structures, many studies addressed various types of mechanic problems, including static, dynamic, stability, and random problems. Authors in [22] used the R-functions method to investigate the critical loads on a sandwich plate subjected to non-uniform loading. Authors in [23] investigated the natural frequencies of functionally graded sandwich beams while considering trigonometric shear deformation. Authors in [24] developed the hybrid stochastic time-domain spectral element method for higher-order sandwich beams using the Karhunen–Loève expansion. The transient responses of a sandwich beam on a spring support under a moving load have also been investigated with FEM [25]. Although various studies have considered random parameters in sandwich beams, all of them used the FEM, and to the best of our knowledge, there are no reports using a semi-analytical approach. Despite numerous contributions to problems involving random parameters, there have been no reported calculations for a sandwich beam using a classical method for determining displacement and considering random fields in the elastic modulus of multiple material layers.

This paper investigates the displacement variability of a statically determined sandwich beam subjected to a distributed load, considering two random fields of elastic modulus, using a semi-analytical method. Calculating variations of displacements and internal forces at an advanced level will be used to evaluate the reliability of the structures.

II. DISPLACEMENT OF A SANDWICH BEAM COMPUTED BY THE CLASSICAL APPROACH

Consider the simply supported sandwich beam subjected to a uniformly distributed load, as shown in Figure 1.

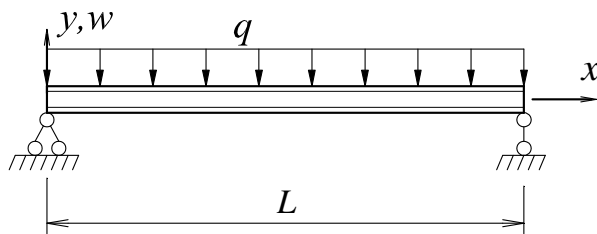


Fig. 1. A simply supported sandwich beam subjected to a uniformly distributed load

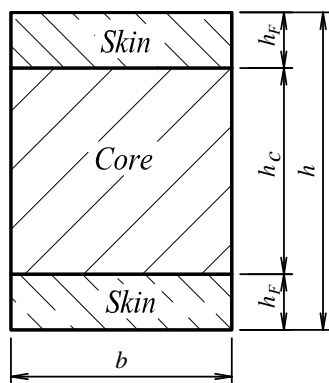


Fig. 2. Section of the sandwich beam.

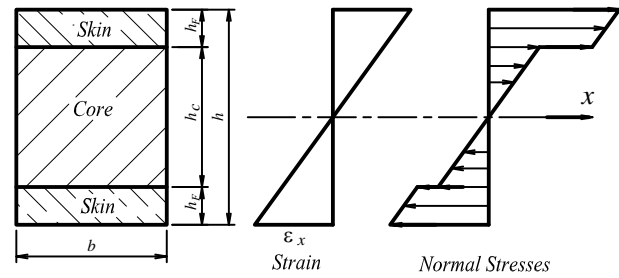


Fig. 3. Strain and normal stress distributions on a sandwich beam.

$$U_e = \frac{1}{2} \int_0^L \left\{ \underbrace{\int_A E_{Face} \left[y \frac{\partial^2 w}{\partial x^2} \right]^2 dA}_{Face} + \underbrace{\int_A E_{Core} \left[y \frac{\partial^2 w}{\partial x^2} \right]^2 dA}_{Core} \right\} dx \quad (1)$$

The displacement at the middle of the beam using Mohr's equation from classical mechanics is expressed as [26]:

$$w_m = \int_0^{L/2} \frac{M(x) \bar{M}(x)}{EI_{eq}} dx + \int_{L/2}^L \frac{M(x) \bar{M}(x)}{EI_{eq}} dx \quad (2)$$

The equivalent flexural rigidity is described by:

$$EI_{eq} = \frac{E_c b h_c^3}{12} + 2E_F b \left[\frac{h_F^3}{12} + \left(\frac{h}{2} - \frac{h_F}{2} \right)^2 \right] \quad (3)$$

The flexural diagram of a simple sandwich beam in both its actual and virtual states is shown in Figure 4.

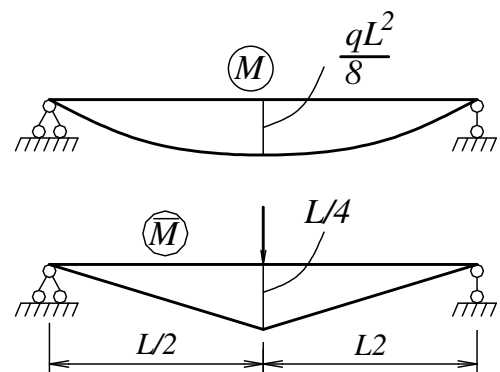


Fig. 4. Bending moment diagrams of a sandwich beam in the actual (top) and virtual (bottom) states.

The moment expressions for a sandwich beam are expressed by:

- Actual states:

$$M = \frac{qL}{2} x - \frac{qx^2}{2} \quad (4)$$

- Virtual states:

$$\bar{M} = \begin{cases} \frac{x}{2} & \text{if } x \leq \frac{L}{2} \\ \frac{L-x}{2} & \text{if } x \geq \frac{L}{2} \end{cases} \quad (5)$$

III. SEMI-ANALYTICAL SOLUTION

The random fields of the elastic moduli in the skin and core are assumed to be one-dimensional and homogeneous:

$$\begin{aligned} E_c(x) &= [1 + r_c(x)] E_{c0} \\ E_F(x) &= [1 + r_F(x)] E_{F0} \end{aligned} \quad (6)$$

where $r_c(x)$, $r_F(x)$ are one-dimensional homogeneous random fields with a mean of zero. Using the coefficients of variation σ_c , σ_F and correlation distances d_c , d_F for the random fields $r_c(x)$, $r_F(x)$, the autocorrelation functions of these random fields are described by:

$$\begin{aligned} R_c(\xi) &= \sigma_c^2 \exp\left(-\frac{|\xi|}{d_c}\right) \\ R_F(\xi) &= \sigma_F^2 \exp\left(-\frac{|\xi|}{d_F}\right) \end{aligned} \quad (7)$$

where $\xi = x_j - x_i$ is the relative distance vector.

The random fields of the elastic moduli in the beam are approximated by random variables r_i at n points in the element:

$$EI_{eq} = EI_0 + \alpha r_c(x) + \beta r_F(x) \quad (8)$$

where:

$$\begin{aligned} EI_0 &= \frac{E_{c0} b h_c^3}{12} + 2E_{F0} b \left[\frac{h_F^3}{12} + h_F \left(\frac{h}{2} - \frac{h_F}{2} \right)^2 \right] \\ \alpha &= \frac{E_{c0} b h_c^3}{12} \\ \beta &= 2E_{F0} b \left[\frac{h_F^3}{12} + h_F \left(\frac{h}{2} - \frac{h_F}{2} \right)^2 \right] \end{aligned} \quad (9)$$

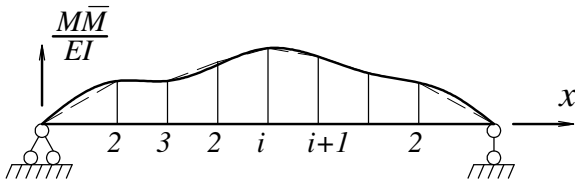


Fig. 5. Randomness in the elastic modulus of the beam.

The displacement is approximated by:

$$w_m = \frac{1}{2} \sum \Delta x \left\{ \frac{M_i(x) \bar{M}_i(x)}{(EI_{eq})_i} + \frac{M_{i+1}(x) \bar{M}_{i+1}(x)}{(EI_{eq})_{i+1}} \right\} \quad (10)$$

Ignoring the infinitesimal term of higher order, displacement is approximated using the following expression:

$$\begin{aligned} w_m &\approx \sum \frac{M_i(x) \bar{M}_i(x)}{(EI_{eq})_i} \Delta x \\ &\approx \sum \frac{M_i(x) \bar{M}_i(x)}{EI_0} \left[1 - \alpha \frac{r_c(x)}{EI_0} - \beta \frac{r_F(x)}{EI_0} \right] \Delta x \end{aligned} \quad (11)$$

Then, the expected displacement is:

$$\begin{aligned} \bar{W} &= E \left\{ \sum \frac{M_i(x) \bar{M}_i(x)}{EI_0} \left[1 - \alpha \frac{r_c(x)}{EI_0} - \beta \frac{r_F(x)}{EI_0} \right] \Delta x \right\} \\ &= \sum \left\{ \frac{M_i(x) \bar{M}_i(x)}{EI_0} \right\} \end{aligned} \quad (12)$$

The variance of the displacement is:

$$\begin{aligned} \text{var}(w_m) &= \\ &= E \left\{ \left(\sum \frac{M_i(x) \bar{M}_i(x)}{EI_0} \left[1 - \alpha \frac{r_c(x)}{EI_0} - \beta \frac{r_F(x)}{EI_0} \right] \Delta x - \bar{W} \right)^2 \right\} \end{aligned} \quad (13)$$

The coefficient of variation (COV) for the displacement is:

$$\text{COV} = \frac{\sqrt{\text{var}(w_m)}}{\bar{W}} \quad (14)$$

IV. MONTE CARLO SIMULATION

The previous section presented the semi-analytical approach used for determining the mean and variance of displacements. However, it is essential to validate the results of this approach. Therefore, Monte Carlo simulation is conducted using the spectral representation method [27]. In this, the random field of the elastic modulus is generated by:

$$\begin{aligned} r_c(x) &= \sqrt{2} \sum_{n=0}^{N-1} A_n^c \cos(\omega_n^c x + \phi_n^c) \\ r_F(x) &= \sqrt{2} \sum_{n=0}^{N-1} A_n^F \cos(\omega_n^F x + \phi_n^F) \end{aligned} \quad (15)$$

$$\begin{aligned} A_n^F &= \sqrt{2S_{ff}^F(\omega_n^F) \Delta \omega^F} \\ A_n^c &= \sqrt{2S_{ff}^c(\omega_n^c) \Delta \omega^c} \\ \Delta \omega^F &= \frac{\omega_u^F}{N}; \quad \Delta \omega^c = \frac{\omega_u^c}{N} \\ \omega_n^F &= n \Delta \omega^F, \quad \omega_n^c = n \Delta \omega^c, \quad n = 0, 1, 2, \dots, N-1 \end{aligned} \quad (16)$$

where ω_n^F, ω_n^c denote the upper cut-off frequency for the skin and core of a sandwich beam, beyond which the power spectral density functions $S_{ff}^F(\omega_n^F), S_{ff}^c(\omega_n^c)$ denote the random field of stiffness for the elastic moduli in the skin and core of a sandwich beam, respectively.

By substituting (15) into (11) the displacement can be evaluated by:

$$w_m \approx \sum \frac{M_i(x) \bar{M}_i(x)}{EI_0} \left[1 - \frac{\alpha}{EI_0} \left\{ \sqrt{2} \sum_{n=0}^{N-1} A_n^c \cos(\omega_n^c x + \phi_n^c) \right\} - \frac{\beta}{EI_0} \left\{ \sqrt{2} \sum_{n=0}^{N-1} A_n^F \cos(\omega_n^F x + \phi_n^F) \right\} \right] \Delta x \quad (17)$$

V. NUMERICAL EXAMPLES

A sandwich beam subjected to uniformly distributed loads, is considered for the evaluation of the semi-analytical approach (Figure 1). The sandwich beam has a length of $L = 10$ m and height of $h = 60$ cm, with $h_F = 2$ cm and $h_c = 56$ cm. The material properties of the beam are given in Table I.

TABLE I. MATERIAL PROPERTIES OF THE SANDWICH BEAM

Fiber	Material	Young's modulus (GPa)	Coefficient of variation
Skin layer	Steel	200	0.1
Core	Concrete	30	0.1

The sandwich beam is divided into 20 segments, and variations in the two random fields are considered while maintaining an equal coefficient of variation $\sigma_c = \sigma_F$ and equal correlation length distance $d_c = d_F = d$ (m). In Figure 6, we compare the coefficient of variation for the displacements of the sandwich beam obtained using the semi-analytical method (the proposed approach) and the Monte Carlo simulation. The variability coefficient results of the semi-analytical method are consistent with the numerical Monte Carlo simulation.

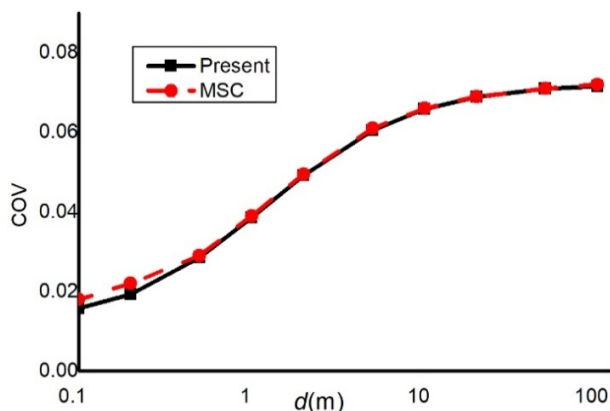


Fig. 6. Coefficients of variation for the beam displacement determined with the proposed method and the Monte Carlo simulation.

As the next step, different cases are employed to investigate the influence of random field parameters for the elastic moduli in the two materials on the displacement variability coefficient. Three cases of the coefficient of variation for the two random fields of the elastic moduli for the skin and core layers are considered, as shown in Table II.

Figure 7 illustrates the variability coefficient of the beam displacement corresponding to three cases of variation in the elastic moduli of the material properties, showing that the coefficient of variation increases with the correlation length. In general, the curve of the coefficient of variation for the displacement in case 1, corresponding to the smallest input value of the coefficient of variation, exhibits the smallest coefficient of variation. Additionally, the curve of the coefficient of variation for the displacement in case 3, corresponding to the largest input value of the coefficient of variation, shows the largest coefficient of variation. Thus, the coefficient of variation for the displacement is directly proportional to the coefficient of variation for random elastic modulus.

TABLE II. COEFFICIENT OF VARIATION FOR THE ELASTIC MODULUS

Case	Skin layer	Core
1	0.05	0.10
2	0.10	0.10
3	0.10	0.05

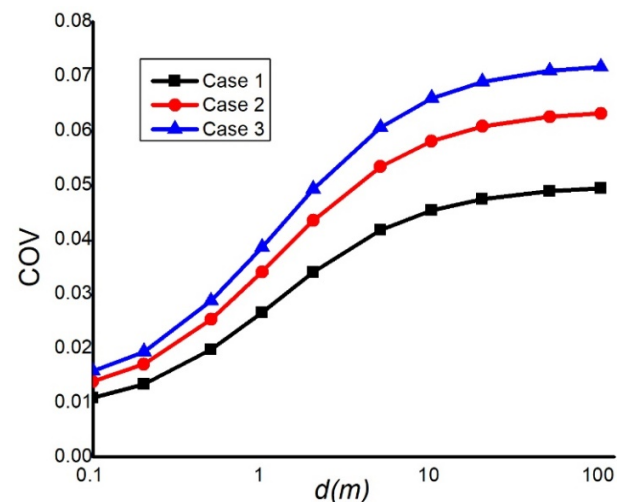


Fig. 7. Coefficients of variation for the beam displacement, determined with the proposed method, considering three cases with different material properties.

VI. CONCLUSIONS

This paper demonstrates a semi-analytical solution for calculating the deflection variation of a statically determined simply supported sandwich beam subjected to a uniformly distributed load. Two random fields of the elastic modulus are discretized by random variables at the beam division nodes. Using Mohr's equation to calculate displacement, an approximate expression for the mean and variance of the displacement is determined. The numerical results are validated

with Monte Carlo simulations, demonstrating the accuracy of the proposed semi-analytical approach. Numerical examples illustrate the significant impact of the correlation length distance on the variability coefficient of the displacement. The coefficient of variation of the displacement when the distance is correlated to infinity is smaller than the largest number of the input coefficient of variation of the elastic modulus in the two material layers.

REFERENCES

- [1] B. Du, Q. Li, C. Zheng, S. Wang, C. Gao, and L. Chen, "Application of Lightweight Structure in Automobile Bumper Beam: A Review," *Materials*, vol. 16, no. 3, Jan. 2023, Art. no. 967, <https://doi.org/10.3390/ma16030967>.
- [2] F. A. Fazzolari, "Sandwich Structures," in *Stability and Vibrations of Thin-Walled Composite Structures*, H. Abramovich, Ed. Cambridge, MA, USA: Woodhead Publishing, 2017, pp. 49–90.
- [3] C. A. Steeves and N. A. Fleck, "Material selection in sandwich beam construction," *Scripta Materialia*, vol. 50, no. 10, pp. 1335–1339, May 2004, <https://doi.org/10.1016/j.scriptamat.2004.02.015>.
- [4] M. Yan, L. Wang, B. Chen, and H. Gao, "Deflection assessment of prestressed steel-concrete-steel sandwich panel: experiment and numerical simulation," *Journal of Sandwich Structures & Materials*, vol. 25, no. 3, pp. 351–371, Mar. 2023, <https://doi.org/10.1177/10996362221140072>.
- [5] D. Wu, Y. Lei, Z. Wang, B. Yu, and D. Zhang, "Free Vibration Analysis of Carbon-Nanotube-Reinforced Beams Resting on a Viscoelastic Pasternak Foundation by the Nonlocal Eshelby–Mori–Tanaka Method," *Mechanics of Composite Materials*, vol. 59, no. 3, pp. 479–494, Jul. 2023, <https://doi.org/10.1007/s11029-023-10110-0>.
- [6] B. Eshmatov, R. Abdikarimov, M. Amabili, and N. Vatin, "Magazine of Civil Engineering Nonlinear vibrations and dynamic stability of viscoelastic anisotropic fiber reinforced plates," *Magazine of Civil Engineering*, vol. 118, no. 1, Apr. 2023, Art. no. 11811, <https://doi.org/10.34910/MCE.118.11>.
- [7] M. Rezaiee-Pajand, A. R. Masoodi, and A. Alepaighambar, "Lateral-Torsional Buckling of a Bidirectional Exponentially Graded Thin-Walled C-Shaped Beam," *Mechanics of Composite Materials*, vol. 58, no. 1, pp. 53–68, Mar. 2022, <https://doi.org/10.1007/s11029-022-10011-8>.
- [8] T. C. T. Ngoc, "Analytical truss model for concrete beams reinforced with FRP bars," *Transport and Communications Science Journal*, vol. 74, no. 4, pp. 456–468, May 2023, <https://doi.org/10.47869/tcsj.74.4.6>.
- [9] H. D. Ta, K. T. Nguyen, T. D. Ngoc, H. T. Do, T. X. Nguyen, and D. D. Nguyen, "Approximation solution for steel concrete beam accounting high-order shear deformation using trigonometric-series," *Journal of Materials and Engineering Structures*, vol. 9, no. 4, pp. 599–605, Dec. 2022.
- [10] A. M. D. de Sousa, L. P. Prado, and M. K. El Debs, "Reliability-based design of reinforced concrete pipes to satisfy the TEBT," *Latin American Journal of Solids and Structures*, vol. 20, Aug. 2023, Art. no. e500, <https://doi.org/10.1590/1679-78257510>.
- [11] J. Liu, B. He, W. Ye, and F. Yang, "High performance model for buckling of functionally graded sandwich beams using a new semi-analytical method," *Composite Structures*, vol. 262, Apr. 2021, Art. no. 113614, <https://doi.org/10.1016/j.compstruct.2021.113614>.
- [12] A. W. de Q. R. Reis, R. B. Burgos, and M. F. F. de Oliveira, "Nonlinear Dynamic Analysis of Plates Subjected to Explosive Loads," *Latin American Journal of Solids and Structures*, vol. 19, Jan. 2022, Art. no. e422, <https://doi.org/10.1590/1679-78256706>.
- [13] S. B. Akhazhanov, N. I. Vatin, S. Akhmediyev, T. B. Akhazhanov, O. Khabidolda, and A. Z. Nurgoziyeva, "Beam on a two-parameter elastic foundation: simplified finite element model," *Magazine of Civil Engineering*, vol. 121, no. 5, 2023, Art. no. 12107, <https://doi.org/10.34910/MCE.121.7>.
- [14] J. Singh and A. Kumar, "Vibration and the Buckling Response of Functionally Graded Plates According to a Refined Hyperbolic Shear Deformation Theory," *Mechanics of Composite Materials*, vol. 59, no. 4, pp. 725–742, Sep. 2023, <https://doi.org/10.1007/s11029-023-10127-5>.
- [15] A. V. Alekseytsev and S. A. Sazonova, "Numerical analysis of the buried fiber concrete slabs dynamics under blast loads," *Magazine of Civil Engineering*, vol. 117, no. 1, 2023, Art. no. 11703, <https://doi.org/10.34910/MCE.117.3>.
- [16] J. Liu, C. Hao, W. Ye, and Q. Zang, "Application of a new semi-analytic method in bending behavior of functionally graded material sandwich beams," *Mechanics Based Design of Structures and Machines*, vol. 51, no. 4, pp. 2130–2153, Apr. 2023, <https://doi.org/10.1080/15397734.2021.1890615>.
- [17] H. V. Quan, T. Canh, and V. P. Le, "Vehicle model dynamic analysis under random excitation of uneven pavement as measured by the international roughness index," *Transport and Communications Science Journal*, vol. 8, pp. 866–880, Oct. 2023, <https://doi.org/10.47869/tcsj.74.8.2>.
- [18] T. D. Hien, "A static analysis of nonuniform column by stochastic finite element method using weighted integration approach," *Transport and Communications Science Journal*, vol. 71, pp. 359–367, May 2020, <https://doi.org/10.25073/tcsj.71.4.5>.
- [19] D. T. Hang, X. T. Nguyen, and D. N. Tien, "Stochastic Buckling Analysis of Non-Uniform Columns Using Stochastic Finite Elements with Discretization Random Field by the Point Method," *Engineering, Technology & Applied Science Research*, vol. 12, no. 2, pp. 8458–8462, Apr. 2022, <https://doi.org/10.48084/etasr.4819>.
- [20] M. L. Larsen, S. Adhikari, and V. Arora, "Analysis of stochastically parameterized prestressed beams and frames," *Engineering Structures*, vol. 249, Dec. 2021, Art. no. 113312, <https://doi.org/10.1016/j.engstruct.2021.113312>.
- [21] H. T. Duy, N. D. Diem, G. V. Tan, V. V. Hiep, and N. V. Thuan, "Stochastic Higher-order Finite Element Model for the Free Vibration of a Continuous Beam resting on Elastic Support with Uncertain Elastic Modulus," *Engineering, Technology & Applied Science Research*, vol. 13, no. 1, pp. 9985–9990, Feb. 2023, <https://doi.org/10.48084/etasr.5456>.
- [22] L. Kurpa, T. Shmatko, and A. Linnik, "Buckling Analysis of Functionally Graded Sandwich Plates Resting on an Elastic Foundation and Subjected to a Nonuniform Loading," *Mechanics of Composite Materials*, vol. 59, no. 4, pp. 645–658, Sep. 2023, <https://doi.org/10.1007/s11029-023-10122-w>.
- [23] R. Kumar, A. Lal, and B. M. Sutaria, "Free Vibration of Porous Functionally Graded Sandwich Plates with Hole," *Journal of Vibration Engineering & Technologies*, vol. 11, no. 8, pp. 4205–4221, Nov. 2023, <https://doi.org/10.1007/s42417-022-00810-7>.
- [24] H. Sharma, S. Mukherjee, and R. Ganguli, "Uncertainty analysis of higher-order sandwich beam using a hybrid stochastic time-domain spectral element method," *International Journal for Computational Methods in Engineering Science and Mechanics*, vol. 21, no. 5, pp. 215–230, Aug. 2020, <https://doi.org/10.1080/15502287.2020.1808912>.
- [25] T. D. Hien, N. D. Hung, N. T. Hiep, G. V. Tan, and N. V. Thuan, "Finite Element Analysis of a Continuous Sandwich Beam resting on Elastic Support and Subjected to Two Degree of Freedom Sprung Vehicles," *Engineering, Technology & Applied Science Research*, vol. 13, no. 2, pp. 10310–10315, Apr. 2023, <https://doi.org/10.48084/etasr.5464>.
- [26] I. A. Karnovsky and O. Lebed, *Advanced Methods of Structural Analysis*, 2nd ed. New York, NY, USA: Springer, 2021.
- [27] M. Shinozuka and G. Deodatis, "Simulation of Stochastic Processes by Spectral Representation," *Applied Mechanics Reviews*, vol. 44, no. 4, pp. 191–204, Apr. 1991, <https://doi.org/10.1115/1.3119501>.

A Metaheuristic Approach of predicting the Dynamic Modulus in Asphalt Concrete

Ilham Yahya Amir

Department of Civil Engineering, Near East University, Cyprus
20227176@std.neu.edu.tr

Abdinasir Mohamed Yusuf

Department of Civil Engineering, Near East University, Cyprus | Department of Civil Engineering, Aden Adde International University, Somalia
20225266@std.neu.edu.tr

Ikenna D. Uwanuakwa

Department of Civil Engineering, Near East University, Cyprus
ikenna.uwanuakwa@neu.edu.tr (corresponding author)

Received: 24 December 2023 | Revised: 7 January 2024 | Accepted: 9 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6808>

ABSTRACT

The prediction of the asphalt dynamic modulus (E^*), which measures the material's ability to withstand changes in shape or structure, is important. Previous studies indicated that the well-known Witczak 1-40D model for E^* is outperformed by machine learning models. Additionally, the application of machine learning algorithms requires manual fine-tuning of their hyperparameters. In this study, the artificial Hummingbird and Harris Hawks optimization algorithms were employed in the automatic calibration of the Random Forest and Gradient Boost algorithms' hyperparameters for modeling E^* using the Witczak 1-40D model and additional parameters. In addition, the model was interpreted using the Shapley value and permutation feature importance. The results indicate that the optimized artificial hummingbird algorithm model performed better, with R^2 reaching 0.97. The interpretability of the model suggests that the binder parameters exhibited the highest effect on the variance of E^* .

Keywords-asphalt dynamic modulus; automatic calibration; modeling; machine learning

I. INTRODUCTION

Road quality is crucial for the economic development of a country and well-designed roads are vital to the economic progress of any region. The construction of a high-quality road requires careful knowledge of road distress and a proper selection of materials to withstand these stresses. Rutting is one of the asphalt pavement distresses prevalent on roads exposed to high temperatures and traffic loads. Asphalt concrete dynamic modulus (E^*), which measures the material's ability to withstand changes in shape or structure [1], is often used to characterize the resistance of a material to rutting. Several models have been developed to predict rutting, the most notable among them being the Witczak model [2-3]. The performance evaluation of predicting E^* using the Witczak model gave an R^2 of 0.94 with Gaussian process regression [4].

Previous studies have shown that machine learning evaluation performance can be improved when an optimization approach is applied to fine-tune model hyperparameters instead of performing manual trial-and-error calibration [5-7]. This is evident in the results reported in [8-12] in the prediction of E^* .

Despite the reported high accuracy, there are significant gaps in the literature:

- The databases used are limited.
- The input parameters do not account for all the parameters that affect the variance of E^* .
- Model interpretability is missing.

This study sought to address the aforementioned gaps using the Witczak database with 346 test cases, which totals 7,400 data points. Furthermore, the Witczak model input parameter was used, since it accounts for the binder, aggregate, void, and test parameters, in addition to five other parameters.

II. METHODOLOGY

This study focuses on the automatic calibration of traditional machine learning models, namely Random Forest (RF) and Gradient Boost Regression (GBR), to predict E^* using Witczak 1-40D parameters. This study aims to improve the prediction accuracy of E^* using the Witczak 1-40D parameters, six additional parameters, and the Witczak 1-40D

database. The parameters of the Witczak 1-40D are: r_{200} is the percentage passing a #200 sieve, r_4 is the cumulative percentage retained on a #4 sieve, r_{38} is the cumulative percentage retained on a $\frac{3}{8}$ " sieve, r_{34} is the cumulative percentage retained on a $\frac{3}{4}$ " sieve, V_a is the air void percentage, V_{beff} is the effective binder content percentage, $|G^*|$ is the binder complex shear modulus in psi, and db is the binder phase angle. The six additional parameters were: Asphalt Content (AC), Temperature (T), frequency of loading (f_s), air voids (A), Volumetric Total Solids (VTS), and phase angle of

the mixture (ϕ). This study also aims to provide model feature interpretability using the Shapley (SHAP) value and Permutation Feature Importance (PFI) analysis to identify the key parameters that significantly influence predictive accuracy. The Witczak 1-40D database used contains 5970 unique test cases and was partitioned into training and testing sets in a 65:35 ratio. The test set was used for the model evaluation after model training was completed. Table I presents the details of the features of the model, assessed with the relevant statistical parameters.

TABLE I. DESCRIPTIVE STATISTICS OF THE MODEL

	Mean	Standard Deviation	Sample Variance	Kurtosis	Skewness	Min	Max	Count
r_{34} (%)	3.96	6.51	42.38	3.23	1.98	0.00	26.10	5970
r_{38} (%)	25.75	11.39	129.83	-0.78	-0.51	2.60	43.00	5970
r_4 (%)	51.17	9.84	96.83	0.20	0.09	30.00	74.00	5970
r_{200} (%)	4.90	1.45	2.10	3.42	1.03	1.80	11.80	5970
AC (%)	5.42	1.17	1.37	1.62	1.20	3.60	8.90	5970
V_a (%)	6.81	2.25	5.06	3.21	0.63	0.70	18.13	5970
V_{beff} (%)	10.48	2.66	7.05	10.58	2.58	6.08	25.09	5970
A	10.52	0.77	0.59	2.28	-1.47	7.58	11.79	5970
VTS	-3.51	0.28	0.08	2.18	1.46	-3.98	-2.44	5970
$ G^* $ (psi)	1105.57	1480.93	2.19E+06	1.24	1.34	0.01	7386.60	5970
db (deg)	57.09	22.30	497.11	-1.00	-0.29	11.86	90.00	5970
T (°F)	70.30	42.09	1771.63	-1.28	0.00	0.00	130.00	5970
f_s (Hz)	1.14	1.43	2.04	-0.09	1.18	0.02	3.98	5970
ϕ (deg)	22.59	11.32	128.18	-0.03	0.21	1.59	123.40	5970
E^* (psi)	1.48E+06	1.55E+06	2.40E+12	3.59E-01	1.05E+00	1.05E+04	8.64E+06	5970

The trained models were evaluated using standard error metrics such as the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R-squared (R^2), Nash-Sutcliffe Efficiency (NSE), and Kling-Gupta efficiency (KGE). The mathematical expressions of these metrics are:

$$R^2 = \frac{SSR}{SST} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Y_{Predi} - Y_{Obi})^2}{N}} \quad (2)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |Y_{Obi} - Y_{Predi}| \quad (3)$$

$$KGE = 1 - \sqrt{(CC - 1)^2 + \left(\frac{cd}{rd} - 1\right)^2 + \left(\frac{cm}{rm} - 1\right)^2} \quad (4)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (Y_{Obi} - Y_{Predi})^2}{\sum_{i=1}^n (Y_{Obi} - \bar{Y})^2} \quad (5)$$

where $SSR = \sum (Y_{Pred,i} - \bar{Y})^2$, $SST = \sum (Y_{Obs,i} - \bar{Y})^2$, \bar{Y} is the mean of y value, N is the number of observed values, Y_{Predi} is the predicted value, Y_{Obi} is the observed value, CC represents the correlation coefficient, Cd is the coefficient of determination, rd denotes the relative deviation, Cm represents the mean error, and rm represents the relative mean error.

A. Model Training

The model was trained with the GBR and RF algorithms from the scikit-learn library in Python. Training these algorithms traditionally requires manual calibration of the algorithm parameters. This could be time-consuming and with a low possibility of reaching the optimum solution. For this reason, two metaheuristic optimization algorithms, Haris Hawks Optimization (HHO) and Artificial Hummingbird

Algorithm (AHA), were used to automatically calibrate the GBR and RF model training hyperparameters.

HHO is a metaheuristic algorithm inspired by hawks' hunting strategies. It demonstrates exceptional performance in many optimization problems, such as mobile robot path planning, and enhanced feature selection performance by integrating chaotic opposition and simulated annealing in an improved version of the algorithm consisting of two main stages, investigation and exploitation, utilizing equations with variables representing hawks' positions, prey location, solution space bounds, and random numbers. Although this algorithm is very effective, it can struggle with population diversity and local optima [13]. AHA [14] is a novel bioinspired design that is classified as a new metaheuristic optimizer, based on intelligent foraging strategies and unique flights of hummingbirds. AHA simulates three foraging behaviors, guided, territorial, and migration foraging, in the optimization process. It uses a visit table to model the flight abilities and clever foraging techniques of hummingbirds. AHA has three basic components: food sources, hummingbirds, and visit tables. Food sources measure nectar quality, nectar replenishment rate, and last visit to flowers. Hummingbirds are assigned to particular food sources and communicate with each other. For each food source, visit levels are recorded in visit tables [14].

For GBR, the key hyperparameters optimized are the number of estimators, learning rate, and maximum depth. These hyperparameters control the trade-off between accuracy and complexity in gradient boost models, where more trees, lower learning rate, and higher depth can improve the fit but also increase the risk of overfitting and the computational cost. The upper and lower bounds of the optimization algorithm

search ranges for these parameters were: 50-200 estimators, 0.01-1 learning rate, and 1-10 maximum depth. The search was carried out over a population of 50 in 180 iterations. RF optimization was applied to the number of estimators, maximum depth, and maximum features. The lower and upper values of the optimization algorithm search range were: 10-200 estimators, 1-50 maximum depth, and maximum feature range from 1 to the maximum number of input variables. The search was carried out over a population of 100 in 180 iterations.

B. Model Feature Selection

Model feature importance can help to understand how the model makes predictions, identifying and removing irrelevant or redundant features, and comparing and contrasting different models or datasets. There are different ways to calculate model feature importance, such as coefficients from linear models, decision trees, permutation importance, and SHAP values. This study used PFI and SHAP analysis to offer a comprehensive, multi-faceted evaluation of feature importance in the predictive models.

The SHAP value is a concept from cooperative game theory that distributes the total pay-off among the players of a game according to their marginal contribution. In this context, the players are the features, the game is the prediction of the model, and the payoff is some measure of the importance or influence of the feature subset. The SHAP value for a feature is calculated by averaging the difference in the model output when the feature is included or excluded from all possible subsets of features [15]. PFI is a technique that measures how much the prediction error of a machine learning model increases when the values of a feature are randomly shuffled. This technique can be used for any fitted model and any tabular data, and it is especially useful for non-linear or opaque models that are hard to explain. PFI can help in understanding which features are more important to the model and how they affect its performance and accuracy [16].

III. RESULTS AND DISCUSSION

This study evaluates the performance potentials of the AHA and HHO-optimized GB and RF models to predict E^* . This study is unique as it used two distinct sets of input features for its predictions. On the one hand, it used the Witczak 1-40D model input features, a conventional method known for its reliability in predictions. On the other hand, the model was also trained with a combination of Witczak 1-40D input features and six additional features: AC , T , f_s , A , VTS , and ϕ . This dual approach in feature evaluation aimed to explore the extent to which the integration of these additional features could enhance the model's accuracy in predicting E^* , thereby providing a comprehensive understanding of the model's predictive capabilities. Furthermore, the interpretability of the model parameters was evaluated using the SHAP value and PFI.

A. Performance using Witczak 1-40D Input Parameters

The evaluation of Witczak parameters using GB and RF models, reveals significant insights into their predictive accuracy and efficiency. While the GB model exhibits excellent performance in the training phase (R^2 ranging from 0.90 to 0.99, NSE from 0.91 to 0.99, KGE from 0.90 to 0.99), it shows a slight decline in the testing phase (R^2 from 0.89 to 0.93, NSE from 0.90 to 0.93, KGE from 0.89 to 0.96), indicating a potential overfitting to the training data. On the other hand, the RF model demonstrates remarkable consistency and robustness, maintaining high scores across all metrics in both training (R^2 , NSE , KGE all around 0.98 to 0.99) and testing phases (R^2 and NSE from 0.92 to 0.94 and KGE from 0.92 to 0.93). Figure 1 shows that RF models are more fitted compared to GB models. This consistent high performance of the RF model, especially in the testing phase, underscores its superior ability to generalize, making it a more reliable choice for applications requiring stable and accurate predictions on new, unseen data.

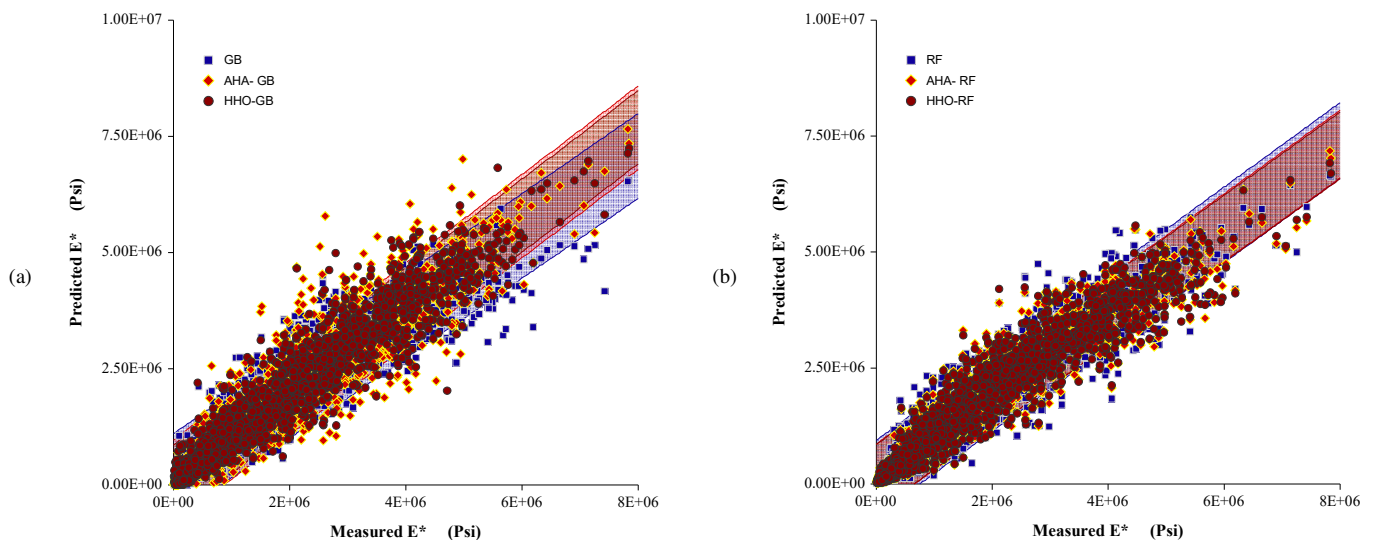


Fig. 1. Correlation of measured vs predicted dynamic modulus using Witczak input parameters: (a) Gradient Boosting Regression, (b) Random Forest.

TABLE II. OUTPUT TRAIN AND TEST USING WITCZAK PARAMETER

Models	Train					Test				
	R^2	RMSE	MAE	NSE	KGE	R^2	RMSE	MAE	NSE	KGE
GB	0.90	4.71E+05	3.06E+05	0.91	0.90	0.89	5.14E+05	3.38E+05	0.90	0.89
AHA-GB	0.99	8.52E+04	1.05E+04	0.99	0.99	0.92	4.64E+05	2.76E+05	0.91	0.96
HHO-GB	0.98	1.91E+05	1.32E+05	0.98	0.99	0.93	4.08E+05	2.72E+05	0.93	0.96
RF	0.98	1.83E+05	1.07E+05	0.99	0.97	0.92	4.39E+05	2.74E+05	0.92	0.93
AHA-RF	0.98	1.69E+05	9.85E+04	0.99	0.97	0.94	4.05E+05	2.53E+05	0.93	0.92
HHO-RF	0.98	1.66E+05	9.64E+04	0.99	0.96	0.94	4.03E+05	2.52E+05	0.94	0.92

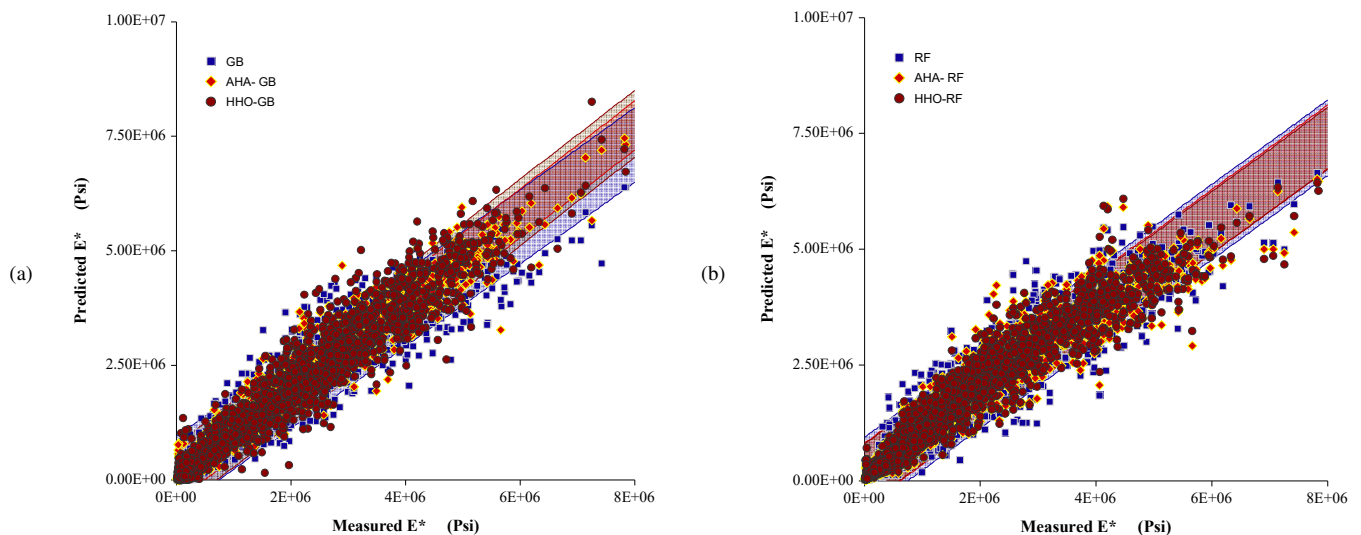


Fig. 2. Correlation of measured vs predicted dynamic modulus using all parameters: (a) GBR, (b) RF.

TABLE III. OUTPUT OF TRAIN AND TEST USING ALL PARAMETERS

Models	Train					Test				
	R^2	RMSE	MAE	NSE	KGE	R^2	RMSE	MAE	NSE	KGE
GB	0.93	3.95E+05	2.54E+05	0.93	0.93	0.92	4.44E+05	2.87E+05	0.92	0.92
AHA-GB	0.99	1.53E+04	1.10E+04	1	0.1	0.97	2.80E+05	1.65E+05	0.97	0.97
HHO-GB	0.99	1.48E+05	1.09E+05	0.99	0.99	0.94	3.75E+05	2.56E+05	0.94	0.97
RF	0.99	1.44E+05	8.56E+04	0.99	0.98	0.94	3.82E+05	2.29E+05	0.94	0.94
AHA-RF	0.99	1.49E+05	8.67E+04	0.99	0.97	0.94	3.77E+05	2.24E+05	0.94	0.93
HHO-RF	0.99	1.46E+05	8.47E+04	0.99	0.97	0.94	3.79E+05	2.26E+05	0.94	0.94

B. Performance using Witczak 1-40D Input and Additional Parameters

For the model with the Witczak 1-40D model parameters and the additional six parameters, the GBR training results were outstanding, particularly for the AHA and HHO variants, with R^2 , NSE , and KGE values reaching perfection. However, a slight decline was observed in the test phase, although the metric values remain high (R^2 , NSE , KGE ranging from 0.92 to 0.97), indicating strong model generalization but a slight overfit in training. In contrast, the RF model exhibited exceptional consistency and slightly higher robustness, with almost perfect scores in training (R^2 , NSE , KGE all around 0.99) and very high scores in testing (R^2 , NSE , KGE all around 0.94), showcasing its remarkable ability to generalize across unseen data. These results indicate that while GBR provides excellent fit and predictive accuracy as shown in Figure 2(b), especially when its parameters were optimized with AHA and HHO, RF stands out for its consistency and robustness, making it exceptionally reliable for practical applications where

generalization and stability across diverse datasets are crucial. Furthermore, the addition of the six parameters shows that the performance of the E^* prediction can be improved when compared to the performance reported in [4, 17].

C. Parameter Interpretability and Optimize Input Features for the Model

For the model interpretability, the E^* measurement of asphalt mixture depends on grouped factors: binder parameters, mixture test parameters, mixture volumetric parameters, and the aggregate distribution properties.

1) Binder Parameters

Figure 3 shows that the binder's complex modulus (G^*) significantly influences the prediction of the E^* of an asphalt mixture. This impact is more pronounced than the effects of the binder's phase angle, which although important, is less influential than the phase angle of the mixture and the test temperature. Additionally, in regions where VTS and A values

for the binder are close to zero, there is an increased importance of these properties in predicting E^* .

2) Mixture Test Parameters

The analysis of a group of mixture test parameters, including mixture, phase angle, test temperature T , and frequency F_s , shows that they are significantly influential in the prediction of the E^* of an asphalt mixture, following the binder parameter group in importance. Specifically, the mixture phase

angle within this group is highlighted for its high permutation importance, as shown in Figure 3(b). According to the SHAP explainer plot shown in Figure 3(a), these parameters generally exhibit a negative distribution, implying a particular kind of influence on E^* predictions. Additionally, it was observed that the test frequency, although a part of this group, has a relatively minor effect on the prediction as indicated by its low permutation importance score.

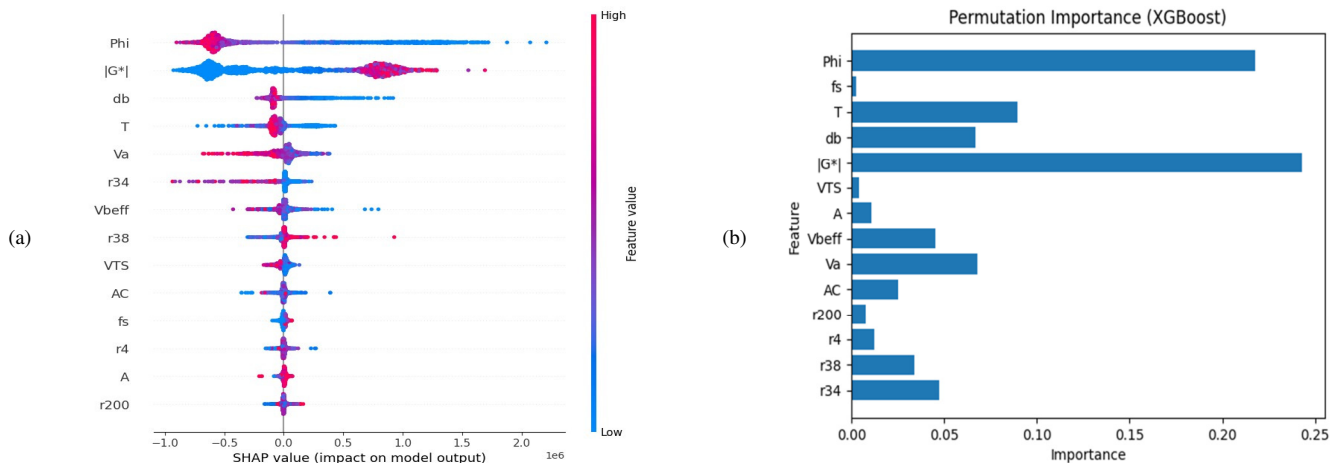


Fig. 3. (a) Shap value, (b) permutation importance.

3) Aggregate Distribution Properties

In evaluating the E^* of asphalt mixtures, the significance of the aggregate gradation parameters progressively decreases. Among these, the percentage of aggregates retained on a 19.05 mm sieve stands out (r_{34}), ranking as the second most influential factor in this category. On the contrary, the percentage of aggregates passing through a 0.13 mm sieve r_{200} is observed to have the least influence on E^* predictions within the same group. This trend underscores hierarchical importance in aggregate sizes, where larger sizes, such as those retained on the 19.05 mm sieve, play a more substantial role, while finer sizes, like those passing through the 0.13 mm sieve, contribute less to the prediction of the mixture's dynamic modulus.

4) Mixture Volumetric Properties

In the analysis of the E^* of asphalt mixtures, the parameters AC , VA , and V_{beff} have a significant influence. These parameters substantially affect the E^* predictions, a fact underscored by their notable permutation importance scores. Despite this, their Shapley effect distributions tend to hover near zero. This implies that while each parameter individually contributes to E^* predictions, their collective impact within the predictive model is more balanced or neutral.

IV. CONCLUSION

This study investigated the novel approach of using the Artificial Hummingbird Algorithm (AHA) and Harris Hawks Optimization (HHO) for the automatic calibration of Gradient Boosting Regression (GBR) and Random Forest (RF) algorithm hyperparameters for the prediction of E^* . Additionally, the prediction of E^* using the Witczak model

parameters was improved by including six additional inputs for an accurate account of its variation. The presented results show that by adding asphalt content, temperature, frequency of loading, air voids, volumetric total solids, and mixture phase angle to the existing Witczak model parameters, the R^2 performance of the model prediction improved from 0.92 to 0.97 in a model trained with a hybrid AHA-optimized Gradient Boost model. The model interpretability presented showed that the binder properties and mixture phase angle had the highest effect on the variance of E^* . The size of the aggregate shows a decreasing effect with a decrease in the aggregate size, indicating the dependency of the mixture stiffness on the aggregate size.

REFERENCES

- [1] Y. R. Kim, Y. Seo, M. King, and M. Momen, "Dynamic Modulus Testing of Asphalt Concrete in Indirect Tension Mode," *Transportation Research Record*, vol. 1891, no. 1, pp. 163–173, Jan. 2004, <https://doi.org/10.3141/1891-19>.
- [2] M. W. Witczak and O. A. Fonseca, "Revised Predictive Model for Dynamic (Complex) Modulus of Asphalt Mixtures," *Transportation Research Record*, vol. 1540, no. 1, pp. 15–23, Jan. 1996, <https://doi.org/10.1177/0361198196154000103>.
- [3] J. F. Shook *et al.*, "Factors Influencing Dynamic Modulus of Asphalt Concrete," in *Association of Asphalt Paving Technologists Proceedings*, Feb. 1969, vol. 38, pp. 140–178.
- [4] I. D. Uwanuakwa, A. Busari, S. I. A. Ali, M. R. Mohd Hasan, A. Sani, and S. I. Abba, "Comparing Machine Learning Models with Witczak NCHRP 1-40D Model for Hot-Mix Asphalt Dynamic Modulus Prediction," *Arabian Journal for Science and Engineering*, vol. 47, no. 10, pp. 13579–13591, Oct. 2022, <https://doi.org/10.1007/s13369-022-06935-x>.

- [5] A. K. Dubey, A. K. Sinhal, and R. Sharma, "An Improved Auto Categorical PSO with ML for Heart Disease Prediction," *Engineering, Technology & Applied Science Research*, vol. 12, no. 3, pp. 8567–8573, Jun. 2022, <https://doi.org/10.48084/etasr.4854>.
- [6] P. Dhaka, R. Sehrawat, and P. Bhutani, "An Innovative Approach to Cardiovascular Disease Prediction: A Hybrid Deep Learning Model," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12396–12403, Dec. 2023, <https://doi.org/10.48084/etasr.6503>.
- [7] S. R. Gopi and M. Karthikeyan, "Effectiveness of Crop Recommendation and Yield Prediction using Hybrid Moth Flame Optimization with Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11360–11365, Aug. 2023, <https://doi.org/10.48084/etasr.6092>.
- [8] T. H. Le, H. L. Nguyen, and C. T. Pham, "Artificial intelligence approach to predict the dynamic modulus of asphalt concrete mixtures," *Journal of Science and Transport Technology*, pp. 22–31, Jun. 2022, <https://doi.org/10.58845/jstt.utt.2022.en.2.2.22-31>.
- [9] A. Behnood and E. Mohammadi Golafshani, "Predicting the dynamic modulus of asphalt mixture using machine learning techniques: An application of multi biogeography-based programming," *Construction and Building Materials*, vol. 266, Jan. 2021, Art. no. 120983, <https://doi.org/10.1016/j.conbuildmat.2020.120983>.
- [10] T. H. Le *et al.*, "Artificial Intelligence-Based Model for the Prediction of Dynamic Modulus of Stone Mastic Asphalt," *Applied Sciences*, vol. 10, no. 15, Jan. 2020, Art. no. 5242, <https://doi.org/10.3390/app10155242>.
- [11] J. Huang, G. Shiva Kumar, J. Ren, J. Zhang, and Y. Sun, "Accurately predicting dynamic modulus of asphalt mixtures in low-temperature regions using hybrid artificial intelligence model," *Construction and Building Materials*, vol. 297, Aug. 2021, Art. no. 123655, <https://doi.org/10.1016/j.conbuildmat.2021.123655>.
- [12] W. Xu, X. Huang, Z. Yang, M. Zhou, and J. Huang, "Developing Hybrid Machine Learning Models to Determine the Dynamic Modulus (E*) of Asphalt Mixtures Using Parameters in Witczak 1-40D Model: A Comparative Study," *Materials*, vol. 15, no. 5, Jan. 2022, Art. no. 1791, <https://doi.org/10.3390/ma15051791>.
- [13] I. Lahmar, A. Zaier, M. Yahia, and R. Boaullegue, "A Novel Improved Binary Harris Hawks Optimization For High dimensionality Feature Selection," *Pattern Recognition Letters*, vol. 171, pp. 170–176, Jul. 2023, <https://doi.org/10.1016/j.patrec.2023.05.007>.
- [14] W. Zhao, L. Wang, and S. Mirjalili, "Artificial hummingbird algorithm: A new bio-inspired optimizer with its engineering applications," *Computer Methods in Applied Mechanics and Engineering*, vol. 388, Jan. 2022, Art. no. 114194, <https://doi.org/10.1016/j.cma.2021.114194>.
- [15] B. Rozemberczki *et al.*, "The Shapley Value in Machine Learning," arXiv, May 26, 2022, <https://doi.org/10.48550/arXiv.2202.05594>.
- [16] A. Altmann, L. Tološi, O. Sander, and T. Lengauer, "Permutation importance: a corrected feature importance measure," *Bioinformatics*, vol. 26, no. 10, pp. 1340–1347, May 2010, <https://doi.org/10.1093/bioinformatics/btq134>.
- [17] S. El-Badawy, R. Abd El-Hakim, and A. Awed, "Comparing Artificial Neural Networks with Regression Models for Hot-Mix Asphalt Dynamic Modulus Prediction," *Journal of Materials in Civil Engineering*, vol. 30, no. 7, Jul. 2018, Art. no. 04018128, [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0002282](https://doi.org/10.1061/(ASCE)MT.1943-5533.0002282).

Extraction of Solar Module Parameters using a Novel Optimization Technique

Hossam E. Ahmed

Department of Electrical Engineering, College of Engineering, Northern Border University, Saudi Arabia
| Electrical Engineering Department, Benha Faculty of Engineering, Benha University, Egypt
hossam.ahmed@nbu.edu.sa

Yehya I. Mesalam

Department of Industrial Engineering, College of Engineering, Northern Border University, Saudi Arabia
| Industrial Engineering Department, Faculty of Engineering, Zagazig University, Egypt
yehya.mesalam@nbu.edu.sa (corresponding author)

Shaaban M. Shaaban

Department of Electrical Engineering, College of Engineering, Northern Border University, Saudi Arabia
| Department of Engineering Basic Science, Faculty of Engineering, Menoufia University, Egypt
shaban.awdallah@nbu.edu.sa

Received: 16 December 2023 | Revised: 1 January 2024 | Accepted: 9 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6760>

ABSTRACT

The parameters of a Photovoltaic (PV) model are pivotal in gauging its efficiency under varying sunlight irradiances, temperatures, and different load scenarios. Determining these PV model parameters poses a complex non-linear optimization challenge. This study is based on a new metaheuristic optimization algorithm called the Pelican Optimization Algorithm (POA) to discern the unknown parameters of the PV model. The suggested POA algorithm underwent testing using a monocrystalline panel, encompassing its single-diode configuration. The objective function is designed to minimize the root of the mean squared errors between the predicted and actual current values, adhering to specific parameter constraints. Various statistical error metrics were utilized to emphasize the performance of the proposed algorithm. A comparative analysis with other well-established algorithms was conducted, indicating that POA stands out as highly competitive since it showcases superior efficiency in parameter identification compared to its counterparts.

Keywords-PV cells; parameter extraction; modeling; POA; optimization

I. INTRODUCTION

Parameter estimation of solar PV (photovoltaic) models is crucial for accurately predicting the performance, efficiency, and behavior of solar panels under various environmental and operating conditions. By determining the precise values of these parameters, researchers and engineers can optimize the design, operation, and integration of solar panels into larger systems, ensuring they produce maximum power output and operate efficiently. Furthermore, accurate parameter estimation aids in system diagnostics and health monitoring, allowing for timely maintenance, longer lifespan, and consistent energy production, which is essential for the viability and cost-effectiveness of solar energy solutions.

PV datasheets typically highlight three primary points concerning the I/V relationship under standard test conditions (STC) – ambient temperature (T) of 25°C, solar irradiance (G) of 1 kW/m², and 1.5 air mass [1]. These are the no-load

terminal voltage (V), short-circuit current (I_{sc}), and both voltage (V_{mp}) and current at peak power (P_{max}). However, these specified data points alone are insufficient for extended PV system studies, as environmental conditions constantly fluctuate. A precise I/V representation across all operating conditions is crucial for an in-depth performance analysis of PV systems. The I/V characteristics of PV cells are typically represented using two prevalent models: the One-Diode Model (ODM) and the Two-Diode Model (TDM) [2-4]. Notably, while the TDM demands more computational resources than the ODM, the performance efficiency difference between them is marginal. The ODM requires the accurate determination of five parameters, whereas the TDM needs seven. There is also a triple diode model, which requests the identification of nine parameters. However, this research primarily focuses on ODM.

The characterization of solar cell behavior typically relies on the ODM usage. For effective enhancement of PV system efficiency through simulation studies, it is imperative to

accurately estimate and identify model parameters [5, 6]. Many researchers focus on refining PV model parameters using diverse design methodologies. These optimization techniques can be primarily categorized as deterministic or heuristic when examined from an algorithmic point of view. Both approaches address parameter extraction by transforming it into an optimization task, drawing on specific reference points from the given I-V characteristic curve. Deterministic techniques, such as the least squares (based on the Newton method [3] and Lambert W-functions [7, 8], place certain constraints on the objective functions, namely continuity, convexity, and differentiability. Moreover, they are prone to influences from initial conditions and gradient details. This makes them susceptible to becoming ensnared in local optima, especially when navigating multifaceted multimodal challenges. These constraints lead deterministic strategies to face obstacles when tasked with solving nonlinear, multimodal parameter extraction issues. On the other hand, heuristic strategies are more versatile, not being bound by strict conditions related to the optimization problem's structure. This allows them to bypass the pitfalls of sensitivity to initial conditions and gradient details. As a result, there has recently been a surge in interest regarding these methods. Some of the successfully employed heuristic techniques for PV model parameter extraction encompass Genetic Algorithms (GAs) [9], Particle Swarm Optimization (PSO) [10], Differential Evolution (DE) [11], Artificial Bee Colony (ABC) optimization [12], Harmony Search (HS) [13], teaching-learning-based optimization [14], Chaotic Whale Optimization Algorithm (CWOA) [15], Lévy flight trajectory-based Whale Optimization Algorithm (LWOA) [16], hybrid PSO-WOA [17], IJAYA [5], and Hybrid Firefly and Pattern Search Algorithm (HFAPS) [18].

In this study, we have integrated the Pelican Optimization Algorithm (POA) for parameter extraction of PV cells [19]. The POA is founded on simulating the natural hunting behaviors of pelicans. In this algorithm, the search agents are depicted as pelicans in search of food sources.

This article presents several notable contributions, which are:

- A novel approach to identifying parameters in solar PV models using the POA is proposed.
- The POA is applied to estimate unknown parameters for ODM.
- The experiments are conducted using monocrystalline PV panels.
- The proposed POA and five additional algorithms (CWOA, LWOA, PSO-WOA, IJAYA, and HFAPS) were simulated and their outcomes were compared. The comparison was primarily based on statistical error and cost function values.

II. MODELING OF A SOLAR PV SYSTEM

The estimation of PV module performance and the design of power systems hinge upon the current-voltage (I-V) electrical characteristics exhibited by the modules across varying solar radiation degrees and diverse temperatures [19]. For simulating PV cells and modules, equations mirroring the

intrinsic attributes of the cells have been suggested. The literature has put forth numerous electrical models to simulate PV cells under different conditions. These models vary in complexity based on the number of parameters to be identified, such as the R_s (series resistance) and R_{sh} (shunt resistance). At their core, these models enhance the fundamental model, which is composed of a diode symbolizing the PN junction and a current source indicative of the received solar energy [20]. To achieve a more accurate representation of PV cell behavior in specific operating conditions, several additional elements can be incorporated. The most utilized models are ODM and TDM.

A. Solar PV System Modeling. The One-Diode Model

The ODM, also known as the single diode model or simply the diode model, is a widely used mathematical representation of the electrical behavior of a solar cell or of a PV cell. This model is applicable to various types of solar cells, including monocrystalline and polycrystalline cells.

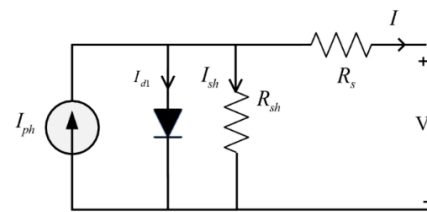


Fig. 1. The one-diode model.

Equation (1) models how the output current (I) of the PV cell varies with voltage (V), considering factors like photocurrent (I_{ph}), diode current (I_{d1}), reverse saturation current (I_{sd1}), diode ideality factor (α_1), temperature (T), the constant of Boltzmann (K), the electron charge (q), and series and shunt resistance (R_s, R_{sh}).

$$I = I_{ph} - I_{d1} - I_{sh} \quad (1)$$

with:

$$I_{d1} = I_{sd1} \left(e^{\left(\frac{q(V + R_s I)}{\alpha_1 K T} \right)} - 1 \right)$$

B. Objective Functions for Extracting PV Parameters

The objective functions in the parameter extraction process serve as mathematical criteria that guide the optimization process. The procedure of obtaining optimal parameter values involves comparing the estimated current values to the actual experimental ones. To achieve this, it is essential to utilize an objective function that aims to minimize the Root Mean Square Error (RMSE) across various measured data points [21, 22]:

$$RMSE = \sqrt{\frac{1}{N} \left(\sum_{i=1}^N (I_m - I_e) \right)^2} \quad (2)$$

where N is the number of samples, which must be big enough to attain the global optimum, while the measured and estimated current are denoted by I_m and I_e .

The objective function is employed with the primary goal of reducing the RMS) across a diverse range of measured data points. In the case of the ODM, the objective function is formulated as depicted in (3):

$$f_{ODM} = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(I_m - I_{ph} + I_{sd1} \left[e^{\left(\frac{q(V+R_s I)}{\alpha_i K T} \right)} - 1 \right] + \frac{V + R_s I}{R_{sh}} \right)^2} \quad (3)$$

III. PELICAN OPTIMIZATION ALGORITHM FOR PV PARAMETER ESTIMATION

In this section, the conceptual foundation and mathematical framework of the proposed swarm-inspired POA are introduced.

A. Hunting Strategies and Characteristics of Pelicans

Pelicans are social birds with elongated beaks and pouches for catching prey. They often live in sizable groups and predominantly feed on fish [23, 24]. These birds hunt collaboratively, diving from heights to corner fish in shallow waters. Upon capturing, they expel excess water from their pouch before consuming it [25]. Their strategic hunting techniques showcase their natural intelligence. This behavior serves as the foundation of the proposed algorithm.

B. Mathematical Modeling

The suggested POA operates as a population-centric algorithm, with pelicans representing its constituents. Within such algorithms, every constituent signifies a potential solution. Each constituent recommends variable values for the optimization challenge based on their location within the search domain. At the outset, these constituents are randomly set based on the problem's lower and upper constraints, as described by (4).

$$x_{i,j} = l_j + rand.(u_j - l_j) \quad (4)$$

where $i = 1, 2, \dots, N$, $j = 1, 2, \dots, m$.

Let $x_{i,j}$ denote the value of the j^{th} variable determined by the i^{th} potential solution. Here, N signifies the total number of population members, m represents the count of problem variables, and $rand$ is a random number between [0, 1], whereas l_j and u_j are the lower and upper boundaries of the j^{th} problem variable, respectively. In POA, the members of the pelican population are depicted using a matrix termed as the population matrix, as shown in (5). In this matrix, each row corresponds to a potential solution, and the columns highlight the suggested values for the variables in question.

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_N \end{bmatrix}_{N \times m} = \begin{bmatrix} x_{1,1} & \dots & x_{1,j} & \dots & x_{1,m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i,1} & \dots & x_{i,j} & \dots & x_{i,m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{N,1} & \dots & x_{N,j} & \dots & x_{N,m} \end{bmatrix}_{N \times m} \quad (5)$$

Let X represent the pelican population matrix, with X_i signifying the i^{th} pelican. In POA, every member of the population corresponds to a pelican, representing a potential solution for the given challenge. As a result, the problem's objective function can be gauged for each potential solution. The outcomes of this function are expressed using a vector, referred to as the objective function vector, depicted in (6):

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix}_{N \times 1} = \begin{bmatrix} F(X_1) \\ \vdots \\ F(X_i) \\ \vdots \\ F(X_N) \end{bmatrix}_{N \times 1} \quad (6)$$

where F represents the vector of the objective function and F_i signifies the computed value of the objective function corresponding to the i^{th} pelican.

The suggested POA emulates the tactics and actions of pelicans during their predatory pursuits to refine potential solutions. The pelicans' hunting approach is replicated through two phases.

- Approaching prey (exploratory phase).
- Gliding over the water's surface (refinement phase).

1) Phase 1: Approaching Prey

During the initial stage, pelicans determine the position of their prey and navigate towards this pinpointed zone. Emulating this behavior enhances the scanning and exploration capacity of the POA across various regions of the search domain. A key aspect in the POA is the random generation of the prey's location within the search area. This amplifies the exploratory efficiency of the POA in thoroughly probing the solution space. The pelican's method in moving to the prey's location is represented mathematically in (7):

$$x_{i,j}^{P_1} = \begin{cases} x_{i,j} + rand(p_j - I.x_{i,j}), & F_p < F_i; \\ x_{i,j} + rand(x_{i,j} - p_j), & \text{else} \end{cases} \quad (7)$$

where $x_{i,j}^{P_1}$ represents the updated state of the i^{th} pelican in the j^{th} dimension according to phase 1, I is a random variable taking values of either 1 or 2, p_j denotes the prey's location in the j^{th} dimension, and F_p represents its corresponding objective function value. The parameter I , is chosen anew for each iteration and member. When this parameter assumes a value of 2, it induces a greater shift for a member, potentially directing that member to uncharted regions of the search domain. Hence, the parameter I significantly influences the explorative capacity of the POA in thoroughly investigating the search domain.

In POA, a pelican's new location is adopted if there is an improvement in the objective function value at that spot. This kind of update, termed as "efficient updating," prevents the algorithm from venturing into sub-optimal regions. This methodology is captured in (8):

$$X_i = \begin{cases} X_i^{P_1}, & F_i^{P_1} < F_i \\ X_i, & \text{else} \end{cases} \quad (8)$$

where $X_i^{P_1}$ represents the newly computed position for the i^{th} pelican, derived during the first phase. Within this context, $x_{i,j}^{P_1}$ signifies the j^{th} dimension of this new position, while $F_i^{P_1}$ represents its corresponding objective function value.

2) Phase 2: Gliding over the Water's Surface

During this phase, once the pelicans arrive at the water surface, they spread their wings across it, pushing fish upwards before scooping them up with their throat pouch. This tactic ensures that more fish in the targeted region are captured. Replicating this activity enables the proposed POA to hone in on superior points within the hunt zone. This step enhances the localized search capability and the refinement efficiency of the POA. Mathematically speaking, the algorithm scrutinizes points around the pelican's position to converge to an optimized solution. This hunting behavior is represented in (9):

$$x_{i,j}^{P_2} = x_{i,j} + R \cdot \left(1 - \frac{t}{T}\right) \cdot (2 \cdot \text{rand} - 1) \cdot x_{i,j} \quad (9)$$

where $x_{i,j}^{P_2}$ denotes the updated position of the i^{th} pelican in the j^{th} dimension. The constant R is set at 0.2. The term $R(1-t/T)$ defines the neighborhood radius around $x_{i,j}$, t stands for the current iteration, and T represents the total number of iterations. This term acts as a measure for localized searching around each member, aiding convergence to a superior solution. In the initial stages, its value is high, implying broader search vicinity for each member. However, as iterations progress, this value shrinks, narrowing the search vicinity. Consequently, the algorithm refines its search, allowing to gravitate towards solutions that are closer, if not exactly, to the global optimum. During this phase, effective updating is utilized to determine whether to accept or decline the new position of the pelican, as represented in (10):

$$X_i = \begin{cases} X_i^{P_2}, & F_i^{P_2} < F_i \\ X_i, & \text{else} \end{cases} \quad (10)$$

where $X_i^{P_2}$ represents the newly computed position of the i^{th} pelican, derived during the second phase. Within this context, $x_{i,j}^{P_2}$ signifies the j^{th} dimension of this new position, while $F_i^{P_2}$ represents its corresponding objective function value.

Once all members of the population have undergone updates from the two phases, the top solution is refreshed considering the current population state and objective function values. The algorithm then proceeds to the next cycle, and the steps outlined based on (7)–(10), are reiterated until the process concludes. At the end, the most favorable solution identified across all iterations is offered as a near-optimal answer to the specified issue. Figure 2 depicts the steps of the POA in flowchart form.

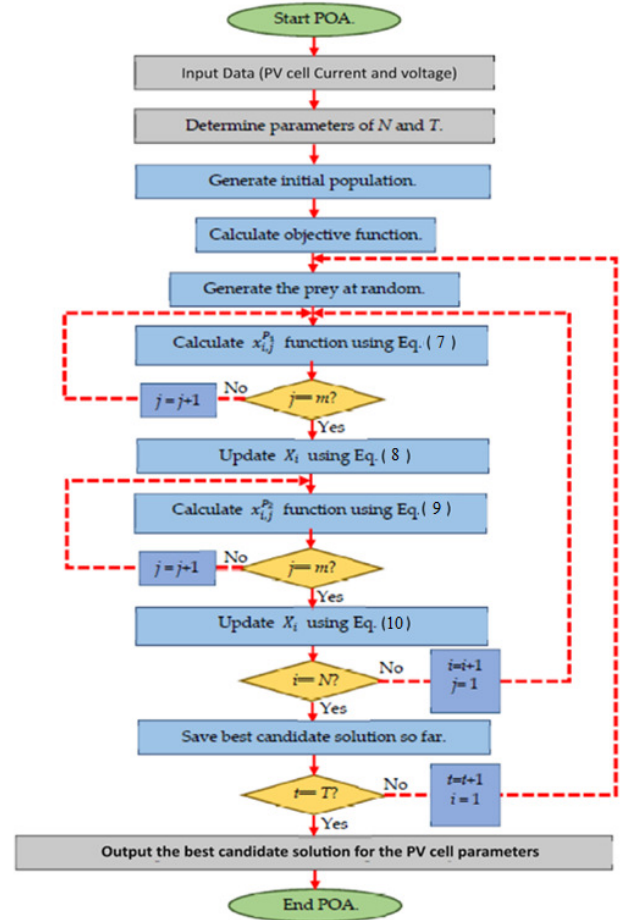


Fig. 2. POA flowchart

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

Utilizing MATLAB Simulink 2021, we conducted simulations with the RTC France Company mono-crystalline module under specific conditions of solar irradiance $G = 1000 \text{ W/m}^2$ and temperature $T = 33 \text{ }^\circ\text{C}$. We contrasted the results from our parameter identification approach with those from other optimization algorithms to assess the precision of POA.

A. Comparative Analysis of ODM Parameter Extraction

The ODM estimated parameters from all considered algorithms are shown in Table I. Figure 3 graphically represents the various outcomes. It is evident that the results achieved from POA, have the smallest error for the majority of the values.

TABLE I. EXTRACTED ODM PARAMETERS

Approach	I_{ph} (A)	I_0 (μA)	R_p (Ω)	R_s (Ω)	α
POA	0.7607	0.3107	52.89	0.036	1.477
CWOA	0.7600	0.2831	62.61	0.0371	0.0371
LWOA	0.7602	0.4607	0.035	75.46	1.5177
PSO-WOA	0.7597	0.314	0.0366	58.80	1.4783
IJAYA	0.7608	0.3228	0.0364	53.75	1.4811
HFAPS	0.7607	0.3226	0.0363	1.481	53.67

TABLE II. STATISTICAL RESULTS

	HFAPS	IJAYA	PSO-WOA	LWOA	CWOA	POA
IAT	0.0217	0.02151	0.02267	0.0222	0.0208	0.01845
RMSE	0.000992	0.00098	0.00101	0.0010	0.00096	0.0008441
SSE	2.56E-05	2.50E-05	2.69E-05	2.93E-05	2.41E-05	1.85E-05
MAE	0.0008346	0.00082	0.00087	0.00085	0.0008	0.0007096

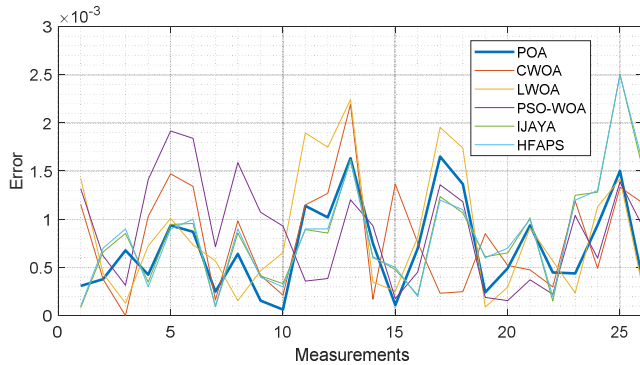


Fig. 3. Calculated errors obtained by POA, CWOA, LWOA, PSO-WOA, IJAYA, and HFAPS algorithms for the ODM case.

To assess the accuracy of the derived parameters, we compare the current-voltage and power-voltage characteristics obtained from the estimated parameters via the POA method with the experimental and estimated data. Figure 4 provides a visual representation of this comparison, focusing on the ODM scenario.

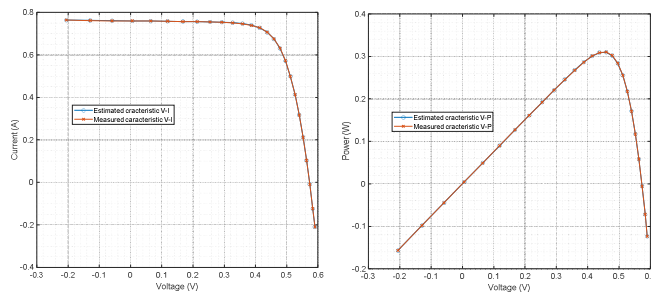


Fig. 4. Experimental and measured I-V characteristics achieved through the proposed POA for the ODM case.

The characteristics results presented in Figure 5 demonstrate a strong agreement between the reconstructed ODM and the measured data. Figure 5 depicts the convergent curves for various PV cell models. Overall, all the models display a satisfactory with POA results showing the most rapid convergence rate. In Figure 5, the illustrated average fitness functions are essentially a representation of the fitness associated with the extraction parameters for PV cells. The standout advantage of the results delivered by POA, especially when juxtaposed against other optimization algorithms, lies in its distinct capability to minimize error and ensure rapid convergence. This is evident particularly with ODM and TDM. The efficient and precise performance of POA makes it an invaluable tool for applications that require high accuracy and quick adaptability.

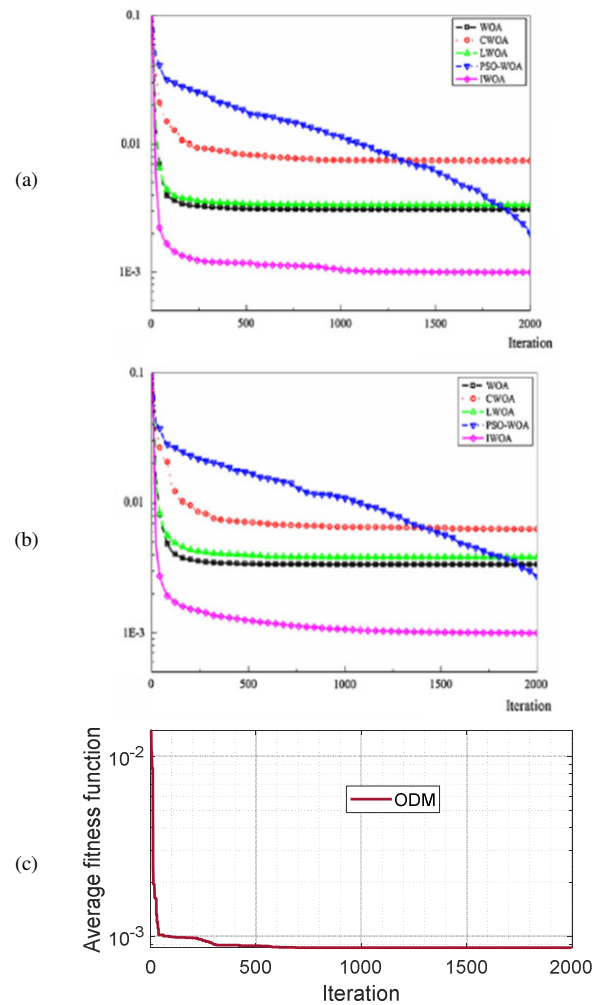


Fig. 5. Average fitness functions. Values obtained for (a) ODM, (b) TDB by various optimization algorithms [16]. (c) Fitness function for ODM using POA.

V. CONCLUSION

The Pelican Optimization Algorithm (POA) was designed to boost the accuracy of parameter extraction in solar cells. Its efficacy was assessed with regard to monocrystalline PV cell types within a single diode model. To ascertain its relative performance, the POA results were compared with the ones of other methods documented in prior studies. Both statistical evaluations and graphical representations indicated the notable precision and stability of POA, positioning it ahead of the other known methods.

ACKNOWLEDGMENT

The authors gratefully acknowledge the approval and support of this study by grant no. ENGA-2022-11-1671 from the Deanship of Scientific Research of the Northern Border University, Arar, Saudi Arabia.

REFERENCES

- [1] G. B. Gharehpetian and S. M. M. Agah, in *Distributed Generation Systems: Design, Operation and Grid Integration*, Oxford, UK: Butterworth-Heinemann, 2017, pp. 69–147.
- [2] H. Kraiem *et al.*, "Parameters Identification of Photovoltaic Cell and Module Models Using Modified Social Group Optimization Algorithm," *Sustainability*, vol. 15, no. 13, Jan. 2023, Art. no. 10510, <https://doi.org/10.3390/su151310510>.
- [3] H. M. Ridha, H. Hizam, S. Mirjalili, M. L. Othman, M. E. Ya'acub, and M. Ahmadipour, "Parameter extraction of single, double, and three diodes photovoltaic model based on guaranteed convergence arithmetic optimization algorithm and modified third order Newton Raphson methods," *Renewable and Sustainable Energy Reviews*, vol. 162, Jul. 2022, Art. no. 112436, <https://doi.org/10.1016/j.rser.2022.112436>.
- [4] I. Tebbal and A. F. Hamida, "Effects of Crossover Operators on Genetic Algorithms for the Extraction of Solar Cell Parameters from Noisy Data," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10630–10637, Jun. 2023, <https://doi.org/10.48084/etasr.5417>.
- [5] K. Yu, J. J. Liang, B. Y. Qu, X. Chen, and H. Wang, "Parameters identification of photovoltaic models using an improved JAYA optimization algorithm," *Energy Conversion and Management*, vol. 150, pp. 742–753, Oct. 2017, <https://doi.org/10.1016/j.enconman.2017.08.063>.
- [6] K. Njeh, M. A. Zdiri, M. B. Ammar, A. Rabhi, and F. B. Salem, "Energy Management of an Autonomous Photovoltaic System under Climatic Variations," *Engineering, Technology & Applied Science Research*, vol. 13, no. 1, pp. 9849–9854, Feb. 2023, <https://doi.org/10.48084/etasr.5375>.
- [7] X. Gao, Y. Cui, J. Hu, G. Xu, and Y. Yu, "Lambert W-function based exact representation for double diode model of solar cells: Comparison on fitness and parameter extraction," *Energy Conversion and Management*, vol. 127, pp. 443–460, Nov. 2016, <https://doi.org/10.1016/j.enconman.2016.09.005>.
- [8] R. M. A. Qasem and S. M. Massadeh, "Solving Cell Placement Problem Using Harmony Search Algorithms," *Engineering, Technology & Applied Science Research*, vol. 8, no. 4, pp. 3172–3176, Aug. 2018, <https://doi.org/10.48084/etasr.2113>.
- [9] J. A. Jervase, H. Bourdouce, and A. Al-Lawati, "Solar cell parameter extraction using genetic algorithms," *Measurement Science and Technology*, vol. 12, no. 11, Jul. 2001, Art. no. 1922, <https://doi.org/10.1088/0957-0233/12/11/322>.
- [10] A. Harrag and S. Messalti, "Three, Five and Seven PV Model Parameters Extraction using PSO," *Energy Procedia*, vol. 119, pp. 767–774, Jul. 2017, <https://doi.org/10.1016/j.egypro.2017.07.104>.
- [11] C. Chellaswamy and R. Ramesh, "Parameter extraction of solar cell models based on adaptive differential evolution algorithm," *Renewable Energy*, vol. 97, pp. 823–837, Nov. 2016, <https://doi.org/10.1016/j.renene.2016.06.024>.
- [12] X. Chen, B. Xu, C. Mei, Y. Ding, and K. Li, "Teaching–learning–based artificial bee colony for solar photovoltaic parameter estimation," *Applied Energy*, vol. 212, pp. 1578–1588, Feb. 2018, <https://doi.org/10.1016/j.apenergy.2017.12.115>.
- [13] W.-Y. Lee, S.-M. Park, and K.-B. Sim, "Optimal hyperparameter tuning of convolutional neural networks based on the parameter-setting-free harmony search algorithm," *Optik*, vol. 172, pp. 359–367, Nov. 2018, <https://doi.org/10.1016/j.ijleo.2018.07.044>.
- [14] Z. Liao, Z. Chen, and S. Li, "Parameters Extraction of Photovoltaic Models Using Triple-Phase Teaching-Learning-Based Optimization," *IEEE Access*, vol. 8, pp. 69937–69952, 2020, <https://doi.org/10.1109/ACCESS.2020.2984728>.
- [15] D. Prasad, A. Mukherjee, G. Shankar, and V. Mukherjee, "Application of chaotic whale optimisation algorithm for transient stability constrained optimal power flow," *IET Science, Measurement & Technology*, vol. 11, no. 8, pp. 1002–1013, 2017, <https://doi.org/10.1049/iet-smt.2017.0015>.
- [16] G. Xiong, J. Zhang, D. Shi, and Y. He, "Parameter extraction of solar photovoltaic models using an improved whale optimization algorithm," *Energy Conversion and Management*, vol. 174, pp. 388–405, Oct. 2018, <https://doi.org/10.1016/j.enconman.2018.08.053>.
- [17] H.-P. Hsu and C.-N. Wang, "Hybridizing Whale Optimization Algorithm With Particle Swarm Optimization for Scheduling a Dual-Command Storage/Retrieval Machine," *IEEE Access*, vol. 11, pp. 21264–21282, 2023, <https://doi.org/10.1109/ACCESS.2023.3246518>.
- [18] A. M. Beigi and A. Maroosi, "Parameter identification for solar cells and module using a Hybrid Firefly and Pattern Search Algorithms," *Solar Energy*, vol. 171, pp. 435–446, Sep. 2018, <https://doi.org/10.1016/j.solener.2018.06.092>.
- [19] P. Trojovský and M. Dehghani, "Pelican Optimization Algorithm: A Novel Nature-Inspired Algorithm for Engineering Applications," *Sensors*, vol. 22, no. 3, Jan. 2022, Art. no. 855, <https://doi.org/10.3390/s22030855>.
- [20] H. Kraiem, F. Aymen, L. Yahya, A. Trivino, M. Alharthi, and S. S. M. Ghoneim, "A Comparison between Particle Swarm and Grey Wolf Optimization Algorithms for Improving the Battery Autonomy in a Photovoltaic System," *Applied Sciences*, vol. 11, no. 16, Jan. 2021, Art. no. 7732, <https://doi.org/10.3390/app11167732>.
- [21] A. A. Z. Diab, A. Al Sumaiti, A. A. Ezzat, A. E. Rafaat, K. A. Denis, and A. G. A. El-Magd, "New objective function of parameters extraction of photovoltaic modules for plummeting execution time complexity," *IET Renewable Power Generation*, pp. 1–13, 2022, <https://doi.org/10.1049/rpg2.12491>.
- [22] D. S. K. Karunasingha, "Root mean square error or mean absolute error? Use their ratio as well," *Information Sciences*, vol. 585, pp. 609–629, Mar. 2022, <https://doi.org/10.1016/j.ins.2021.11.036>.
- [23] A. Louchart, N. Tourment, and J. Carrier, "The earliest known pelican reveals 30 million years of evolutionary stasis in beak morphology," *Journal of Ornithology*, vol. 152, no. 1, pp. 15–20, Jan. 2011, <https://doi.org/10.1007/s10336-010-0537-5>.
- [24] P. J. Higgins, J. M. Peter, and S. J. Cowling, *Handbook of Australian, New Zealand, and Antarctic Birds: Volume 7: Boatbill to Starlings*. Oxford, UK: Oxford University Press, 2006.
- [25] C. M. Perrins, *New Encyclopedia of Birds*. Oxford, UK: Oxford University Press, 2003.

An FPGA Accelerator for Real Time Hyperspectral Image Compression based on JPEG2000 Standard

Refka Ghodhbani

Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia | Laboratory of Electronic and Microelectronics, Faculty of Sciences, University of Monastir, Tunisia
refka.ghodhbani@nbu.edu.sa (corresponding author)

Taoufik Saidani

Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
taoufik.saidan@nbu.edu.sa

Layla Horigue

Laboratory of Electronic and Microelectronics, Faculty of Sciences, University of Monastir, Tunisia
layla.k-12@hotmail.com

Asaad M. Algarni

Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
asaad.algarni@nbu.edu.sa

Muteb Alshammari

Department of Information Technology, Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
muteb.alshammari@nbu.edu.sa

Received: 1 January 2024 | Revised: 14 January 2024 | Accepted: 15 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6853>

ABSTRACT

Lossless hyperspectral images have the advantage of reducing the data size, hence saving on storage and transmission costs. This study presents a dynamic pipeline hardware design for compressing and decompressing images using the Joint Photographic Experts Group-Lossless (JPEG2000) algorithm. The proposed architecture was specifically tailored for implementation on a Field Programmable Gate Array (FPGA) to accomplish efficient image processing. The introduction of a pipeline pause mechanism effectively resolves the issue of coding errors deriving from parameter modifications. Bit-plane coding was employed to enhance the efficacy of image coding calculations, leading to a reduction of parameter update delays. However, the context and decision creation procedure were streamlined, resulting in a significant enhancement in throughput. A hardware module utilizing the parallel block compression architecture was developed for JPEG2000 compression/decompression, allowing for configurable block size and bringing about enhanced image, compression/decompression, throughput, and reduced times. Verification results were obtained by implementing the proposed JPEG 2000 compression on a Zynq-7000 system-on-chip. The purpose of this system was to enable on-board satellite processing of hyperspectral image cubes with a specific focus on achieving lossless compression. The proposed architecture outperformed previous approaches by using fewer resources and achieving a higher compression ratio and clock frequency.

Keywords-hyperspectral image compression; JPEG2000; EBCOT; MQ coder; FPGA; `

I. INTRODUCTION

Hyperspectral imaging (HSI) is used in a variety of remote sensing applications, including intelligence gathering, commerce, agriculture, the military, and humanitarian efforts. Among these uses are environmental monitoring [1], field observation in agriculture [2], geological mapping [3], and mineral exploration [4], while its growth in recent years has been steady. In 2018-2023, the global market for HSI was expected to expand at a Compound Annual Growth Rate (CAGR) of 14.7%, according to the Business Communications Company, USA. The forecasts indicated that by 2023, the market would have grown from \$104.0 million in 2018 to \$206.2 million [4].

Hyperspectral image compression is the creation of a smaller version of data from a larger one, usually to make storage or transmission easier. Lossless and lossy compression are the two main types of compression. Whether or not the compressed version can be used to reconstruct the original datastream in its entirety determines this classification [5]. The DC program employs two different image resolutions: 4096x2160 (4K) and 2048x1080 (2K). On top of that, the 2K image can support 24 or 48 fps. However, the 4K image can only support 24 fps. So, for a three-hour film, the amount of raw data is nearly 9 GB. For that reason, a fast JPEG 2000 encoder is required for the real-time compression of these data. This study goes deep into image compression, focusing on the JPEG 2000 standard in particular.

The specific standard was produced by ISO/IEC and introduced in 2000 by the Joint Photographic Experts Group (JPEG) committee. The former uses the Discrete Wavelet Transform (DWT) instead of the Discrete Cosine Transform (DCT) [6]. The incorporation of many features in the JPEG 2000 standard, in addition to its improved compression ratios [7], is what differentiates it from earlier generations of compression methods. Its scalability stands out among the rest. Parts of the compressed JPEG 2000 bitstream can be extracted, together with a configurable spatial locality value between 1 and 5 [6, 8]. Due to this, the possible uses of JPEG 2000 are considerably broadened. Figure 1 shows a schematic of the JPEG 2000 encoding algorithm. Separating the original picture into smaller, more manageable pieces, called tiles and code blocks, allows for more efficient processing and reduces memory utilization. In [8], the techniques implemented in JPEG 2000, which include Embedded Block Coding with Optimum Truncation (EBCOT) and DWT [9], were described. Two processing stages make up the EBCOT Tier-1 encoding scheme: Matrix Quantizer (MQ) and Context Modeling (CM). To generate Context Decision (CX/D) pairings, CM processes the data contained in a CB based on algorithm evaluation. The MQ coder utilizes these CX/Ds pairings to create an embedded bitstream.

II. JPEG2000 ALGORITHM

The JPEG2000 standard employs some picture properties to perform compression sequentially. Its principal use is photo compression, but it has found other uses, such as in electroencephalography, video, and hyperspectral pictures.

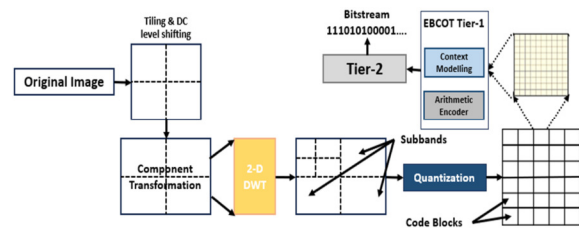


Fig. 1. JPEG 2000 encoder architecture.

An RGB input color space is transformed into a luminance and chrominance model via a pixel-wise color transformation because changes in brightness are more perceptible than color changes. Downsampling the color channels significantly reduces the number of data bits, but there is no noticeable loss in quality [8, 10]. After that, each channel should undergo a wavelet transform [10]. A high-pass filter is applied horizontally to all rows and columns, and a low-pass filter is applied to all columns, to implement a wavelet transform. This can be accomplished adopting either irreversible (lossy) or reversible (lossless) methods. In the end, there is a partitioned channel with zones that compress the data better than the originals due to their distinct patterns. The values that emerge from the wavelet transform are then quantized to produce integers. The lossy wavelet transform is useful but results in some data being lost [11]. The values are encoded at the end. Each block in the image can hold up to 4096 samples. The wavelet transform's patterns and local redundancies are utilized to encode each block independently.

Since hyperspectral compression decorrelates the spectral dimension during dimensionality reduction, the color transform is superfluous. Encoding and wavelet transform operations are the only ones carried out. In JPEG2000, encoding takes up around 70% of the total execution time, while the wavelet transform is much faster [12-14]. Furthermore, it has a major impact on the execution time within JYPEC [15]. This study aims to reduce the method's total execution times by focusing on an FPGA implementation that speeds up the bottleneck and greatly improves encoding.

III. JPEG2000 PROPOSED ARCHITECTURE

The BPC and MQ coders, when combined, provide the complete tier-1 coder. Figure 2 shows the basic organizational structure of the tier-1 coder. The bit plane coder creates CX/D pairs after retrieving data from memory. The MQ coder codes them, and an output stream is produced.

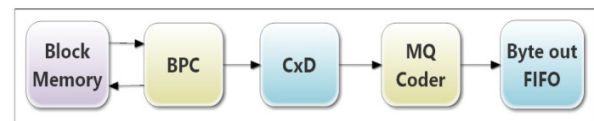


Fig. 2. Tier-1 coder architecture.

A. Bit-Plane Coding

Figure 3 displays the proposed EBCOT architecture. The DWT coefficients are kept in a codebook (CB) with a dimension of 32x9x32, including ten bits. Following the

initialization of state variables (r , $r0$, and g), the memory controller retrieves bit plane data and sign matrix data from the code block memory, as well as state variable data. These are then written into the stripe generator, which uses the column generator module to construct the columns. The stripe controller carefully picks the stripe to be processed, while considering the boundary criteria. Every column is sequentially encoded using the basic coding in each coding step. The pass detection logic identifies the pass to be executed and activates the required primitive modules. To expedite the intense block EBCOT process and reduce its complexity, all samples in a column are coded simultaneously inside a single clock cycle. In addition, the column of state variables is updated and both of its neighboring variables are made accessible simultaneously

[8]. This has the potential to produce anywhere from one to eight CX/D pairs concurrently. Therefore, to enhance the speed of design and minimize the memory needed for BPC, a novel architecture for bit-plane coding is proposed that incorporates a more efficient approach to data organization and memory arrangement. In this proposed design, MRP and CUP are processed concurrently. The BPC controller determines the appropriate pass to execute, chooses a specific coding primitive, and saves context data in the context sequencer based on the conditions generated during runtime. The purpose of the last step is to segregate and store these pairs in the appropriate sequence before they are transmitted to the MQ coder module.

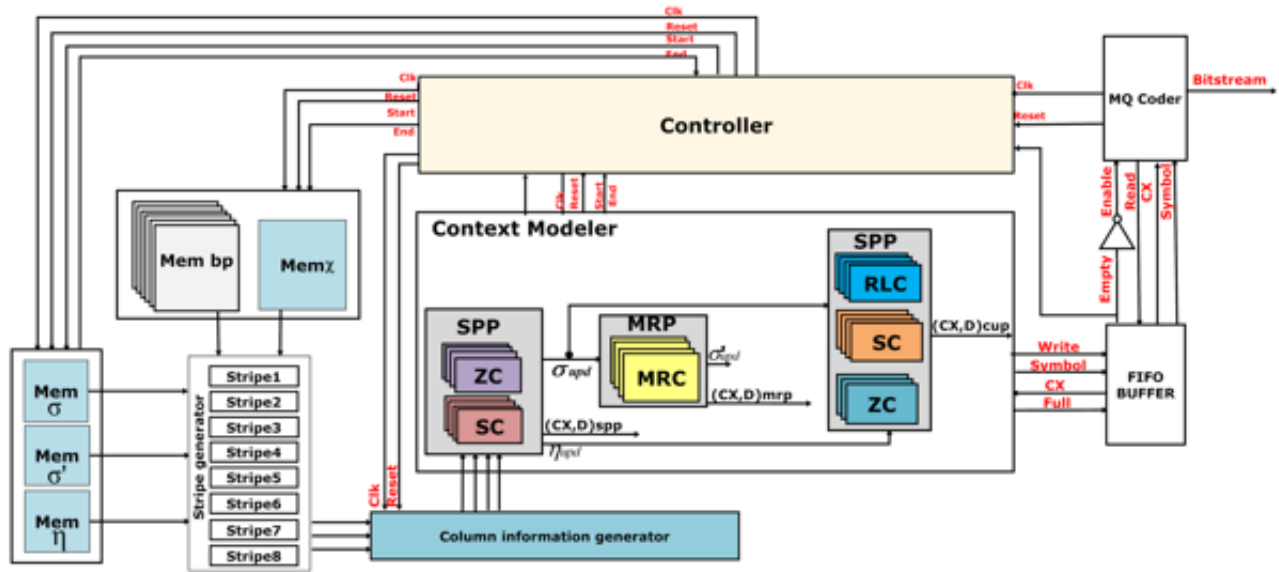


Fig. 3. Top module of the proposed column-parallel EBCOT context modeling.

B. MQ-Coder

The block diagram portrayed in Figure 4 illustrates the proposed design of the encoder architecture. The MQ coder receives the pairs (C, D) as input and produces a sequence of bytes named *ByteOutReg* as output. This design comprises two components: the prediction component, which uses two RAMs (ICX, MPS) and four ROMs (NMPS, NLPS, Switch, Qe) to calculate the likelihood of the symbol to be coded, and the coding component which consists of a state machine [12]. The four ROMs remain unaltered during the coding process. The pairs (CX, D) are initially read sequentially. Afterward, the CX context will be communicated via the bus address of both the ICX RAM and the MPS RAM. Subsequently, the values of $I(CX)$ and $MPS(CX)$ will be retrieved. Subsequently, the $I(CX)$ index will be transmitted to the four ROMs. The m_{ps_D} will be executed with signal D , generating the LPS_en signal. If the signal is equal to one, the CODELPS state will be executed. Otherwise, the CODEMPS state will be executed. The ICX RAM is primarily dependent on the Ren_out signal for updating. This signal will be set to one when renormalization is performed. The MPS RAM will only update if the LPS_SW signal is set to one. Figure 4 depicts the architecture for

estimating probabilities. The block diagram in Figure 5 illustrates the proposed encoder design. The MQ coder receives the pairs (C, D) as input and generates a sequence of b . This study focuses on the coding aspect of controlling the process of coding finite states. This involves replacing different subalgorithms with states. The outputs depend on the present state and the inputs, and they immediately respond to any alterations in the inputs. The MQ encoder process is described by a set of fourteen states.

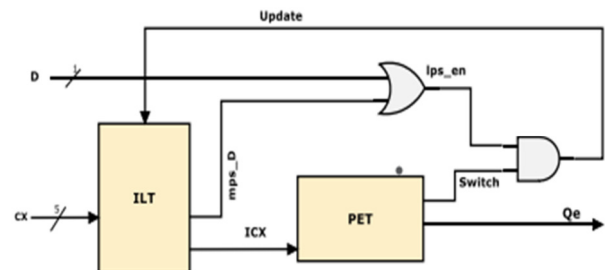


Fig. 4. Architecture for estimating probabilities.

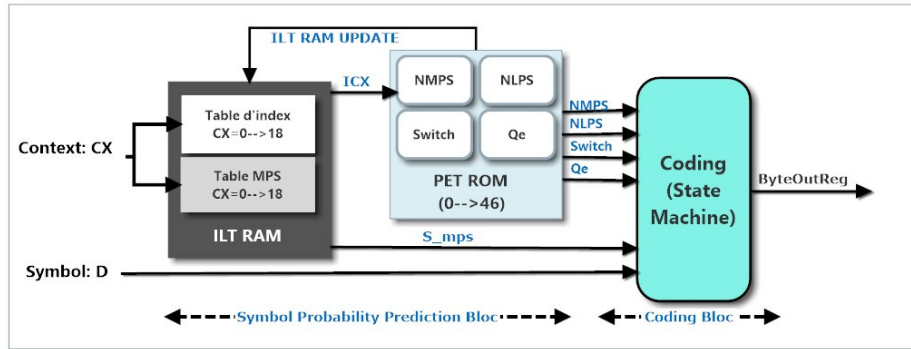


Fig. 5. MQ-coder architecture. The interval updates are fused when possible, having fewer bound updates which could stall the pipeline.

IV. FPGA IMPLEMENTATION

The compression/decompression system was developed on the Zynq7020 platform. The compression and decompression modules were packaged separately as tailored AXI IP cores. The top level of the IP core consists of the instantiation of the compression/decompression module and two BRAMs for storing the original/compressed and compressed/decompressed data. If the size of the compressed block is larger than the size of the original block, the original block is used as compressed data, and the block size is stored in the block header information. If the size of a block exceeds the original block size, it will not undergo decompression in the decompression process. The IP core should incorporate the AXI logic to perform read and write operations in DDR memory, using a burst length of 16 and a size of 4. Furthermore, it should adaptively adjust the number of burst transfers according to the size of each block. The header file of the compression/decompression module allows modification of the block size options. The CPU is tasked with getting the original image data from the SD card and executing the block function, before writing it into DDR. In addition, the CPU needs to access and configure registers to control the compression/decompression module.

The proposed hardware accelerator was developed utilizing the Vivado Design suite to evaluate its performance in compressing hyperspectral photos, based on an optimized version of JPEG2000. Xilinx offers a toolchain that encompasses Vivado HLS, a specialized tool aimed at optimizing the development of Intellectual Property (IP) components for FPGA-based solutions on Xilinx devices. The initial prototype employed the EBCOT algorithm and the MQ coder, which were specifically tailored for the variant XC7Z020-CLG484 of the Xilinx Zynq-7000 SoC. The FPGA was chosen for its cost-effectiveness, lightweight design, and great adaptability, making it an attractive option for incorporation into aerial vehicles such as drones. The purpose of this early prototype was to evaluate the efficiency of a mid-range reconfigurable FPGA for JPEG2000 compression. FPGA was used to implement the JPEG preprocessing accelerator specifically developed for picture categorization. This study engaged Verilog for the implementation of the Register-Transfer Level (RTL) and used Xilinx Vivado v2017.4 to compile the source code into the bitstream. The ImageNet Large Scale Visual Recognition Challenge 2012

(ILSVRC2012) was utilized as an input picture dataset. This dataset consists of 50,000 JPEG photos with various dimensions, ranging from 80×60 to 3657×2357, accompanied by multiple sampling compression algorithms. The efficiency of the proposed preprocessing accelerator was evaluated by employing a cost-effective 16 nm XCZU7EV chip. The preprocessing accelerator was synthesized with a clock frequency of 250 MHz.

A. Performance Analysis

Table I presents the performance comparison results of the JPEG2000 preprocessing techniques between the proposed hardware design and the 12 nm Nvidia GeForce RTX 2080Ti GPU with the Nvidia DALI library. This study evaluated and analyzed the power consumption, energy efficiency, and throughput of each of the three individual implementations. Thermal Design Power (TDP) and utilization data were used to evaluate the computational capability of a system based on a CPU. The Vivado Toolset and NVIDIA System Management Interface offer insight into the advantages of using a hardware accelerator that relies on FPGAs and GPU for implementation.

TABLE I. PERFORMANCE COMPARISON OF JPEG2000 PREPROCESSING

Metrics	Nvidia GeForce RTX 2080Ti 12 nm	Xilinx XCZU7EV 16 nm (proposed)
FPS	984.45	875.67
Speedup	2.84×	2.52×
Power (W)	58	12.35
Joule per frame (J/F)	0.059	0.014
Energy efficiency	4.21×	1×

The results indicated that both the proposed FPGA hardware accelerator and GPU-based implementation surpassed the optimized CPU-based architecture by a factor of 2.52 and 2.84, respectively. The energy efficiency of the two options was evaluated and compared. The proposed FPGA-based hardware accelerator reduced the energy utilization for each photo preprocessing frame by a factor of 23.07 and 4.21, respectively, compared to the CPU and GPU. The performance findings illustrate that the proposed FPGS hardware accelerator effectively performs photo preprocessing.

B. Resource Utilization

Look-Up Table (LUT) resources and slice resources are distinct categories of logic resources found in FPGA devices,

each serving a specific purpose in the implementation of various functionalities. A LUT is a data structure that may contain truth tables, allowing for the implementation of combinational logic or distributed memory. A slice is a logical unit consisting of several LUTs and flip-flops and is capable of implementing sequential logic, arithmetic operations, data selectors, and other functions. In general, a lower utilization of resources indicates that the system has a greater amount of idle resources available to fulfill more requests or loads [16-17]. The XC7Z020-CLG484-1 Zynq7000 development board used in this study features 53200 LUTs, 106400 Flip-Flops (FFs), 4.9 Mb of Block RAM (BRAM), 512 MB of DDR memory integrated on the board, and an SD card port.

The cores of the processing system were completely integrated into the Zynq 7000 SoC XC7Z020-CLG-484-1 [18]. The entire system was synthesized and implemented at an ideal frequency of approximately 100 MHz. When examining performance and flexibility, it is possible to identify some physical constraints, such as the exhaustion of DSP blocks. The distribution of resources within the SoC depends on the complexity of the architecture being implemented. Three separate filter designs are used for the input resolution of

1920×1080 photos. Table II presents a concise overview of the system's resource utilization. The results highlight the advantages of using the Vivado 2017.4 HLS tool on the versatile SoC platform [18]. Approximately 6% of all logic cells were currently in use. The application of logic cells varies depending on the specific location and routing techniques. The frequency of the proposed architecture was 320 MHz.

TABLE II. ZYNQ DEVICE RESOURCE UTILIZATION REPORT OF THE PROPOSED IMPLEMENTATION

Resource	Utilized	Total(%)
No. of programmable logic cells	5217	6.05
Look Up Tables (LUTs)	3,692	4.28
Flip Flops (FFs)	4,218	4.89
Block RAMs (BRAMs)	21	15
Frequency (MHz)	320	

The IP created by the HLS must be integrated into an HW/SW architecture for the FPGA SoC, which was realized by the Xilinx Vivado Design Suite. After importing the exported "IP Catalog" as a new repository file into the Vivado catalog, the block diagram of the design was created and the accelerator design was added as a new IP element, as shown in Figure 6.

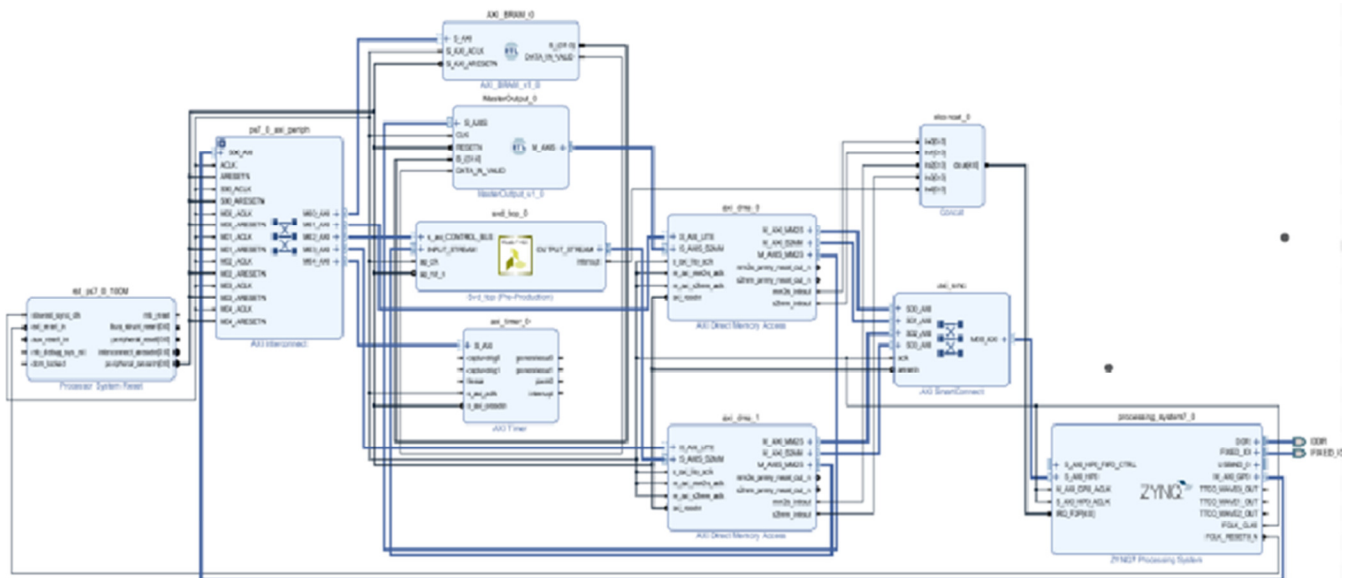


Fig. 6. RTL design for JPEG2000 architecture.

V. CONCLUSION

JPEG2000 is an advanced method that requires robust technology to achieve rapid real-time execution. The tier 1 coder in JPEG2000 is the most costly component because of the difficulty in optimizing code with unpredictable branching for ordinary processors. The algorithm's arithmetic and logic operations are very compatible with FPGA implementation because of their inherent simplicity. A swift and efficient design was created specifically for the highly skilled tier-1 coder in JPEG 2000, based on two fundamental concepts. The bit plane coder operates by simultaneously processing bits in batches of four, resulting in substantial performance

improvement. A system comprising First-In-First-Out (FIFO) queues and buffers ensures a seamless stream of CxD pairs to the MQ-coder. Furthermore, the coder itself was painstakingly optimized using a pipelined technique. To avoid pipeline stalling, a strategy of consolidating multiple updates was utilized whenever possible, addressing a previous concern in design. This study presented an implementation of the proposed JPEG 2000 architecture on a Zynq7000 board and compared it with an NVIDIA GeForce RTX 2080TI demonstrating approximately 4 times better energy efficiency. Within the field of hyperspectral imaging, the use of advanced lossy compression techniques allows real-time performance within the specified limits of the AVIRIS-ng sensor (30-72

MS/s, corresponding to a total of 491.52 Mb/s). This allows rapid compression of data at high speeds for immediate and long-lasting storage while preserving exceptional quality for future analysis.

ACKNOWLEDGMENT

The authors gratefully acknowledge the approval and support of this study by grant no. CSCR-2023-12-2062 from the Deanship of Scientific Research at Northern Border University, Arar, Saudi Arabia.

REFERENCES

- [1] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, "Modern Trends in Hyperspectral Image Analysis: A Review," *IEEE Access*, vol. 6, pp. 14118–14129, 2018, <https://doi.org/10.1109/ACCESS.2018.2812999>.
- [2] P. Ghamisi *et al.*, "Advances in Hyperspectral Image and Signal Processing: A Comprehensive Overview of the State of the Art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 37–78, Sep. 2017, <https://doi.org/10.1109/MGRS.2017.2762087>.
- [3] M. Hernández-Cabronero *et al.*, "The CCSDS 123.0-B-2 'Low-Complexity Lossless and Near-Lossless Multispectral and Hyperspectral Image Compression' Standard: A comprehensive review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 9, no. 4, pp. 102–119, Sep. 2021, <https://doi.org/10.1109/MGRS.2020.3048443>.
- [4] O. Barkovska, I. Filippenko, I. Semenenko, V. Korniienko, and P. Sedlaček, "Adaptation of FPGA architecture for accelerated image preprocessing," *Radioelectronic and Computer Systems*, no. 2, pp. 94–106, May 2023, <https://doi.org/10.32620/reks.2023.2.08>.
- [5] D. Báscones, C. González, and D. Mozos, "An FPGA Accelerator for Real-Time Lossy Compression of Hyperspectral Images," *Remote Sensing*, vol. 12, no. 16, Jan. 2020, Art. no. 2563, <https://doi.org/10.3390/rs12162563>.
- [6] C. Fu, Y. Yi, and F. Luo, "Hyperspectral image compression based on simultaneous sparse representation and general-pixels," *Pattern Recognition Letters*, vol. 116, pp. 65–71, Dec. 2018, <https://doi.org/10.1016/j.patrec.2018.09.013>.
- [7] P. K. Nath and S. Banerjee, "A high throughput pass parallel block decoder architecture for JPEG 2000 that prevents stalling in the decoding process," *Integration*, vol. 71, pp. 170–182, Mar. 2020, <https://doi.org/10.1016/j.vlsi.2019.11.013>.
- [8] D. Taubman and M. Marcellin, *JPEG2000 Image Compression Fundamentals, Standards and Practice: Image Compression Fundamentals, Standards and Practice*. New York, NY, USA: Springer Science & Business Media, 2012.
- [9] R. Ghodhban, T. Saidani, L. Horrigue, and M. Atri, "An efficient pass-parallel architecture for embedded block coder in JPEG 2000," *Journal of Real-Time Image Processing*, vol. 16, no. 5, pp. 1595–1606, Oct. 2019, <https://doi.org/10.1007/s11554-017-0666-7>.
- [10] B. Penna, T. Tillo, E. Magli, and G. Olmo, "Progressive 3-D coding of hyperspectral images based on JPEG 2000," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 125–129, Jan. 2006, <https://doi.org/10.1109/LGRS.2005.859942>.
- [11] Y. Kang and X. Xu, "A System and Its Implementation Based on FPGA for Video JPEG2000 Codec and Network Transmission," in *Proceedings of the 2021 5th International Conference on Digital Signal Processing*, Chengdu, China, Jun. 2021, pp. 260–265, <https://doi.org/10.1145/3458380.3458425>.
- [12] L. Horrigue, T. Saidani, R. Ghodhban, J. Dubois, J. Miteran, and M. Atri, "An efficient hardware implementation of MQ decoder of the JPEG2000," *Microprocessors and Microsystems*, vol. 38, no. 7, pp. 659–668, Oct. 2014, <https://doi.org/10.1016/j.micpro.2014.06.005>.
- [13] T. Saidani and R. Ghodhban, "Hardware Acceleration of Video Edge Detection with High Level Synthesis on the Xilinx Zynq Platform," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 8007–8012, Feb. 2022, <https://doi.org/10.48084/etasr.4615>.
- [14] L. Kechiche, L. Touil, M. Jemai, and B. Ouni, "A Power-Aware Real-Time System for Multi-Video Treatment on FPGA with Dynamic Partial Reconfiguration and Voltage Scaling," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 8996–9004, Aug. 2022, <https://doi.org/10.48084/etasr.5099>.
- [15] "ISO/IEC 15444-1:2000 - JPEG 2000 image coding system." International Standards Organization, Geneva, Switzerland, 2000.
- [16] R. Ghodhban, T. Saidani, A. Alhomoud, A. Alshammari, and R. Ahmed, "Real Time FPGA Implementation of an Efficient High Speed Harris Corner Detection Algorithm Based on High-Level Synthesis," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12169–12174, Dec. 2023, <https://doi.org/10.48084/etasr.6406>.
- [17] L. A. Aranda, A. Sánchez, F. Garcia-Herrero, Y. Barrios, R. Sarmiento, and J. A. Maestro, "Reliability Analysis of the SHyLoC CCSDS123 IP Core for Lossless Hyperspectral Image Compression Using COTS FPGAs," *Electronics*, vol. 9, no. 10, Oct. 2020, Art. no. 1681, <https://doi.org/10.3390/electronics9101681>.
- [18] J. Caba, M. Díaz, J. Barba, R. Guerra, J. A. de la Torre, and S. López, "FPGA-Based On-Board Hyperspectral Imaging Compression: Benchmarking Performance and Energy Efficiency against GPU Implementations," *Remote Sensing*, vol. 12, no. 22, Jan. 2020, Art. no. 3741, <https://doi.org/10.3390/rs12223741>.

Transformer Encoder with Protein Language Model for Protein Secondary Structure Prediction

Ammar Kazm

School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Malaysia | College of Education for Pure Sciences, Wasit University, Iraq
awadkazm@graduate.utm.my (corresponding author)

Aida Ali

School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Malaysia
aida@utm.my

Haslina Hashim

School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Malaysia
haslinah@utm.my

Received: 2 January 2024 | Revised: 15 January 2024 | Accepted: 17 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6855>

ABSTRACT

In bioinformatics, protein secondary structure prediction plays a significant role in understanding protein function and interactions. This study presents the TE_SS approach, which uses a transformer encoder-based model and the Ankh protein language model to predict protein secondary structures. The research focuses on the prediction of nine classes of structures, according to the Dictionary of Secondary Structure of Proteins (DSSP) version 4. The model's performance was rigorously evaluated using various datasets. Additionally, this study compares the model with the state-of-the-art methods in the prediction of eight structure classes. The findings reveal that TE_SS excels in nine- and three-class structure predictions while also showing remarkable proficiency in the eight-class category. This is underscored by its performance in Qs and SOV evaluation metrics, demonstrating its capability to discern complex protein sequence patterns. This advancement provides a significant tool for protein structure analysis, thereby enriching the field of bioinformatics.

Keywords-protein secondary structure prediction; bioinformatics; nine-class protein prediction; transformer model; Ankh protein language model

I. INTRODUCTION

Proteins are made up of chains of amino acids. By altering their arrangement, 20 different types of acids can create a wide range of proteins. The primary structure of a protein is represented by one sequence, comprising the specific order in which the amino acids are arranged [1], and is referred to as 1D structure. Tertiary structures, often referred to as three shapes, are formed in living organisms through the interactions, among amino acids. These interactions play a crucial role in determining the function of proteins [2]. To fully comprehend the relationship between the tertiary structures of a protein, it is important to predict its secondary structure [3]. The use of efficient techniques to forecast protein structures has become essential in closing the disparity between the number of recognized protein sequences and the determined structures due to the limitations that experimental procedures entail, such as

time requirements and the substantial costs involved [4]. These predictive models are instrumental in enhancing our comprehension of protein functions and may be utilized in applications like drug development and disease control [5].

Secondary structure in proteins refers to the folded patterns that occur within a chain of acids as a result of forces like hydrogen and van der Waals bonds. To precisely define structure, the Dictionary of Secondary Structure of Proteins (DSSP) was devised [6]. This program analyzes the coordinates of proteins with known structures to identify patterns of hydrogen bonding and geometric characteristics. DSSP assigns a type of secondary structure to each residue in the protein. The original classification consisted of eight classes: G (310 helix), H (α -helix), I (π -helix), B (isolated β -strand), E (extended strand), S (bend), T (turn), and L (irregular structure). These categories are often grouped into three group classes: H, G, I to

helix (H), (B, E) to strand (E), and T, S, L to coil (C). The latest iteration of DSSP, version 4.0, which was released in 2021, marks a significant update in the field of protein secondary structure classification. This version extends the conventional eight types of secondary structures to include a ninth type, known as the poly-proline helix (P) [7]. The task of Protein Secondary Structure Prediction (PSSP) involves assigning classes of structures, such as alpha helices, beta sheets, and coils to each individual amino acid in a protein chain. For computational methods to predict a structure, it is necessary to represent acids as numeric vectors. One-hot vector approach uses 21-encoding for each amino acid in protein sequence, which includes the 20 standard amino acids that make up the proteins and 1 non-standard amino acid represented by X to indicate an unknown or unspecified amino acid. However, this method has shown limited accuracy, in prediction. Another used technique involves utilizing PSSM profile features [8] or HHM profile features [9]. These profile features incorporate information derived from analyzing sequence alignments obtained from a large protein sequence database.

Creating Hidden Markov Models (HMMs) or Position Specific Scoring Matrices (PSSMs) for each template sequence can be a time-consuming process, especially when dealing with proteins that have no sequences. To overcome this hurdle, recent advancements have introduced novel protein representation techniques inspired by methods used in natural language processing [10-13]. These techniques involve the usage of pretrained protein language models, followed by fine tuning for specific tasks. These models can achieve performance even with limited task specific data available, where embedding from a language model pretrained on a large corpus of protein sequences effectively replaces evolutionary information. The implementation of this approach has demonstrated encouraging outcomes in several protein-related subsequent studies [4, 7, 14-19]. In the early stages of PSSP research statistical approaches were predominantly used. These methods focused on determining the likelihood of amino acids, in protein structures [20]. Initially these predictors were designed for a three-class secondary structure prediction due to training data availability and computational constraints of that time. However, the particular methods encountered challenges in achieving high Q3 accuracy because they struggled to extract information from primary protein structure sequences. To overcome this limitation and improve their performance, researchers started incorporating information and position specific scoring matrices into the prediction process. This advancement proved significant, leading to a Q3 accuracy exceeding 70% [21].

Various machine learning techniques have been used for performing coarse-grained prediction, including decision trees [22], support vector machines [23], Neural Networks (NN) [24], HMMs [25], probabilistic graph models [26], and k-nearest neighbors [27]. The methodologies in this area primarily utilize a fixed-size sliding window approach. This method was employed to forecast the secondary structure category of the essential amino acid residue in a given sequence. JPred4 [28] and PSIPRED V3.0 [29] were notable among the initial prediction algorithms. These techniques laid

the foundation for further progress in the field, demonstrating the effectiveness of machine learning in understanding and predicting protein structures. The increased availability of data has led to the dominance of sequence-to-sequence deep model predictions, which have achieved state-of-the-art performance. Innovations in this area include DCRNN [30], which uses cascaded Convolutional and Recursive NN to extract both multiscale local and global contextual features. Other significant contributions include multiscale chained convolutional architecture for improved eight-state prediction [31]. SPIDER3 [32] uses LSTM BRNNs to capture complex amino acid interactions, DeepACLSTM [33] integrates networks with LSTM units and utilizes specific dimensions in protein sequence feature vectors. MUFOLD SS [34] and SAINT [35] both employ Deep inception-inside-inception networks with MUFOLD SS emphasizing inception modules while SAINT incorporates self-attention mechanisms. SPOD 1D [36] combines LSTM BRNN and ResNet models with residue contact maps for its predictions. NetSurfP 2.0 [37] employs convolutional and LSTM networks, while ShuffleNet_SS [5] focuses on a lightweight convolutional NN. Another important development is the introduction of the protein encoder [38]. This method employs a two-step process, beginning with an unsupervised autoencoder for feature extraction, followed by an ensemble of feature selection methods. A common element in earlier prediction models is their reliance on profile features, which are primarily obtained from Multiple Sequence Alignments (MSA). Nevertheless, the specific process, especially considering the rapidly expanding protein sequence databases, poses a significant time constraint. In response to this challenge, recent research has shifted toward leveraging embedding features extracted from pretrained protein language models. For instance, DML_SS [4] applies learning through a deep centroid model, for its predictions. SPOT 1D LM [19] synergizes embeddings from language models with one hot encoding techniques. LIFT SS [7] focuses on tuning pretrained protein language models.

In the field of predicting protein secondary structures, it is interesting to note that most current predictors apart from LIFT_SS [7] rely on the eight class assignments of structures from the previous version of the DSSP program for their training and evaluation data. Notably, even subsequent studies published after the introduction of DSSP 4.0 have continued to rely on eight-class secondary structure information, rather than adopting the more recent nine-class secondary structure classification. This trend indicates that these methodologies are being trained and evaluated using potentially outdated labeling information. In this study, the latest edition DSSP 4, a comprehensive database for secondary structure sequences, was utilized. The use of DSSP 4 ensured the training and evaluation data were based on a detailed classification of protein structures. Additionally, the Ankh protein language model [39] was adopted for obtaining protein embeddings, leveraging its capability to accurately represent protein sequences and replace the need for more computationally intensive evolutionary information. TE_SS, a deep transformer-based model [40], specifically designed to discern complex relationships between distant and proximal amino acid sequences in proteins, is proposed. This model is specifically

designed to discern complex relationships between local and nonlocal amino acid sequences in proteins, processing sequential features in parallel, in contrast to existing models that extract features sequentially. The architecture of this model enables it to capture patterns and interactions within protein structures, enhancing the prediction of secondary structure.

II. METHODS

A. Dataset

In this research, a collection of protein training data was generated using the PISCES server [38]. The latter is well known for creating curated lists of sequence subsets from the Protein Data Bank (PDB). To assess protein structure prediction algorithms, criteria and parameters related to sequence identity were applied. The PISCES server utilizes a filter based on predefined protein parameters. Subsequently it sends the resulting lists and sequence files directly to the email address provided. To ensure the dataset's reliability and usefulness, the PISCES server was configured with settings, including a maximum resolution of 2.0 Å, an upper limit R value of 2.0 and a requirement that there will be no more than 50% sequence identity between any pair of protein sequences. Initially the dataset suggested consisted of 16,225 proteins. However, 188 proteins from this collection were excluded because they lacked corresponding information. In addition, to maintain the integrity of the performed analysis and avoid data contamination, any proteins that overlapped with the proposed test dataset were eliminated. Following these rigorous filtering criteria, a refined dataset, labeled as 16,037, consisting of 16,037 proteins was successfully curated. This dataset was strategically partitioned, with 15,037 proteins designated for the training set and the remaining 1000 proteins allocated for validation purposes.

This study involves datasets for secondary structure analysis of protein based on the 9-class classification provided by DSSP4. The DSSP software generates a DSSP file for each protein with an established structure, which contains detailed secondary structure information derived from the protein's three-dimensional structural data recorded in the PDB database. In the performed methodology, the Biopython library was initially employed to retrieve the PDB file corresponding to a given protein chain. This file is accessed from the PDB website using the specific PDB ID and the chain ID of the protein chain. The occurrence of nonstandard amino acids in these files, including modified residues was observed. A notable example includes the representation of methionine (MET) and selenomethionine (MSE) by the one-letter code M. To address this issue, a conversion process in which the three-letter amino acid codes in the PDB file were translated to their one-letter equivalents, with nonstandard amino acids denoted as X, was implemented. This modified sequence is referred to as the target primary sequence. Subsequently, the identical DSSP file was acquired on the basis of the PDB file. Contiguous fragments of amino acid residues and their associated secondary structures were extracted from this file, guided by the chain ID and the residue sequence number. However, extracting a primary sequence from the DSSP file that exactly matches the target primary sequence in terms of sequence composition or length is often not feasible [6]. To accurately

represent the sequence of protein structure, the primary sequence of interest is matched with the sequence obtained from the DSSP file. During this alignment process, any gaps that occur are filled with the letter X to indicate unassigned types of structures. To achieve this alignment, the Pairwise2 alignment algorithm from the BioPython package was utilized [39].

Performance metrics for nine-class prediction were assessed on diverse datasets. Three editions of the CASP competition, namely CASP12, CASP13, and CASP14, were utilized. These datasets encompass a selection of 47, 41, and 33 protein chains, respectively, carefully chosen to represent real-world challenges in protein structure prediction. Additionally, the CB433 test data [4], a curated and filtered subset of the widely used CB513 dataset, comprising 433 protein structures was considered. Evaluating the proposed model fairly against existing models necessitates the use of datasets that adhere to the same 8-class system. For this purpose, two well-known test datasets, TEST2016 and TEST2018 [34], containing 1213 and 250 protein sequences accordingly, were chosen. These are in line with the training and validation sets, which include 10029 and 983 proteins, respectively. Across all four datasets, the maximum length of any protein sequence does not exceed 700. Additionally, the primary and secondary structure sequence data, which is the standard for these datasets, is also utilized.

B. Embedding

Pretrained models that focus on protein language (Protein Language Models-PLMs) have become a tool, in biological applications serving as a strong foundation for modeling protein related tasks. While most approaches rely on these models for extracting features, this study takes an approach by utilizing the Ankh model [39], which is a large unsupervised PLM. Ankh, built on a transformer-based architecture and has been trained on the BFD [41] and UniRef50 [42] dataset. It achieves state of the art performance while using less than 10% of the parameters compared to models. This impressive efficiency opens up possibilities for accessible and scalable protein modeling applications. One of the strengths of the Ankh model lies in its ability to extract high quality embedding features that represent proteins accurately. These features are representations of protein sequences that capture information about their structure, function and evolutionary relationships. To acquire the embedding feature of a specific protein chain using Ankh we input its sequence into the model encoder and retrieve its output. Each amino acid, in a protein sequence is assigned a 1536-feature vector through the output embedding, which captures its information. For every protein sequence L this model generates an embedding vector of size (L*1536). The embeddings acquired from the embedding process were used as the input for the model.

C. Model Architecture

A novel model has been developed to predict protein structures. It heavily relies on transformer architectures. These transformers are great, at identifying both distant relationships within protein sequences by utilizing self-attention mechanisms and feed forward layers. The initial input for this model is a two-dimensional embedding, with (L, 1536) sequence

dimensions. This embedding is generated using the training Ankh model. Additionally, the model incorporates encoding for each amino acid in the sequence to enhance information representation. After that, a series of N transformer encoders process the input as shown in Figure 1. This approach demonstrates how effective the model is at capturing patterns, within protein sequences leading to accurate predictions of protein structures.

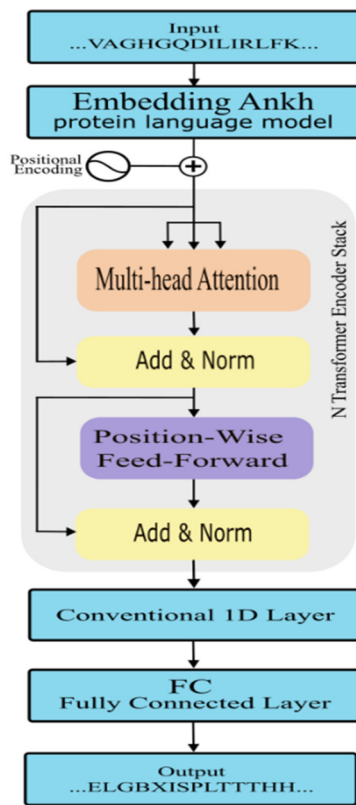


Fig. 1. Transformer-based model architecture for protein secondary structure prediction.

1) Positional Encoding

To ensure the proposed model effectively takes into account the nature of protein sequences, positional encoding was incorporated. This approach produces data on the precise locations of amino acids throughout the protein sequence. By combining positional encoding with amino acid embeddings, not only can this model comprehend the unique characteristics of each amino acid, but also their contextual relationships within the sequence. This approach is crucial for capturing the spatial details of amino acids, which are essential for accurately predicting protein secondary structure. Positional encoding (PESS) is defined as follows:

$$\text{PESS}_{(\text{psn}, 2i)} = \sin(\text{psn}/10000^{2i/\text{dim}_{\text{model}}}) \quad (1)$$

$$\text{PESS}_{(\text{psn}, 2i+1)} = \cos(\text{psn}/10000^{2i/\text{dim}_{\text{model}}}) \quad (2)$$

where psn represents the position of an amino acid in the sequence and i is its dimension in the encoding space, whereas $\text{dim}_{\text{model}}$ refers to the dimensionality of the model [43]. In (1)

the encoding for odd sequence positions is addressed, while (2) pertains to the encoding for even sequence positions.

In (3) the positional encoding obtained from (1) and (2) is illustrated and added into the input embeddings.

$$Xemb'_i = Xemp_i + \text{PESS} \quad (3)$$

where $Xemb'_i$ is positionally encoded embedding for i amino acid and $Xemp_i$ is i amino acid's embedding obtained from the Ankh model.

By including these data the suggested model acquires a comprehension of the protein's arrangement, which improves its predictive abilities, for the secondary structure.

2) Transformer Encoder in Protein Secondary Structure Prediction

The transformer encoder is a component of the transformer architecture [43] used for processing sequences in parallel. It is composed of layers, each of which has two sublayers: the Position-Wise Feed Forward Network and the Multi Head Self Attention Mechanism.

The key elements of the Transformer Encoder are:

a) Multi-head Attention

The multi head self-attention mechanism plays a role, in the encoder by allowing the model to evaluate and adjust the importance of segments within an input sequence. It creates three vector representations, i.e. query (Q), key (K), and value (V) for each input element. By measuring the similarity between Q and K, the attention scores are calculated to determine a sum of V vectors highlighting most relevant information. This process is performed across multiple heads enabling focus on different aspects of the sequence. The mathematical formulation, for this process is [43]:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_K}}\right)V \quad (4)$$

where $\sqrt{d_K}$ serves as a scaling factor to ensure values for sequences.

b) Feed-Forward Networks

After the self-attention mechanism, the data pass through a feed-forward NN, which is applied to each position separately and identically. This network consists of fully connected layers with activation functions and is responsible for further transforming the representation.

c) Layer Normalization

Each sublayer of both self-attention and feed-forward networks in the transformer encoder has a residual connection around it, followed by layer normalization. The residual connections help mitigate the vanishing gradient problem, enabling the training of very deep models.

d) Stacking of Layers

The self-attention, multi-head attention, and feed-forward layers are stacked together, forming multiple encoder layers. Each layer builds upon the previous one, gradually extracting increasingly complex and higher-level representations of the sequence.

3) Convolution 1D Layer

To further augment the model's capability to extract informative features, a 1D convolutional layer follows the transformer encoder. This layer operates along the feature dimension, applying learnable filters to capture local patterns and dependencies within the feature space. Mathematically, the expression of the one-dimensional convolution operation can be formally articulated as:

$$Y = f(w \otimes X + b) \quad (5)$$

where Y is the output feature map, X is the input feature map, w is the convolutional filter, b is the bias term, \otimes is the convolution operator, and f is the activation function.

4) Final Fully Connected Layer

The architecture concludes with a fully connected layer, which serves as the classification component of the model. This layer translates the processed features into predictions of the protein's secondary structure.

D. Evaluation Metrics

To evaluate the effectiveness of the proposed approach, two employed metrics were utilized; Q_s accuracy and Segment Overlap (SOV) [44]. Q_s accuracy measures how the predicted secondary structure aligns, with the determined secondary structure specifically looking at the proportion of residues that match. Meanwhile, SOV assesses how closely the predicted and experimentally determined secondary structure segments resemble each other. In addition to these metrics, F1, Precision, and Recall were also employed to evaluate the proposed model's performance on the selected test dataset. Q_s accuracy quantifies the proportion of residues where the predicted secondary structure aligns with the findings. This metric plays a role in assessing a model's ability to accurately classify types of secondary structures found in proteins. Precision is indicated by how residues are correctly predicted for their corresponding secondary structures. It expands on the conventional Q_s accuracy measure $S = (H, E, C)$ by categorizing secondary structures into nine categories: $S = (H, G, I, P, B, E, T, S, L)$. To compute Q_s we divide the number of correctly predicted residues, in state s (n_s) by the total number of residues actually in state s (N_s), with s representing each state within the set S . This is formally represented in (6):

$$Q_s = \frac{n_s}{N_s} \times 100, s \in S \quad (6)$$

To calculate the overall accuracy for per residue prediction all (n_s) values for each state s in set S are summed up and divided by the sum of all (N_s) values for each state s in set S [4]:

$$Q_{|S|} = \frac{\sum_{s \in S} n_s}{\sum_{s \in S} N_s} \times 100 \quad (7)$$

The SOV metric is crucial when evaluating the precision of protein secondary structure predictions. Unlike accuracy measures, SOV provides a detailed evaluation by considering both length and overlap between the predicted and actual segments. This metric is useful when assessing predictions for structure elements like alpha helices and beta sheets which can vary significantly in length. SOV compares how well predicted

segments align with segments in terms of length and overlap. It takes into account variations, in segment size making it a comprehensive and realistic measure to assess prediction performance for complex proteins that exhibit diverse secondary structures.

E. Implementation Details

PyTorch framework was used as it offers a graph, imperative execution style and a wide range of tools and libraries. To ensure training and avoid overfitting to data patterns, the minibatch size was set to 8 and random sampling was employed to create minibatches. For optimizing the suggested models, the AdamW optimizer was used with a weight decay value of 0.0001. Throughout the training process a fixed learning rate of 0.00005 was maintained. To enhance the proposed model's performance, a custom cross loss function that handles class imbalances by allowing optional weights, for different classes was implemented. This function calculates the loss for each instance without reduction, and then averages it across the minibatch while considering the provided class weights. This approach ensures an impact of each class on the models learning process. Moreover, a stopping criterion was implemented. The particular criterion halts training if there is no improvement in Q_s accuracy, on the validation set for 5 consecutive epochs. This study experiments were conducted using an NVIDIA Tesla V100 GPU with 16 GB VRAM and 32 GB system memory. The transformer encoder architecture used in this study consisted of 5 layers, each equipped with 8 attention heads. In these layers, a dropout rate of 0.2 was incorporated. The dimension of the feed forward network was set to 2048. The model's convolutional layers produced an output with 1024 channels.

III. RESULTS

A. Ablation Study

We comprehensively evaluated the performance of the proposed method through a series of experiments on the CB433 test set and our validation set. These experiments were meticulously designed to analyze the influence of key hyperparameters, specifically the number of transformer encoder layers, the number of attention heads, and the learning rate, on the model's effectiveness.

1) Number of Encoder Layers

To examine the effect of encoder layer depth on model performance, this study experimented with architectures ranging from 1 to 7 layers, each coupled with a fixed configuration of 8 attention heads. The validation and test results on the CB433 dataset, as depicted in Figure 2, indicate that the architecture with 5 encoder layers achieved the highest 9-class accuracy (Q9). Deeper models can learn more complex contextual representations and better capture long-range dependencies in protein sequences, performance plateaus, but this ability slightly declines beyond 5 layers. This could be indicative of overfitting or vanishing gradients, affecting the model's generalizability and learning efficacy.

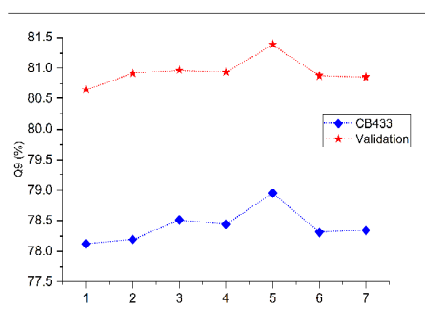


Fig. 2. Performance evaluation of transformer models with varying numbers of encoder layers.

2) Number of Attention Heads

To investigate the optimal configuration for protein secondary structure prediction, there was a focus on the number of attention heads in the transformer encoder layer. Configurations with 1, 2, 3, 4, 6, 8, 12, and 16 attention heads were tested for their impact on the model's performance to be determined. As shown in Figure 3, the model with 8 attention heads proved most effective in both validation and testing sets, particularly on the CB433 dataset, achieving significant improvements in 9-class accuracy, pointing to the optimal balance between the granularity and breadth of attention mechanisms. While increasing the number of attention heads generally improves model's ability to discern intricate relationships within protein structures, a threshold exists beyond which additional heads may not enhance or could even reduce predictive accuracy. This highlights the importance of fine-tuning attention mechanisms in transformer models for specialized bioinformatics tasks.

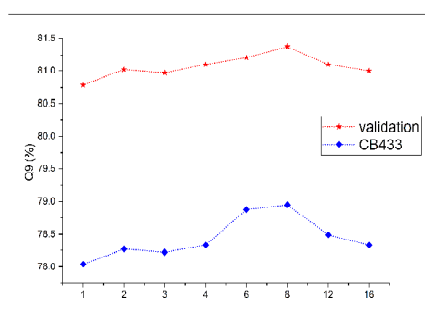


Fig. 3. Comparative analysis of prediction accuracy across different numbers of attention heads.

3) Learning Rate

The study examined the best hyperparameters for protein secondary structure prediction and found that the learning rate had a significant impact on model accuracy. The former rigorously assessed the model's performance throughout a range of learning rates, as shown in Figure 4: 0.001, 0.0005, 0.0001, 0.00005, 0.00001, and 0.000005. In the testing and validation stages, a learning rate of 0.00005 produced the best 9-class accuracy, especially when employing the CB433

dataset. Interestingly, there was a clear trend in the model's performance: the accuracy decreased dramatically at higher learning rates (0.001 and 0.0005) pointing the detrimental effect of rapid weight adjustments. However, as the learning rate was gradually reduced, a notable improvement in accuracy was observed, culminating in the optimal performance at 0.00005.

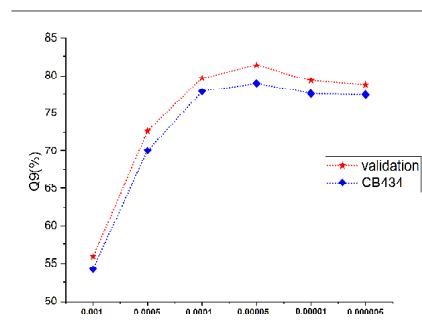


Fig. 4. Impact of learning rate on model accuracy.

B. Comparative Analysis on Eight-State Prediction

This section provides a comparative analysis of the proposed method against a selection of state-of-the-art predictors methods, specifically focusing on two distinct test data called, TEST2016 and TEST2018. To ensure an equitable comparison, data for existing predictors were sourced from the literature [4, 7, 9]. The comparison encompasses a variety of methods: 10 predictors based on profile features and 3 predictors based on embedding features (Table I). For the LIFT_SS method, the most accurate results were selected from three lightweight fine-tuning approaches. In the conducted analysis the TE_SS method was evaluated, against these 13 predictors through the employment of different metrics on two test datasets. These metrics include Q8 and SOV8 for predictions in 8 classes and Q3 and SOV3 for predictions in 3 classes. The detailed results can be found in Table I, which showcases the performance of the proposed method alongside the 13 methods for each metric. Table I clearly demonstrates that the TE_SS method outperforms the others in predicting protein secondary structures in both 8- and 3- class formats. Not only does this comprehensive analysis reveal the strength of the TE_SS model, but also its advancement over existing state of the art methods setting a new standard, in protein secondary structure prediction.

C. Comparative Analysis on Nine-State Prediction

To assess the effectiveness of the TE_SS framework experiments were conducted using four widely used benchmark datasets, in the field of protein structure analysis. TE_SS was compared against two leading methods for predicting 9 class protein structures; DML_SSembed and LIFT_SS. Both these methods utilize embeddings derived from ProtTrans, a trained PL). These two methods were selected based on their utilization of the 9 class predictor from DSSP4.

DML_SSembed employs a centroid model for sequence to sequence prediction. It assigns a centroid in the embedding

space to each structure category and aims to maximize the similarity between each amino acid and its corresponding centroid. This approach enhances the accuracy of secondary structure prediction. In contrast, LIFT_SS utilizes a fine tuning strategy on the pre trained PLM by employing 7 state of the art fine tuning techniques. This enables LIFT_SS to predict structures accurately by introducing new parameters during the embedding process. The results of these comparisons including

predictions, for both 9- and 3- class scenarios are presented in Tables II and III. Notably, the highest metric values were taken from the 7 fine-tuning techniques used by LIFT_SS. The data for existing predictors were obtained from [7]. It is worth mentioning that the TE_SS model consistently outperformed both DML_SSEmbed and LIFT_SS exhibiting its accuracy and effectiveness, in predicting protein structure.

TABLE I. COMPARISON (Q8, Q3, SOV8, AND SOV3 ACCURACY) WITH STATE-OF-THE-ART METHODS

Method	TEST2016				TEST2018			
	Q8	SOV8	Q3	SOV3	Q8	SOV8	Q3	SOV3
CNN_BIGRU [45]	73.91	70.92	85.04	81.61	72.78	68.75	84.17	79.41
DeepACLSTM [33]	75.19	73.67	85.62	82.6	73.42	71.32	84.66	80.05
DCRNN [30]	72.19	68.63	83.72	78.39	70.6	65.82	82.75	75.1
DeepCNN [31]	74.54	71.56	85.14	79.31	72.75	69.18	84.16	76.83
MUFold-SS [34]	76.03	73.67	85.97	81.98	74.29	71	84.63	79.53
NetSurfP-2.0 [37]	-	-	-	-	73.81	71.14	85.31	78.58
SPOD-1D [36]	76.03	73.88	86.67	79.52	74.26	71.45	85.66	78.77
SAINT [35]	76.23	-	-	-	74.48	-	-	-
SPIDER-3 [32]	-	-	84.66	75.62	-	-	83.84	73.89
DML_SS [4]	76.62	74.6	86.1	82.72	74.82	72.23	84.83	80.5
SPOT-1D-LM [19]	-	-	-	-	76.47	-	86.74	-
DML_SSEmbed [4]	78.03	75.9	87.41	84.51	76.48	73.44	86.82	82.43
LIFT_SS [7]	78.7	76.79	87.84	84.76	76.86	74.24	87.13	82.32
TE_SS	79.08	77.02	87.99	84.81	77.57	74.60	87.31	82.47

TABLE II. COMPARATIVE 9-CLASS PSSP RESULTS ON THE TEST DATASETS

Method	CASP12		CASP13		CASP14		CB433	
	Q9	SOV9	Q9	SOV9	Q9	SOV9	Q9	SOV9
DML_SSEmbed [4]	74.81	70.07	72.32	67.25	65.4	58.43	75.59	73.35
LIFT_SS [7]	75.82	70.81	73.07	68.06	66.59	59.71	76.87	74.36
TE_SS	76.44	71.03	75.62	71.51	67.47	60.29	78.95	76.37

TABLE III. COMPARATIVE 3-CLASS PSSP RESULTS ON THE TEST DATASETS

Methods	CASP12		CASP13		CASP14		CB433	
	Q3	SOV3	Q3	SOV3	Q3	SOV3	Q3	SOV3
DML_SSEmbed [4]	84.56	78.48	82.75	74.99	77.7	68.82	85.9	80.63
LIFT_SS [7]	85.24	79.38	83.56	77.3	78.38	68.23	86.69	81.28
TE_SS	85.55	79.84	84.69	77.50	78.56	69.71	87.35	81.64

D. Multi-Metric Evaluation

To thoroughly evaluate models performance, an approach was adopted by considering evaluation metrics, such as F1 score, Precision, and Recall. These metrics were applied to CB433, CASP12, CASP13, and CASP14 datasets. This rigorous evaluation strategy ensured that the model's effectiveness in predicting protein structure was reliable and applicable to a range of protein sequences. Table IV provides a summary of the models performance on these metrics highlighting its accuracy in predicting protein structure. The proposed model consistently performed satisfactorily across all datasets indicating its potential, for various protein structure prediction tasks.

TABLE IV. PERFORMANCE OF TE_SS ON TEST DATASETS

Dataset	F1	Precision	Recall
CASP12	61.28	68.66	57.45
CASP13	58.95	64.24	56.56
CASP14	45.11	53.06	42.94
CB433	67.72	73.16	65.55

IV. CONCLUSIONS

In this study, the effectiveness of the transformer-based TE_SS model in predicting protein structures has been demonstrated. Utilizing the Ankh protein language model for feature embedding, the TE_SS model achieves accurate predictions of protein structures in both nine and eight classification systems. The model's performance in predicting 9-class structures was evaluated on CASP12, CASP13, CASP14, and CB433 test datasets. Also, the model, trained on data containing 8 classes, was evaluated on two publicly available test datasets, TEST2016 and TEST2018. The experimental results indicate improved accuracy compared to the other models. A notable advancement of TE_SS is its adeptness in capturing both short-range and long-range dependencies among residues in proteins. The ability of this transformer-based model to process sequence data in parallel demonstrates its efficiency and effectiveness in analyzing complex protein structures. However, it is worth noting that the proposed method has limitations in terms of its demanding resources and GPU memory requirements. Moreover, the

model currently lacks the ability to provide information regarding the reliability or confidence level of its predictions. This shortcoming is especially evident when the model encounters specific types of proteins or disordered regions within proteins, where its predictions may be less accurate or reliable. For future work, it is imperative to address these limitations, potentially by developing methods to estimate prediction reliability and optimizing the model for reduced resource consumption.

REFERENCES

- [1] S. Damodaran and K. L. Parkin, Eds., "Amino Acids, Peptides, and Proteins," in *Fennema's Food Chemistry*, 5th ed., Boca Raton, FL, USA: CRC Press, 2017.
- [2] S. Tahzeeb and S. Hasan, "A Neural Network-Based Multi-Label Classifier for Protein Function Prediction," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 7974–7981, Feb. 2022, <https://doi.org/10.48084/etasr.4597>.
- [3] M. Zubair *et al.*, "A Deep Learning Approach for Prediction of Protein Secondary Structure," *Computers, Materials & Continua*, vol. 72, no. 2, pp. 3705–3718, Mar. 2022, <https://doi.org/10.32604/cmc.2022.026408>.
- [4] W. Yang, Y. Liu, and C. Xiao, "Deep metric learning for accurate protein secondary structure prediction," *Knowledge-Based Systems*, vol. 242, Apr. 2022, Art. no. 108356, <https://doi.org/10.1016/j.knsys.2022.108356>.
- [5] W. Yang, Z. Hu, L. Zhou, and Y. Jin, "Protein secondary structure prediction using a lightweight convolutional network and label distribution aware margin loss," *Knowledge-Based Systems*, vol. 237, Feb. 2022, Art. no. 107771, <https://doi.org/10.1016/j.knsys.2021.107771>.
- [6] W. Kabsch and C. Sander, "Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features," *Biopolymers*, vol. 22, no. 12, pp. 2577–2637, 1983, <https://doi.org/10.1002/bip.360221211>.
- [7] W. Yang, C. Liu, and Z. Li, "Lightweight Fine-tuning a Pretrained Protein Language Model for Protein Secondary Structure Prediction," *bioRxiv*, Mar. 23, 2023, <https://doi.org/10.1101/2023.03.22.530066>.
- [8] D. T. Jones, "Protein secondary structure prediction based on position-specific scoring matrices," Edited by G. Von Heijne, *Journal of Molecular Biology*, vol. 292, no. 2, pp. 195–202, Sep. 1999, <https://doi.org/10.1006/jmbi.1999.3091>.
- [9] S. R. Eddy, "Profile hidden Markov models," *Bioinformatics*, vol. 14, no. 9, pp. 755–763, Jan. 1998, <https://doi.org/10.1093/bioinformatics/14.9.755>.
- [10] A. Rives *et al.*, "Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences," *Proceedings of the National Academy of Sciences*, vol. 118, no. 15, Apr. 2021, Art. no. e2016239118, <https://doi.org/10.1073/pnas.2016239118>.
- [11] A. Elnaggar *et al.*, "ProtTrans: Toward Understanding the Language of Life Through Self-Supervised Learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 7112–7127, Jul. 2022, <https://doi.org/10.1109/TPAMI.2021.3095381>.
- [12] Z. Lin *et al.*, "Language models of protein sequences at the scale of evolution enable accurate structure prediction," *bioRxiv*, Jul. 21, 2022, <https://doi.org/10.1101/2022.07.20.500902>.
- [13] B. Ahmed, G. Ali, A. Hussain, A. Baseer, and J. Ahmed, "Analysis of Text Feature Extractors using Deep Learning on Fake News," *Engineering, Technology & Applied Science Research*, vol. 11, no. 2, pp. 7001–7005, Apr. 2021, <https://doi.org/10.48084/etasr.4069>.
- [14] J. Singh, T. Litfin, J. Singh, K. Paliwal, and Y. Zhou, "SPOT-Contact-LM: improving single-sequence-based prediction of protein contact map using a transformer language model," *Bioinformatics*, vol. 38, no. 7, pp. 1888–1894, Mar. 2022, <https://doi.org/10.1093/bioinformatics/btac053>.
- [15] H. Stark, C. Dallago, M. Heinzinger, and B. Rost, "Light attention predicts protein location from the language of life," *Bioinformatics Advances*, vol. 1, no. 1, Jan. 2021, Art. no. vbab035, <https://doi.org/10.1093/bioadv/vbab035>.
- [16] S. Pokharel, P. Pratyush, M. Heinzinger, R. H. Newman, and D. B. Kc, "Improving protein succinylation sites prediction using embeddings from protein language model," *Scientific Reports*, vol. 12, no. 1, Oct. 2022, Art. no. 16933, <https://doi.org/10.1038/s41598-022-21366-2>.
- [17] A. Villegas-Morcillo, A. M. Gomez, and V. Sanchez, "An analysis of protein language model embeddings for fold prediction," *Briefings in Bioinformatics*, vol. 23, no. 3, May 2022, Art. no. bbac142, <https://doi.org/10.1093/bib/bbac142>.
- [18] M. H. Hoie *et al.*, "NetSurfP-3.0: accurate and fast prediction of protein structural features by protein language models and deep learning," *Nucleic Acids Research*, vol. 50, no. W1, pp. W510–W515, Jul. 2022, <https://doi.org/10.1093/nar/gkac439>.
- [19] J. Singh, K. Paliwal, T. Litfin, J. Singh, and Y. Zhou, "Reaching alignment-profile-based accuracy in predicting protein secondary and tertiary structural properties without alignment," *Scientific Reports*, vol. 12, no. 1, May 2022, Art. no. 7607, <https://doi.org/10.1038/s41598-022-11684-w>.
- [20] M. Levitt and C. Chothia, "Structural patterns in globular proteins," *Nature*, vol. 261, no. 5561, pp. 552–558, Jun. 1976, <https://doi.org/10.1038/261552a0>.
- [21] P. Kumar, S. Bankapur, and N. Patil, "An enhanced protein secondary structure prediction using deep learning framework on hybrid profile based features," *Applied Soft Computing*, vol. 86, Jan. 2020, Art. no. 105926, <https://doi.org/10.1016/j.asoc.2019.105926>.
- [22] J. Selbig, T. Mevissen, and T. Lengauer, "Decision tree-based formation of consensus protein secondary structure prediction," *Bioinformatics*, vol. 15, no. 12, pp. 1039–1046, Dec. 1999, <https://doi.org/10.1093/bioinformatics/15.12.1039>.
- [23] B. Yang, Q. Wu, Z. Ying, and H. Sui, "Predicting protein secondary structure using a mixed-modal SVM method in a compound pyramid model," *Knowledge-Based Systems*, vol. 24, no. 2, pp. 304–313, Mar. 2011, <https://doi.org/10.1016/j.knsys.2010.10.002>.
- [24] M. H. Zangooei and S. Jalili, "PSSP with dynamic weighted kernel fusion based on SVM-PHGS," *Knowledge-Based Systems*, vol. 27, pp. 424–442, Mar. 2012, <https://doi.org/10.1016/j.knsys.2011.11.002>.
- [25] Z. Aydin, Y. Altunbasak, and M. Borodovsky, "Protein secondary structure prediction for a single-sequence using hidden semi-Markov models," *BMC Bioinformatics*, vol. 7, no. 1, Mar. 2006, Art. no. 178, <https://doi.org/10.1186/1471-2105-7-178>.
- [26] J. Martin, J.-F. Gibrat, and F. Rodolphe, "Analysis of an optimal hidden Markov model for secondary structure prediction," *BMC Structural Biology*, vol. 6, no. 1, Dec. 2006, Art. no. 25, <https://doi.org/10.1186/1472-6807-6-25>.
- [27] W. Yang, K. Wang, and W. Zuo, "Prediction of protein secondary structure using large margin nearest neighbour classification," *International Journal of Bioinformatics Research and Applications*, vol. 9, no. 2, pp. 207–219, Jan. 2013, <https://doi.org/10.1504/IJBRA.2013.052445>.
- [28] A. Drozdetskiy, C. Cole, J. Procter, and G. J. Barton, "JPred4: a protein secondary structure prediction server," *Nucleic Acids Research*, vol. 43, no. W1, pp. W389–W394, Jul. 2015, <https://doi.org/10.1093/nar/gkv332>.
- [29] D. W. A. Buchan, S. M. Ward, A. E. Lobley, T. C. O. Nugent, K. Bryson, and D. T. Jones, "Protein annotation and modelling servers at University College London," *Nucleic Acids Research*, vol. 38, no. suppl_2, pp. W563–W568, Jul. 2010, <https://doi.org/10.1093/nar/gkq427>.
- [30] Z. Li and Y. Yu, "Protein Secondary Structure Prediction Using Cascaded Convolutional and Recurrent Neural Networks," *arXiv*, Apr. 25, 2016, <https://doi.org/10.48550/arXiv.1604.07176>.
- [31] A. Busia and N. Jaitly, "Next-Step Conditioned Deep Convolutional Neural Networks Improve Protein Secondary Structure Prediction," *arXiv*, Feb. 13, 2017, <https://doi.org/10.48550/arXiv.1702.03865>.
- [32] R. Heffernan, Y. Yang, K. Paliwal, and Y. Zhou, "Capturing non-local interactions by long short-term memory bidirectional recurrent neural networks for improving prediction of protein secondary structure, backbone angles, contact numbers and solvent accessibility,"

- Bioinformatics*, vol. 33, no. 18, pp. 2842–2849, Sep. 2017, <https://doi.org/10.1093/bioinformatics/btx218>.
- [33] Y. Guo, W. Li, B. Wang, H. Liu, and D. Zhou, "DeepACLSTM: deep asymmetric convolutional long short-term memory neural models for protein secondary structure prediction," *BMC Bioinformatics*, vol. 20, no. 1, Jun. 2019, Art. no. 341, <https://doi.org/10.1186/s12859-019-2940-0>.
- [34] C. Fang, Y. Shang, and D. Xu, "MUFOLD-SS: New deep inception-inside-inception networks for protein secondary structure prediction," *Proteins: Structure, Function, and Bioinformatics*, vol. 86, no. 5, pp. 592–598, 2018, <https://doi.org/10.1002/prot.25487>.
- [35] M. R. Uddin, S. Mahbub, M. S. Rahman, and M. S. Bayzid, "SAINT: self-attention augmented inception-inside-inception network improves protein secondary structure prediction," *Bioinformatics*, vol. 36, no. 17, pp. 4599–4608, Nov. 2020, <https://doi.org/10.1093/bioinformatics/btaa531>.
- [36] J. Hanson, K. Paliwal, T. Litfin, Y. Yang, and Y. Zhou, "Improving prediction of protein secondary structure, backbone angles, solvent accessibility and contact numbers by using predicted contact maps and an ensemble of recurrent and residual convolutional neural networks," *Bioinformatics*, vol. 35, no. 14, pp. 2403–2410, Jul. 2019, <https://doi.org/10.1093/bioinformatics/bty1006>.
- [37] M. S. Klausen *et al.*, "NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning," *Proteins: Structure, Function, and Bioinformatics*, vol. 87, no. 6, pp. 520–527, 2019, <https://doi.org/10.1002/prot.25674>.
- [38] Uzma, U. Manzoor, and Z. Halim, "Protein encoder: An autoencoder-based ensemble feature selection scheme to predict protein secondary structure," *Expert Systems with Applications*, vol. 213, Mar. 2023, Art. no. 119081, <https://doi.org/10.1016/j.eswa.2022.119081>.
- [39] A. Elnaggar *et al.*, "Ankh Φ : Optimized Protein Language Model Unlocks General-Purpose Modelling," *bioRxiv*, Jan. 18, 2023, <https://doi.org/10.1101/2023.01.16.524265>.
- [40] T. S. Mian, "Evaluation of Stock Closing Prices using Transformer Learning," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11635–11642, Oct. 2023, <https://doi.org/10.48084/etasr.6017>.
- [41] M. Steinegger and J. Soding, "Clustering huge protein sequence sets in linear time," *Nature Communications*, vol. 9, no. 1, Jun. 2018, Art. no. 2542, <https://doi.org/10.1038/s41467-018-04964-5>.
- [42] B. E. Suzek, Y. Wang, H. Huang, P. B. McGarvey, and C. H. Wu, "UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches," *Bioinformatics*, vol. 31, no. 6, pp. 926–932, Mar. 2015, <https://doi.org/10.1093/bioinformatics/btu739>.
- [43] A. Vaswani *et al.*, "Attention is All you Need," in *31st Conference on Neural Information Processing Systems*, Long Beach, CA, USA, Dec. 2017, vol. 30, pp. 1–15.
- [44] A. Zemla, C. Venclovas, K. Fidelis, and B. Rost, "A modified definition of Sov, a segment-based measure for protein secondary structure prediction assessment," *Proteins: Structure, Function, and Bioinformatics*, vol. 34, no. 2, pp. 220–223, 1999, [https://doi.org/10.1002/\(SICI\)1097-0134\(19990201\)34:2<220::AID-PROT7>3.0.CO;2-K](https://doi.org/10.1002/(SICI)1097-0134(19990201)34:2<220::AID-PROT7>3.0.CO;2-K).
- [45] I. Drori *et al.*, "High Quality Prediction of Protein Q8 Secondary Structure by Diverse Neural Network Architectures," *arXiv*, Nov. 17, 2018, <https://doi.org/10.48550/arXiv.1811.07143>.

Building Information Modeling (BIM) for Construction Project Schedule Management: A Review

Tuan Anh Nguyen

Ho Chi Minh City University of Transport, Vietnam
tuanna@ut.edu.vn (corresponding author)

Tu Anh Nguyen

Ho Chi Minh City University of Transport, Vietnam
nguyenanhtugts@gmail.com

The Van Tran

DEOCA Research and Training Institute, DEOCA Group, Vietnam | Ho Chi Minh City University of Transport, Vietnam
tran.van.the@deoca.vn

Received: 29 December 2023 | Revised: 16 January 2024 | Accepted: 18 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6834>

ABSTRACT

Nowadays, the implementation and application of the BIM process in construction project management is a pressing need. This aligns with the global development trends in the construction sector and project information management in general. Numerous scholars and companies are actively engaged in learning, understanding, and investigating various aspects of BIM to stay up-to-date and meet the inevitable developmental requirements. This study focuses on the role and application of BIM, intending to identify limitations that hinder its fulfillment of expectations in project schedule management. In addition, it explores studies that show how other countries have effectively employed BIM in project management and progress tracking throughout the project lifecycle. The study aims to address three main objectives: (a) comprehensively examine and provide evidence related to the concept of BIM in project schedule management, (b) present the benefits of applying BIM in comparison to traditional methods in project management and operation, and (c) identify limitations stemming from various factors that may pose challenges in the application of BIM in project schedule management.

Keywords-*BIM; schedule management; project; model; technology*

I. INTRODUCTION

Building Information Modeling (BIM) is an information management process that encompasses the creation and administration of digital elements throughout various phases, including design, construction, and commissioning, as well as the operation of works that can encompass construction projects or industrial products. Information is categorized into two primary groups: geometric information, covering dimensions and positioning of building components, such as columns, beams, floors, pipes, bathtubs, light bulbs, cabinets, and tables, and non-geometric information (data), which provide additional details about components, such as manufacturer, maintenance time, cost, and supplier. BIM serves as a comprehensive model that contains project information and encourages collaborative utilization between

departments and stakeholders from the conceptual design phase through the construction and operational stages.

Not only does BIM act as a repository for storing and delivering project information throughout the design, construction, and operation management phases, but also as a dynamic process for generating and utilizing information sources to construct a virtual model of the building. This approach aims to optimize the design, construction, operation, and overall management of the project [1-3]. BIM provides users with a comprehensive project overview, helping project stakeholders make informed decisions, mitigate risks, and improve work efficiency while minimizing design changes during construction. The BIM model's content encompasses parameter information that can be adjusted at the user's discretion. When a party requires a design revision, the model and related elements, like volume and drawings, are promptly

updated. This feature is highly effective for designs that undergo substantial modifications during the approval process and the actual construction. BIM technology offers participants early exposure to the project, allowing users to easily identify inconsistent design points in different disciplines. This capability facilitates accurate design adjustments, minimizing conflicts and enhancing harmonization during project implementation.

Implementing BIM produces significant economic benefits for all stakeholders involved. The design team can reduce design and production costs, while contractors can coordinate more effectively, resulting in savings in estimating and supply management. Investors can reliably calculate project-related results, including costs, time, and operation and maintenance. Investors benefit from minimized management and supervision costs, optimized designs, reduced construction time, and minimized additional costs. The ability of BIM to enable accurate design and anticipate difficulties during construction without wasting time allows contractors and investors to reduce unnecessary costs.

BIM management technology significantly simplifies Design & Build work, with the incorporation of 4D BIM as a supplementary technique to enhance project management. The primary goal of 4D BIM is to ameliorate the visualization of the construction progress by linking the 3D model of the building with a pre-established schedule. This study outlines key aspects for visual comparison to underscore the effectiveness of using 4D BIM for project management in contrast to traditional methods: (a) Planning and scheduling, (b) Baseline vs actual progress comparison, (c) Construction progress control, (c) Lookahead plan, (d) Claim analysis and dispute resolution, (e) Construction methods presentation, (f) Work safety planning and management, (g) Tower crane location research and safety, (h) Total floor management, and (i) Construction drawings management. Embracing general project management with a unified BIM shared model represents a pivotal turning point in the era of applying new technologies to construction project management. Undoubtedly, it will soon see widespread use and become an indispensable tool for engineers, architects, and construction managers in Vietnam.

II. OVERVIEW OF BIM APPLICATION IN CONSTRUCTION PROJECT SCHEDULE MANAGEMENT

The adoption of information and communication technology in the construction industry remains relatively low, contributing to the emergence of certain complications, such as low productivity, frequent time and cost overruns in projects, as well as challenges with incident management processes, communication, and process automation [4]. Manual creation and updates of project progress persist, leading to time-consuming and error-prone practices that negatively impact planning and scheduling, a significant factor in project delays [5]. Numerous known and unknown risks underlie schedule delays, prompting the development of various theories and models. However, addressing this aspect still requires considerable effort due to the ongoing time failures in projects [6]. In 2004, the National Institute of Standards and Technology (NIST) reported that interoperability issues and

inadequate data management cost the construction industry an estimated \$15.8 billion annually, equivalent to approximately 3-4% of total industry costs [7].

The advent of BIM technology has presented robust technical capabilities to achieve informatization and digitization of the engineering construction sector, significantly accelerating the transformation and improvement of management practices in engineering construction projects [8-10]. The development of information management aligned with BIM for construction data enhances BIM's capacity to integrate modeling, communication, collaboration, as well as harmonization of design and construction requirements throughout all phases of sustainable construction projects [11]. Consequently, BIM plays a pivotal role in advancing automation within the construction industry and its corresponding management systems. However, the successful implementation of such initiatives depends on the comprehensive skills, capacity, and enthusiasm of builders and contractors, recognizing them as critical factors for the future success of these endeavors [12].

BIM proves advantageous in the creation of schedules, estimation, change tracking, and on-site logistics management [13]. In recent years, there has been an increasing focus on the use of BIM for project and cost management, becoming a prominent area of interest for project owners [14]. Storing a facility management plan within a BIM model helps facility managers comprehend task schedules, durations, and maintenance personnel, thus improving efficiency and convenience in facility management. Not only does this BIM input method streamline facility management processes, but also mitigates scheduling issues arising from human factors [15]. Visual representations of project schedules, coupled with ongoing project visuals, can aid in tasks, such as retention, record management, and identification of effective construction strategies for project time management [16]. Throughout the planning, design, and implementation phases, various tasks heavily rely on diverse location data, including worker and equipment locations for safety planning and management, as well as material location data for progress tracking. Given that a significant proportion of construction activities occur indoors, indoor location data for workers, materials, and other construction resources significantly affect quality, project safety, and productivity [17].

The concept of BIM theoretically originated and evolved at the Georgia Institute of Technology in the late 1970s and rapidly developed thereafter. This growth was driven by increasing attention from construction teams and companies that recognized the value of integrating and managing construction project processes using BIM. The term "Building Information Modeling" was coined in 2002 to describe virtual design, construction, and facility management [18-21]. Since BIM is perceived as a management process rather than just a model, a term gaining broader usage in recent times is "Building Information Management" [22]. This includes 3D/4D/5D BIM tools as innovative approaches to constructing visual models, performing collision detection, volume dissection, simulating construction, and graphically scheduling to ensure timely project completion [23]. BIM has emerged as

a prominent technological trend and a promising tool for various functions in planning, design, construction, and facility management [24]. In construction planning, BIM allows for the detection of conflicts during the design phase, enhancing project schedules [25]. Regulations should be flexible enough to allow the reorganization of project teams even during implementation, and a well-established BIM-enabled project structure should be transferable to other projects, ultimately reducing construction costs and improving communication among participants [26-27]. BIM offers the advantage of easily inserting, extracting, updating, or modifying digital data by various stakeholders, including owners, customers, engineers, architects, contractors, suppliers, and building owners [28].

Integrating project schedules with BIM elements can produce visual representations aligned with goals that involve the interpretation and production of construction drawings using computer-aided drafting, reasoning, and application of BIM software [29-30]. In the United States, integrated project delivery has become the preferred system for major BIM-related projects [31]. The government actively encourages and facilitates the application of BIM to increase productivity, conserve quality resources in construction activities, and improve project operation management [32]. A BIM model has significant potential to streamline the data collection and storage process for a project, serving as a comprehensive source for all project-related data. The overarching purpose of using BIM in facility management is to leverage facility data throughout the lifecycle, ensuring a safe, healthy, efficient, and productive work environment [33].

BIM revolutionizes how information is generated, managed, and communicated within project teams [34]. Scheduling is a crucial factor influencing the success of construction projects, guiding the when and where of tasks for all team members. Despite its importance, scheduling is often based on approximate timelines, leading to frequent delays and additional costs [35]. Many 4D models have been created by linking the 3D components of a BIM to project scheduling activities. However, traditional 4D models have limitations, as BIM primarily provides information about 3D components. In [36], an interface system was introduced to enhance the benefits of BIM in 4D applications, take advantage of its capabilities to reduce material requirements (e.g., steel, molds, concrete), and support site-level activity simulation, ultimately contributing to project scheduling. The proposed system facilitates information exchange among different software packages, allowing for improved operation simulations that consider uncertain task durations and competing resource needs. This enables the evaluation of various resource allocation strategies to create an optimal construction plan. The resulting project schedule is then linked to the 3D BIM components, creating an enhanced BIM-based 4D model [36].

To address safety concerns in construction schedules, the risk of accidental falls can be identified using automated safety checks in a 4D simulation application [37]. During BIM modeling, the tracking and management of various versions of the BIM model are crucial for all BIM engineers. Regular meetings are essential to discuss collaborative work, where BIM engineers submit progress and unfinished work to the

BIM manager before each meeting. The BIM manager updates and communicates the latest integrated collaborative work, addressing issues, such as corrupted BIM model files on the server side. To mitigate this problem, BIM engineers should daily backup the BIM model on the client side and save new BIM model files exclusively on the server side [38].

Achieving active control and dynamic optimization of project objectives through integrated information management is a key benefit of BIM technology-based cost management. This approach enables professional coordination in design, pipeline collision testing, energy-saving calculations, sustainability analyses, program simulations, construction process simulations, scaffold support system testing, schedule control, cost control, and safety management. Such dynamic optimization upgrades the quality of engineering design, construction, and maintenance management, improving work efficiency and the scientific management level, reflecting the dynamic optimization of information for project services [39]. During design development, challenges related to data collection and management often arise along with multidisciplinary collaboration issues. Design team members, even from the same discipline, often use different software tools and work simultaneously [40]. Adopting BIM improves design reviews, accelerates processes, and enhances conflict detection and collaboration, leading to more efficient work. The application of BIM facilitates the early detection of potential conflicts during the design phase.

Effective communication of progress and schedule variations is deemed a project management strategy to enable timely corrective actions. The 4D model in construction control changes colors to represent progressing activities over time, aiding understanding [42]. Construction progress directly impacts speed, cost, and quality, and the construction schedule is determined based on the main circuit, considering factors, such as time, space, technology, management, and resources [43]. Semantic search, utilizing ontological mapping, automates conflict resolution by identifying and negotiating solutions, reducing delays. Mobile applications play a crucial role in providing instant feedback and accelerating decision-making processes [44]. To improve the quality, schedule, and cost performance of off-site construction projects, BIM incorporates as-built information, pairing it with design information for project control. As-built information studies focus on quality control and evaluation of structural members [45]. In [46], a BIM-based early warning model was proposed for construction project cost and risk management, while BIM functions, namely "construction process simulation" and "4D description of construction progress" are widely used in prefabricated construction project management for post-production planning, resource utilization planning, and workspace congestion analysis [47]. BIM's knowledge management and storage function effectively address discrete information challenges during green building construction, applying construction information modeling to manage quality, schedule, cost, and safety in project supervision [48]. The former strives to realize the partial application of the supervision audit system in actual engineering projects [49].

Design consulting for phased implementation necessitates progress tracking and output deliverables for deployment. In the realm of BIM implementation, the tangible success of the application is significantly influenced by the seamless coordination among organizations and departments engaged in construction projects. The Specialized Construction Investment Project Management Board within the Ministry of Construction organizes weekly meetings with stakeholders. These gatherings serve as a platform to share insights, foster coordination, and align on the implementation process, progress expectations, and overarching goals. In addition, regular online meetings are held to ensure timely updates and maintain alignment with the project's progress and objectives [50].

BIM applications are used for bridge projects, emphasizing their utility in cost and time analysis and management. The cost estimation application accommodates both approximate and detailed cost estimates, offering flexibility with default or user-defined values. Diverse impact metrics are used to monitor project costs and progress throughout the construction implementation phase [51]. Cost and schedule represent two crucial pivotal objectives for construction projects, with risk management serving as the linchpin for their realization. In [52], a novel risk early warning method was introduced based on emerging BIM technology. In the face of design changes, construction managers traditionally spend time reviewing updated drawings and revising schedules, a process highly dependent on the manager's interpretation of the drawings. The introduction of BIM aims to improve efficiency through shared databases and precision through 4D model visualization. A 4D model is generated, combining the BIM model and the construction schedule, using 4D tools, such as Autodesk NavisWorks, offering a 3D model integrated with time-based geometry [53]. This 4D BIM acts as a tool for safety management and monitoring, contributing to the reduction of safety-related hazards. Secure processes are implemented to address the limited automation in 4D modeling [54]. The construction process is simulated, followed by adjustments and optimizations of the construction sequence and method, refining the construction process to obtain an optimal plan. Construction simulation encompasses progress, program simulation, site layout, and more, ensuring not only control over construction progress but also quality assurance to prevent delays [55].

Project management plays a key role in minimizing or eliminating risks associated with a project. Risks can be due to human activities or natural causes, some being foreseeable, while others are unforeseen. Project risks can significantly impact the project if not competently managed. Collaborative risk reduction is crucial for project teams, and various approaches are employed. BIM has emerged as a vital approach to mitigate project risks, playing a critical role in the project lifecycle stages, including planning, design, and construction management. BIM ensures comprehensive project monitoring through information management, facilitating efficient execution and milestone achievement. Despite BIM's effectiveness in various projects, its specific application in the construction industry remains somewhat unclear. In [56], the significance of BIM as a highly suitable method to reduce project risk was investigated. The substantial potential of BIM

for quality management in infrastructure projects has been recognized, showcasing successes in error detection and mitigation, obstacle anticipation and resolution, improved design solutions, improved communication with stakeholders, and optimized budget and schedule options [7, 57]. In the construction industry, the project management team's core responsibility is to complete projects within the specified budget, time, and quality identified in the contract documents [58]. The integration of BIM with advanced digital technologies offers the prospect of automating modern construction project management, improving productivity, and controlling waste and resource consumption [59]. In [60], the application of 4D visualization in construction was explored, bridging the gap between 4D research and BIM principles, particularly addressing the level of development [60].

Construction management leverages BIM for visual communication among key stakeholders, knowledge integration to enhance management expertise, and automation for improved efficiency. Despite these applications, BIM data in construction management are often utilized for simple and localized purposes. Traditional construction progress management is susceptible to disruptions from natural, objective, and subjective environments, leading to interruptions and blockages in the construction process. However, with the support of BIM models and BIM5D software, construction progress management anticipates future steps, tracks real-time progress, and provides insight into resource requirements, equipment needs, and capital utilization. It also facilitates the formation of comprehensive construction progress management modes, ensuring timely monitoring of quality and safety issues, defect recording, data integration, model linking, and prompt corrective actions [61]. Using the BIM model, defect managers examine information relevant to the management of reinforced concrete work defects, such as building geometry, materials, and project schedules. BIM technologies have revolutionized the construction industry by organizing relationships among stakeholders, offering benefits in design, procurement, construction, and maintenance processes, along with change control, repetition reduction, energy efficiency, health and safety, risk management, and quality [62]. BIM spans beyond 3D modeling, encompassing dimensions, such as 4D (time), 5D (cost), and even 6D (as-built activity). It integrates information and cost data and creates an as-built model for operational phases [63].

Addressing limiting factors in management involves adapting working processes to new technologies, exploring BIM's applications in cost, schedule, and safety management, developing guidelines and standards, and adopting integrated project delivery. Integrated project delivery reduces fragmentation and is more effective when combined with BIM than traditional methods [64]. Owner motivation remains a crucial factor in BIM application, emphasizing the importance of aligning with owner requirements [65]. BIM's applications in building design and construction planning are evolving rapidly, enabling the storage of physical and functional data in multiple dimensions for in-depth analysis and prediction [66-68]. Basic construction management practices use advanced computer tools for budgeting, scheduling, risk analysis, and buildability studies, providing accurate and high-speed

solutions [69]. The theoretical framework for BIM data in facility management assumes a seamless data exchange between BIM and facility management software [70]. Case studies are essential to understand the adoption of BIM and its impact on remote construction projects [71]. Testing and inspection schedules benefit from detailed planning using stored database information [72].

BIM enhances engineering work during the design stage by creating 3D models that integrate all building features. These models can be further ameliorated by linking to schedules (4D) and costs (5D), allowing better planning during the design phase and improved time and cost control, which is crucial for any construction organization [73]. Integrating quality information into BIM facilitates the creation of a shared quality model among stakeholders. Augmented Reality (AR) technology supports on-site inspections based on this quality model, providing a simplified integration with 4D BIM [74]. Construction schedules undergo modifications, and BIM provides a valuable tool for managing changes efficiently [75]. Visualizing risk management information improves the efficiency of dynamic risk management practitioners in terms of schedule, cost, and safety in challenging construction projects [76]. Traditional site inspection methods and manual progress tracking are being transformed, with quality management focusing on ensuring that construction specifications align with the design [77].

A. Quality Assurance

BIM serves as a repository for quality-related design and construction information within its models. Integration of mobile computing with BIM offers practical solutions for field workers to address real-time construction workmanship quality issues. This includes supervising on-site activities, tracking work progress, and checking project milestones. BIM models also facilitate the capture and management of quality checklists for each task [78]. Ongoing research aims to improve process efficiency in BIM-based quality control and inspection for safety management, focusing on areas, likee data interoperability and decision-making [79].

B. Definition and Applications

BIM is defined as a set of construction project management tools enabled by new technology. It encompasses the entire construction process, providing a 3D simulation of the project components linked to relevant information for planning, construction, operation, and decommissioning [80]. BIM acts as a complementary model to a project's technical information database, storing architectural designs, geometry, and engineering information. Infrastructure-integrated applications include earthworks simulation, schedule management, construction equipment management, walkability assessment, and facility management [81]. Quality control is a crucial element in project management, with various countries incorporating quality control requirements into BIM guidelines [82].

C. Schedule Management

Effective platforms for visually enhancing completion schedules during the construction phase are essential for general contractors. As the adoption of BIM becomes more

widespread, integrated schedule management with BIM becomes necessary for visual construction management. A BIM information model for project cost management, based on BIM platform communication, enables real-time information sharing through parameterized material and component information. The use of BIM from pre-construction to post-construction stages in building construction processes has been discussed, emphasizing its potential benefits in improving project quality, lifecycle data management, collaboration optimization, and progress management [49, 83-86].

D. Construction Progress Management

Traditional construction schedule management faces challenges due to computational demands and a lack of an intuitive platform, leading to inefficiencies in progress control. BIM-4D technology, coupled with the key chain method, is proposed to build an informative, intuitive, and dynamic construction progress control system. This system includes modules for 4D realistic model management, key chain monitoring of buffer zones, and early warning of progress. BIM aims to address the weaknesses in the overall schedule management process, providing a solid foundation for construction progress management and control [47].

E. Life Cycle Representation

4D BIM extends beyond geometry and material information, incorporating progress and cost-related data for construction projects. The building information model encapsulates geometries, geospatial relationships, quantities, characteristics of building elements, material inventory, cost estimates, and implementation schedules. This comprehensive model represents the entire life cycle of a building, ensuring consistency and facilitating easy data updates [87-89].

F. Demand Forecasting

BIM modules are used to calculate order due dates and forecast demand based on construction schedules. However, uncertainties in activity dates and times require advanced information and communication technologies to support cost management and the overall management of construction projects [90].

G. Challenges and Opportunities

A detailed analysis of current scheduling practices reveals weaknesses and unmet information needs. Integration of construction schedule planning within a coordinated BIM-based design process is proposed to address these challenges, eliminating redundancies and improving change management. However, existing software packages require adaptation to the collaborative, model-based scheduling approach, lacking adequate access to underlying data and efficient link mechanisms between product and process models [91].

Project time control involves monitoring the status of project activities, updating progress, and managing changes to the baseline schedule to ensure adherence to the plan. This process is essential to recognize deviations from the actual plan, allowing corrective and preventive measures to mitigate risks. The timely execution of the project is crucial for stakeholders, but project delays are prevalent worldwide, particularly in developing and underdeveloped countries [92].

The application of BIM extends to cost, quality, safety, engineering, and field management, significantly enhancing communication efficiency among project participants and allowing more effective cost and schedule control [93]. In [94], a BIM application was proposed for the cost and time management of bridge infrastructure using a 5D approach. The building information model encompasses geometry, spatial relationships, geographic information, quantity, characterization of building elements, cost estimates, material inventory, and project schedule. This model represents the entire life cycle of a building [95].

Succar's definition of BIM underscores its comprehensive complexity, including a software that goes beyond numerical modeling and data input to include tools and processes related to project management. BIM, viewed in its entirety, aligns with the field of structural project management, offering potential benefits to structural project operators by reducing documentation time and delivering favorable project outcomes. Specific documentation of BIM use in notable project scenarios, such as Heathrow Terminal 5 and the Walt Disney Concert Hall, is an aspect of BIM documentation [96]. Challenges in design management practices, coupled with the introduction of BIM, require new solutions. BIM technology addresses these challenges by creating component libraries to improve modeling efficiency, simulating the 4D process to improve site construction accuracy, and implementing quality management systems for analyzing problematic components [97-98]. The BIM model enables construction managers to promptly identify and address issues by marking their location within the model. Tracking verification and statistical analysis play a crucial role in effective control [99]. The introduction of BIM-based project management tools improves the development of realistic project-based classroom assignments for educators. BIM allows the design of class projects that simulate real-world project conditions, helping students learn various project management methods and optimize project plans [100]. BIM's data-rich nature facilitates model-based quantity reduction, cost estimation, scheduling simulation, and design coordination. Therefore, beyond teaching BIM in design education, it is crucial to educate students on applying BIM throughout the project life cycle and how to effectively manipulate and manage the model [101].

The Construction BIM-enabled Schedule Management (ConBIM-SM) system improves the visibility of updated 'Built' schedules in real-time for prefabricated construction, improving on-site efficiency [102]. Facility maintenance-related information can be updated in the facility model post-maintenance, preventing management errors related to staff verification [15]. The negative impacts of BIM integration are often attributed to project teams that do not optimize BIM usage, resistance from stakeholders, and frustrations with underutilizing the potential of BIM [103]. A continuously updated 4D building information model helps logistics coordinators determine the exact material requirements in each project stage, supporting efficient project planning [104]. These advances suggest a shift from the document model to the project-integrated database model, combining 3D CAD models with construction operations to visualize construction progress over time [105]. The use of 4D models, linking components

with activities, allows stakeholders to review the construction plan and the 3D CAD model for any day, week, or month of the project, providing visual intelligence for construction management [106].

III. DISCUSSION

Drawing on the synthesis of research findings presented in the overview of BIM applications in construction project progress management, it becomes evident that the utilization of BIM in managing project schedules offers numerous advantages. Insights from various studies underscore the significant potential of BIM evaluation, highlighting its widespread adoption in the application process. In Vietnam, BIM is gradually finding applications in project progress management, particularly in items associated with foreign-invested projects. Multinational companies and their branches, as well as various agencies and organizations, have demonstrated commendable proficiency in using BIM for project information management. The integration of the BIM process into project scheduling proves instrumental in mitigating the shortcomings associated with traditional manual scheduling methods. The time-consuming nature of manual creation and updating, error-prone tendencies, and resultant sub-optimal planning and progress, which are major contributors to project delays, can be resolved through BIM application. The pervasive issue of schedule delays, fraught with known and unknown risks, has prompted the development of multiple theories and models. However, addressing this persistent challenge demands continued efforts in the face of ongoing project time failures.

Beyond the rudimentary role of BIM in information modeling, its significance extends to considerable dimensions, including 4D (time), 5D (cost), and even 6D (as-built operations). In 4D, BIM establishes links between information and data in 3D object models, facilitating simulation analyses of construction operations. In 5D, BIM integrates this information with cost data, encompassing quantity, schedule, and price. In 6D, BIM represents an as-built model applicable in the operational phases of a facility. BIM's application enables the seamless update of information, ensuring consistency with all alterations. Enhanced interaction among project management stakeholders, coupled with regular conflict checks, guarantees continuous monitoring and updates from project initiation. This streamlined process aids in error reporting and progress tracking while it ensures project transparency during evaluations.

Despite the commendable aspects of BIM, certain limitations and risks persist due to various factors. A crucial factor is the national context, where each country's trends, policies, and regulatory frameworks impact the management landscape, including the construction sector. Implementing the BIM model requires extensive updates, exchanges, and changes in the approval and implementation processes. On the business front, adopting the BIM process incurs operating costs, equipment expenses, training fees, and adjustments in work procedures and organizational structures. The high costs associated with BIM software and skills training pose risks to businesses, particularly concerning mass adoption, where experienced individuals in this field are scarce. The human

factor emerges as a fundamental prerequisite, demanding continuous learning and knowledge exchange between workers. With technology advancing each year, there is an increasing challenge in meeting the demand for work experience. New graduates may struggle to keep up with evolving technologies, while established employees may resist embracing change and updating their skills. This human factor poses a significant consideration when contemplating the application of the BIM process in construction project schedule management.

IV. CONCLUSIONS AND RECOMMENDATIONS

The delineation of advantages and drawbacks underscores the global proliferation of BIM as a ubiquitous language in construction management. However, certain factors, such as the country's level of development, applicable standards, operational methodologies, and technological considerations, continue to pose challenges and elicit dissatisfaction. To effectively address these issues, greater support is imperative from entities, namely the state, government, and businesses. Simultaneously, facing these challenges requires concerted efforts from both emerging talents and seasoned professionals well-versed in traditional management and application methods. Currently, enhancing the efficiency of BIM in construction management is highly dependent on the proactive involvement of government and state regulatory bodies. Their crucial role involves vigilant monitoring and the formulation of guidelines tailored to the specific prerequisites in Vietnam. These standardized processes are subsequently introduced and disseminated to authorities, serving as a foundational database for the broader implementation of the BIM process model, encompassing project progress management.

Integration of BIM software application training programs into university and college curricula is essential, offering advanced guidance on analytical thinking and the perspective required for handling digital data and project information at the graduate level. To optimize the effectiveness of BIM in managing the progress of construction projects, it is imperative to expand the dissemination of shared resources. These resources include standard documents, books, newspapers, and other guidelines on the BIM application process. Ensuring accessibility to high-quality and reputable resources for students, businesses, and implementation agencies is crucial. In addition to the common applications mentioned above, investors and design units should gradually promote the adoption of project information modeling. This involves adjusting and upgrading the equipment to ensure full integration and storage of the updated BIM application processes and project information management. Enterprises are encouraged to conduct proactive training sessions, widely disseminating guidance on the application of the BIM process to all personnel at various levels. Accessing insights from standard sources and journal articles or organizing seminars with the participation of esteemed domestic and foreign experts in BIM project management contributes to the comprehensive understanding and implementation of BIM processes.

REFERENCES

- [1] I. A. Bhatti, A. H. Abdullah, S. Nagapan, N. B. Bhatti, S. Sohu, and A. A. Jhatial, "Implementation of Building Information Modeling (BIM) in Pakistan Construction Industry," *Engineering, Technology & Applied Science Research*, vol. 8, no. 4, pp. 3199–3202, Aug. 2018, <https://doi.org/10.48084/etasr.2145>.
- [2] W. A. Hatem, A. M. Abd, and N. N. Abbas, "Motivation Factors for Adopting Building Information Modeling (BIM) in Iraq," *Engineering, Technology & Applied Science Research*, vol. 8, no. 2, pp. 2668–2672, Apr. 2018, <https://doi.org/10.48084/etasr.1860>.
- [3] T. W. Alvarenga, E. N. da Silva, and L. C. B. de B. Mello, "BIM and Lean Construction: The Evolution Obstacle in the Brazilian Civil Construction Industry," *Engineering, Technology & Applied Science Research*, vol. 7, no. 5, pp. 1904–1908, Oct. 2017, <https://doi.org/10.48084/etasr.1278>.
- [4] J. Ratajczak, M. Riedl, and D. T. Matt, "BIM-based and AR Application Combined with Location-Based Management System for the Improvement of the Construction Performance," *Buildings*, vol. 9, no. 5, May 2019, Art. no. 118, <https://doi.org/10.3390/buildings9050118>.
- [5] A. K. Singh, A. Pal, P. Kumar, J. J. Lin, and S.-H. Hsieh, "Prospects of Integrating BIM and NLP for Automatic Construction Schedule Modelling," presented at the 40th International Symposium on Automation and Robotics in Construction, Chennai, India, Jul. 2023, <https://doi.org/10.22260/ISARC2023/0034>.
- [6] M. Sami Ur Rehman, M. J. Thaheem, A. R. Nasir, and K. I. A. Khan, "Project schedule risk management through building information modelling," *International Journal of Construction Management*, vol. 22, no. 8, pp. 1489–1499, May 2022, <https://doi.org/10.1080/15623599.2020.1728606>.
- [7] P. C. Suermann and R. R. A. Issa, "Evaluating industry perceptions of building information modelling (BIM) impact on construction," *Journal of Information Technology in Construction (ITcon)*, vol. 14, no. 37, pp. 574–594, Aug. 2009.
- [8] Z. Song, G. Shi, J. Wang, H. Wei, T. Wang, and G. Zhou, "Research on management and application of tunnel engineering based on BIM technology," *Journal of Civil Engineering and Management*, vol. 25, no. 8, pp. 785–797, Sep. 2019, <https://doi.org/10.3846/jcem.2019.11056>.
- [9] J. Zou, R. Du, F. Yang, and F. Zhang, "BIM Information Technology Application in Construction Management Research," *Hydraulic and Civil Engineering Technology VIII*, pp. 376–387, 2023, <https://doi.org/10.3233/ATDE230738>.
- [10] A. Behúnová, T. Mandičák, M. Behún, and P. Mésároš, "The Building Information Modelling Through Information Technology and Impacts on Selected Circular Economy Performance Indicators of Construction Projects," *Mobile Networks and Applications*, Nov. 2023, <https://doi.org/10.1007/s11036-023-02268-7>.
- [11] D. T. Hai, N. N. Tu, N. V. Anh, and N. T. T. Ha, "BIM Application in Construction Waste Risk Management," in *E3S Web of Conferences - XII International Scientific and Practical Forum "Environmentally Sustainable Cities and Settlements: Problems and Solutions" (ESCP-2023)*, 2023, vol. 403, Art. no. 02004, <https://doi.org/10.1051/e3sconf/202340302004>.
- [12] A. GhaffarianHoseini *et al.*, "Application of nD BIM Integrated Knowledge-based Building Management System (BIM-IBBMS) for inspecting post-construction energy efficiency," *Renewable and Sustainable Energy Reviews*, vol. 72, pp. 935–949, May 2017, <https://doi.org/10.1016/j.rser.2016.12.061>.
- [13] L. Chen and H. Luo, "A BIM-based construction quality management model and its applications," *Automation in Construction*, vol. 46, pp. 64–73, Oct. 2014, <https://doi.org/10.1016/j.autcon.2014.05.009>.
- [14] J. Zhang, Y. Long, S. Lv, and Y. Xiang, "BIM-enabled Modular and Industrialized Construction in China," *Procedia Engineering*, vol. 145, pp. 1456–1461, Jan. 2016, <https://doi.org/10.1016/j.proeng.2016.04.183>.
- [15] Y. C. Su, Y. C. Lee, and Y. C. Lin, "Enhancing maintenance management using building information modeling in facilities management: 28th International Symposium on Automation and Robotics in Construction, ISARC 2011," Seoul, Korea (South), 2011, pp. 752–757, <https://doi.org/10.22260/isarc2011/0140>.

- [16] Y. Vacanas, K. Themistocleous, A. Agapiou, and D. Hadjimitsis, "Building Information Modelling (BIM) and Unmanned Aerial Vehicle (UAV) technologies in infrastructure construction project management and delay and disruption analysis," in *Third International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2015)*, Paphos, Cyprus, Jun. 2015, vol. 9535, pp. 93–103, <https://doi.org/10.1117/12.2192723>.
- [17] Y. Fang, Y. K. Cho, S. Zhang, and E. Perez, "Case Study of BIM and Cloud-Enabled Real-Time RFID Indoor Localization for Construction Management Applications," *Journal of Construction Engineering and Management*, vol. 142, no. 7, Jul. 2016, Art. no. 05016003, [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001125](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001125).
- [18] Á. Moreno Bazán, M. G. Alberti, A. Arcos Álvarez, and J. A. Trigueros, "New Perspectives for BIM Usage in Transportation Infrastructure Projects," *Applied Sciences*, vol. 10, no. 20, Jan. 2020, Art. no. 7072, <https://doi.org/10.3390/app10207072>.
- [19] Z. Song, G. Shi, J. Wang, H. Wei, T. Wang, and G. Zhou, "Research on management and application of tunnel engineering based on BIM technology," *Journal of Civil Engineering and Management*, vol. 25, no. 8, pp. 785–797, Sep. 2019, <https://doi.org/10.3846/jcem.2019.11056>.
- [20] F. Abbondati, S. A. Biancardo, S. Palazzo, F. S. Capaldo, and N. Viscione, "I-BIM for existing airport infrastructures," *Transportation Research Procedia*, vol. 45, pp. 596–603, Jan. 2020, <https://doi.org/10.1016/j.trpro.2020.03.052>.
- [21] S. Rokooei, "Building Information Modeling in Project Management: Necessities, Challenges and Outcomes," *Procedia - Social and Behavioral Sciences*, vol. 210, pp. 87–95, Dec. 2015, <https://doi.org/10.1016/j.sbspro.2015.11.332>.
- [22] M. N. Kocakaya, E. Namli, and Ü. Işıkdağ, "Building Information Management (BIM), A New Approach to Project Management," *Journal of Sustainable Construction Materials and Technologies*, vol. 4, no. 1, pp. 323–332, Oct. 2019.
- [23] Y. Rui, L. Yaik-Wah, and T. C. Siang, "Construction Project Management Based on Building Information Modeling (BIM)," *Civil Engineering and Architecture*, vol. 9, no. 6, pp. 2055–2061, Oct. 2021, <https://doi.org/10.13189/cea.2021.090633>.
- [24] J. Wong and J. Yang, "Research and application of building information modelling (BIM) in the architecture, engineering and construction (AEC) industry: a review and direction for future research," in *Proceedings of the 6th International Conference on Innovation in Architecture, Engineering and Construction (AEC)*, Laoughborough, UK, 2010, pp. 356–365.
- [25] S. Mohd and A. A. Latiffi, "Building Information Modeling (BIM) Application in Construction Planning," presented at the Seventh International Conference on Construction in the 21st Century (CITC-VII) "Challenges in Innovation, Integration and Collaboration in Construction & Engineering," Bangkok, Thailand, Dec. 2013.
- [26] M. Scheffer, H. Mattern, and M. König, "BIM Project Management," in *Building Information Modeling: Technology Foundations and Industry Practice*, A. Borrmann, M. König, C. Koch, and J. Beetz, Eds. Cham, Switzerland: Springer International Publishing, 2018, pp. 235–249.
- [27] A. A. Latiffi, S. Mohd, N. Kasim, and M. S. Fathi, "Building Information Modeling (BIM) Application in Malaysian Construction Industry," *International Journal of Construction Engineering and Management*, vol. 2, no. 4A, pp. 1–6, 2013.
- [28] J. D. Goedert and P. Meadati, "Integrating Construction Process Documentation into Building Information Modeling," *Journal of Construction Engineering and Management*, vol. 134, no. 7, pp. 509–516, Jul. 2008, [https://doi.org/10.1061/\(ASCE\)0733-9364\(2008\)134:7\(509\)](https://doi.org/10.1061/(ASCE)0733-9364(2008)134:7(509)).
- [29] H. Y. Chou and J. B. Yang, "Preliminary Evaluation of BIM-based Approaches for Schedule Delay Analysis," *IOP Conference Series: Materials Science and Engineering*, vol. 245, no. 6, Jul. 2017, Art. no. 062048, <https://doi.org/10.1088/1757-899X/245/6/062048>.
- [30] N. Lee and D. A. Hollar, "Probing BIM Education in Construction Engineering and Management Programs Using Industry Perceptions," in *49th ASC Annual International Conference Proceedings*, San Louis Obispo, CA, USA, 2013.
- [31] S. Azhar, M. Khalfan, and T. Maqsood, "Building Information Modeling (BIM): Now and beyond," *The Australasian Journal of Construction Economics and Building*, vol. 12, no. 4, pp. 15–28, Aug. 2020, <https://doi.org/10.3316/informit.013120167780649>.
- [32] P. T. Nguyen, T. A. Nguyen, Q. L. H. T. T. Nguyen, and V. D. B. Huynh, "Application of SWOT for construction company quality management using building information modelling," *Journal of Mechanics of Continua and Mathematical Sciences*, vol. 13, no. 5, pp. 25–33, Dec. 2018, <https://doi.org/10.26782/jmms.2018.12.00003>.
- [33] R. Grys, "Implementation of Building Information Modelling (BIM) on Public Infrastructure and Building Projects in Qatar," in *International Conference on Civil Infrastructure and Construction (CIC 2023)*, 2023, pp. 180–188, <https://doi.org/10.29117/cic.2023.0028>.
- [34] D.-J. Gibbs, S. Emmitt, W. Lord, and K. Ruikar, "BIM and construction contracts – CPC 2013's approach," *Proceedings of the Institution of Civil Engineers - Management, Procurement and Law*, vol. 168, no. 6, pp. 285–293, Dec. 2015, <https://doi.org/10.1680/jmapl.14.00045>.
- [35] G. Malacarne, G. Toller, C. Marcher, M. Riedl, and D. T. Matt, "Investigating benefits and criticisms of BIM for construction scheduling in SMES: An Italian case study," *International Journal of Sustainable Development and Planning*, vol. 13, no. 01, pp. 139–150, Jan. 2018, <https://doi.org/10.2495/SDP-V13-N1-139-150>.
- [36] W. C. Wang, S. W. Weng, S. H. Wang, and C. Y. Chen, "Integrating building information models with construction process simulations for project scheduling support," *Automation in Construction*, vol. 37, pp. 68–80, Jan. 2014, <https://doi.org/10.1016/j.autcon.2013.10.009>.
- [37] E. M. Wetzel and W. Y. Thabet, "The use of a BIM-based framework to support safe facility management processes," *Automation in Construction*, vol. 60, pp. 12–24, Dec. 2015, <https://doi.org/10.1016/j.autcon.2015.09.004>.
- [38] Y. C. Lin and H.-H. Yang, "A Framework for Collaboration Management of BIM Model Creation in Architectural Projects," *Journal of Asian Architecture and Building Engineering*, vol. 17, no. 1, pp. 39–46, Jan. 2018, <https://doi.org/10.3130/jaabe.17.39>.
- [39] Y. L. Yin and K. Qian, "Construction Project Cost Management Based on BIM Technology," *Applied Mechanics and Materials*, vol. 357–360, pp. 2147–2152, 2013, <https://doi.org/10.4028/www.scientific.net/AMM.357-360.2147>.
- [40] Y. Arayici and G. F. Aouad, "Building information modelling (BIM) for construction lifecycle management," in *Construction and Building: Design, Materials, and Techniques*, New York, NY, USA: Nova Science Publishers, 2010, pp. 99–118.
- [41] S. N. A. M. Noor, S. R. Junaidi, and M. K. A. Ramly, "Adoption of Building Information Modelling (BIM): Factors Contribution and Benefits," *Information System and Technology Management*, vol. 3, no. 10, pp. 47–63, Dec. 2018.
- [42] D. M. Brito and E. A. M. Ferreira, "Strategies for Representation and Analyses of 4D Modeling Applied to Construction Project Management," *Procedia Economics and Finance*, vol. 21, pp. 374–382, Jan. 2015, [https://doi.org/10.1016/S2212-5671\(15\)00189-6](https://doi.org/10.1016/S2212-5671(15)00189-6).
- [43] Z. H. Han, Z. K. Wang, C. Gao, M. X. Wang, and S. T. Li, "Application of GIS and BIM Integration Technology in Construction Management," *IOP Conference Series: Earth and Environmental Science*, vol. 526, no. 1, Mar. 2020, Art. no. 012161, <https://doi.org/10.1088/1755-1315/526/1/012161>.
- [44] S. Lijun and D. K. H. Chua, "Application of Building Information Modeling (BIM) and Information Technology (IT) for Project Collaboration," in *Proceedings of the 2011 (2nd) International Conference on Engineering, Project, and Production Management*, Sep. 2011, pp. 67–76, <https://doi.org/10.32738/CEPPM.201109.0008>.
- [45] X. Yin, H. Liu, Y. Chen, and M. Al-Hussein, "Building information modelling for off-site construction: Review and future directions," *Automation in Construction*, vol. 101, pp. 72–91, May 2019, <https://doi.org/10.1016/j.autcon.2019.01.010>.
- [46] F. H. Abanda, A. M. Musa, P. Clermont, J. H. M. Tah, and A. H. Oti, "A BIM-based framework for construction project scheduling risk management," *International Journal of Computer Aided Engineering and Technology*, vol. 12, no. 2, pp. 182–218, Jan. 2020, <https://doi.org/10.1504/IJCAET.2020.105575>.

- [47] S. Zhang, Z. Li, T. Li, and M. Yuan, "A holistic literature review of building information modeling for prefabricated construction," *Journal of Civil Engineering and Management*, vol. 27, no. 7, pp. 485–499, Oct. 2021, <https://doi.org/10.3846/jcem.2021.15600>.
- [48] Y. Cao, S. N. Kamaruzzaman, and N. M. Aziz, "Green Building Construction: A Systematic Review of BIM Utilization," *Buildings*, vol. 12, no. 8, Aug. 2022, Art. no. 1205, <https://doi.org/10.3390/buildings12081205>.
- [49] H. Yan, L. Wang, and X. Jiang, "Dynamic Control of Integrated Project Management System based on Engineering Projects," *Computer-Aided Design and Applications*, vol. 19, no. S4, pp. 90–101, Jul. 2021, <https://doi.org/10.14733/cadaps.2022.S4.90-101>.
- [50] T.-Q. Nguyen and Q. V. Dao, "A Case Study of BIM Application in a Public Construction Project Management Unit in Vietnam: Lessons Learned and Organizational Changes," *Engineering Journal*, vol. 25, no. 7, pp. 177–192, Jul. 2021, <https://doi.org/10.4186/ej.2021.25.7.177>.
- [51] S. M. E. Sepasgozar, A. M. Costin, R. Karimi, S. Shirowzhan, E. Abbasian, and J. Li, "BIM and Digital Tools for State-of-the-Art Construction Cost Management," *Buildings*, vol. 12, no. 4, Apr. 2022, Art. no. 396, <https://doi.org/10.3390/buildings12040396>.
- [52] C. Sun, Q. Man, and Y. Wang, "Study on BIM-based construction project cost and schedule risk early warning," *Journal of Intelligent & Fuzzy Systems*, vol. 29, no. 2, pp. 469–477, Jan. 2015, <https://doi.org/10.3233/IFS-141178>.
- [53] A. Nadeem, A. K. D. Wong, G. Akhanova, S. Azhar, and S. N. Wong, "Application of Building Information Modeling (BIM) in Site Management—Material and Progress Control," in *Proceedings of the 21st International Symposium on Advancement of Construction Management and Real Estate*, 2018, pp. 289–297, https://doi.org/10.1007/978-981-10-6190-5_26.
- [54] J. Jupp, "4D BIM for Environmental Planning and Management," *Procedia Engineering*, vol. 180, pp. 190–201, Jan. 2017, <https://doi.org/10.1016/j.proeng.2017.04.178>.
- [55] J. Xu, "Research on Application of BIM 5D Technology in Central Grand Project," *Procedia Engineering*, vol. 174, pp. 600–610, Jan. 2017, <https://doi.org/10.1016/j.proeng.2017.01.194>.
- [56] H. M. Alzoubi, "BIM as a tool to optimize and manage project risk management," *International Journal of Mechanical Engineering*, vol. 7, no. 1, 2022.
- [57] N. Lee, T. Salama, and G. Wang, "Building Information Modeling for Quality Management in Infrastructure Construction Projects," pp. 65–72, Jun. 2014, <https://doi.org/10.1061/9780784413616.009>.
- [58] M. Khanzadi, M. Sheikhhoshkar, and S. Banihashemi, "BIM applications toward key performance indicators of construction projects in Iran," *International Journal of Construction Management*, vol. 20, no. 4, pp. 305–320, Jul. 2020, <https://doi.org/10.1080/15623599.2018.1484852>.
- [59] C. Chen and L. Tang, "BIM-based integrated management workflow design for schedule and cost planning of building fabric maintenance," *Automation in Construction*, vol. 107, Nov. 2019, Art. no. 102944, <https://doi.org/10.1016/j.autcon.2019.102944>.
- [60] A. Guerriero, S. Kubicki, F. Berroir, and C. Lemaire, "BIM-enhanced collaborative smart technologies for LEAN construction processes," in *2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, Madeira, Portugal, Jun. 2017, pp. 1023–1030, <https://doi.org/10.1109/ICE.2017.8279994>.
- [61] W. Jeong, S. Chang, J. Son, and J.-S. Yi, "BIM-Integrated Construction Operation Simulation for Just-In-Time Production Management," *Sustainability*, vol. 8, no. 11, Nov. 2016, Art. no. 1106, <https://doi.org/10.3390/su8111106>.
- [62] X. Li, J. Xu, and Q. Zhang, "Research on Construction Schedule Management Based on BIM Technology," *Procedia Engineering*, vol. 174, pp. 657–667, Jan. 2017, <https://doi.org/10.1016/j.proeng.2017.01.214>.
- [63] K. Oh-Seong and C. Park, "Application utilizing BIM, Mobile-Computing and Augmented Reality Technology," presented at the Australasian Conference on Innovative Technologies in Construction, Wuhan, China, Jan. 2012.
- [64] H. Aladag, G. Demirdögen, and Z. Isık, "Building Information Modeling (BIM) Use in Turkish Construction Industry," *Procedia Engineering*, vol. 161, pp. 174–179, Jan. 2016, <https://doi.org/10.1016/j.proeng.2016.08.520>.
- [65] P. Smith, "BIM & the 5D Project Cost Manager," *Procedia - Social and Behavioral Sciences*, vol. 119, pp. 475–484, Mar. 2014, <https://doi.org/10.1016/j.sbspro.2014.03.053>.
- [66] C. Sun, S. Jiang, M. J. Skibniewski, Q. Man, and L. Shen, "A literature review of the factors limiting the application of BIM in the construction industry," *Technological and Economic Development of Economy*, vol. 23, no. 5, pp. 764–779, Sep. 2017, <https://doi.org/10.3846/20294913.2015.1087071>.
- [67] G. Tekbas and G. Guven, "BIM-Based Automated Safety Review for Fall Prevention," in *Advances in Building Information Modeling*, Istanbul, Turkey, 2020, pp. 80–90, https://doi.org/10.1007/978-3-030-42852-5_7.
- [68] M. Parsamehr, U. S. Perera, T. C. Dodanwala, P. Perera, and R. Ruparathna, "A review of construction management challenges and BIM-based solutions: perspectives from the schedule, cost, quality, and safety management," *Asian Journal of Civil Engineering*, vol. 24, no. 1, pp. 353–389, Jan. 2023, <https://doi.org/10.1007/s42107-022-00501-4>.
- [69] M. Yalcinkaya and D. Ardit, "Building Information Modeling (BIM) and the Construction Management Body of Knowledge," in *Product Lifecycle Management for Society*, Nantes, France, 2013, pp. 619–629, https://doi.org/10.1007/978-3-642-41501-2_61.
- [70] K. Kensek, "BIM Guidelines Inform Facilities Management Databases: A Case Study over Time," *Buildings*, vol. 5, no. 3, pp. 899–916, Sep. 2015, <https://doi.org/10.3390/buildings5030899>.
- [71] Y. Arayici, C. Egbu, and P. Coates, "Building information modelling (BIM) implementation and remote construction projects: Issues, Challenges and Critiques," *Journal of Information Technology in Construction (ITcon)*, vol. 17, no. 5, pp. 75–92, May 2012.
- [72] Z. Ma, S. Cai, N. Mao, Q. Yang, J. Feng, and P. Wang, "Construction quality management based on a collaborative system using BIM and indoor positioning," *Automation in Construction*, vol. 92, pp. 35–45, Aug. 2018, <https://doi.org/10.1016/j.autcon.2018.03.027>.
- [73] A. Jrade and J. Lessard, "An Integrated BIM System to Track the Time and Cost of Construction Projects: A Case Study," *Journal of Construction Engineering*, vol. 2015, Nov. 2015, Art. no. e579486, <https://doi.org/10.1155/2015/579486>.
- [74] M. Mirshokraei, C. I. De Gaetani, and F. Migliaccio, "A Web-Based BIM–AR Quality Management System for Structural Elements," *Applied Sciences*, vol. 9, no. 19, Jan. 2019, Art. no. 3984, <https://doi.org/10.3390/app9193984>.
- [75] B. Dasović, M. Galić, and U. Klanšek, "A Survey on Integration of Optimization and Project Management Tools for Sustainable Construction Scheduling," *Sustainability*, vol. 12, no. 8, Jan. 2020, Art. no. 3405, <https://doi.org/10.3390/su12083405>.
- [76] Y. Zou, A. Kiviniemi, and S. W. Jones, "A review of risk management through BIM and BIM-related technologies," *Safety Science*, vol. 97, pp. 88–98, Aug. 2017, <https://doi.org/10.1016/j.ssci.2015.12.027>.
- [77] M. Al-Adhami, "Extended Reality Approach for Construction Quality Control," presented at the CIB World Building Congress 2019, Hong Kong, China, Jun. 2019.
- [78] H. Y. Chong, R. Lopez, J. Wang, X. Wang, and Z. Zhao, "Comparative Analysis on the Adoption and Use of BIM in Road Infrastructure Projects," *Journal of Management in Engineering*, vol. 32, no. 6, Nov. 2016, Art. no. 05016021, [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000460](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000460).
- [79] S. Park and I. Kim, "BIM-Based quality control for safety issues in the design and construction phases," *Archnet-IJAR*, vol. 9, no. 3, pp. 111–129, 2015, <https://doi.org/10.26687/archnet-ijar.v9i3.881>.
- [80] K. Hannele, M. Reijo, M. Tarja, P. Sami, K. Jenni, and R. Teija, "Expanding uses of building information modeling in life-cycle construction projects," *Work*, vol. 41, no. Suppl. 1, pp. 114–119, Jan. 2012, <https://doi.org/10.3233/WOR-2012-0144-114>.
- [81] Y. C. Lin, J. X. Chang, and Y. C. Su, "Developing construction defect management system using BIM technology in quality inspection,"

- Journal of Civil Engineering and Management*, vol. 22, no. 7, pp. 903–914, Oct. 2016, <https://doi.org/10.3846/13923730.2014.928362>.
- [82] Z. Ma and Y. Ren, "Integrated Application of BIM and GIS: An Overview," *Procedia Engineering*, vol. 196, pp. 1072–1079, Jan. 2017, <https://doi.org/10.1016/j.proeng.2017.08.064>.
- [83] J. Choi, S. Lee, and I. Kim, "Development of Quality Control Requirements for Improving the Quality of Architectural Design Based on BIM," *Applied Sciences*, vol. 10, no. 20, Jan. 2020, Art. no. 7074, <https://doi.org/10.3390/app10207074>.
- [84] H. P. Tserng, S. P. Ho, and S. H. Jan, "Developing BIM-assisted as-built schedule management system for general contractors," *Journal of Civil Engineering and Management*, vol. 20, no. 1, pp. 47–58, Jan. 2014, <https://doi.org/10.3846/13923730.2013.851112>.
- [85] Y. C. Lin, Y. P. Chen, W. T. Huang, and C. C. Hong, "Development of BIM Execution Plan for BIM Model Management during the Pre-Operation Phase: A Case Study," *Buildings*, vol. 6, no. 1, Mar. 2016, Art. no. 8, <https://doi.org/10.3390/buildings6010008>.
- [86] S. Zhang and C. Liang, "Research on Construction Schedule Control Based on Critical Chain Method and BIM," *Journal of Applied Science and Engineering Inno*, vol. 5, no. 2, pp. 47–50, 2018.
- [87] Y. Deng, V. J. L. Gan, M. Das, J. C. P. Cheng, and C. Anumba, "Integrating 4D BIM and GIS for Construction Supply Chain Management," *Journal of Construction Engineering and Management*, vol. 145, no. 4, Apr. 2019, Art. no. 04019016, [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001633](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001633).
- [88] R. Masood, M. K. N. Kharal, and A. R. Nasir, "Is BIM Adoption Advantageous for Construction Industry of Pakistan?," *Procedia Engineering*, vol. 77, pp. 229–238, Jan. 2014, <https://doi.org/10.1016/j.proeng.2014.07.021>.
- [89] W. N. F. Wan Abdul Basir, Z. Majid, U. Ujang, and A. Chong, "Integration of GIS and BIM Techniques in Construction Project Management - A Review," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-4-W9, pp. 307–316, Oct. 2018, <https://doi.org/10.5194/isprs-archives-XLII-4-W9-307-2018>.
- [90] P. Mesároš and T. Mandičák, "Exploitation and Benefits of BIM in Construction Project Management," *IOP Conference Series: Materials Science and Engineering*, vol. 245, no. 6, Jul. 2017, Art. no. 062056, <https://doi.org/10.1088/1757-899X/245/6/062056>.
- [91] J. Tulke, M. Nour, and K. Beucke, "A Dynamic Framework for Construction Scheduling based on BIM using IFC," in *17th IABSE Congress: Creating and Renewing Urban Structures – Tall Buildings, Bridges and Infrastructure*, Chicago, IL, USA, 2008, pp. 158–159, <https://doi.org/10.2749/222137908796292119>.
- [92] M. M. Tahir, A. N. Haron, N. A. Alias, I. B. Mohammed, and D. L. Baba, "Improving Cost and Time Control in Construction using Building Information Model (BIM): A Review," *Pertanika Journal of Science and Technology*, vol. 26, no. 1, 2018.
- [93] X. Pan, B. Zhong, X. Wang, and R. Xiang, "Text mining-based patent analysis of BIM application in construction," *Journal of Civil Engineering and Management*, vol. 27, no. 5, pp. 303–315, Jun. 2021, <https://doi.org/10.3846/jcem.2021.14907>.
- [94] S. Kaewunruen, J. Sresakoolchai, and Z. Zhou, "Sustainability-Based Lifecycle Management for Bridge Infrastructure Using 6D BIM," *Sustainability*, vol. 12, no. 6, Jan. 2020, Art. no. 2436, <https://doi.org/10.3390/su12062436>.
- [95] S. M. Khoshnava, A. Ahankoob, R. Rostami, and C. Preece, "Potential Application of BIM in Construction Dispute and Conflict," presented at the Management in Construction Research Association (MiCRA) Postgraduate Conference, Dec. 2012.
- [96] M. Abazid, H. Gökçekuş, and T. Çelik, "Study of the Quality Concepts Implementation in the Construction of Projects in Saudi Arabia by using Building Information Modelling (BIM)," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, no. 3, pp. 84–87, 2019.
- [97] M. Tauriainen, P. Marttinen, B. Dave, and L. Koskela, "The Effects of BIM and Lean Construction on Design Management Practices," *Procedia Engineering*, vol. 164, pp. 567–574, Jan. 2016, <https://doi.org/10.1016/j.proeng.2016.11.659>.
- [98] S. Chao, Z. Haoli, Y. Guogang, W. Shuangshuang, and X. Yidong, "Research on construction quality control of prefabricated concrete bridge based on BIM technology," *IOP Conference Series: Earth and Environmental Science*, vol. 643, no. 1, Jan. 2021, Art. no. 012063, <https://doi.org/10.1088/1755-1315/643/1/012063>.
- [99] R.-R. Dong, "The Application of BIM Technology in Building Construction Quality Management and Talent Training," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 13, no. 7, pp. 4311–4317, Jun. 2017, <https://doi.org/10.12973/eurasia.2017.00860a>.
- [100] F. Peterson, T. Hartmann, R. Fruchter, and M. Fischer, "Teaching construction project management with BIM support: Experience and lessons learned," *Automation in Construction*, vol. 20, no. 2, pp. 115–125, Mar. 2011, <https://doi.org/10.1016/j.autcon.2010.09.009>.
- [101] R. Klinc, "Project-based learning in a building information modeling for construction management course," *Journal of Information Technology in Construction (ITcon)*, vol. 21, no. 11, pp. 164–176, Jul. 2016.
- [102] J. Matthews, P. E. D. Love, S. Heinemann, R. Chandler, C. Rumsey, and O. Olatunji, "Real time progress management: Re-engineering processes for cloud-based BIM in construction," *Automation in Construction*, vol. 58, pp. 38–47, Oct. 2015, <https://doi.org/10.1016/j.autcon.2015.07.004>.
- [103] D. Bryde, M. Broquetas, and J. M. Volm, "The project benefits of Building Information Modelling (BIM)," *International Journal of Project Management*, vol. 31, no. 7, pp. 971–980, Oct. 2013, <https://doi.org/10.1016/j.ijproman.2012.12.001>.
- [104] K. Whitlock, F. H. Abanda, M. B. Manjia, C. Pettang, and G. E. Nkeng, "4D BIM for Construction Logistics Management," *CivilEng*, vol. 2, no. 2, pp. 325–348, Jun. 2021, <https://doi.org/10.3390/civileng2020018>.
- [105] A. Fazli, S. Fathi, M. H. Enferadi, M. Fazli, and B. Fathi, "Appraising Effectiveness of Building Information Management (BIM) in Project Management," *Procedia Technology*, vol. 16, pp. 1116–1125, Jan. 2014, <https://doi.org/10.1016/j.protcy.2014.10.126>.
- [106] J. Nissilä, R. Heikkilä, I. Romo, M. Malaska, and T. Aho, "BIM based schedule control for precast concrete supply chain," in *31st International Symposium on Automation and Robotics in Construction and Mining, ISARC 2014 - Proceedings*, 2014, pp. 667–671.

Efficient Energy Management with Emphasis on EV Charging/Discharging Strategy

Habib Kraiem

Department of Electrical Engineering, College of Engineering, Northern Border University, Saudi Arabia
| Processes, Energy, Environment and Electrical Systems (Code: LR18ES34), National Engineering
School of Gabes, University of Gabes, Tunisia
alhabeeb.kareem@nbu.edu.sa (corresponding author)

Wiem Gadri

Department of Mathematics, College of Science, Northern Border University, Saudi Arabia
wiem.gadri@nbu.edu.sa

Aymen Flah

Processes, Energy, Environment, and Electrical Systems (code: LR18ES34), National Engineering
School of Gabes, University of Gabes, Tunisia | MEU Research Unit, Middle East University, Amman,
Jordan | College of Engineering, University of Business and Technology (UBT), Jeddah 21448, Saudi
Arabia | Private Higher School of Applied Sciences and Technologies of Gabes, University of Gabes,
Tunisia | Applied Science Research Center, Applied Science Private University, Amman, 11931, Jordan
flahaymening@yahoo.fr

Received: 24 December 2023 | Revised: 8 January 2024 | Accepted: 14 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6807>

ABSTRACT

Leveraging the Vehicle-to-Grid (V2G) concept, this research explores how a decentralized energy reserve from hybrid electric vehicles can enhance the power system, particularly in large-scale implementations. The study introduces a V2G solution designed for effective microgrid frequency control over a full day. Targeting a scenario with minimal usage, typically in spring or fall, the microgrid is scaled to represent a community of 2000 homes. This is exemplified by integrating 500 Electric Vehicles (EVs) based on a 1:4 vehicle-to-household ratio, reflecting a plausible future scenario. The research conducts a comprehensive examination of the microgrid's voltage, current, and active power. By synchronizing the management of diesel and Renewable Energy Source (RES) generation, power transactions, and EV generation, the microgrid's frequency is effectively regulated through V2G devices adjusting load demand. The implemented V2G-enriched microgrid demonstrates improved energy management and mitigates the inconsistencies and fluctuations inherent in RES power generation, showing notable performance enhancements. In various operational contexts, system parameter fluctuations have been analyzed, revealing that deviations are maintained below a 5% threshold.

Keywords-Vehicle-to-Grid (V2G); power system enhancement; energy management optimization; Renewable Energy Source (RES); microgrid

I. INTRODUCTION

EVs are increasingly recognized for their potential to address climate change. While they emit no greenhouse gases directly, their charging largely depends on an energy mix that includes fossil fuel-based power plants [1]. The energy consumed in EV manufacturing also factors into their overall CO₂ emissions. This dependence means the net CO₂ reduction potential of EVs varies with the type of power generation, especially in coal-reliant systems [2]. Consequently, the integration of RESs alongside EVs is critical for genuinely diminishing CO₂ emissions. In line with this, the Saudi

Standards, Metrology, and Quality Organization (SASO) introduced regulations for EV usage in January 2018, and by June 2020, the Saudi Press Agency confirmed the official approval for importing and using EVs as per these guidelines. Vision 2030 in Saudi Arabia includes an objective to "Reduce All Types of Pollution," aiming to decrease air pollution from various residential and industrial activities [3]. CO₂ emissions in Saudi Arabia have notably increased, rising from 200 million tons in 1996 to 600 million tons in 2014, with the transportation sector contributing 20% of this amount [4]. EVs, emitting zero direct emissions, are poised to play a significant role in reducing transportation emissions, thereby enhancing air

quality and public health, and lessening environmental impacts. Research by KAPSARC has shown that realistic scenarios of EV integration could lead to a significant decrease in CO₂ emissions [4]. As Plug-in Hybrid Electric Vehicles (PHEV) increasingly integrate into the power grid, a potential rise in peak electricity demands is observed. However, PHEVs also present an opportunity for implementing energy storage and peak-shaving technologies.

The proliferation of EVs, while reducing oil consumption in transportation, leads to a surge in electricity demand for EV charging [5-6], increasing the daily load profile significantly, creating additional peak demands, intensifying existing peak loads, and escalating overall electricity requirements. A notable consequence of high EV penetration includes transformer overloads and distribution feeder congestions. Addressing these challenges, Demand Response (DR) strategies are employed in residential distribution networks to manage load effectively. This approach allows consumers to selectively control their loads, enhancing their comfort. A case study conducted in the Virginia Tech Electric Service area, considering varying levels and types of EVs along with diverse charging profiles, illustrates an increase in peak loads and potential violations of consumer comfort indices with increased EV penetration [7-8].

By 2030, it is expected that the EV fleet will grow by 10%, adding further strain to the electric system and necessitating increased power generation [9]. Projections indicate that global EV charging could demand about 2500 TWh of electricity by 2050 [10], with the National Renewable Energy Lab forecasting a 38% increase in electricity demand attributable to EVs by the same year [11-12]. In 2018, SASO established regulations for EV usage in Saudi Arabia [12-13] and in 2019, the Saudi Press Agency announced the official permission for EV importation and usage in the Kingdom as per these regulations. In 2021, Saudi Arabia launched its green initiative to combat the effects of climate change, including a project to increase Riyadh's EV presence to 30% of total vehicles [13-14].

The aim of this paper is to provide a detailed overview and simulation of an integrated microgrid system that incorporates V2G technology, specifically designed for a 24-hour MATLAB simulation catering to a community of 2000 households. It encompasses a diesel generator for reliable backup power, renewable energy contributions from wind and photovoltaic farms, and the innovative V2G system, which allows EVs to both draw from and contribute to the grid. The simulation showcases a modern energy management approach that leverages a mix of traditional and renewable sources, ensuring efficient power distribution and promoting energy independence. The system demonstrates the potential for localized energy production and consumption, aiming to reduce transmission losses and carbon footprint, while providing a resilient and flexible power network capable of adapting to varying load demands of residential and industrial consumers. This comprehensive microgrid model emphasizes advanced grid management and the integration of renewable energy sources, reflecting a progressive shift towards sustainable and self-sufficient energy ecosystems.

II. STRUCTURE OF THE PROPOSED V2G INTEGRATED MICROGRID SYSTEM

Figure 1 offers a concise overview of a projected microgrid system that integrates a V2G technology. This system is designed for simulation in MATLAB, focusing on a 24-hour operation. Key components include a diesel generator, a combination of photovoltaic (PV) and wind power sources, and the V2G setup. It is tailored for a community with 2000 households, with a significant emphasis on the integration of renewable energy sources and advanced grid management, ensuring balance and efficiency in power distribution and utilization. At the core, a Diesel Generator with a 15 MW capacity acts as the primary power source, flanked by a 8 MW PV farm and a 4.5 MW wind farm. The microgrid is designed to cater to both residential and industrial loads, with a substantial 10 MW allocated for residential consumption and 0.16 MVA for industrial use. In this energy ecosystem, the V2G component is dual-functional, contributing 4 MW as either an additional load or a power source, depending on the grid's demand. The system's energy balance is regulated through a phasor model operating at 60 Hz, ensuring a stable supply against the variable load demands over time, which is tracked by an integrated clock marking the simulation hours. The microgrid acts as a nexus point, distributing generated electricity to meet the demands of different load profiles, specifically residential and industrial loads. Residential load refers to the energy consumption by households, which typically includes lighting, heating, cooling, and electronics. On the other hand, the industrial load represents the power drawn by factories and industrial plants, which is generally significant and constant. At the forefront of this energy ecosystem is the V2G component, where EVs are depicted as both consumers and potential suppliers of energy. V2G technology allows for bidirectional energy flow, enabling not only the charging of the EVs, but also the provision of the stored energy back to the grid when needed, e.g. during peak demand or when the generation from other sources is low. This integrated system showcases a modern approach to energy management, where flexibility and sustainability are paramount. It leverages both traditional and renewable energy sources, alongside innovative technologies like V2G, to create a resilient and adaptive power network.

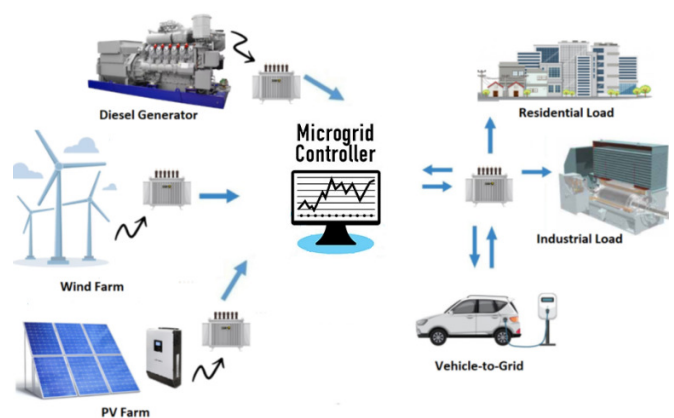


Fig. 1. Integrated microgrid system with V2G connectivity.

The infrastructure displayed in Figure 1 highlights the potential for localized energy production and consumption, minimizing transmission losses and promoting energy independence.

The flowchart in Figure 2 describes a process for managing V2G interactions, focusing on charge conditions and plug states. It differentiates between charging and regulation modes based on whether the cars are plugged in. For cars in the charge state, the V2G charging power is divided by the number of cars charging. For regulation, the V2G power is divided by the number of cars available for regulation. The plug state determines if a constant value is set to "1" (plugged in) or "0" (not plugged in), which affects the efficiency adjustments for discharging or charging. The State of Charge (SoC) is calculated by integrating the initial values, and based on this, the system either enters regulation or charge mode. This flowchart is instrumental in estimating the percentage state of charge for the vehicles in profile 1, optimizing the microgrid's performance.

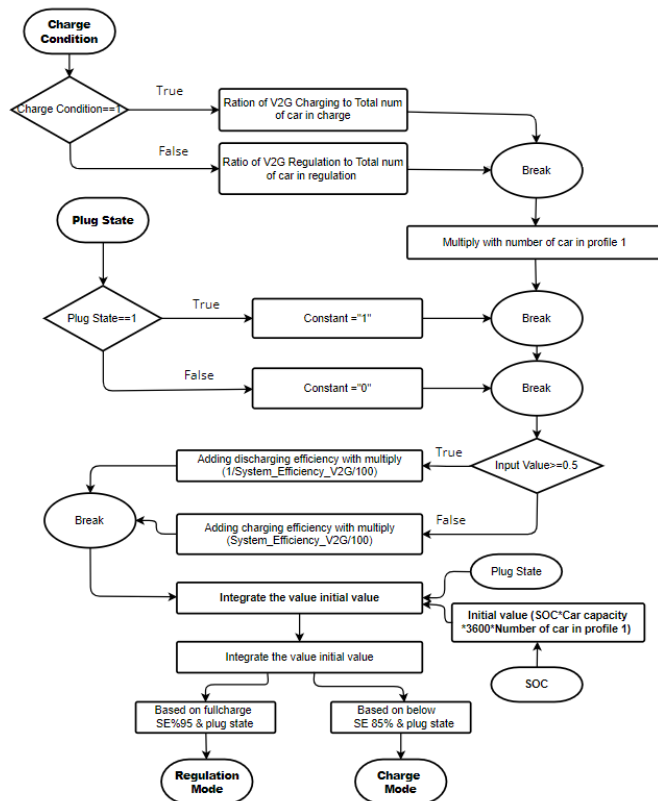


Fig. 2. Flowchart for V2G charge and regulation.

Figure 3 encapsulates the operation of V2G technology explored in this research, delineating its functional architecture. Central to managing the frequency of PHEVs is an aggregator that diligently oversees the EV network integrated with the grid. The potential for energy provision to this network hinges on the number of V2G-capable vehicles. Operators communicate with the aggregator, who then orchestrates strategic capacity

regulation and optimal bid selection. The V2G system, through its control unit, manages charging and discharging activities based on battery status and grid demands, with the Battery Management System (BMS) ensuring the battery's optimal SoC. Control mechanisms detailed in [20] underpin the study's operational analysis. The EVs' charging behavior is characterized using SoC and connection duration. Figure 3 delineates the methodology to discern EVs' charging or grid support engagement. Within the SoC control framework, dual constraints are applied to mitigate fluctuations caused by connection status and SoC at the onset. Figure 3 details the operational flow for V2G's charging and regulation actions. Notably, the charger control operates with distinctive criteria: it ceases charging beyond 95% SoC to maintain power quality and activates charging below 85% SOC, with the range in-between reserved for regulation activities.

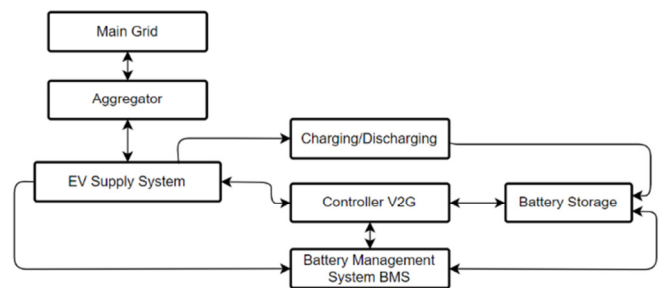


Fig. 3. V2G system operations.

III. RESULTS AND DISCUSSION

The simulation results were generated using Matlab/Simulink 2022 across a full day, totaling 86,400 s. The simulation adopted a variable step size and employed the Ode23tb solver, which is well-suited for stiff problems. This method effectively managed the complex behavior of a microgrid system comprising diverse power generation units, namely a diesel generator, a wind farm, a PV farm, and the bidirectional energy flows characteristic of V2G technology, in addition to meeting significant residential and industrial load demands. Figure 4 provides insight into the active power output of the PV RES throughout the day. It clearly illustrates the relationship between solar irradiance and power production, with a peak corresponding to the highest sun intensity. The graph presents a pronounced dip at around 12.1 hours, attributed to partial shading on the solar panels, effectively demonstrating the sensitivity of PV output to changes in solar conditions. This visualization serves as an effective tool for understanding the dynamics of solar energy contribution to the grid. Figure 5 presents the active power contribution of the wind farm to the electrical grid, with output levels fluctuating in tandem with wind speed variations. Higher wind speeds result in the wind turbines generating an increased amount of energy for grid supply. Conversely, a decline in wind speed leads to reduced capacity for active power transmission from the wind turbines to the grid. Therefore, wind speed variations are pivotal in determining the active power output from the wind generators.

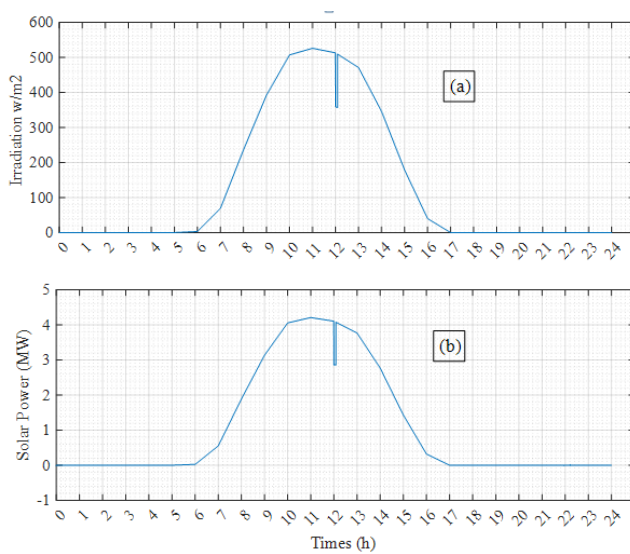


Fig. 4. PV power generator.

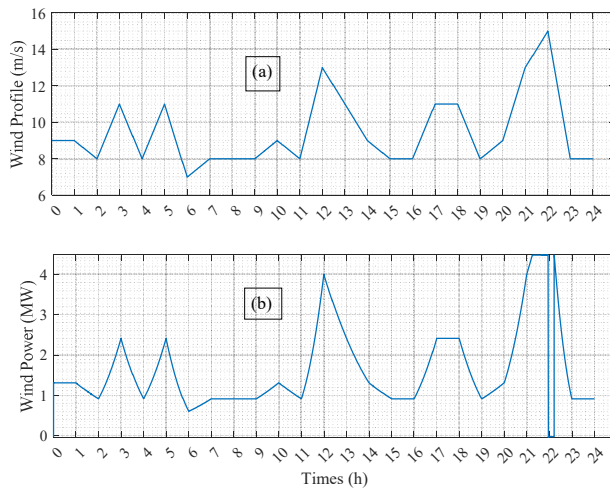


Fig. 5. Wind power generator.

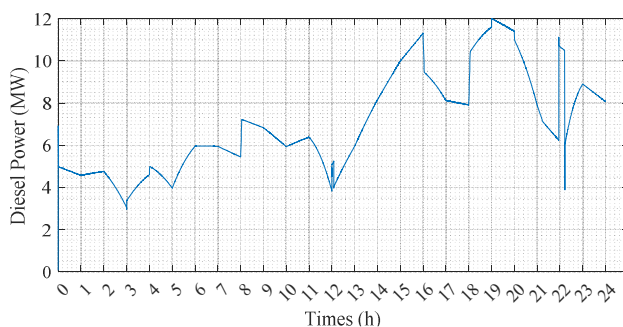


Fig. 6. Active power of the diesel generator.

Figure 6 illustrates the distribution of active power from the Diesel Generator to the grid. The active power output is contingent upon the generation levels from wind and PV farms, as well as the power transacted via the V2G system. Figure 7 depicts the active power consumption by the residential load

from the microgrid. It is observed that the load draws active power ranging between 5 MW and 10 MW. Notably, there are significant variations in load demand during specific intervals (11.9 to 12.1 and 21.8 to 22.4), indicating periods of abrupt changes in power requirements. Figure 8 showcases the active power consumption of the Asynchronous Machine (ASM) from the grid. Upon activation, the ASM demonstrates a significant initial draw of active power, followed by a consistent consumption of 0.321 MW. Additionally, there are notable fluctuations in active power usage, particularly during the intervals of 11.9 to 12.1 hours and 21.9 to 22.3 hours.

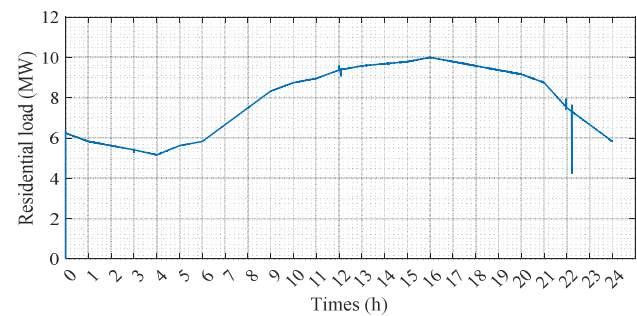


Fig. 7. Active power of residential load.

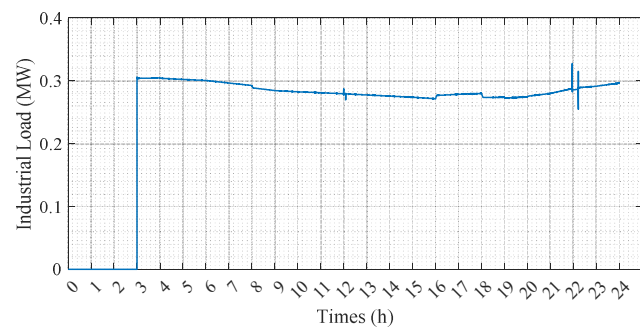


Fig. 8. Active power of industrial load.

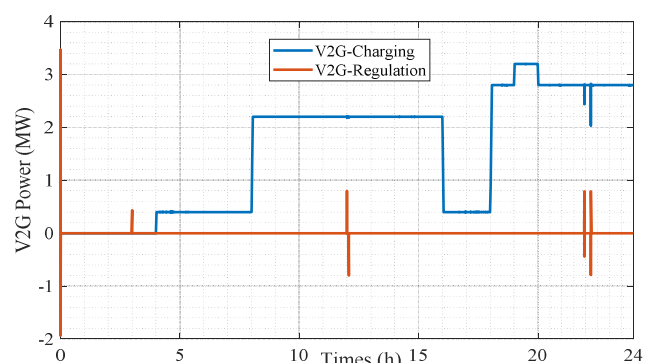


Fig. 9. Power flow characteristic of V2G.

Figure 9 illustrates the active power measurements during the charging and regulation mode of vehicles via the V2G system. Power consumption from the grid is denoted as negative, indicating that during four specific time periods, the batteries are utilizing grid power for charging. During the intervals of 11.9 to 12.1 hours and 21.9 to 22.3 hours, the

vehicles are observed to draw reactive power from the grid for battery charging purposes. Figures 10 and 11 exhibit the microgrid's active and reactive power distributions. Figure 10 indicates the active power with a significant event at 22:00, where the total power sharply drops from around 10 MW due to the wind generation disconnecting, triggering the diesel generator to supply upwards 5 MW to compensate for the shortfall, highlighting the grid's adaptive response mechanisms. Figure 11 shows the reactive power fluctuating, peaking just above 6 MVar, which could suggest increased demand or decreased supply during peak solar hours, possibly due to partial shading.

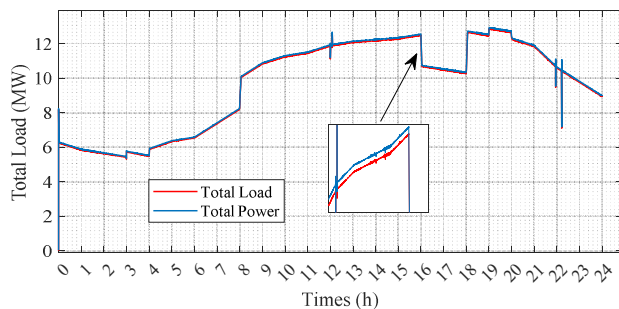


Fig. 10. Total power flow of the micro grid.

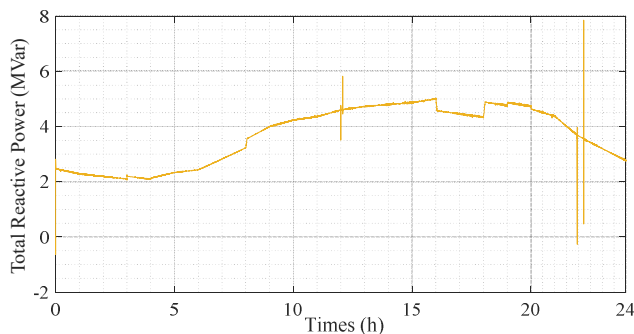


Fig. 11. Total power flow of the micro grid.

IV. CONCLUSION

This research significantly advances sustainable energy systems, showcasing an innovative microgrid model that integrates Vehicle-to-Grid (V2G) technology with traditional and renewable sources. It highlights the microgrid's capability in balancing energy through Matlab/Simulink simulations, under various conditions. The study emphasizes V2G's role in grid efficiency and stability, offering insights for future microgrid designs, especially in managing fluctuating renewable sources. Overall, it sets a foundational framework for integrated microgrids, pivotal in evolving energy management towards resilience and efficiency in the renewable energy era.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the approval and support of this research study by the grant no. ENGA-2023-12-2214 from the Deanship of Scientific Research at Northern Border University, Arar, K.S.A.

REFERENCES

- [1] A. M. Elshurafa and N. Peerbocus, "Electric vehicle deployment and carbon emissions in Saudi Arabia: A power system perspective," *The Electricity Journal*, vol. 33, no. 6, Jul. 2020, Art. no. 106774, <https://doi.org/10.1016/j.tej.2020.106774>.
- [2] H. Zakaria, M. Hamid, E. M. Abdellatif, and A. Imane, "Recent Advancements and Developments for Electric Vehicle Technology," in *2019 International Conference of Computer Science and Renewable Energies (ICCSRE)*, Agadir, Morocco, Jul. 2019, pp. 1–6, <https://doi.org/10.1109/ICCSRE.2019.8807726>.
- [3] "Saudi Vision 2030." <http://www.vision2030.gov.sa/en/>.
- [4] B. Wogan, E. Carey, and D. Cooke, *Policy Pathways to Meet Saudi Arabia's Contribution to the Paris Agreement*. KAPSARC, 2019.
- [5] Y. Kassem, H. Gokcekus, and A. Aljatlawe, "Utilization of Solar Energy for Electric Vehicle Charging and the Energy Consumption of Residential Buildings in Northern Cyprus: A Case Study," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11598–11607, Oct. 2023, <https://doi.org/10.48084/etasr.6142>.
- [6] N. T. Diep, T. D. Hiep, and N. K. Trung, "Constant Current Charging and Transfer Efficiency Improvements for a Dynamic Wireless Charging System," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12320–12326, Dec. 2023, <https://doi.org/10.48084/etasr.6315>.
- [7] S. A. Almohaimeed, "Electric Vehicle Deployment and Integration in the Saudi Electric Power System," *World Electric Vehicle Journal*, vol. 13, no. 5, May 2022, Art. no. 84, <https://doi.org/10.3390/wevj13050084>.
- [8] V. K. B. Ponnamp and K. Swarnasri, "Multi-Objective Optimal Allocation of Electric Vehicle Charging Stations and Distributed Generators in Radial Distribution Systems using Metaheuristic Optimization Algorithms," *Engineering, Technology & Applied Science Research*, vol. 10, no. 3, pp. 5837–5844, Jun. 2020, <https://doi.org/10.48084/etasr.3517>.
- [9] A. Aljanad, A. Mohamed, and H. Shareef, "Impact study of plug-in electric vehicles on electric power distribution system," in *2015 IEEE Student Conference on Research and Development (SCoReD)*, Kuala Lumpur, Malaysia, Sep. 2015, pp. 339–344, <https://doi.org/10.1109/SCoRED.2015.7449352>.
- [10] *Modelling Load Shifting Using Electric Vehicles in a Smart Grid Environment*, vol. 2010/07. Paris, France: International Energy Agency, 2010, <https://doi.org/10.1787/5km7v0bsmw7k-en>.
- [11] T. Mai *et al.*, "Electrification Futures Study: Scenarios of Electric Technology Adoption and Power Consumption for the United States," National Renewable Energy Lab. (NREL), Golden, CO, USA, Technical Report NREL/TP-6A20-71500, Jan. 2018. <https://doi.org/10.2172/1459351>.
- [12] "Saudi Arabia Corporate Average Fuel Economy (CAFE) standards – Policies," IEA. <https://www.iea.org/policies/8395-saudi-arabia-corporate-average-fuel-economy-cafe-standards>.
- [13] "Green Riyadh Project," *Royal Commission for Riyadh City*. <https://www.rcrc.gov.sa/ar/projects/green-riyadh>.

Optimal Artificial Neural Network-based Fabric Defect Detection and Classification

Nesamony Sajitha

Department of Computer and Information Science, Faculty of Science, Annamalai University, India
sajithap06@gmail.com (corresponding author)

Srinivasan Prasanna Priya

Thiru A. Govindasamy Govt Arts College, India
prasannapriyatdm@gmail.com

Received: 19 December 2023 | Revised: 4 January 2024 | Accepted: 8 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6773>

ABSTRACT

Automated Fabric Defect (FD) detection plays a crucial role in industrial automation within fabric production. Traditionally, the identification of FDs heavily relies on manual assessment, facilitating prompt repairs of minor defects. However, the efficiency of manual recognition diminishes significantly as labor working hours increase. Consequently, there is a pressing need to introduce an automated analysis method for FD recognition to reduce labor costs, minimize errors, and improve fabric quality. Many researchers have devised defect detection systems utilizing Machine Learning (ML) approaches, enabling swift, accurate, and efficient identification of defects. This study presents the Optimal Artificial Neural Network-based Fabric Defect Detection and Classification (OANN-FDDC) technique. The OANN-FDDC technique exploits handcrafted features with a parameter-tuning strategy for effectively detecting the FD process. To obtain this, the OANN-FDDC technique employs CLAHE and Bilateral Filtering (BF) model-based contrast augmentation and noise removal. Besides, the OANN-FDDC technique extracts shape, texture, and color features. For FD detection, the ANN method is utilized. To improve the detection results of the ANN method, the Root Mean Square Propagation (RMSProp) optimization technique is used for the parameter selection process. The simulation outputs of the OANN-FDDC technique were examined on an open fabric image database. The experimental results of the OANN-FDDC technique implied a better outcome than the 96.97% of other recent approaches.

Keywords-textile industry; fabric defect; machine learning; automation; feature extraction

I. INTRODUCTION

Textile manufacturing is a widespread and complex industrial field [1]. The process of textile manufacturing contains multiple complicated and ordered processes, which mostly comprise dyeing, spinning, printing, weaving, finishing, and manufacturing of garments [2]. The quality and stability of textile fabric manufactured by the entire production lines are essential to some enterprises. FD identification is a quality control process that has to ensure the recognition of defects existing in the textile fabric [3]. These defects can decrease the textile fabric cost by approximately 45-65%. A conventional investigation model is the collection of manual operators and their work is to identify the defects while the machines transfer the fabric [4]. Conventionally, these motored machines unroll the fabric rolls. Consequently, the fabric is stretched and presented to the workers without differences in thickness and folding. This process depends on human attention and visual capability. These tasks can be more time-consuming and tedious which can result in fatigue and human errors [5]. Hence, conventional models usually obtain an accuracy of 60-75%, despite their extremely slow speed when compared to the

rate of production. Automated visual investigation models for ensuring the higher quality of products in manufacturing lines are improving requirements [6]. The major benefits of automated defect identification methods are stability, dependability, and higher efficiency [7].

Several researchers have utilized DL methods to FD identification issues and achieved satisfactory outcomes for improving textile production efficiency and product quality [8]. Despite DL techniques have proved to be strong when dealing with classification and segmentation difficulties, there are still a few issues in real-time applications of particular factories [9]. Foremost, the real textile production line needs a higher real-time technique of performance, which is the need for great implementation efficiency [10]. Additionally, when comparing standard defect-free instances, information on defective images is complicated to acquire, which gives challenges to the DL training procedure [11]. Deep Convolutional Neural Network (DCNN)-based techniques have accomplished fitting outputs on visual responsibilities and are broadly applied in the industry.

This study presents the Optimal Artificial Neural Network-based Fabric Defect Detection and Classification (OANN-FDDC) technique. The OANN-FDDC technique exploits handcrafted features with a parameter-tuning strategy for effectively detecting the FD process. To obtain this, the OANN-FDDC technique employs CLAHE and Bilateral Filtering (BF), model-based contrast augmentation, and noise removal. Also, the OANN-FDDC technique extracts shape, texture, and color features. For FD detection, the ANN method was used. To improve the detection results of the ANN method, the Root Mean Square Propagation (RMSProp) optimization technique was used for the parameter selection process. The simulation outputs of the OANN-FDDC technique were examined on a fabric image database in the Kaggle repository.

II. RELATED WORKS

Authors in [12] established the Hybrid Mutation Moth Flame Optimizer with a DL Based Smart FD Detection (HMFODL-FDD) approach to maintainable manufacturing. To realize this purpose, their approach utilized a contrast enhancement procedure for boosting the image quality. For the extracting process, it utilizes the Inceptionv3 approach with an HMFO technique-based hyper-parameter optimizer. The FD classifier utilizes the BPNN approach. Authors in [13] presented a fabric design defect detection approach depended on a vision-based tactile sensor. The robotic arm equipped with the tactile sensor can be utilized for automating and standardizing the data gathered procedure and creating fabric databases. Also, a CNN combined with an attention process from the channel domain was established for detecting fabric types. Authors in [14] examined a learning-dependent structure for the automatic recognition of FDs whereas the Inception-V1 approach was utilized for forecasting the existence of errors in the local regions. Eventually, the authors execute the LeNet-5 approach which roles a vital play in voting, for identifying the FD type. Authors in [15] introduced an extremely effectual DL-based approach for a pixel-level FD classifier based on CNNs. Authors in [16] proposed a CNN-based textile recognition method called Faster R-CNN. Initially, a textile defect-mapping feature is extracted by the ResNet101 DCNNs. Faster R-CNN feature extraction in the last layer of feature mapping generates a loss of low-level place data.

Authors in [17] considered the Attention-Gate based on U-shaped Reconstruction Networking (AGUR-Net) and a dual-threshold segmentation post-processed approach. Authors in [18] introduced the Attention based Feature Fusion GAN (AFFGAN) method for unsupervised recognition of defects of yarn-dyed fabrics. This structure employs an adapted FPN to fuse multilevel data and employs an attention process for enhancing the model feature illustration abilities. This method utilizes a patch-level discriminator and an attention fusion generator. Authors in [19] introduced the FN-Net, a lightweight CNN-based architecture for FD detection. In contrast to the state-of-the-art models, FN-Net demonstrates 3 to 33 times faster training speed with reduced GPU and memory usage. In [20], the presented model utilizes the Bayesian optimization algorithm for network pruning. The training and detection phase utilize the pruned network, followed by employing the

image processing flow outlined for the conclusive assessment in FD recognition.

III. THE PROPOSED MODEL

The novel OANN-FDDC technique for the automatic detection and classification of FDs is presented in this study. The OANN-FDDC technique exploits handcrafted features with a parameter-tuning strategy for an effective FD detection process. The OANN-FDDC approach comprises the stages of pre-processing, handcrafted feature extraction, ANN, and RMSProp-based classification and tuning. Figure 1 exhibits the workflow of the OANN-FDDC approach.

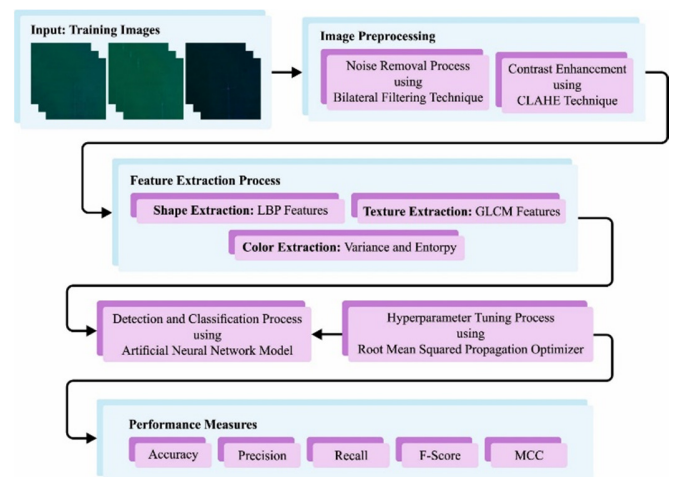


Fig. 1. Workflow of the OANN-FDDC approach.

A. Preprocessing

The preprocessing of the input images is conducted in two approaches: BF and CLAHE-based noise removal and contrast augmentation. BF is a nonlinear filtering model that intends to preserve edges while reducing noise in an image. To achieve this, it considers the spatial distance and intensity differences between adjacent pixels. The filter uses a weighted average for the pixels within the local neighborhood, where the weight depends on the spatial and intensity differences. The underlying concept behind BF is that neighboring pixels with related intensity must contribute more to the filtering method, whereas pixels with considerably different intensities must have a lesser impact. BF efficiently smoothens the image while preserving edges by integrating spatial and intensity information. CLAHE is a contrast enhancement method that enhances the image's local contrast by reallocating pixel intensity. It adapts histogram equalization to work on smaller regions of the images, which allows enhancing the localized contrast. The major objective of CLAHE is to improve the details in an image while avoiding artifacts and excessive noise amplification. It can be accomplished by limiting the maximum amplification possible for all the regions, thus preventing over-amplification of the noise.

B. Feature Extraction

At this stage, three sets of features, namely shape, color, and texture are extracted from the pre-processed fabric images.

1) LBP Features

LBP features contain 2 bitwise transitions, in [0-1] and [1-0]. LBP evaluates the variance and means of the entire intensity of the pixel employing a greyscale imagery as input [21]. The mathematical formula of LBP is:

$$\text{Texture Features}_{LBP}(\mathfrak{I}, \mathcal{R}) = \sum_{\mathfrak{I}=0}^{\mathfrak{I}-1} S(\mathfrak{U}_{\mathfrak{I}} - \mathfrak{U}_{\mathfrak{C}}) 2^{\mathfrak{I}} \quad (1)$$

In (1), $\mathfrak{U}_{\mathfrak{C}}$ shows the intensity contrast, \mathfrak{I} denotes the neighborhood intensity count, \mathcal{R} represents the radius, and $\mathfrak{U}_{\mathfrak{I}}$ indicates the variance of adjacent pixel intensity derived from $(\mathfrak{I}, \mathcal{R})$.

$$S_n(\mathfrak{I}) = \begin{cases} 1, & \text{if } \mathfrak{I} \geq t \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

In (2), the central pixel t is related to the neighboring pixels $S_n(\mathfrak{I})$. It produces a 1×59 feature vector for a single image and $N \times 59$ for N images.

2) GLCM Features

GLCM is employed for the extraction of the texture features assuming the 2nd-order relationships among neighboring and reference pixels. GLCM devises the co-occurrence matrix by relating the adjacent pixel values. The row count and column count of matrices are equivalent to the gray level counts [22]. Calculating GLCM for the k^{th} channel of data, G_k , with L levels, 4 features are taken out, correlation (r_k), contrast (c_k), homogeneity (h), and energy (e_k). c_k and r_k are calculated by:

$$\begin{aligned} c_k &= \sum_{i=1}^L \sum_{j=1}^L (i-j)^2 G_k(i-j) \\ r_k &= \sum_{i=1}^L \sum_{j=1}^L G_k(i-j) \left(\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{\sigma_i^2 \sigma_j^2}} \right) \end{aligned} \quad (3)$$

where:

$$\begin{aligned} \mu_i &= \sum_{j=1}^L \sum_{i=1}^L i G_k(i, j) \\ \mu_j &= \sum_{j=1}^L \sum_{i=1}^L j G_k(i, j) \\ \sigma_i^2 &= \sum_{j=1}^L \sum_{i=1}^L (i - \mu_i)^2 G_k(i, j) \\ \sigma_j^2 &= \sum_{j=1}^L \sum_{i=1}^L (j - \mu_j)^2 G_k(i, j) \end{aligned} \quad (4)$$

where the 4×1 vector $v_k = [c_k, r_k, e_k, h_k]^T$ denotes the feature vector for all the channels. Therefore, the feature vector of epochs considering every channel with 60 features is attained using:

$$f_{60 \times 1} = [v_1^T, \dots, v_{15}^T]^T \quad (5)$$

3) Variance and Entropy Features

Color features including variance and entropy are widely applied to calculate the complexity and distribution of colors in an image. Variance measures the variability or spread of color values in an image. It indicates how much the colors deviate from the mean or average color. In terms of color images, variance is separately measured for combinations of channels or all the color channels (i.e. red, green, and blue). High variance values imply a large range of color values, which indicates potentially more diverse vibrant colors in the image.

At the same time, a low variance value suggests a more limited range of colors, possibly leading to a subdued or more monochromatic appearance. Entropy is a measure of the quantity of uncertainty or information present in a random variable. In terms of color images, color entropy measures the randomness or complexity of the color distribution. It shows the uniformity and diversity of colors in an image. Entropy is frequently used in computer vision and image analysis tasks to describe the texture, uniqueness, or complexity of an image.

C. Optimal ANN-based Classification

To detect the FDs, an ANN model is used [23]. Generally, the ANN mechanism comprises 3 types of layers, i.e. Hidden Layers (HLs), input layer, and output layer. The hidden and input layers have an unobservable and predictor node and employ a non-linear conversion to the input layer's linear group. The resultant layer has the effect that is any function of a hidden unit. In output and hidden layers, the accurate procedure of function relies on the user definition and network type.

The activation transfer function (g_h) is a sigmoid function from HLs and the output layer (g_0). The i^{th} response y_i for the predictor value H_{ih} is a non-linear function as:

$$\begin{aligned} y_i &= g_0(H'_i \beta) + \varepsilon_i \\ H_{ih} &= g_h(X'_i \alpha_h) \end{aligned} \quad (6)$$

where β indicates the weighted vector of hidden to output units, X'_i denotes the i^{th} row of the input data matrix X , α indicates the weighted matrix of input to hidden units, and H_{ih} is a non-linear function of linear integration of the input dataset:

$$\begin{aligned} y_i &= g_0(H'_i \beta) + \varepsilon_i = \\ g_0[\beta_0 + \sum_{h=1}^{H-1} \beta_h g_h(X'_i \alpha_h)] + \varepsilon_i \end{aligned} \quad (7)$$

Using the sigmoid function, (7) is expressed as:

$$\begin{aligned} y_i &= [1 + \exp(-H'_i \beta)]^{-1} + \varepsilon_i \\ &= [1 + \exp(-\beta_0 - \sum_{h=1}^{H-1} \beta_h [1 + \exp(-X'_i \alpha_h)]^{-1})]^{-1} + \varepsilon_i \\ &= g(X_i, \beta, \alpha_1, \alpha_2, \dots, \alpha_H) + \varepsilon_i \end{aligned} \quad (8)$$

In (6), $\beta, \alpha_1, \alpha_2, \dots, \alpha_H$ are unknown parameter vectors, X_i denotes the vector of identified constants, and ε_i represents the residuals. The weights (parameters) are evaluated by enhancing some criterion functions like minimising the sum of squared errors or maximizing the log-likelihood function.

In such problems, overfitting is one of the most serious problems. To overcome it, a penalty term was considered as an optimizer condition. The penalized least square condition is used for estimating parameters:

$$\begin{aligned} E^* &= \sum_{i=1}^n (y_i - \hat{y})^2 + p_\lambda(\beta, \alpha_1, \alpha_2, \dots, \alpha_H) \\ &= \sum_{i=1}^n (y_i - g(X_i, \beta, \alpha_1, \alpha_2, \dots, \alpha_H))^2 \\ &\quad + p_\lambda(\beta, \alpha_1, \alpha_2, \dots, \alpha_H) \end{aligned} \quad (9)$$

where the penalty term is: $p_\lambda(\beta, \alpha_1, \alpha_2, \dots, \alpha_H) = \lambda(\sum \beta_i^2 + \sum \alpha_{ij}^2)$.

Finally, the RMSProp optimizer is employed for the optimal selection of the parameters related to the ANN model [24]. It utilizes a decaying average of partial gradient from the step size adaptation for all the parameters. In the RMSprop, every upgrade was completed based on (10). Upgrading was completed individually for all parameters.

$$v_t = \delta * v_{t-1} - (1 - \delta) * g_t^2$$

$$\Delta \omega_t = -\frac{\eta}{\sqrt{v_t + \epsilon}} * g_t$$

$$\omega_{t+1} = \omega_t + \Delta \omega_t \quad (10)$$

where g_t shows the gradient at t time along w_j , η denotes the rate of learning, and v_t indicates the gradient square's exponential average. In the training process, the weight in the network can be upgraded by shifting the value it can save for all the neurons every time a novel batch of trained data is obtained.

IV. RESULTS AND DISCUSSION

In this section, the simulation outcome of the OANN-FDDC method is investigated by implementing the FD dataset from [25] and the ZJU-Leaper dataset [26]. The features involved are 640 LBP features, GLCM Features (Energy, Contrast, Homogeneity, Correlation, Dissimilarity, Active Shape Models), and Texture Features (Variance, Entropy). Figure 2 depicts sample images of the two databases.

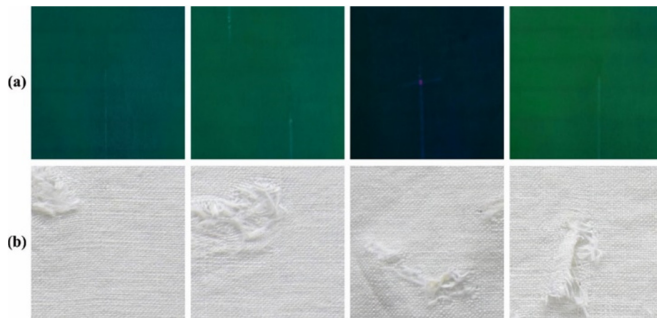


Fig. 2. Sample images: (a) [25], (b) [26].

The suggested technique is put under simulation by employing Python 3.6.5 tool on PC i5-8600k, 250GB SSD, GeForce 1050Ti 4GB, 16GB RAM, and 1TB HDD. The parameter set up is: learning rate: 0.01, activation: ReLU, epoch count: 50, dropout: 0.5, and size of batch: 5.

In Table I, the complete outputs of the OANN-FDDC technique on the dataset from [25] are shown. The outputs indicate that the OANN-FDDC technique accomplishes enhanced performance. On TRP, the OANN-FDDC technique achieves an $accu_y$ of 96.97%, $prec_n$ of 99.02%, $reca_l$ of 96.97%, F_{score} of 97.92%, and MCC of 97.19%. Also, on TRP, the OANN-FDDC method accomplishes $accu_y$ of 96.97%, $prec_n$ of 99.02%, $reca_l$ of 96.97%, F_{score} of 97.92%, and MCC of 97.19%.

Table II shows the results of the total analysis of the OANN-FDDC method on ZJU-Leaper dataset [26]. The

resultant represented the OANN-FDDC method to achieve improved performance. On TRP, the OANN-FDDC method obtains $accu_y$ of 94.78%, $prec_n$ of 95.17%, $reca_l$ of 94.78%, F_{score} of 94.72%, and MCC of 89.96%. Moreover, on TRP, the OANN-FDDC method obtains $accu_y$ of 95.46%, $prec_n$ of 96.02%, $reca_l$ of 95.46%, F_{score} of 95.55%, and MCC of 91.48%.

In Table III, the relative outputs of the OANN-FDDC models are compared with those of recent approaches [27, 28]. The simulation outcome indicates that the CNN, ResNet50v2, and DenseNet169v2 models gave poor performance, whereas FPN, Bi-FPN, NAS-FPN, and Dense-FPN models reported moderately greater outputs. The i-FPN model accomplishes considerable performance with an $accu_y$ of 94.20%, $prec_n$ of 89.16%, $reca_l$ of 90.05%, and $F1_{score}$ of 92.34%, but the OANN-FDDC method reaches maximum results with $accu_y$ of 96.97%, $prec_n$ of 99.02%, $reca_l$ of 96.97%, and $F1_{score}$ of 97.92%. These results illustrate the supreme accomplishment of the OANN-FDDC method in terms of various performance metrics.

TABLE I. CLASSIFICATION OUTPUT OF OANN-FDDC METHODOLOGY ON [25]

Metric	Training Phase	Testing Phase
Accuracy	96.97	95.83
Precision	99.02	97.62
Recall	96.97	95.83
F-Score	97.92	96.54
MCC	97.19	94.99

TABLE II. CLASSIFICATION OUTPUT OF OANN-FDDC METHODOLOGY ON [26]

Metrics	Training phase	Testing phase
Accuracy	94.78	95.46
Precision	95.17	96.02
Recall	94.78	95.46
F-Score	94.72	95.55
MCC	89.96	91.48

TABLE III. RELATIVE OUTPUT OF OANN-FDDC METHODOLOGY WITH CURRENT MODELS

Techniques	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{score}$
CNN	67.00	70.00	66.00	79.00
RestNet50v2	56.00	56.00	55.00	55.00
DenseNet169v2	56.00	59.00	55.00	57.00
FPN	83.40	90.44	92.43	89.72
Bi-FPN	85.20	89.95	89.75	89.71
NAS-FPN	87.90	89.03	90.26	91.49
Dense-FPN	90.20	91.92	91.95	91.38
i-FPN	94.20	89.16	90.05	92.34
OANN-FDDC	96.97	99.02	96.97	97.92

V. CONCLUSION

This study introduces the novel OANN-FDDC methodology for automated FD detection and classification. The OANN-FDDC methodology utilizes handcrafted features with a hyperparameter-tuning procedure for effectually detecting FDs. To manage this, the OANN-FDDC technique comprises pre-processing, handcrafted extraction, and ANN

and RMSProp classification and tuning. In order to enhance the recognition outputs of the ANN method, the RMSProp optimizer was used for the parameter selection process. The experimental results of the OANN-FDDC technique produced an outcome of 96.97%, which surpasses other known approaches.

The OANN-FDDC model excels in FD detection but may struggle with intricate patterns and lighting variations, necessitating enhanced robustness for diverse operational settings. In the future, DL methods will improve the performance of the OANN-FDDC method and adaptive learning mechanisms for handling diverse fabric patterns will be included and its robustness in dynamic lighting conditions will be refined.

REFERENCES

- [1] Y. Kahraman and A. Durmusoglu, "Deep learning-based fabric defect detection: A review," *Textile Research Journal*, vol. 93, no. 5–6, pp. 1485–1503, Mar. 2023, <https://doi.org/10.1177/00405175221130773>.
- [2] M. Chen *et al.*, "Improved faster R-CNN for fabric defect detection based on Gabor filter with Genetic Algorithm optimization," *Computers in Industry*, vol. 134, Jan. 2022, Art. no. 103551, <https://doi.org/10.1016/j.compind.2021.103551>.
- [3] L. Cheng, J. Yi, A. Chen, and Y. Zhang, "Fabric defect detection based on separate convolutional UNet," *Multimedia Tools and Applications*, vol. 82, no. 2, pp. 3101–3122, Jan. 2023, <https://doi.org/10.1007/s11042-022-13568-7>.
- [4] Q. Liu, C. Wang, Y. Li, M. Gao, and J. Li, "A Fabric Defect Detection Method Based on Deep Learning," *IEEE Access*, vol. 10, pp. 4284–4296, 2022, <https://doi.org/10.1109/ACCESS.2021.3140118>.
- [5] T. Meeradevi, S. Sasikala, S. Gomathi, and K. Prabakaran, "An analytical survey of textile fabric defect and shade variation detection system using image processing," *Multimedia Tools and Applications*, vol. 82, no. 4, pp. 6167–6196, Feb. 2023, <https://doi.org/10.1007/s11042-022-13575-8>.
- [6] J. Jing, Z. Wang, M. Rättsch, and H. Zhang, "Mobile-Unet: An efficient convolutional neural network for fabric defect detection," *Textile Research Journal*, vol. 92, no. 1–2, pp. 30–42, Jan. 2022, <https://doi.org/10.1177/0040517520928604>.
- [7] H. Xie and W. Xu, "Effective fabric defect detection using contrastive learning and layered fusion network," in *Fourteenth International Conference on Graphics and Image Processing*, Nanjing, China, Oct. 2022, vol. 12705, pp. 199–208, <https://doi.org/10.1117/12.2680579>.
- [8] K. Pandya, D. Dabhi, P. Mochi, and V. Rajput, "Levy Enhanced Cross Entropy-based Optimized Training of Feedforward Neural Networks," *Engineering, Technology & Applied Science Research*, vol. 12, no. 5, pp. 9196–9202, Oct. 2022, <https://doi.org/10.48084/etasr.5190>.
- [9] G. Anuradha and D. N. Jamal, "Classification of Dementia in EEG with a Two-Layered Feed Forward Artificial Neural Network," *Engineering, Technology & Applied Science Research*, vol. 11, no. 3, pp. 7135–7139, Jun. 2021, <https://doi.org/10.48084/etasr.4112>.
- [10] Y. Ivanova, "Damage Detection in Free-Free Glass Fiber Fabric Composite Beams by measuring Flexural and Longitudinal Vibrations," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10685–10690, Jun. 2023, <https://doi.org/10.48084/etasr.5788>.
- [11] H. Uzen, M. Turkoglu, and D. Hanbay, "Texture defect classification with multiple pooling and filter ensemble based on deep neural network," *Expert Systems with Applications*, vol. 175, Aug. 2021, Art. no. 114838, <https://doi.org/10.1016/j.eswa.2021.114838>.
- [12] N. Alruwais *et al.*, "Hybrid mutation moth flame optimization with deep learning-based smart fabric defect detection," *Computers and Electrical Engineering*, vol. 108, May 2023, Art. no. 108706, <https://doi.org/10.1016/j.compeleceng.2023.108706>.
- [13] B. Fang, X. Long, F. Sun, H. Liu, S. Zhang, and C. Fang, "Tactile-Based Fabric Defect Detection Using Convolutional Neural Network With Attention Mechanism," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–9, 2022, <https://doi.org/10.1109/TIM.2022.3165254>.
- [14] X. Jun, J. Wang, J. Zhou, S. Meng, R. Pan, and W. Gao, "Fabric defect detection based on a deep convolutional neural network using a two-stage strategy," *Textile Research Journal*, vol. 91, no. 1–2, pp. 130–142, Jan. 2021, <https://doi.org/10.1177/0040517520935984>.
- [15] Z. Wang, J. Junfeng, H. Zhang, and Y. Zhao, "Real-Time Fabric Defect Segmentation Based on Convolutional Neural Network," *AATCC Journal of Research*, vol. 8, no. 1_suppl, pp. 91–96, Sep. 2021, <https://doi.org/10.14504/ajr.8.S1.12>.
- [16] D. He, J. Wen, and Z. Lai, "Textile Fabric Defect Detection Based on Improved Faster R-CNN," *AATCC Journal of Research*, vol. 8, no. 1_suppl, pp. 82–90, Sep. 2021, <https://doi.org/10.14504/ajr.8.S1.11>.
- [17] H. Zhang, S. Wang, S. Lu, L. Yao, and Y. Hu, "Attention-Gate-based U-shaped Reconstruction Network (AGUR-Net) for color-patterned fabric defect detection," *Textile Research Journal*, vol. 93, no. 15–16, pp. 3459–3477, Aug. 2023, <https://doi.org/10.1177/00405175221149450>.
- [18] H. Zhang, G. Qiao, S. Lu, L. Yao, and X. Chen, "Attention-based Feature Fusion Generative Adversarial Network for yarn-dyed fabric defect detection," *Textile Research Journal*, vol. 93, no. 5–6, pp. 1178–1195, Mar. 2023, <https://doi.org/10.1177/00405175221129654>.
- [19] A. Suryarasmii, C.-C. Chang, R. Akhmalia, M. Marshallia, W.-J. Wang, and D. Liang, "FN-Net: A lightweight CNN-based architecture for fabric defect detection with adaptive threshold-based class determination," *Displays*, vol. 73, Jul. 2022, Art. no. 102241, <https://doi.org/10.1016/j.displa.2022.102241>.
- [20] C. C. Ho, W. C. Chou, and E. Su, "Deep convolutional neural network optimization for defect detection in fabric inspection," *Sensors*, vol. 21, no. 21, Nov. 2021, Art. no. 7074, <https://doi.org/10.3390/s21217074>.
- [21] U. Ishtiaq, E. R. M. F. Abdullah, and Z. Ishtiaque, "A Hybrid Technique for Diabetic Retinopathy Detection Based on Ensemble-Optimized CNN and Texture Features," *Diagnostics*, vol. 13, no. 10, Art. no. 1816, May 2023, <https://doi.org/10.3390/diagnostics13101816>.
- [22] S. M. Mousavi, A. Asgharzadeh-Bonab, and R. Ranjbarzadeh, "Time-Frequency Analysis of EEG Signals and GLCM Features for Depth of Anesthesia Monitoring," *Computational Intelligence and Neuroscience*, vol. 2021, Aug. 2021, Art. no. e8430565, <https://doi.org/10.1155/2021/8430565>.
- [23] A. Biglarian, E. Bakhshi, A. R. Baghestani, M. R. Gohari, M. Rahgozar, and M. Karimloo, "Nonlinear Survival Regression Using Artificial Neural Network," *Journal of Probability and Statistics*, vol. 2013, Feb. 2013, Art. no. e753930, <https://doi.org/10.1155/2013/753930>.
- [24] F. Mehmood, S. Ahmad, and T. K. Whangbo, "An Efficient Optimization Technique for Training Deep Neural Networks," *Mathematics*, vol. 11, no. 6, Jan. 2023, Art. no. 1360, <https://doi.org/10.3390/math11061360>.
- [25] S. Ranathunga, "Fabric Defect Dataset." [Online]. Available: <https://www.kaggle.com/datasets/rmshashi/fabric-defect-dataset>.
- [26] C. Zhang, S. Feng, X. Wang, and Y. Wang, "ZJU-Leaper: A Benchmark Dataset for Fabric Defect Detection and a Comparative Study," Dec. 2020, [Online]. Available: <http://www.qaas.zju.edu.cn/zju-leaper/>.
- [27] J. Xiang, R. Pan, and W. Gao, "Online Detection of Fabric Defects Based on Improved CenterNet with Deformable Convolution," *Sensors*, vol. 22, no. 13, Jun. 2022, Art. no. 4718, <https://doi.org/10.3390/s22134718>.
- [28] A. Amelio *et al.*, "Defining a deep neural network ensemble for identifying fabric colors," *Applied Soft Computing*, vol. 130, Nov. 2022, Art. no. 109687, <https://doi.org/10.1016/j.asoc.2022.109687>.

Thermal and Mechanical Properties Enhancement of Cement Mortar using Phosphogypsum Waste: Experimental and Modeling Study

Ehab M. Ragab

Department of Civil Engineering, Engineering College, Northern Border University, Saudi Arabia
ehabhosny_3@hotmail.com (corresponding author)

Tarek M. Awwad

Department of Civil Engineering, Engineering College, Northern Border University, Saudi Arabia |
Department of Civil Engineering, Engineering College, Al-Azhar University, Egypt
tarek.awwad@nbu.edu.sa

Nidhal Becheikh

Department of Chemical and Materials Engineering, Engineering College, Northern Border University, Saudi Arabia
nidhal.becheikh@nbu.edu.sa

Received: 6 January 2024 | Revised: 18 January 2024 | Accepted: 20 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6875>

ABSTRACT

This research presents an in-depth investigation into the application of phosphogypsum (PG), a by-product of phosphate fertilizer plants and chemical industries, as a replacement material for cement in mortar, with a focus on enhancing its thermal and mechanical properties. The influence of PG as a partial replacement for cement on the compressive strength of mortar after 3, 7, and 28 days is investigated. Utilizing the Box-Behnken design within Response Surface Methodology, this study analyzed factors, such as sulfuric acid concentration, washing time, calcination temperature, and PG to cement ratio. Results indicate that optimal PG levels enhance mortar strength, particularly at 28 days, through sustained ettringite formation and microstructure optimization. Sulfuric acid concentration and calcination temperature were identified as the most significant elements influencing compressive strength, with the latter improving PG quality and reactivity. A PG to cement ratio up to 10% was found beneficial, while washing time had a negligible effect. The research highlights a critical synergy between the sulfuric acid concentration applied during the purification of PG and the calcination temperature. A significant improvement of 21% in compressive strength was achieved, underscoring the combined effect of chemical and thermal treatment on PG's efficacy in mortar. The increased sulfuric acid concentration is presumed to purify the PG by removing impurities, thus improving its reactivity. Concurrently, calcination alters the PG's crystalline structure and diminishes its organic composition. This interdependent optimization is instrumental in enhancing the structural integrity of PG-modified mortar. The potential for raw PG to be used as an insulating material is more pronounced at higher replacement rates (10%), while sulphuric acid treated PG (SCPG) and heat treated PG (HTPG) seem to be unable to provide a clear insulative advantage.

Keywords-waste management; phosphogypsum; thermo-mechanical improvement; cement mortar; RSM modeling

I. INTRODUCTION

The use of solid waste as a partial replacement in cement mortar has emerged as an interesting subject, not only regarding the construction industry, but also for environmental

reasons [1-3]. The global concrete market size was valued almost half a trillion dollars in 2020 and should grow about 5% each year until 2028 based on forecasts, strongly indicating a huge possibility for utilizing solid waste as a partial replacement in cement mortar [4]. Among all solid wastes, fly

ash is one of those that have received wide attention [5]. It has been established through several investigations that fly ash can improve workability and durability along with decreasing permeability and CO₂ emission potential of cement mortar [6-8]. Silica fume is a by-product of silicon alloy manufacture that improves strength and durability within cement mortar since it contains highly reactive low size pozzolanic particles [9-11]. Industrial slags, such as blast furnace slag and steel slag, have also been investigated [12]. These materials can enhance the mechanical properties and durability of cement mortar, especially when finely ground [13]. Recycled Concrete Aggregate (RCA), obtained from demolished concrete structures, has been used as a replacement for natural aggregate in cement mortar [14, 15]. It is estimated that the construction industry globally generates around 1.3 billion tons of Construction and Demolition Waste (CDW) per year [16]. Other types of CDW, such as crushed bricks and tiles, have also been used as partial replacements in cement mortar, contributing to waste reduction and recycling in the construction industry, and providing satisfactory performance [17]. The use of solid waste in cement mortar can lead to the development of sustainable and environmentally friendly construction materials, but the properties and performance of the resulting mortars depend on the type and proportion of the waste used, as well as the specific mix design and curing conditions. Further research is needed to optimize these factors and to assess the long-term performance and environmental impact of these mortars.

The exploration of phosphogypsum (PG), a by-product of the phosphorous fertilizer industry, as a potential substitute for a portion of cement in concrete has garnered interest in recent years [18-20]. Each year, the production of PG by the fertilizer industry surpasses 100 million tons globally [21]. Its physical and chemical characteristics are similar to those of natural gypsum, which makes it a promising option for industrial uses. Early research indicated that substituting 5% of cement with PG could enhance the setting time and increase the compressive strength of cement by nearly 10%, thereby establishing a basis for future studies [22]. Further building on this premise, a more detailed study investigated the effects of substituting 10% - 40% of cement with PG, and found that a 20% replacement could enhance the compressive strength of concrete by as much as 15% [23]. PG may serve as a viable substitute for cement mortar, potentially improving the structural properties of concrete. However, the radioactivity of PG, due to radionuclides, such as Ra-226 and Th-232, has been a significant concern, limiting its broader usage. Authors in [24] sought to establish the safe levels of PG in cement, concluding that radioactivity could be kept within safe limits — typically less than 370 Bq/kg for Ra-226 and less than 200 Bq/kg for Th-232 — with a replacement level of up to 10%. This conclusion paved the way for the safer application of PG in concrete.

As cement production contributes to roughly 7% of global CO₂ emissions, any decrease in the usage of cement could have a substantial environmental benefit [25, 26]. Authors in [27] found that substituting cement with PG could lead to a reduction in these emissions by as much as 10%, suggesting a more eco-friendly approach to concrete production. Presently,

ongoing research aims to further enhance the application of PG in mortar. In conclusion, existing studies indicate that PG holds promise as a viable substitute for a portion of cement in mortar, contributing to both improved structural characteristics and significant reductions in CO₂ emissions. However, the radioactivity associated with PG needs to be thoroughly understood, and additional research is needed to refine its application, ensuring safety and efficiency. Regardless of the noteworthy number of research studies on the PG use to ameliorate mechanical properties, few studies have focused on experimental condition optimization of the partial replacement of cement mortar employing PG. In this context, Response Surface Methodology (RSM) design modeling stands out as a valuable analytical technique for investigating the complex effects of various PG processing methods, namely chemical treatment, washing, and calcination and how they interact to improve compressive strength. By conducting a detailed investigation of the individual and combined effects of chemical, washing, and calcination treatments, this study will pinpoint the most effective conditions for significantly boosting compressive strength.

II. EXPERIMENTAL PART

The methodology encompasses the collection and preparation of PG samples and the formulation of various mortar mixtures. The mortar's compressive strength is measured, with the data collected serving as a basis for analysis and comparison against traditional mortar mixtures. RSM is used to refine the processing of PG for cement mortar enhancement. It investigates three PG variants: raw, untreated (RPG), water-washed and acid-treated (CPG), and heat-treated (HTPG). Ordinary Portland Cement (OPC) and well graded sand are used to ensure uniformity in the mortar mixtures with varying ratios of PG to cement.

A. Materials and Methods

PG is the main substance of the current investigation. Table I exhibits the chemical composition of natural gypsum and PG.

TABLE I. CHEMICAL COMPOSITION OF NATURAL GYPSUM AND PHOSPHOGYPSUM.

	Gypsum	PG
P ₂ O ₅ (%)	0.08	1.2
Fe ₂ O ₃ (%)	0.29	0.13
MgO (%)	8.23	0.26
Al ₂ O ₃ (%)	0.89	0.21
SO ₃ (%)	31.26	43.8
CaO (%)	34.2	30.7
K ₂ O (%)	0.15	0.08
Na ₂ O (%)	0.09	0.03
SiO ₂ (%)	5.1	3.5
F (%)	0.09	1.9
TOC (%)	0	0.64
Cl (ppm)	269	399

The study employs PG in three forms. RPG, is used in the condition it is found within industrial disposal areas. In contrast, the cleaned form CPG is subjected to an intensive purification process including water washing (WCPG) and sulphuric acid treatment (SCPG). In the third form, PG

undertakes heat treatment (HTPG) that eliminates contaminants and moisture.

The protocol, presented in Figure 1, is designed to explore a series of treatment methods aimed at purifying and enhancing the quality of PG. It includes procedures for acid treatment using sulfuric acid (SCPG) at different concentrations, water washing (WCPG) to remove soluble impurities, and calcination (HTPG) at varying temperatures to alter the physical and chemical properties of PG.

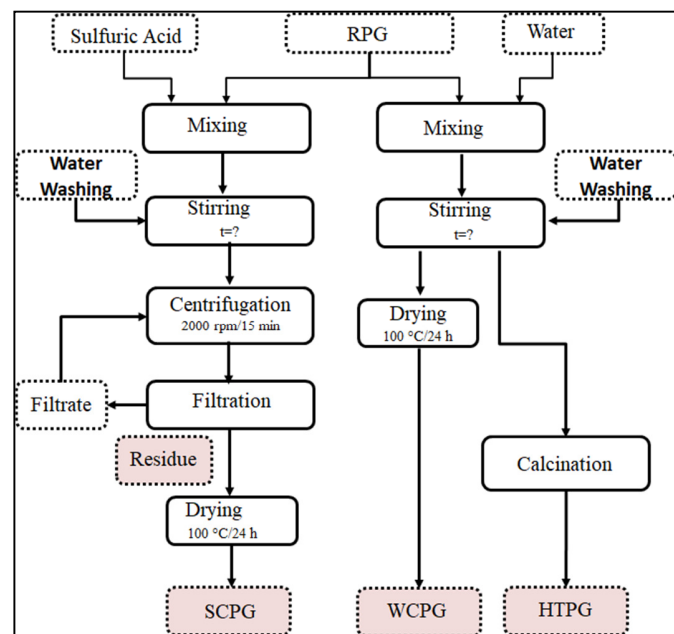


Fig. 1. Protocol of PG treatments.

- Acid Treatment with H_2SO_4 : The treatment of raw PG with sulfuric acid is performed at varying concentrations to remove impurities. PG is soaked, stirred, filtered, rinsed until neutral, and dried at $100^\circ C$.
- Water Washing: The Basic Water Washing (BWW) method consists of agitating 100 grams of raw PG in 1 Lt of distilled water for 1, 2, and 3 hr to dissolve soluble impurities. Post-washing, the PG is filtered, repeatedly rinsed until the pH is neutral, and then dried at $100^\circ C$ to remove all water content.
- Calcination: Calcination treatments are applied to PG post-treatment to further modify its properties. The Low-Temperature Calcination (LTC) is conducted at $100^\circ C$ for 1 hr, focusing on dehydrating the PG without altering its chemical structure significantly. Medium-Temperature Calcination (MTC) is performed at $400^\circ C$ for 2 hr. The temperature is higher to boost dehydration and possible phase changes. Finally, High-Temperature Calcination (HTC) at $700^\circ C$ for 3 hr is the most intensive calcination method, designed to complete the thermal treatment and potentially enhance the purity and reactivity of the PG.

Type I OPC is the chosen binding agent due to its widespread use and dependable properties in the field of construction, with a specific gravity of 3.15 to ensure standardized mix ratios. The sand selected for this study is well graded, with a fineness modulus of 2.92, and having removed any dust, clay, or organic debris, thereby assuring its neutral impact on the mortar's performance characteristics. Six mortar mixtures with varying PG-to-cement ratios were prepared according to the proportions detailed in Table II. The preparation process was carefully controlled. Each batch was mixed with precise amounts of cement, sand, PG, and water until homogenous consistency was reached.

TABLE II. MIXTURE PROPORTIONS OF PG - CEMENT MORTAR (FOR 9 CUBES)

Mixture	Cement (g)	Sand (g)	Water (mL)	PG (g)	PG/Cement (% wt.)
M1	740	2035	359	0	0
M2	725.2	2035	359	14.8	2
M3	703	2035	359	37	5
M4	684.5	2035	359	55.5	7.5
M5	666	2035	359	74	10
M6	647.5	2035	359	92.5	12.5

B. Response Surface Methodology

RSM is a powerful statistical approach used to model and analyze the effects of multiple variables on a response variable. The Box-Behnken design, a subset of RSM, is particularly efficient for experiments involving three levels of each factor, minimizing the number of experimental runs required while still providing a comprehensive understanding of the system. In our specific application of a Box-Behnken design, we consider 4 factors:

- $[H_2SO_4]$ (M): This continuous factor represents the concentration of sulfuric acid, with 3 levels: 0, 2, and 4 M.
- Washing Time (hr): Another continuous factor, washing time, also with 3 levels: 1, 2, and 3 hr.
- Calcination Temperature ($^\circ C$): The third continuous factor is the temperature at which calcination occurs, with 3 levels set at 100, 400, and $700^\circ C$.
- PG/Cement Weight Percentage: The categorical factor in this design is the ratio of PG to cement, measured in weight percentage. It has 5 levels: 2, 5, 7.5, 10, and 12.5%.

This particular design allowed us to investigate the individual effects of each factor and the interactions between them on the response variable: the Compressive Strength (C.S) of a cement mortar. By using 3 levels of the continuous factors, the design will enable the fitting of a full quadratic model, which includes linear, quadratic, and two-factor interaction effects. The addition of 5 levels for the categorical factor, although not standard for a Box-Behnken design, provides additional support in understanding how the replacement of cement with PG affects the response. The unique combination of continuous and categorical variables within this Box-Behnken design offers a robust framework for optimizing the material properties of cement mortar. In a typical Box-Behnken design for 3-level factors, the number of required experiments (N) is calculated by:

$$N = 2^k (k - 1) + C_0 \quad (1)$$

where k is the number of factors and C_0 is the number of center points. Since we have a categorical factor with 5 levels, separate Box-Behnken design for each level of the categorical factor was conducted. For a single level of the categorical factor, the total number of runs is: $12 + 3 = 15$. Since the categorical factor has 5 levels, the total number of experiments required is: $15 * 5 = 75$.

C. Thermal Conductivity

To determine the thermal conductivity of cement mortar, a cubic sample with 5 cm thickness was prepared and fully cured. This sample was positioned within a heat flow meter setup, inserted between a heated plate and a cooled plate, while thermocouples monitored the temperatures on each side. The cold plate was set to 20°C, and the hot plate to 30°C, generating a steady-state temperature gradient of 10°C across the mortar. Upon reaching equilibrium, the heat flow meter registered a heat transfer rate of 0.6 W through the sample. Thermal conductivity was then calculated using the steady-state heat flow, the sample's cross-sectional area, and the established temperature difference, resulting in a thermal conductivity value indicative of the mortar's ability to conduct heat. This value is crucial for evaluating the material's effectiveness as an insulator in building structures.

III. RESULTS AND DISCUSSION

A. Age Influence on Compressive Strength

When PG is used as a partial replacement for cement in mortar, its impact on the C.S varies with different ages (3, 7, and 28 days) as presented in Figure 2.

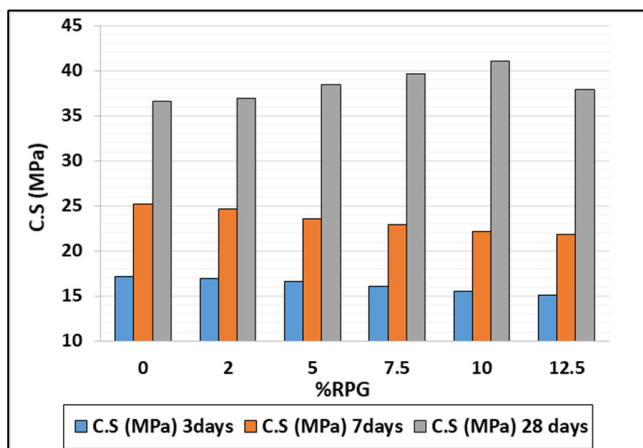


Fig. 2. Compressive strength of all mixtures after 3, 7, and 28 days.

During the first 3 days, the hydration process of mortar is particularly sensitive to PG. Low PG % can be beneficial, aiding in ettringite formation and thus, early strength, but higher PG levels might introduce detrimental substances, such as phosphates and organics that hinder this process. As the time progresses to 7 days, the ongoing hydration generally results in strength increase due to the presence of more calcium silicate hydrate (C-S-H) and ettringite. For higher PG levels the

negative effects of contaminants appear. After 28 days, the mortar's hydration is largely complete, and the full impact of PG is evident: optimal PG levels can lead to strength enhancement via continued ettringite formation and microstructure optimization, whereas excessive PG (12.5%) can undermine strength through mechanisms, like Delayed Ettringite Formation (DEF), phosphate complexation, and the creation of non-beneficial secondary products.

B. Results of the Experimental Design

The Box-Behnken design, a component of RSM, investigates the effects of four factors across various levels. This influence of sulfuric acid concentration (0, 2, 4 M), washing time (1, 2, 3 hr), calcination temperature (100, 400, 700 °C), and the PG to cement ratio (2, 5, 7.5, 10, 12.5%) were assessed through a thorough analysis of their impact on compressive strength (C.S). The p-value of each variable is compared against the alpha level of 0.05 to determine statistical significance with respect to the dependent variable. A p-value below 0.05 indicates a significant relationship. The null hypothesis assumes that variables have a zero coefficient, implying no effect on the outcome. This study establishes the alpha at 0.05 to identify significant variables. Regression analysis results are illustrated in Figure 3.

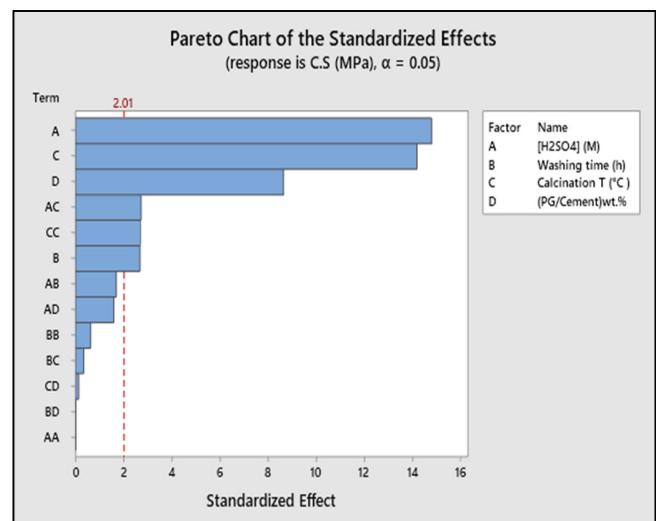


Fig. 3. Pareto chart for C.S response.

The standardized effect Pareto chart of Figure 3 reveals that sulfuric acid concentration, washing time, calcination temperature, and the ratio of PG to cement have essential impacts. Additionally, the interaction between sulfuric acid concentration and calcination temperature has considerable effects on C.S. In addition, the chart indicates that the sulfuric acid concentration and calcination temperature are notably more influential than PG to cement ratio and washing time. All other interactions between factors appear to be insignificant. These interpretations are verified by discrete analysis of the impact of each factor, as illustrated in Figure 4, which shows the individual contribution of each variable to the C.S. A higher concentration of sulfuric acid during the purification process of PG can lead to the removal of impurities, such as soluble salts,

heavy metals, and residual phosphates, resulting in a more reactive and pure form of PG. This purified PG, when used as a partial replacement of traditional cement in mortar, has the potential to increase C.S due to its improved bonding and hydration characteristics.

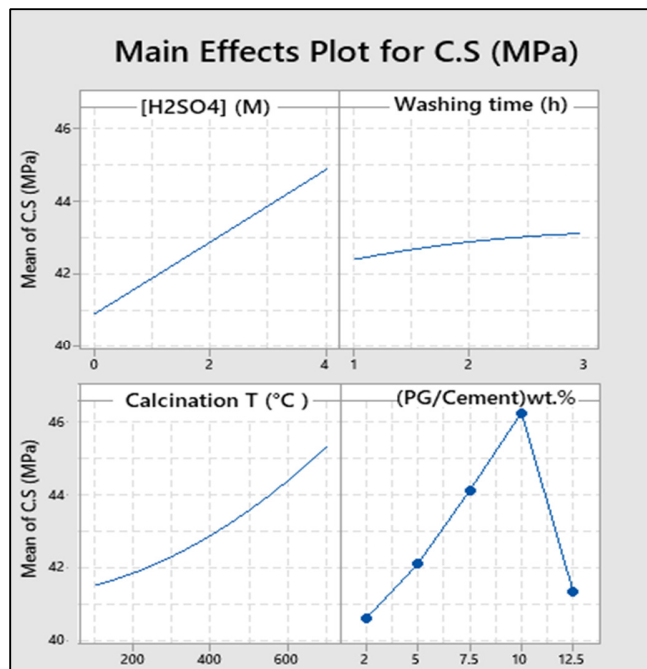


Fig. 4. Main effect plots on C.S.

Calcination temperature also plays a crucial role in determining the quality of PG. Higher calcination temperatures enhance the removal of moisture and organic materials, which in turn can transform the crystal structure of the gypsum. This process creates a calcined product that, when mixed with cement, leads to an increase in the C.S of the mortar. The calcined PG is more chemically active, promoting a stronger bond within the mortar matrix and contributing to the formation of a denser, harder structure. The ratio of PG to cement is another critical factor in achieving optimal compressive strength in mortar applications. An increased PG ratio can improve compressive strength up to 10% by contributing sulfate ions that facilitate the formation of ettringite, a mineral that can positively affect the microstructure and strength of the mortar. Finally, washing time of PG appears to have a marginal effect on the C.S of the mortar, indicating that the main contaminants within PG do not significantly impact the hydration process or are not readily washed away within the tested time frames. This suggests that while washing is a part of the PG purification process, it is less critical than the other factors, like the concentration of sulfuric acid used and the calcination temperature in determining the overall quality and effectiveness of PG as a cement replacement in mortar. These results are confirmed by the influence of washing time on C.S for different levels of acid concentration and calcination temperature as shown in Figures 5 and 6. On the other hand,

the interaction between sulfuric acid and calcination temperature, illustrated in Figure 7, is significant according to the Pareto diagram as it shows that C.S is more enhanced when increasing both these operational variables.

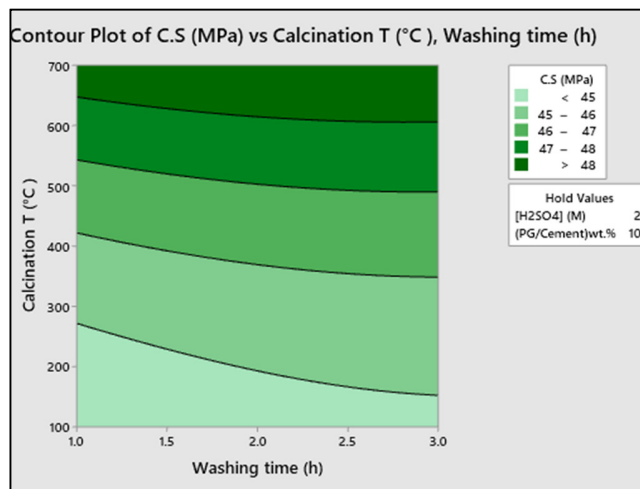


Fig. 5. Washing time influence on C.S for different calcination temperatures.

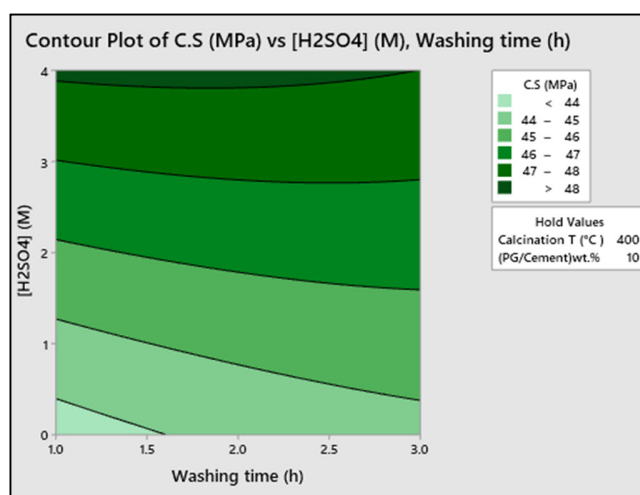


Fig. 6. Washing time influence on C.S for different [H₂SO₄] values.

Finally, the use of PG to cement ratio of 10% and a washing time of 2 hr with extensive chemical and thermal treatment ([H₂SO₄] from 3 to 4 M and calcination temperature of 600 to 700°C), improves the mechanical behavior of mortar samples with TPG. An upgrade in C.S of 21% was achieved in comparison with mortar samples that utilized RPG as a replacement material. In fact, increasing the sulfuric acid concentration causes chemical changes that contribute to the strength of the material, such as creating more cross-linkages or improving the packing density of the material's microstructure. Likewise, increasing the calcination temperature causes physical changes, like the removal of any volatile components, the enhancement of crystallinity, or the sintering of particles, which can also increase the material's strength.

C. Thermal Study Results

A comparative analysis of the thermal conductivity of mortar when different types of PG are used as partial replacement of cement is shown in Figure 8. Three types of PG were considered: RPG, SCPG, and HTPG, with varying replacement percentages, from 0% to 12.5%. The control group's thermal conductivity remains constant at 1.20 W/(m·K) as it represents the standard mortar without any PG replacement, with the consideration that a decrease in thermal conductivity less than or equal to 3% may fall within the experimental error. RPG remains the only PG type which has a potential decrease in thermal conductivity that could be considered significant and not attributable to experimental error, but only at replacement percentages of 7.5% and higher. SCPG and HTPG do not display decreases that would be considered significant. Therefore, the potential for RPG to be used as an insulating material is more pronounced at higher replacement rates, while SCPG and HTPG seem to be unable to provide a clear insulative advantage.

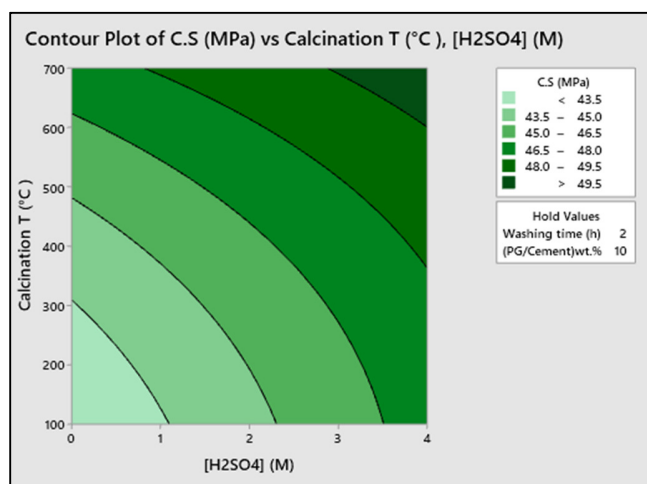


Fig. 7. Calcination temperature influence on C.S for different $[H_2SO_4]$ values.

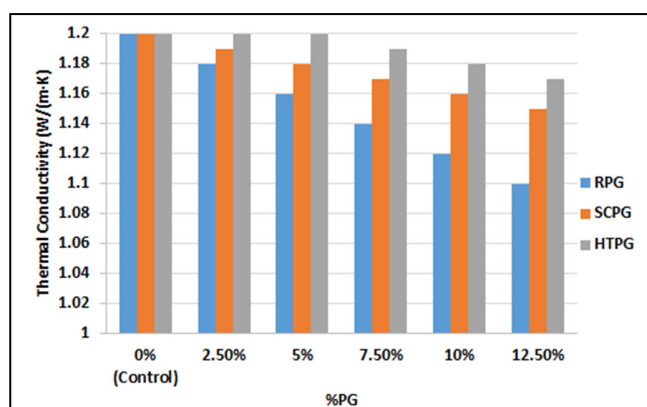


Fig. 8. Comparative analysis of the thermal conductivity of mortar with different types of PG treatments.

IV. CONCLUSION

In conclusion, this study provides valuable insights into the potential use of PG as a partial replacement of cement in mortar. By exploring unprocessed, water-washed, sulphuric acid treated, and heat-treated forms of PG, the research identifies various treatment methods that can optimize the PG quality. The use of Response Surface Methodology, employing the Box-Behnken design, allowed an efficient and structured investigation to be conducted. The latter explored the effects of sulfuric acid concentration, washing time, calcination temperature, and PG to cement ratio on mortar compressive strength.

The results clearly indicate that the treatment of PG with sulfuric acid and the application of appropriate calcination temperatures significantly enhance the compressive strength of the mortar. These treatments effectively remove impurities and improve the microstructure of PG. Furthermore, the study confirms that an optimal ratio of PG to cement exists, at which the benefits of sulfate contributions to mortar strength are maximized without the adverse effects of contaminants.

Washing time, while a part of the purification process, has a less pronounced impact on compressive strength, suggesting that the primary contaminants in PG are either not significantly affecting the hydration process or are not easily removed within the tested time frame. However, it serves as a necessary step in the overall enhancement of the PG quality.

The interaction between sulfuric acid concentration and calcination temperature is particularly noteworthy, demonstrating that a synergistic effect exists when both parameters are increased simultaneously, leading to substantial improvements in compressive strength. This combination of chemical and thermal treatments, particularly at an acid concentration range of 3-4 M and calcination temperatures between 600 and 700 °C, has been shown to significantly improve the mechanical properties of the mortar.

The potential for RPG to be used as an insulating material is more pronounced at higher replacement rates, while SCPG and HTPG seem to be unable to provide a clear insulative advantage.

The study's findings are promising for the construction industry, suggesting that with proper treatment, PG can be a viable partial replacement of cement. This could lead to more sustainable practices by recycling a byproduct of the fertilizer industry and reducing the carbon footprint associated with cement production.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the approval and support of this research study by the grant no. ENGA-2022-11-1864 from the Deanship of Scientific Research at Northern Border University, Arar, K.S.A.

REFERENCES

- [1] C. Maraveas, "Production of Sustainable Construction Materials Using Agro-Wastes," *Materials*, vol. 13, no. 2, Jan. 2020, Art. no. 262, <https://doi.org/10.3390/ma13020262>.

- [2] K. Srivastava, A. Srivastava, P. Singh, R. S. Jagadish, R. Verma, and V. Jaiswal, "Role of Eco-Friendly Materials in Construction for Making Cities Smart: A Case Study of Noida and Greater Noida," in *Making Cities Resilient*, V. R. Sharma and Chandrakanta, Eds. New York, NY, USA: Springer, 2019, pp. 235–252.
- [3] A. Nowotna, B. Pietruszka, and P. Lisowski, "Eco-Friendly Building Materials," *IOP Conference Series: Earth and Environmental Science*, vol. 290, no. 1, Mar. 2019, Art. no. 012024, <https://doi.org/10.1088/1755-1315/290/1/012024>.
- [4] "Precast Concrete Market Size, Share & Trends Analysis Report By Product (Structural Building Components, Transportation Products), By Application (Residential, Commercial, Infrastructure), And Segment Forecasts, 2024 - 2030," Grand View Research, GVR-2-68038-360-7.
- [5] H. Vu, T. Frydl, T. Bastl, P. Dvorak, E. Kristianova, and T. Tomasko, "Recent Development in Metal Extraction from Coal Fly Ash," in *Clean Coal Technologies: Beneficiation, Utilization, Transport Phenomena and Prospective*, R. K. Jyothi and P. K. Parhi, Eds. New York, NY, USA: Springer, 2021, pp. 575–603.
- [6] N. Bheel, M. A. Jokhio, J. A. Abbasi, H. B. Lashari, M. I. Qureshi, and A. S. Qureshi, "Rice Husk Ash and Fly Ash Effects on the Mechanical Properties of Concrete," *Engineering, Technology & Applied Science Research*, vol. 10, no. 2, pp. 5402–5405, Apr. 2020, <https://doi.org/10.48084/etasr.3363>.
- [7] C. L. Wong, K. H. Mo, U. J. Alengaram, and S. P. Yap, "Mechanical strength and permeation properties of high calcium fly ash-based geopolymer containing recycled brick powder," *Journal of Building Engineering*, vol. 32, Nov. 2020, Art. no. 101655, <https://doi.org/10.1016/j.jobe.2020.101655>.
- [8] M. Amran, S. Debbarma, and T. Ozbakkaloglu, "Fly ash-based eco-friendly geopolymer concrete: A critical review of the long-term durability properties," *Construction and Building Materials*, vol. 270, Feb. 2021, Art. no. 121857, <https://doi.org/10.1016/j.conbuildmat.2020.121857>.
- [9] B. Uzbas and A. C. Aydin, "Microstructural Analysis of Silica Fume Concrete with Scanning Electron Microscopy and X-Ray Diffraction," *Engineering, Technology & Applied Science Research*, vol. 10, no. 3, pp. 5845–5850, Jun. 2020, <https://doi.org/10.48084/etasr.3288>.
- [10] M. A. Rahman, M. F. S. Zawad, and S. N. Priyom, "Potential use of microsilica in concrete: A critical review," in *5th International Conference on Advances in Civil Engineering*, Chittagong, Bangladesh, Mar. 2021.
- [11] M. T. S. V. Saxena and A. Verma, "Evaluation of Physical and Chemical Properties Concrete Incorporating Silica Fume and Nano-SiO₂," *International Journal of Scientific Research & Engineering Trends*, vol. 6, no. 4, pp. 2169–2172, 2020.
- [12] Z. Peng, W. fan, G. Qian, Z. Tao, W. You-tuan, and Q. Xueyuan, "Use of coal-fired slag in filling bodies with early strength for mining applications," *Journal of Cleaner Production*, vol. 414, Aug. 2023, Art. no. 137465, <https://doi.org/10.1016/j.jclepro.2023.137465>.
- [13] M. A. Saafan, Z. A. Etman, and D. M. El lakany, "Microstructure and Durability of Ground Granulated Blast Furnace Slag Cement Mortars," *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, vol. 45, no. 3, pp. 1457–1465, Sep. 2021, <https://doi.org/10.1007/s40996-020-00533-3>.
- [14] V. Francioso, C. Moro, and M. Velay-Lizancos, "Effect of recycled concrete aggregate (RCA) on mortar's thermal conductivity susceptibility to variations of moisture content and ambient temperature," *Journal of Building Engineering*, vol. 43, Nov. 2021, Art. no. 103208, <https://doi.org/10.1016/j.jobe.2021.103208>.
- [15] S. A. Memon, Z. Bekzhanova, and A. Murzakarimova, "A Review of Improvement of Interfacial Transition Zone and Adherent Mortar in Recycled Concrete Aggregate," *Buildings*, vol. 12, no. 10, Oct. 2022, Art. no. 1600, <https://doi.org/10.3390/buildings12101600>.
- [16] K. V. Kumar and N. V. Rao, "Experimental study on recycling and reuse of municipal solid waste in construction industry," *Journal of Engineering Sciences*, vol. 14, no. 1, 2023.
- [17] A. Juan-Valdes *et al.*, "Mechanical and microstructural properties of recycled concretes mixed with ceramic recycled cement and secondary recycled aggregates. A viable option for future concrete," *Construction and Building Materials*, vol. 270, Feb. 2021, Art. no. 121455, <https://doi.org/10.1016/j.conbuildmat.2020.121455>.
- [18] Y. Liu *et al.*, "Enhancing the sustainable immobilization of phosphogypsum by cemented paste backfill with the activation of γ -Al₂O₃," *Construction and Building Materials*, vol. 347, Sep. 2022, Art. no. 128624, <https://doi.org/10.1016/j.conbuildmat.2022.128624>.
- [19] A. R. D. Costa, S. R. C. Matos, G. Camarini, and J. P. Gonçalves, "Hydration of sustainable ternary cements containing phosphogypsum," *Sustainable Materials and Technologies*, vol. 28, Jul. 2021, Art. no. e00280, <https://doi.org/10.1016/j.susmat.2021.e00280>.
- [20] J. Rosales *et al.*, "Treated phosphogypsum as an alternative set regulator and mineral addition in cement production," *Journal of Cleaner Production*, vol. 244, Jan. 2020, Art. no. 118752, <https://doi.org/10.1016/j.jclepro.2019.118752>.
- [21] M. Pliaka and G. Gaidajis, "Potential uses of phosphogypsum: A review," *Journal of Environmental Science and Health, Part A*, vol. 57, no. 9, pp. 746–763, Jul. 2022, <https://doi.org/10.1080/10934529.2022.2105632>.
- [22] H. Tayibi, M. Choura, F. A. Lopez, F. J. Alguacil, and A. Lopez-Delgado, "Environmental impact and management of phosphogypsum," *Journal of Environmental Management*, vol. 90, no. 8, pp. 2377–2386, Jun. 2009, <https://doi.org/10.1016/j.jenvman.2009.03.007>.
- [23] C. Ma, G. Chen, L. Cao, H. Zhou, and W. Ren, "Effects and mechanisms of waste gypsum influencing the mechanical properties and durability of magnesium oxychloride cement," *Journal of Cleaner Production*, vol. 339, Mar. 2022, Art. no. 130679, <https://doi.org/10.1016/j.jclepro.2022.130679>.
- [24] Y. Huang *et al.*, "Phosphogypsum as a component of calcium sulfoaluminate cement: Hazardous elements immobilization, radioactivity and performances," *Journal of Cleaner Production*, vol. 248, Mar. 2020, Art. no. 119287, <https://doi.org/10.1016/j.jclepro.2019.119287>.
- [25] M. B. Ali, R. Saidur, and M. S. Hossain, "A review on emission analysis in cement industries," *Renewable and Sustainable Energy Reviews*, vol. 15, no. 5, pp. 2252–2261, Jun. 2011, <https://doi.org/10.1016/j.rser.2011.02.014>.
- [26] A. S. Jaafar, Z. K. Abbas, and A. A. Allawi, "Investigating the Ability of producing Sustainable Blocks using Recycled Waste," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12006–12011, Dec. 2023, <https://doi.org/10.48084/etasr.6357>.
- [27] I. H. Shah, S. A. Miller, D. Jiang, and R. J. Myers, "Cement substitution with secondary materials can reduce annual global CO₂ emissions by up to 1.3 gigatons," *Nature Communications*, vol. 13, no. 1, Sep. 2022, Art. no. 5758, <https://doi.org/10.1038/s41467-022-33289-7>.

Rational Pile Design using Computer-based Program Coding in Matlab: A Case Study

Cao Van Hoa

Department of Construction Technology, Faculty of Civil Engineering, Ho Chi Minh City University of Architecture, Vietnam

hoa.caovan@uah.edu.vn (corresponding author)

Received: 5 January 2024 | Revised: 17 January 2024 | Accepted: 19 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6867>

ABSTRACT

Scientific approaches to pile design have made significant progress in recent years. However, despite these advancements, estimating the axial resistance and settlement of piles still heavily relies on empirical correlations. The design of resource-efficient and environmentally friendly piles is a pressing need. Yet, there is no explicit theoretical or practical experience to guide pile design rationally. Typically, determining a pile's resistance and settlement is treated as separate problems without considering the interactions between the pile and the soil. Additionally, soil data are inconsistent due to the heterogeneous and isotropic character of the soil in the half-space under the foundation. In this study, the modified Fellenius Unified method was coded in Matlab and applied to analyze pile behavior, considering the resistance and settlement of each pile, as well as interactions between piles and the soil simultaneously. The results showed that this approach is promising for practical applications. Moreover, its implementation in the evaluation of pile design for an apartment project in Binh Duong, Vietnam, suggests that the pile's length can be reduced even further than it currently is.

Keywords-pile design; computer-based program; rational design; resistance; pile settlement

I. INTRODUCTION

To rationally design piles, it is often necessary to determine the accuracy and representativeness of soil parameters through a large number of tests and the best selection of testing methods, including sampling methods, laboratory test types, and onsite tests. A pile design method should simultaneously estimate the resistance and settlement at the head, toe, and along the pile body. In [1], 63 static pile load tests were examined to validate and refine well-known pile design methods based on published soil properties. Many studies utilized statistical approaches in safety selection to suggest improvements to the existing standards for pile design [2-3]. This approach distinguishes between the variability of pile resistance within individual sites and the global variability on which model correlations are based.

Currently, there are many methods available to estimate the resistance of a pile [4]. This article uses the resistance calculation methods proposed by the Vietnamese National Standard (VNS) [5] to increase practical applicability. The VNS considers the effect of negative friction using an impact factor on the resistance at the pile head. However, the unified method approach for negative friction estimation is also utilized. In [6], it was shown that when dealing with negative skin friction, pile design codes treat the drag force as an unfavorable design load that should be imposed on the pile. These codes increase the value of the drag force while ignoring the shaft resistance above the neutral plane and decreasing the positive shaft resistance below the neutral plane. This means

that pile design in deep, soft clays, such as those found in Singapore and Asian coastal plains, will result in excessively long piles to meet the code requirements. In [7], the level of confidence that can be placed (a) in the conceptual and analytical frameworks to estimate pile resistance and (b) in the quantitative parameters required for design were discussed. From a practical point of view, it is necessary to design approaches that are less sensitive to the estimated pile resistance. In general, resistance is defined by shaft friction resistance and toe resistance. However, several studies considered both resistance and load along the entire length of the pile, from the pile toe to the pile head, to overcome these limitations [4, 8-10].

For settlement estimation, design codes typically estimate pile settlement by considering the settlement of the equivalent raft for the entire pile group [5, 11-12]. This is then assumed to be the settlement of the foundation at the bottom of the raft, which is also at the head of the piles. However, various studies [4, 8-10, 13-14] indicated that pile displacement and deformation can differ from pile head to toe. Traditional settlement calculation methods do not consider the load-settlement behavior at the pile toe or soil-pile and pile-pile interactions. In [15], the philosophy of using piles as settlement reducers was discussed along with the conditions under which such an approach can be successful. In [16], designs for the undrained behavior of a piled raft system criteria were developed based on examining the average and differential settlements, the raft bending moment, and the load ratio to be born by the pile.

In [17-18], an optimal design method was presented to determine the pile length for piled raft foundations. The method considered raft, piles, and soil interaction to evaluate foundation settlement. Settlement analysis was simplified using Steinbrenner's equation, and the total pile length was minimized while considering the settlement constraint. On the other hand, in [19], it was argued that a conventional piled foundation is designed to provide sufficient resistance and limit overall settlement, including controlling differential settlement within acceptable limits. As such, piles are often the same length and size. In [20], the design parameters for the CFG pile compound foundation were empirically investigated using PLAXIS software to evaluate different pile lengths and analyze their relationship with various factors. The pile length was optimized and found that it could be 2 m shorter than the conventional design, reducing costs. Construction practice did not show any problems during operation.

In [21], the optimal conceptual design of pile foundations was studied during the initial design stage, developing a minimum-cost optimization model that considered multiple design constraints based on the Chinese code and a cardinality constraint. This model aimed to achieve the concurrent optimization of pile size and layout. In [22], an optimization problem was proposed to achieve the most economical design of the pile foundation layout. This study compared the performance of two different design procedures in assessing each candidate design obtained during the optimization process. The proposed formulation was validated on a real-world structure. Furthermore, in [23], a user guide was presented for a computer program called PILEOPT, which can help to analyze and design pile foundation layouts. In [24], an automated optimal design method was introduced that used a hybrid genetic algorithm for pile group foundation design. This method aimed to optimize the size and pile arrangement and minimize the volume of the material utilized in the foundation. To achieve this, the piles' configuration, number, and cross-sectional dimensions, along with the thickness of the pile cap, were considered as design variables. The proposed hybrid genetic algorithm successfully minimized the volume of material consumption, and the result aligned with the engineering expectations. In [25], a study on dynamic soil-pile interaction was performed, proposing a model that considered the strong nonlinearity near the pile shaft during dynamic loading. The model engaged the Winkler's hypothesis and paid particular attention to the gap that forms at the soil-pile interface. The former successfully predicted the dynamic pile response observed in field dynamic pile load tests. On the other hand, in [26], different objective procedures were evaluated to reduce uncertainty in the design process. This study focused on three factors: adopting a pile resistance model, selecting a soil strength and soil profile for an ultimate limit state check, and estimating pile head settlement for a serviceability limit state check.

Numerous studies have investigated rational pile design with the primary objective of reducing the amount of materials used. Several constraints, such as resistance, pile settlement, pile arrangement in the foundation, and soil deformation, were considered. Many studies simulated pile behavior using the Finite Element Method (FEM) or elastic solutions combined

with Winkler springs, considering both pile-soil and pile-pile interactions. This study employs the modified unified pile design method that considers down drag - Pile Design with Consideration of Down-Drag (PDWDD) [8-10]. The resistance of the pile was determined using the formula guided by VNS [5], while the settlement of the soil around the pile and the settlement of the pile were determined by the hybrid method, which combines FEM and theory of elasticity.

II. RESEARCH SIGNIFICANCE

This study uses the pile design of Connect 2 Apartment, Binh Duong Province, Vietnam, to verify four objective functions: the load function that acts on each pile element considering negative friction, the resistance function of the pile, the allowable displacement function of the pile, and the settlement function of the soil causing negative friction. To obtain high applicability for design practice, the resistance function of the pile was determined using the VNS-recommended formulas. This study determines the displacement of each pile instead of calculating the settlement of the entire pile group according to the equivalent raft method, which has no relation to the pile's resistance calculation. The displacement of each pile was determined through the elastic solution pile-soil interaction, in which the pile is simulated by bar elements and the soil by springs attached to pile nodes.

III. MATERIALS AND METHODS

A. Load, Resistance, Pile Displacement, and Soil Settlement

The function below defines the load acting along pile length, including down-drag load:

$$f_1(z_i) = P_i + U_{i,n} \Delta \sum_0^{n_{zi}} (6.25N_{i,n}C + 3.33N_{i,n}(1 - C)) \quad (1)$$

The function below defines the resistance of a pile along its length.

$$f_2(z_i) = Q_{ult,i} - U_{i,n} \Delta \sum_0^{n_{zi}} (6.25N_{i,n}C + 3.33N_{i,n}(1 - C)) \quad (2)$$

The ultimate resistance can be calculated by:

$$Q_{ult,i} = \alpha N_{iL} A_{i,n} + U_{i,n} \Delta \sum_0^{n_{Li}} (6.25N_{i,n}C + 3.33N_{i,n}(1 - C)) \quad (3)$$

where, i is the pile number, n is the number of soil layers in the range from pile top to pile toe (pile segment), n_{zi} is the number of soil layers to depth z_i , n_{Li} is the number of soil layers to pile toe of pile i (to the depth of pile length of L), α is the coefficient depending on pile type (30 for driven, press-in piles, or 15 for bore piles), N_{iL} is the SPT value at the toe of the pile, $N_{i,n}$ is the SPT value of n -th soil layer at pile i , $C = 1$ if it is a clayey layer and $C = 0$ if it is a sandy layer, A_i is the cross-section of pile i , and U_i is the perimeter of pile i . Cumulative pile settlement along its length can be estimated by:

$$f_3(Z_i) = S_{allow} - 2\Delta \sum_0^{n_{zi}} (P_{i,n} n / 3E_{i,n} A_{i,n}) \quad (4)$$

Soil settlement along pile length is estimated by:

$$f_4(z_i) = \sum_0^{z_i} w_i \quad (5)$$

where settlement at any point in the subsoil is a sum of settlement caused by load acting at all pile's nodes.

$$w_i = \sum_1^j a_{ij} P_j \quad (6)$$

where j is the node number, S_{allow} is the allowable settlement depending on national codes, $P_{i,n}$ is the external load acting at the pile node, $E_{i,n}$ is Young's modulus of pile material, \bar{G} is the average shear modulus, and r_{ij} , c_j , $R_{1,ij}$, and $R_{2,ij}$ are defined in the Mindlin formula.

$$a_{ij} = \frac{1}{16\pi\bar{G}(1-\nu)} \left[\frac{3-4\nu}{R_{1,ij}} + \frac{8(1-\nu)^2-(3-4\nu)}{R_{2,ij}} + \frac{(z_i-c_j)^2}{R_{1,ij}^3} + \frac{(3-4\nu)(z_i+c_j)^2-2c_jz_k}{R_{2,ij}^3} + \frac{6c_jz_i(z_i+c_j)^2}{R_{2,ij}^5} \right] \quad (7)$$

Functions (1), (2), (4), and (5) are used to simulate pile behavior, both settlement and resistance.

B. Mathematical Model

The goal is to determine the length L of the piles such that functions (1) and (2) and functions (3) and (4) are equal.

$$f_1(z_i) = f_2(z_i) \quad (8)$$

$$f_3(z_i) = f_4(z_i) \quad (9)$$

Equation (8) can be solved to find $z_{(1,2)}$, and (9) can be solved to find $z_{(3,4)}$. Adjust so that $z_{(1,2)} \geq z_{(3,4)}$. In case $z_{(1,2)} = z_{(3,4)}$, the pile is considered to have the optimal length.

C. Matlab Modeling

The Matlab model was based on the one described in [10]. The computation of functions (1), (2), (4), and (5) were coded in Matlab. For simplification, the soil was characterized by a single parameter, N-SPT. The N-SPT value can be calibrated by the results of static loading and strain tests conducted on two test piles, TP01 and TP02, and should be compared and adjusted with the settlement monitoring findings during the calculation. The input data required for this analysis include the coordinates of the pile, its length and diameter, the number of piles and pile elements, the N-SPT number of each soil layer, the N-SPT value under the pile tip, Poisson's index of the soil, allowable settlement of the pile, and elastic modulus of the pile material. The pile head load can be calculated using the SAFE software. This study used the load in the designer calculation sheets for comparison. The friction resistance (negative and positive) was calculated using the VNS formula, which is similar to that proposed by the Japanese Architectural Institute. The load and down-drag force due to negative friction were calculated for each pile node according to (1). The pile resistance at each node was determined using (2). The displacement at the pile head was taken as the allowable displacement according to VNS. The displacement at the pile tip was calculated from the pile head displacement minus the elastic deformation, according to (4). The soil settlement at any node was computed using the first Mindlin solution. The soil

settlement at any analyzed node along the pile equals the sum of the settlement caused by loads acting at all other nodes in the pile group through the Mindlin solution. The cumulative settlement of the analyzed node was calculated using (5), counting from the pile tip.

IV. THE APARTMENT BUILDING CONNECT 2

A. The Connect 2

Connect 2 consists of two blocks. Drilling of boreholes of test piles started in May 2020, followed by basement excavation and foundation construction in October 2020. The superstructure construction began in November 2021, and all construction work ended in April 2023. During construction, no accidents occurred, and to this day, the apartment building is in regular operation. During the design period, after reviewing the design, the owner requested to minimize the pile's cost. The building is a 30-story tower that is 97.25 m high. Its foundation was designed as a pile group consisting of 177 piles. Among these piles, 5 are 1 m in diameter and 45 m long, 122 are 1 m in diameter and 52 m long, and 50 are 1.2 m in diameter and 57 m long. The piles with a length of 45 m have their toes located in coarse sand, which is dense to very dense, with an N-SPT of about 36. The piles of 50 or 57 m long have their toes in medium-sized sand, which is yellow to white-grey and dense, with an N-SPT value of about 37. The underground water level is at a depth of 11 m. Figure 1 shows the arrangement plan for the piles. The red marked piles, which are 57 m long and have a diameter of 1.2 m, were placed under the stairs and elevator case. The green and blue marked piles, which were 50 m long and had a diameter of 1 m, were arranged under the column raft and the shear wall. Five uncolored piles were placed under the swimming pool, measuring 45 m long and 1 m in diameter. Table I displays the N-SPT value by depth.

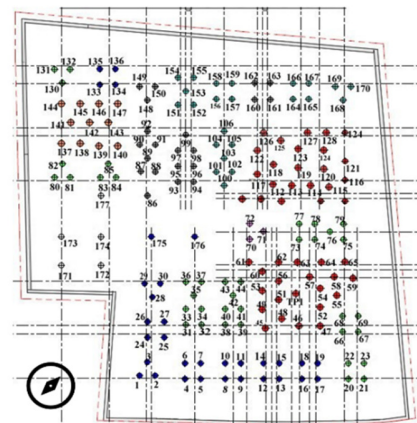


Fig. 1. Pile arrangement.

Static loading and strain tests were performed on two piles, TP01 and TP02, by the Vibrobis company, which has its headquarters in Ho Chi Minh City. The tests started from the end of May to the beginning of June 2020. The load was applied to the pile in two cycles, 100%, and 200% design resistance, complying with VNS. The strain gauges were mounted on the test piles and conducted in parallel with the

static loading test. There were 21 strain gauges mounted at seven points along the pile body, with a distance of 7-10 m between each point. The test results are summarized as follows:

- Test pile TP01 had a diameter of 1.2 m and a length of 62.3 m. Static loading testing, combined with strain measurement, showed that when the pile was loaded with 27,000 kN (equivalent to 200% design load), the load distributed for the pile tip was only 1,010 kN, about 3.7% of the pile head load. During the test, the pile head moved 34.12 millimeters. After comparing the test results with the calculations, the design engineer concluded that the resistance of piles in groups should range between 12,500 and 13,000 kN.
- Test pile TP02 had a diameter of 1.0 m and a length of 53.3 m. The static loading and strain measurement test results showed that with a test load of 18,000 kN (equivalent to 200% design load), the distributed load at the pile tip was only about 680 kN. The displacement at the pile head was 27.7 mm. The designer evaluated and determined the pile's resistance in groups ranging from 9,000 to 9,500 kN.

TABLE I. SOIL PROPERTIES

No.	Soil layer	Depth (m)	N-SPT
1	Sandy clay, medium stiff	10-16	13.5
2	Fine sand, medium-dense	17-32	16.1
3	Clay, stiff	33-36	20.25
4	Fine sand, medium-dense	37-39	17.5
5	Clay, red-brown, stiff	40-50	32.7
6	Sandy clay, medium stiff	51-53	32
7	Coarse sand, very dense	54-56	41.3
8	Clay, stiff	57-58	64
9	Medium sand, dense	59-67	40

B. Results Analyzed by Matlab

The analysis using Matlab was conducted on a total of five piles. It was found that piles No. 1, No. 21, No. 131, and No. 170, located at the corners of the building, had displacements of 3.4 mm to approximately four mm. Figures 2, 3, 4, 5, and 6 visually represent the relationship between load, resistance, pile displacement, and soil settlement of the piles. A pile is considered capable of supporting the load and meeting the allowable settlement requirement when the curve of allowable settlement (y_3) of the pile intersects with the curve of soil settlement 1 (y_4) at a depth smaller than the intersection of the load curve (y_1) and pile resistance curve (y_2).

Figure 2 shows the correlation between load, resistance, pile displacement, and soil settlement for pile No. 1. The load and resistance equilibrium plane is located at an elevation of -38.1 m, while the pile displacement and soil settlement equilibrium plane is at an elevation of -36.8 m. The results show that the pile is capable of bearing the load and that the settlement of the pile is approximately 3.4 mm. Figure 3 shows the correlation between load, resistance, pile displacement, and soil settlement for pile No. 21. The results indicate that the pile is capable of bearing the load (the load and resistance curves intersect at -38.3 m deeper than the intersection of the pile and soil settlement curves at -25.7 m). The soil settlement is approximately 3.6 mm.

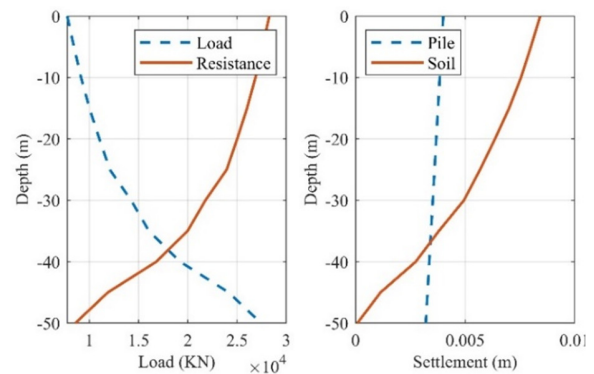


Fig. 2. Pile No. 1: load-resistant and pile-soil settlement relationship.

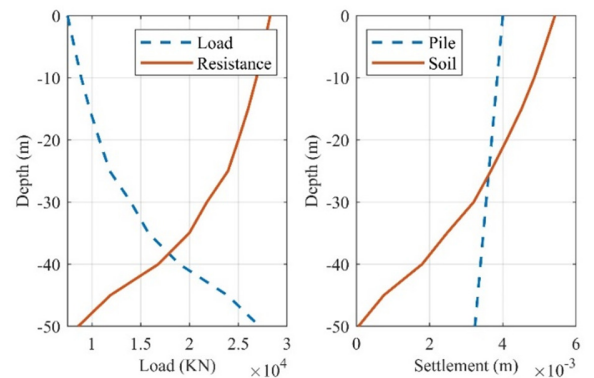


Fig. 3. Pile No.2: load-resistant and pile-soil settlement relationship.

Figures 4 and 5 show that the piles located at the remaining two corners, No. 131 and 170, respectively, are also capable of supporting the load (38.7 vs. 39.7 m and -39.4 vs. -30.1 m), with settlements of 3.5 and 3.7 mm, respectively. For pile No. 131, if the allowable settlement is approximately 4.0 mm, the pile resistance will be greater than the pile load (the load and resistance curves intersect at -38.8 m at a lower elevation than the intersection of pile and soil settlement curves at -37.7 m), indicating it is capable of bearing load. Figure 6 displays the behavior of pile No. 57, which is located beneath the raft of the elevator core. The pile can bear the load because the load and resistance curves intersect at -38.3 m deeper than the intersection of the pile and soil settlement curves at -25.7 m. The settlement of this pile is only 3.5 mm.

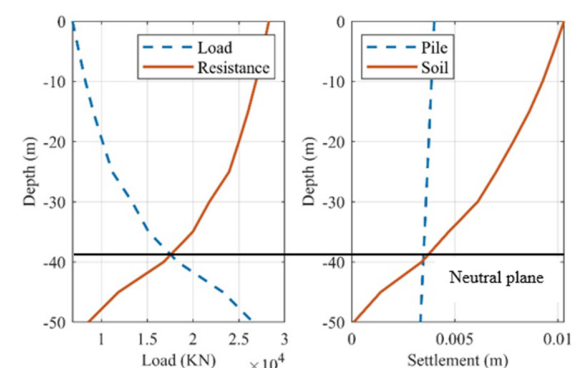


Fig. 4. Pile No. 131: load-resistant and pile-soil settlement relationship.

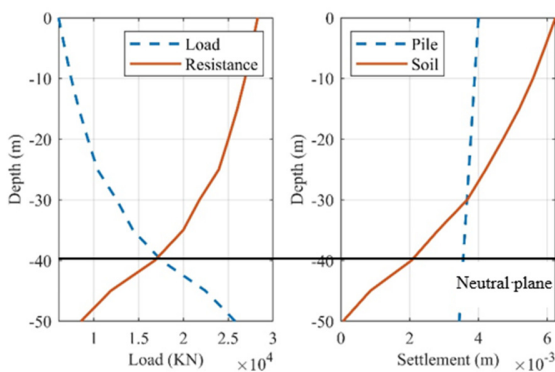


Fig. 5. Pile No. 170: load-resistant, pile-soil settlement relationship.

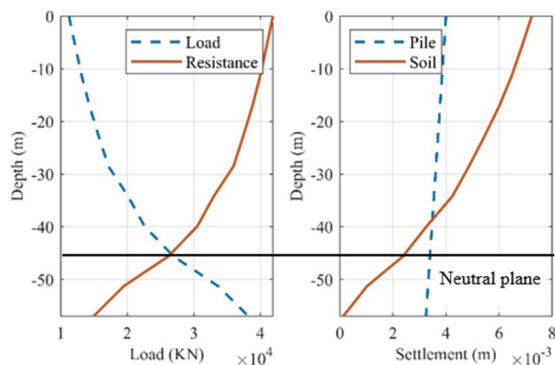


Fig. 6. Pile No. 57: load-resistant and pile-soil settlement relationship.

The results of analyzing 177 piles in the studied pile group using Matlab showed that the average settlement value was determined to be 3.56 mm, and the differential settlement was approximately 1 mm. The settlement monitored during the 16th measurement period on April 19, 2023, was recorded between 1.49 and 2.39 mm. It should also be noted that this is the monitoring result right after the project was completed and handed over for use, so the settlement is still not completed. Over time, the settlement of the soil can become much more. In addition, the settlement calculated from the program is elastic so that the actual soil settlement may reach a more significant value in the long run. However, it can be said that the pile group was designed safely and met all VNS requirements. Comparing the computed results of 3.6 mm with the static loading test results of about 10 mm at 100% design load and the settlement at the monitoring period at the end of construction of about 2 mm, it can be concluded that they are relatively consistent. Although the final monitoring results are expected to be delivered six months and one year later, it is evident that this research direction shows promise.

C. Discussion

This study used the statistical N-SPT value, as suggested by experimental engineers, to determine the behavior of the soil. However, it should be noted that the calculated N-SPT value may vary significantly due to factors, such as energy loss caused by friction between the penetrating rod and the soil, the experimenter's experience, and friction between the falling hammer and the shaft. The graphs in Figures 2-6 show that the curves of functions (1), (2), (4), and (5) do not intersect in a

single neutral plane. Therefore, an optimal pile design analysis should be further studied.

This analysis estimates soil settlement based on the elastic pile-soil interaction. The two factors that affect soil settlement are the load applied at the pile nodes and the elastic modulus of the soil. The load distributed to pile nodes was determined according to Coyle and Reese (1966). The value of the load acting at the pile tip was calculated from tip settlement and soil modulus, which is calibrated with the results of loading tests and strain tests at two test piles, TP01 and TP02, for problem simulation. The settlement of soil under the raft that caused negative friction was estimated based on Mindlin's first solution. Although the estimation of soil settlement under the raft was based on elastic modulus, it does not reflect the actual settlement of the soil. However, according to many studies, the elastic simulation of soil by springs fully represents the pile-soil interaction. Based on a preliminary analysis of pile No. 57, it was observed that the elastic Young modulus variation resulted in different soil settlement values. If the value of Young's modulus equaled the value of the modulus of deformation, the estimated soil settlement at ground level was approximately 82 mm and the pile displacement was 4.4 mm. However, if the elastic modulus used is calculated by $E_{elas.} = 1000 \times N$, the pile displacement is only 3.7 mm, and the ground surface settlement is 7.3 mm.

A pilot study was carried out for a single pile cap, PC09, which consists of 11 piles with a diameter of 1.0 m and a length of 52 m, of the same project currently under study. The analysis conducted in pile No. 35 showed that the load equilibrium plane was at an elevation of -45 m, the settlement equilibrium plane was at a level of -15 m, and the settlement was found only 2 mm. However, this analysis only considered 7 of the total 177 piles in the foundation, leading to incorrect results. When considering the interaction of all 177 piles, the load equilibrium plane was at an elevation of -45 m, the settlement equilibrium plane was at a level of -38 m, and the settlement was 3.9 mm. This indicates that the pile should consider the entire pile-soil interaction of all piles to produce results closer to reality. Therefore, it is essential to consider the radius of influence of pile-soil interaction when designing piles for large foundations, which will be incorporated into the computer program being developed.

A project verification company suggested shortening the piles to reduce the total length from 9,669 m (according to the initial design) to 8,765 m, saving 904 meters. After considering the proposal, the design company decided on a total pile length of 9,175 m, offering significant cost savings. It is important to note that there is no established theoretical basis for reducing pile length before construction, even though actual construction later proves the feasibility. The computed results showed that the load and settlement equilibrium planes of the 177 piles were at various elevations. Therefore, it is necessary to match the equilibrium planes to optimize the pile lengths. Using the pile design method, considering PDWDD combined with Matlab, engineers can design piles more reasonably, as it can analyze pile resistance and settlement simultaneously and provide a theoretical basis for designing piles. This method can also help optimize pile lengths.

V. CONCLUSION

Using the Matlab-coded PDWDD method to analyze the resistance and settlement of piles simultaneously while also considering long-term negative friction provides engineers with a sense of security when designing piles. This study investigated the behavior of the 177 piles in the foundation system of the selected construction project, drawing the relationships between the resistance and load and between the pile's allowable displacement and soil settlement. The loads at one pile node cause the settlement of the soil around other nodes in the pile group, which in turn causes negative friction to the nearby piles. This negative friction is the main reason for the long-term displacement of the pile. Based on the initial findings, the settlement calculated by the program (~3.56 mm) matches well with the monitoring results (~2.1 mm). The computing time to calculate the resistance and settlement of the project's pile group was approximately two minutes. Therefore, the proposed computer-based program promises to have practical applications.

According to the findings, the total length of the pile system used in the project can save more than 494 m compared to the initial design that followed VNS. The Matlab technique demonstrates that the length of piles in the pile group can be reduced even more. Most importantly, the technique has provided engineers with a theoretical basis for pile behavior to select suitable economical design options. Shortening the pile length caused the elevation of the load and the settlement equilibrium planes to change. For these two planes to coincide, the settlement of the pile must be increased within the requirements of the codes. Therefore, accepting a significant settlement can result in a reasonable pile length. The elastic settlement of the soil is usually small, so engineers use their experience to predict long-term settlement. This study provides a theoretical foundation for designing appropriate piles and predicting long-term settlement. In future studies, research will be carried out to enhance the Matlab program, calibrate soil characteristic parameters, and verify the program through actual projects.

ACKNOWLEDGMENT

The author is grateful to Dr. D. N. Truong for his invaluable support and guidance while coding in Matlab.

REFERENCES

- [1] Y. Robert, "A few comments on pile design," *Canadian Geotechnical Journal*, vol. 34, no. 4, pp. 560–567, Aug. 1997, <https://doi.org/10.1139/t97-024>.
- [2] G. B. Baeher and R. Rackwitz, "Factors of safety and pile load tests," *International Journal for Numerical and Analytical Methods in Geomechanics*, vol. 6, no. 4, pp. 409–424, 1982, <https://doi.org/10.1002/nag.1610060404>.
- [3] M. A. Soomro, A. S. Brohi, M. A. Soomro, D. K. Bangwar, and S. A. Bhatti, "3D Numerical Modeling of Pile Group Responses to Excavation-Induced Stress Release in Silty Clay," *Engineering, Technology & Applied Science Research*, vol. 8, no. 1, pp. 2577–2584, Feb. 2018, <https://doi.org/10.48084/etasr.1748>.
- [4] B. H. Fellenius, "Unified Design of Piled Foundations with Emphasis on Settlement Analysis," in *Current Practices and Future Trends in Deep Foundations - Contributions in Honor of George G. Gobel*, Los Angeles, CA, USA, Apr. 2012, pp. 253–275, [https://doi.org/10.1061/40743\(142\)15](https://doi.org/10.1061/40743(142)15).
- [5] "TCVN 10304: Pile Foundation - Design Standard." Vietnam Ministry of Science and Technology, 2014.
- [6] S. A. Tan and B. H. Fellenius, "Negative skin friction pile concepts with soil-structure interaction," *Geotechnical Research*, vol. 3, no. 4, pp. 137–147, Dec. 2016, <https://doi.org/10.1680/jgere.16.00006>.
- [7] M. F. Randolph, "Science and empiricism in pile foundation design," *Géotechnique*, vol. 53, no. 10, pp. 847–875, Dec. 2003, <https://doi.org/10.1680/geot.2003.53.10.847>.
- [8] C. V. Hoa, "Pile design with consideration of down drag," in *Geotechnics for Sustainable Infrastructure Development*, 2020, pp. 145–151, https://doi.org/10.1007/978-981-15-2184-3_18.
- [9] H. V. Cao and T. A. Nguyen, "Verification and Validation of the Pile Design Method with Consideration of Down Drag," *International Journal of Geomate*, pp. 145–152, Aug. 2022.
- [10] H. C. Van and T. A. Nguyen, "Numerical Simulation of Pile Design Method that Considers Negative Friction," *Civil Engineering and Architecture*, vol. 11, no. 5, pp. 2285–2292, Sep. 2023, <https://doi.org/10.13189/cea.2023.110503>.
- [11] M. Tomlinson and J. Woodward, *Pile Design and Construction Practice*, 5th ed. London, UK: CRC Press, 2007.
- [12] M. A. Soomro, D. K. Bangwar, M. A. Soomro, and M. A. Keerio, "3D Numerical Analysis of the Effects of an Advancing Tunnel on an Existing Loaded Pile Group," *Engineering, Technology & Applied Science Research*, vol. 8, no. 1, pp. 2520–2525, Feb. 2018, <https://doi.org/10.48084/etasr.1693>.
- [13] P. Kitiyodom and T. Matsumoto, "A simplified analysis method for piled raft and pile group foundations with batter piles," *International Journal for Numerical and Analytical Methods in Geomechanics*, vol. 26, no. 13, pp. 1349–1369, 2002, <https://doi.org/10.1002/nag.248>.
- [14] P. Kitiyodom and T. Matsumoto, "A simplified analysis method for piled raft foundations in non-homogeneous soils," *International Journal for Numerical and Analytical Methods in Geomechanics*, vol. 27, no. 2, pp. 85–109, 2003, <https://doi.org/10.1002/nag.264>.
- [15] H. G. Poulos, "Piled raft foundations: design and applications," *Géotechnique*, vol. 51, no. 2, pp. 95–113, Mar. 2001, <https://doi.org/10.1680/geot.2001.51.2.95>.
- [16] A. Taghavi Ghalesari, A. Barari, P. Fardad Amini, and L. B. Ibsen, "Development of optimum design from static response of pile-raft interaction," *Journal of Marine Science and Technology*, vol. 20, no. 2, pp. 331–343, Jun. 2015, <https://doi.org/10.1007/s00773-014-0286-x>.
- [17] K. Nakanishi and I. Takewaki, "Optimum pile arrangement in piled raft foundation by using simplified settlement analysis and adaptive step-length algorithm," *Geomechanics and Engineering*, vol. 5, no. 6, pp. 519–540, 2013, <https://doi.org/10.12989/gae.2013.5.6.519>.
- [18] A. Firoozfar, A. Rostami, H. Ghaderi, H. Zamani, and A. Rostamkhani, "Assessing the Effects of Length, Slope and Distance between Piles on the Bearing Capacity of a Pile Group under Axial Loading in Granular Soil," *Engineering, Technology & Applied Science Research*, vol. 7, no. 5, pp. 1894–1899, Oct. 2017, <https://doi.org/10.48084/etasr.1352>.
- [19] Y. C. Tan, C. M. Chow, and S. S. Gue, "Piled raft with different pile length for medium-rise buildings on very soft clay," in *Proceedings of the 16th International Conference on Soil Mechanics and Geotechnical Engineering*, Osaka, Japan, 2005, pp. 2045–2048, <https://doi.org/10.3233/978-1-61499-656-9-2045>.
- [20] H. Y. Cao and Y. F. Liu, "Optimum Design of CFG Pile Compound Foundation Based on Numerical Simulation Method," *Applied Mechanics and Materials*, vol. 578–579, pp. 346–350, 2014, <https://doi.org/10.4028/www.scientific.net/AMM.578-579.346>.
- [21] X. Liu, G. Cheng, B. Wang, and S. Lin, "Optimum Design of Pile Foundation by Automatic Grouping Genetic Algorithms," *International Scholarly Research Notices*, vol. 2012, Art. no. e678329, Oct. 2012, <https://doi.org/10.5402/2012/678329>.
- [22] Ch. Letsios, N. D. Lagaros, and M. Papadarakakis, "Optimum Design of Pile Foundations," presented at the The Sixth International Conference on Engineering Computational Technology, Athens, Greece, 2008, Art. no. 180, <https://doi.org/10.4203/ccp.89.180>.

-
- [23] "User's Guide: Computer Program for Optimal Design and Analysis of Pile Foundations (PILEOPT)," Army Engineer Waterways Experiment Station, ADA099122, 1979.
- [24] J. T. M. Ng, C. M. Chan, and L. M. Zhang, "Optimum Design of Pile Groups in Nonlinear Soil using Genetic Algorithms," presented at the The Eighth International Conference on the Application of Artificial Intelligence to Civil, Structural and Environmental Engineering, Rome, Italy, 2009, Art. no. 35, <https://doi.org/10.4203/ccp.82.35>.
- [25] T. Nogami, J. Otani, K. Konagai, and H.-L. Chen, "Nonlinear Soil-Pile Interaction Model for Dynamic Lateral Motion," *Journal of Geotechnical Engineering*, vol. 118, no. 1, pp. 89–106, Jan. 1992, [https://doi.org/10.1061/\(ASCE\)0733-9410\(1992\)118:1\(89\)](https://doi.org/10.1061/(ASCE)0733-9410(1992)118:1(89)).
- [26] P. J. Vardanega, M. G. Williamson, and M. D. Bolton, "Bored pile design in stiff clay II: mechanisms and uncertainty," *Proceedings of the Institution of Civil Engineers - Geotechnical Engineering*, vol. 165, no. 4, pp. 233–246, Aug. 2012, <https://doi.org/10.1680/geng.11.00063>.

An IoT Smart System for Cold Supply Chain Storage and Transportation Management

Abdulrahman Alshdadi

Department of Information Systems and Technology, College of Computer Science and Engineering,
University of Jeddah, Saudi Arabia
alshdadi@uj.edu.sa

Souad Kamel

Department of Computer and Network Engineering, College of Computer Science and Engineering,
University of Jeddah, Saudi Arabia
skamel@uj.edu.sa (corresponding author)

Eesa Alsolami

Department of Cyber Security, College of Computer Science and Engineering, University of Jeddah,
Saudi Arabia
eaalsolami@uj.edu.sa

Miltiadis D. Lytras

Management of Information Systems Department, Deree College, The American College of Greece,
Greece
miltiadis.lytras@gmail.com

Sahbi Boubaker

Department of Computer and Network Engineering, College of Computer Science and Engineering,
University of Jeddah, Saudi Arabia
sboubaker@uj.edu.sa

Received: 2 January 2024 | Revised: 19 January 2024 | Accepted: 21 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6857>

ABSTRACT

Cold supply chains are becoming more and more attractive due to the high demand induced by increased consumption. To fulfill standards and customers' requirements regarding the conditions under which cold supply chain products (mainly foods and pharmaceuticals) are stored (in warehouses) and transported to the end-users, tracking those conditions is a necessity. To ensure a high level of visibility, fostering emerging technologies can improve the quality of service in supply chains in terms of delivery time, cost, and quality. In this paper, a global framework for monitoring the conditions of storage and transportation of cold products across the whole supply chain is designed and implemented practically. The proposed solution is built around low-cost and low-energy consumption devices such as sensors and microcontrollers which are connected to cloud storage to allow a high level of visibility in the supply chain allowing all parties, including the end-consumers, to follow the products during their transfer, providing a conceptual framework that monitors the performance on a real-time basis and enhances decision making. A prototype using an embedded temperature/humidity sensor, a tiny microcontroller equipped with a Wi-Fi connectivity device, and a mobile 4G/5G network is designed and implemented. The proposed system is connected to a cloud-storage platform continuously accessible by the main parties of the cold supply chain including the provider, the transporter, and the end-consumer. The proposed framework may be handled as a smart contract during which any party can assume its responsibility for the assurance of the best conditions of the supply chain operation. A small-scale real-life scenario conducted in Jeddah City, Saudi Arabia is introduced to show the feasibility of the proposed framework.

Keywords-cold supply chain; tracking; monitoring; microcontroller; sensors; cloud storage

I. INTRODUCTION

Internet of Things (IoT) has attracted much attention over the last few years in various domains, including the supply-chain. The main applications that contributed to the development of the supply-chain domain include traceability and tracking of the conditions under which various goods are stored and transported [1]. Food and pharmaceutical supply-chain involves products that require careful attention during their storage and transportation. The concept of food supply-chain virtualization refers to the opportunities provided by the IoT technology to the cold product sector decision-makers helping them to remotely control and monitor products that are prone to unpredictable events that may affect their safety [2]. While being transported from the source to the destination or warehouses, foods and pharmaceutical products should be kept at prescribed levels of temperature and humidity [3]. Warehouse management systems are in the heart of supply-chain operations to ensure compliance with Industry 4.0 requirements which are nowadays revolutionizing the logistics domain towards efficiency and sustainability [4]. As reported in [5], emerging logistics may include coupling between IoT and blockchain to ensure high efficiency of logistics and transportation operations. Logistics systems are progressively integrating Cyber Physical Systems (CPSs) and information technology as well as communication technologies like 5G. At the design stage of a supply-chain management system, the integration of cloud storage platforms and CPSs has emerged to cut with the traditional logistics systems that focused on vehicles, distribution centers and the characteristics of relevant distribution tasks [6]. In this direction, smart (intelligent) logistics (also called logistics 4.0) has attracted researchers and practitioners and has shown a high potential to ensure efficient tracking, control and automation of the supply-chain [7].

In this paper, an IoT-based storage and transportation system applied to cold supply chain products (including foods and pharmaceuticals), will be designed and implemented under the specific conditions of Jeddah City, Saudi Arabia. The main contributions of this study are:

- The design and implementation of a cost-effective management system for cold supply-chain allowing tracking the conditions under which cold products are stored and transported.
- The designed solutions are based on low-power consumption sensors and microcontrollers which makes them deployable in various and multiple supply-chain operations.
- A major feature of the proposed system is the possibility to extend it to implement smart contracts, where the violation of the contract conditions may lead to canceling the contract or any related operations.

II. LITERATURE REVIEW

Cold supply-chain management systems have attracted much attention due to their high ability to involve/integrate various technologies including IoT, blockchain, information technology, and cyber physical systems in unified frameworks. Authors in [8] integrated IoT and blockchain for efficient traceability in a frozen aquatic product. The developed system was reported to overcome

difficulties faced by traditional systems involving security, decentralized data management, and easy tampering of the exchanged data relevant to the logistics operations. As part of an efficient supply-chain, packaging related issues were considered in [9] using tools from the Model-Based Engineering (MBE). The IoT technology components, including sensors, were implemented in the pallets used to handle the products along the supply-chain. The collected information is expected to improve the control capabilities of the whole system. High-price shipment management and tracking systems in the smart city context were considered in [10]. To ensure a high level of transparency, a completely automated smart-contract was implemented involving all supply-chain parties and including all operations such as contract creation, money deposit, and the possibility of cancellation if any of the agreed conditions is violated. To prevent perishable products spoilage which accounts for high amounts of waste worldwide, real-time tracking and monitoring was considered in [11]. The main aim of the developed system was to detect anomalies in order to allow involved parties to react in order to limit waste and prevent dangers to consumers. The developed system architecture included four layers, namely perception, network, processing, and application. Additional applications of IoT in cold supply-chain included vaccine supply-chain during COVID-19 pandemic [12] and sensor agricultural for Chinese medicine logistics [13]. From a managerial perspective, the IoT combined with other technologies such as blockchain, has contributed to improve the management of supply-chain various operations including warehousing, manufacturing and delivery [14].

Table I summarizes the relative works and emphasizes on the controlled ambient parameters, involved technologies including datalogging, communication and related applications.

III. METHODOLOGY

A. System Design

Based on the literature review, it is seen that the trend in cold supply chain applied research focuses on the application of emerging technologies (IoT and blockchain) to two of the main components of a logistics framework, storage and transportation. Although the cold supply chain structure includes product manufacturing and transportation from the factory to the storage warehouse (Figure 1) and since the temperature is the main factor that may affect the quality of cold products, this paper will focus only on the temperature tracking and product management during the storage and transportation phases since they are more prone to difficulties. Factories are usually taking care about the conditions of their products manufacturing. In this section, the IoT-based system for tracking and monitoring the temperature in cold supply-chain is presented in detail. To design the system, a structured process including five steps is devised. The choice of the system components is based on the users' requirements, determined from those proposed in previous studies, our own observations, direct meetings with the stakeholders, or electronic surveys. Once the requirements are well-understood, the next step consists of choosing the system components that are able to fulfil the requirements while considering several criteria and constraints such as functionality, ease of use, technology, ease of implementation, component availability, cost, etc.

TABLE I. SUMMARY OF SELECTED STUDIES ON IOT IN COLD SUPPLY-CHAIN

Reference	Ambient parameter(s)	Technologies	Datalogging	Communication	Applications
[15]	Temperature, Relative Humidity (RH), O ₂ , CO ₂ , shock and vibration	IoT	Sensor nodes	Wireless communication networks	Fresh fruit and vegetables
[16]	Temperature (low values) for frozen products	IoT, Radio-Frequency Identification (RFID), blockchain	Sensor nodes	Wireless Sensor Networks (WSNs)	Fishery traceability and safety
[17]	Temperature, humidity, luminosity, and gas concentration	Real time monitoring and notification based on IoT + Artificial Neural Networks (ANNs) for classification	Multi-sensor system	Bluetooth, Internet	Management of cold product storage
[18]	Pressure	IoT, ANNs	Microcontrollers	WSNs in the distribution service layer	Coordination in the distribution of cold products based on IoT
[19]	Temperature, biosensors, gas sensors, freshness indicators	IoT and Communication	Not included	Various communication tools	Management of food supply chain
[20]	Freezing temperature	IoT, 5 G communication technology	FPGA	5 G networks, GPS system	Cold supply chain logistics system
[21]	Temperature, location, humidity, pressure, light exposure	Blockchain, IoT	Not included	RFID, Ethereum wallet, cloud messaging services	Smart contract based on IoT and blockchain

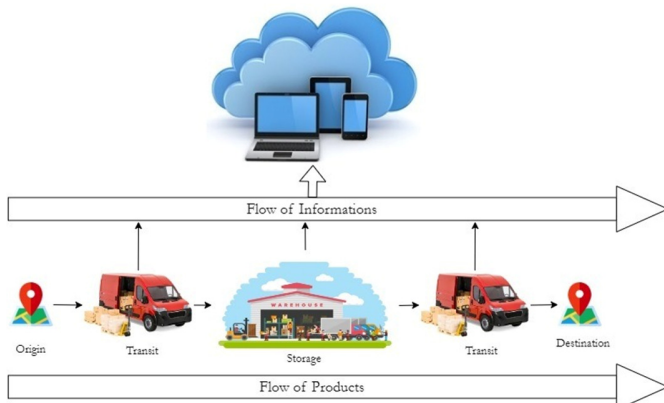


Fig. 1. Block Diagram of an IoT-enabled cold supply-chain management system.

B. Hardware Specifications

Once all components are chosen, the next step is to allocate to the prototype development which will be the base for later testing the efficiency of the system at small-scale before deploying it. The system architecture is provided as a block diagram in Figure 2. The chosen components (hardware), their technical specifications, and their related requirements are provided below.

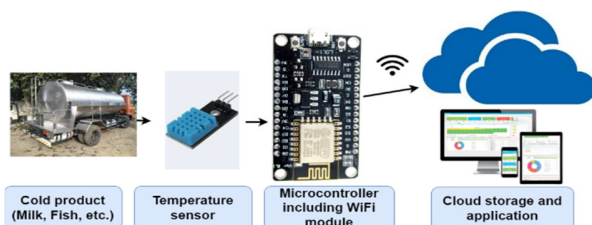


Fig. 2. System architecture.

1) Temperature Sensor

DHT22 was chosen as the temperature sensor for many reasons. First, the sensor provides the temperature in a reasonable range (-40°C to 50°C) which corresponds to a variety of cold and frozen products. As indicated in the foods' transportation regulations, the majority of fresh vegetables, foodstuff, fresh fruit and chilled dairy and dairy products transportation temperatures are covered by the DHT22. It is affordable, particularly for small-scale prototyping which is the case of our study. The sensor accuracy is defined as $\pm 0.5^{\circ}\text{C}$ (for temperature) and $\pm 1\%$ (for humidity) which can be considered as good values, since even the regulations are not so strict in terms of accuracy. In addition, the operating voltage (3.5 V to 5.5 V) and the operating current (0.3 mA (measuring) and 60 uA (standby)) are easily provided by small batteries that can be embedded on the transportation mean.

2) NodeMcu (ESP8266)

The NodeMCU (Node Microcontroller Unit) is an open-source hardware that can be programmed by the Arduino Integrated Development Environment (IDE). The microcontroller is built around a low-cost device called ESP8266. The ESP8266 is a 32-bit microcontroller that includes the crucial small-scale elements of a mini-computer such as a microprocessor and RAM. In addition, the microcontroller contains a networking (Wi-Fi) device that allows connectivity to the Internet. Those specifications made this microcontroller an excellent choice for the IoT-based tracking and monitoring system for supply-chain management developed in this study. The NodeMcu was chosen against other microcontrollers such as the Raspberry Pi for its affordable price, being open-source and the availability of a lot of materials and applications in various fields including online monitoring.

3) Battery

The battery is highly required for feeding the system, particularly during the product transportation. A 3.7 V Lithium-Ion battery was chosen for this system due to its reasonable capacity (2600 mAh) and affordable price (around 7 USD). In

addition, this battery is able to feed the NodeMcu microcontroller as well as the temperature sensors for acceptable periods of time.

C. Software Specifications

For the software components, the Arduino IDE to develop the application code and upload it in the microcontroller flash memory and the free cloud platform Adafruit were utilized. Adafruit is an IoT platform that provides cloud services including analytical tools for data collection, visualization, and analysis. At some extent, the provided services are free of charge which allows simple prototyping mainly at small scale. Although Adafruit can provide large-scale applications (paid services), this study will be limited to the free available services for cost limitation reasons.

IV. RESULTS AND DISCUSSION

A. Prototype Development

After completion of the system's various hardware and software components, the preliminary prototype was implemented. The prototype is depicted in Figure 3. The designed system consists of three layers including perception, network, and application [22-24].

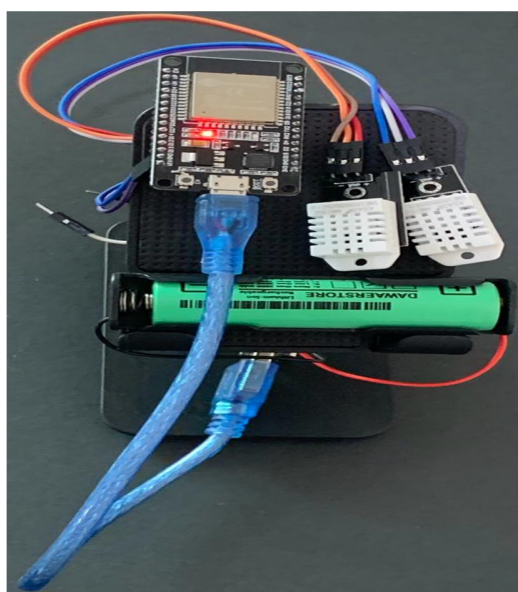


Fig. 3. The prototype of the cold supply-chain management system based on IoT.

The perception layer is composed of two sensor nodes (DHT22) that allow measuring the temperature as a key parameter in the cold supply-chain. It can be recalled here that the temperature is used to illustrate the concept of monitoring in cold supply-chain and that other sensors can be included (RFID, GPS, vibration sensor, etc.) if there are additional standard requirements. In this study, the DHT22 sensor is only used as a temperature sensor although it can also measure humidity. To overcome the issue of sensor calibration, two temperature sensors were used. The network layer includes all devices and protocols that facilitate the data transmission. As depicted in Figure 3, the NodeMcu microcontroller is equipped with a Wi-Fi module

integrated to the microcontroller board to wirelessly send the sensed data (temperature in this case) to the cloud storage. Through the microcontroller particular edge computing of the data can be performed to improve system performance. Among those computing actions, the storage of redundant data can be prevented (such as the case when the temperature doesn't change from one sample to another). The application layer consists of the features supported by the cloud such as storage and visualization. To illustrate the usability of the designed system, several use-case scenarios were practically implemented under Jeddah City, Saudi Arabia conditions. Jeddah climate is characterized by a relatively high temperature accompanied with a high level of humidity due to its proximity to the Red Sea. Under those conditions, cold supply-chains may face several problems during the product storage or transportation. In what follows, the results of a case scenario will be detailed.

At the beginning, a cold supply-chain monitoring channel has been created on the Adafruit platform. As general settings, the channel ID and the Application Programming Interface (API) key are unique references to the channel. They are later required for programming purpose. Two fields mapped to temperature and humidity (the two parameters measured by the DHT22 sensor) are set up, even though only the temperature was considered in this study. Several parameters related to the data visualization can be modified via the platform. The private view mode has been chosen in this scenario. However, sharing the channel view with specific users is possible and this will be implemented in practice to support smart contract terms of the cold supply-chain. In fact, in real-life scenarios, only the allowed parties/stakeholders such as the product supplier, transporter, retailer, and end-user can view the conditions under which the product is stored/transported. The channel settings include the location (latitude, longitude, and elevation) of the sensor node of the variable being monitored. If a Global Positioning System (GPS) module is used, the tracking of the transported cold product will be possible. A specific field will then be created.

B. Experimental Setup

On November 25, 2023, an experimental test of the designed cold supply-chain system based on IoT was conducted under Jeddah City conditions. The experiment was implemented in two phases. First, the NodeMcu microcontroller was connected to the WiFi router and then it was connected to an Android smartphone to simulate the real conditions of a cold product being transported. In real conditions, the smartphone can belong to the truck driver or can be a dedicated one. The values of the temperature during the test phase are shown in Figure 4. Two scenarios simulating the storage and the transportation of a cold product are implemented as shown in Table II. As the entries are being collected, alerts on thresholds were implemented depending on the transported cold product temperature as required by the related standards.

TABLE II. IOT COLD SUPPLY CHAIN MANAGEMENT SCENARIOS: STORAGE AND TRANSPORTATION

Scenario	Start location	End location	Start time	End time
Storage	Alsulaymaniyyah	Alsulaymaniyyah	8:02 PM	8:56 PM
Transportation	Alsulaymaniyyah	Alfaysaliyyah	8:56 PM	9:43 PM

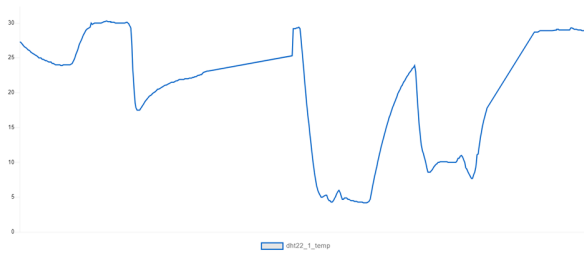


Fig. 4. Records of the temperature of the cold supply-chain channel collected on November 25, 2023 (Jeddah, Saudi Arabia), starting at 8:02 PM.

C. Limitations

During the cold supply-chain monitoring system development, several limitations were faced. Those limitations are summarized as follows:

- Before using any sensor (including the temperature sensor DHT22), its calibration should be performed with a calibrating sensor. One of the difficulties faced in this study is the unavailability of such a sensor. As a solution, two sensors were used almost in the same location [25].
- When connected to a smartphone WiFi network, particularly in the case of long-distance product transportation, the system may face energy consumption issues which can be solved using appropriate power banks/batteries.
- The IoT free channel on Adafruit allows a relatively limited number of entries [26]. In order to provide more storage space, a paid channel may be required depending on the number of the required entries and on the number of used sensor nodes.

D. Cost Analysis

The cost of the cold supply-chain system based on IoT is a primary key of its usefulness. The costs of the system components are shown in Table III.

TABLE III. IOT COLD SUPPLY CHAIN PROTOTYPE COST ANALYSIS

Component	Price (SAR)
NodeMCU microcontroller	70
Temperature sensor DHT22 (2 items)	90
Battery	15
Cable and wires	5
Total cost	180

As can be seen, the cost is relatively affordable and the system is therefore deployable and extendable, especially to high price shipments.

V. CONCLUSION

In this paper, a cold supply-chain monitoring system based on Internet of Things was designed and tested experimentally. The system was based on low-cost devices and free cloud platform services for prototyping purpose. The prototype was tested under Jeddah (Saudi Arabia) conditions and was found to be efficient in monitoring the temperature as the key parameter of cold products (food and pharmaceutical) storage and transportation. This system can be adopted as a basis for smart contracts in supply chain where the various parties, including the supplier,

transporter, retailer, and end-user can track/monitor the product during its journey from the factory/store to the consumer. The designed system is found to be extendable to other applications in logistics and supply-chain by including more sensors and adding other features, particularly in the application layer. As future work, more features can be added to the system, e.g. the required cloud storage space could be minimized by pre-analysis of the collected data as soon as they are received in order to store only significant changes or the use of artificial intelligence techniques to extract only the most significant data features. By implementing those improvements, the designed system can support smart contract implementation as a part of the cold supply chain.

ACKNOWLEDGMENT

This work was funded by the University of Jeddah, Jeddah, Saudi Arabia, under grant No. (UJ-21-ICL-1). The authors, therefore, acknowledge with thanks the University of Jeddah technical and financial support.

REFERENCES

- [1] S. Jagtap *et al.*, "Chapter 5 - IoT technologies in the food supply chain," in *Food Technology Disruptions*, C. M. Galanakis, Ed. Cambridge, MA, USA: Academic Press, 2021, pp. 175–211.
- [2] C. N. Verdouw, J. Wolfert, A. J. M. Beulens, and A. Rialland, "Virtualization of food supply chains with the internet of things," *Journal of Food Engineering*, vol. 176, pp. 128–136, May 2016, <https://doi.org/10.1016/j.jfoodeng.2015.11.009>.
- [3] A. M. Zaenurrohman and S. Alifiah, "Temperature and Humidity Monitoring on IoT Based Shipment Tracking," *Journal of Telematics and Informatics*, vol. 6, no. 1, pp. 27–36, Feb. 2018, <https://doi.org/10.12928/jti.v6i1>.
- [4] C. K. M. Lee, Y. Lv, K. K. H. Ng, W. Ho, and K. L. Choy, "Design and application of Internet of things-based warehouse management system for smart logistics," *International Journal of Production Research*, vol. 56, no. 8, pp. 2753–2768, Apr. 2018, <https://doi.org/10.1080/00207543.2017.1394592>.
- [5] M. Humayun, N. Jhanjhi, B. Hamid, and G. Ahmed, "Emerging Smart Logistics and Transportation Using IoT and Blockchain," *IEEE Internet of Things Magazine*, vol. 3, no. 2, pp. 58–62, Jun. 2020, <https://doi.org/10.1109/IOTM.0001.1900097>.
- [6] N. Zhang, "Smart Logistics Path for Cyber-Physical Systems With Internet of Things," *IEEE Access*, vol. 6, pp. 70808–70819, 2018, <https://doi.org/10.1109/ACCESS.2018.2879966>.
- [7] B. Feng and Q. Ye, "Operations management of smart logistics: A literature review and future research," *Frontiers of Engineering Management*, vol. 8, no. 3, pp. 344–355, Sep. 2021, <https://doi.org/10.1007/s42524-021-0156-2>.
- [8] Y. Zhang, Y. Liu, Z. Jiong, X. Zhang, B. Li, and E. Chen, "Development and assessment of blockchain-IoT-based traceability system for frozen aquatic product," *Journal of Food Process Engineering*, vol. 44, no. 5, 2021, Art. no. e13669, <https://doi.org/10.1111/jfpe.13669>.
- [9] N. Navarro, L. Horvath, and A. Salado, "Design of an IoT System for the Palletized Distribution Supply Chain with Model-Based Systems Engineering Tools," *Systems*, vol. 10, no. 1, Feb. 2022, Art. no. 4, <https://doi.org/10.3390/systems10010004>.
- [10] M. Balfagih, Z. Balfagih, M. D. Lytras, K. M. Alfawaz, A. A. Alshdadi, and E. Alsolami, "A Blockchain-Enabled IoT Logistics System for Efficient Tracking and Management of High-Price Shipments: A Resilient, Scalable and Sustainable Approach to Smart Cities," *Sustainability*, vol. 15, no. 18, Jan. 2023, Art. no. 13971, <https://doi.org/10.3390/su151813971>.
- [11] J. Gillespie *et al.*, "Real-Time Anomaly Detection in Cold Chain Transportation Using IoT Technology," *Sustainability*, vol. 15, no. 3, Jan. 2023, Art. no. 2255, <https://doi.org/10.3390/su15032255>.

- [12] F. Goodarzian, A. Navaei, B. Ehsani, P. Ghasemi, and J. Muñuzuri, "Designing an integrated responsive-green-cold vaccine supply chain network using Internet-of-Things: artificial intelligence-based solutions," *Annals of Operations Research*, vol. 328, no. 1, pp. 531–575, Sep. 2023, <https://doi.org/10.1007/s10479-022-04713-4>.
- [13] M. He and J. Shi, "Circulation traceability system of Chinese herbal medicine supply chain based on internet of things agricultural sensor," *Sustainable Computing: Informatics and Systems*, vol. 30, Jun. 2021, Art. no. 100518, <https://doi.org/10.1016/j.suscom.2021.100518>.
- [14] A. Rejeb, J. G. Keogh, and H. Treiblmaier, "Leveraging the Internet of Things and Blockchain Technology in Supply Chain Management," *Future Internet*, vol. 11, no. 7, Jul. 2019, Art. no. 161, <https://doi.org/10.3390/fi11070161>.
- [15] A. Lamberty and J. Kreyenschmidt, "Ambient Parameter Monitoring in Fresh Fruit and Vegetable Supply Chains Using Internet of Things-Enabled Sensor and Communication Technology," *Foods*, vol. 11, no. 12, Jan. 2022, Art. no. 1777, <https://doi.org/10.3390/foods11121777>.
- [16] L. F. Rahman, L. Alam, M. Marufuzzaman, and U. R. Sumaila, "Traceability of Sustainability and Safety in Fishery Supply Chain Management Systems Using Radio Frequency Identification Technology," *Foods*, vol. 10, no. 10, Oct. 2021, Art. no. 2265, <https://doi.org/10.3390/foods10102265>.
- [17] H. Afreen and I. S. Bajwa, "An IoT-Based Real-Time Intelligent Monitoring and Notification System of Cold Storage," *IEEE Access*, vol. 9, pp. 38236–38253, 2021, <https://doi.org/10.1109/ACCESS.2021.3056672>.
- [18] H. Cui, "Intelligent Coordination Distribution of the Whole Supply Chain Based on the Internet of Things," *Complexity*, vol. 2021, Mar. 2021, Art. no. e5555264, <https://doi.org/10.1155/2021/5555264>.
- [19] M. Ben-Daya, E. Hassini, Z. Bahroun, and B. H. Banimfreg, "The role of internet of things in food supply chain quality management: A review," *Quality Management Journal*, vol. 28, no. 1, pp. 17–40, Dec. 2020, <https://doi.org/10.1080/10686967.2020.1838978>.
- [20] G. Li, "RETRACTED: Development of cold chain logistics transportation system based on 5G network and Internet of things system," *Microprocessors and Microsystems*, vol. 80, Feb. 2021, Art. no. 103565, <https://doi.org/10.1016/j.micpro.2020.103565>.
- [21] H. Hasan, E. AlHadhrani, A. AlDhaheri, K. Salah, and R. Jayaraman, "Smart contract-based approach for efficient shipment management," *Computers & Industrial Engineering*, vol. 136, pp. 149–159, Oct. 2019, <https://doi.org/10.1016/j.cie.2019.07.022>.
- [22] S. Zafar, G. Miraj, R. Baloch, D. Murtaza, and K. Arshad, "An IoT Based Real-Time Environmental Monitoring System Using Arduino and Cloud Service," *Engineering, Technology & Applied Science Research*, vol. 8, no. 4, pp. 3238–3242, Aug. 2018, <https://doi.org/10.48084/etasr.2144>.
- [23] B. F. Alshammari and M. T. Chughtai, "IoT Gas Leakage Detector and Warning Generator," *Engineering, Technology & Applied Science Research*, vol. 10, no. 4, pp. 6142–6146, Aug. 2020, <https://doi.org/10.48084/etasr.3712>.
- [24] M. Hamdani, M. Youcefi, A. Rabehi, B. Nail, and A. Douara, "Design and Implementation of a Medical TeleMonitoring System based on IoT," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 8949–8953, Aug. 2022, <https://doi.org/10.48084/etasr.5040>.
- [25] W. Jiang, "An Intelligent Supply Chain Information Collaboration Model Based on Internet of Things and Big Data," *IEEE Access*, vol. 7, pp. 58324–58335, 2019, <https://doi.org/10.1109/ACCESS.2019.2913192>.
- [26] H. J. Jara Ochoa, R. Pena, Y. Ledo Mezquita, E. Gonzalez, and S. Camacho-Leon, "Comparative Analysis of Power Consumption between MQTT and HTTP Protocols in an IoT Platform Designed and Implemented for Remote Real-Time Monitoring of Long-Term Cold Chain Transport Operations," *Sensors*, vol. 23, no. 10, Jan. 2023, Art. no. 4896, <https://doi.org/10.3390/s23104896>.

Comparative Evaluation of Different Hybrid Intelligent Load-Frequency Controllers for Interconnected Electric Power Grids

Diem-Vuong Doan

Faculty of Control and Automation, Electric Power University, Vietnam
vuongdd@epu.edu.vn

Ngoc-Khoat Nguyen

Faculty of Control and Automation, Electric Power University, Vietnam
khoatnn@epu.edu.vn (corresponding author)

Received: 5 December 2023 | Revised: 28 December 2023 and 13 January 2024 | Accepted: 23 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6706>

ABSTRACT

Network frequency is considered to be one of the most crucial parameters that strongly affect the stability and economic achievements of interconnected electric power grids. System frequency usually fluctuates and deviates from the nominal values due to random and continuous load changes over time, affecting the electric equipment to significantly decrease efficiency and increase instability. A Load-Frequency Control (LFC) strategy has been proposed to solve this problem. This study compared several different control strategies, namely Fuzzy Particle Swarm Optimization (Fuzzy-PSO), Proportional Integral Derivative (PID), Fuzzy-PID, Fuzzy- Proportional Integral (PI), PSO-PI, and FPID to investigate the effectiveness of intelligent hybrid LFC controllers. The above controllers were simulated on a three-area interconnected power network with the participation of renewable energy sources. Taking into account different load cases, the Fuzzy-PSO-PID controller obtained frequency deviations in the range of 0.0015 to 0.002 Hz. The settling time was about 10 s to reach zero frequency error in each area. With the above controller quality parameters, the Fuzzy-PSO-PID controller provided better quality than the other controllers. A comparative numerical simulation in MATLAB/Simulink for various load change scenarios revealed the effectiveness of hybrid smart controllers, such as the Fuzzy-PID-PSO, outperforming the traditional ones.

Keywords-Load-Frequency Control (LFC); Fuzzy-PSO-PID; Fuzzy-PID; Fuzzy-PI; PSO-PI; FPID; RE sources

I. INTRODUCTION

Interconnected power networks are typically complex systems with many parameters that need to be observed and controlled. Among them, system frequency and voltage are the most vital parameters. Frequency deviation from the rated one occurs due to the time-by-time imbalance between the production and consumption of electricity. This negatively affects consumers, especially those dealing with electrical equipment that is designed and optimized to work at rated frequency. When the frequency changes, it leads to a decrease in the performance of the equipment, reducing operation performance. When operating below 49.5 Hz, some steam turbines and rotors vibrate excessively. Excessive frequency attenuation in the system will increase the magnetization current in induction machines and transformers. Therefore, it is highly necessary to design frequency-stabilizing controllers in interconnected power systems. Many studies have used different algorithms for frequency stabilization in interconnected power systems.

In [1], an online controller that combined BIA and fuzzy algorithms to control the frequency stability of four areas was presented. In [2], a Linear Controller (LQ) was proposed based on the Genetic Algorithm (GA) to control frequency stability for a system consisting of two non-reheat turbine areas. In [3], the design of a neural controller for a system consisting of three areas with non-reheat, reheat, and hydraulic turbines was studied. In [4], the ANFIS controller combined neural network and internal fuzzy control to ensure the frequency stability of a system consisting of two areas of reheat and hydraulic turbines with frequency overshoots ranging from 0.01 to 0.08 Hz. The BBO controller in [5] was designed for frequency stabilization in a system of two areas, each consisting of three power sources. In [6], the model predictive-design controller aiming to improve the frequency stability of a four-area system was examined, considering factors for nonlinear systems, such as Governor Dead Band (GDB) and Generation Rate Constraint (GRC). In [7-9], focus was given to frequency and power stability studies in multi-connected power systems. The power stability of the line is one of the additional requirements for

frequency stability [10-13]. In [14], the design of a Load Frequency Control (LFC) component was investigated in a four-area interconnected power system.

The PID controller has been optimized using the Particle Swarm Optimization (PSO) and JAYA optimization methods. In [15], a fuzzy logic controller based on the PID principle was proposed to completely solve the LFC problem in a two-area interconnected power system. Currently, due to high electricity demand, renewable energy sources are being increasingly used in power systems. In addition to the advantages of renewable energy sources, there are also inherent disadvantages. Renewable energy sources in interconnected power systems increase the non-linearity of the system. Solar power sources have no rotational inertia, but wind power sources have low rotational inertia and have a dynamic response that depends on the characteristics of the inverter. Therefore, when solar and wind power account for a high proportion of power production, they will reduce the inertia and vibration resistance coefficient of the grid, leading to an increased risk of destabilizing the power system. Therefore, increasing the frequency deviation in each region, reduces the working capacity of the system equipment, changes the power flow, and augments the losses on the power transmission lines.

In [16], an adaptive model predictive controller for systems with renewable energy sources of wind power was explored. In [17], a hybrid controller (DEPSO) was proposed to optimize the parameters of the fuzzy-PID controller. In [18], the frequency of a power system was stabilized using a combination of PSO and ANFIS hybrid controllers to adjust the PID parameters. In [19], the PSO control method was used to optimize the FOC design for a system that integrated additional wind power sources. In [20], an algorithm based on the cuckoo search method was employed to ensure the frequency stability of power systems with nonlinear factors. The cuckoo search algorithm was utilized in [21] for frequency stabilization of thermal and wind power systems. In [22], a fuzzy PID controller, whose parameters were optimized by the Cuckoo Search algorithm, was proposed to stabilize frequency. In [23], the BFOA method was used to optimize the PID controller parameters and thus stabilize the frequency of two regions with renewable energy. In many studies, the power networks embedded with renewable energy sources are highly difficult to control in a successful operation and distribution. An interconnected power system with nonlinear elements and renewable energy sources is an extremely complex system. The system's frequency stability control will not be effective if classical controllers are used. This study investigated and compared the ability to stabilize the frequency of several efficient load-frequency controllers, including Fuzzy-PSO-PID, Fuzzy-PID, Fuzzy-PI, PSO-PID, and FPID regulators. A system can be improved with intelligent hybrid controllers. The three-area power system was studied, having a non-reheat turbine, a reheat turbine, and a hydraulic turbine. This study considered nonlinear factors and the participation of renewable energy sources, such as wind and solar energy.

II. MODELING OF INTERCONNECTED POWER SYSTEM IN LFC

A single power system consists of the following basic components: a speed governor, a turbine of three types (non-reheat turbine, reheat turbine, and hydraulic turbine), and a generator. Figure 1 shows the structure of a basic single-area power system. A typical three-area electrical system comprises three single-area systems connected via interconnection lines. Each area has its own generating unit(s) that provide electricity to consumers. The link offers information about the scheduled exchange of electricity between the three areas. Figure 2 shows a model of a typical three-area power system. The power exchange deviation between two areas is calculated by:

$$\Delta P_{tie,ij}(s) = \frac{2\pi T_{ij}}{s} (\Delta f_i(s) - \Delta f_j(s)) \quad (1)$$

where $\Delta P_{tie,ij}$, Δf_i , and T_{ij} denote the tie-line power change, frequency variation, and constant coefficient between two areas i and j , respectively.

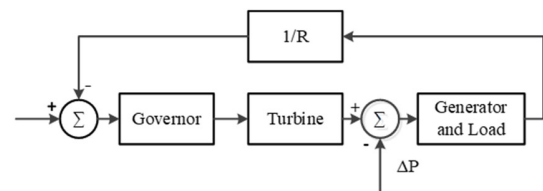


Fig. 1. The basic structure of a single-area power system.

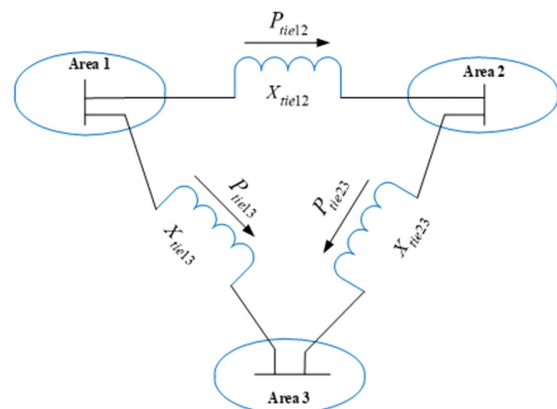


Fig. 2. The structure of three-area power systems.

A. Turbine

A turbine unit is used to convert primary energy, such as energy from steam or water, into mechanical energy (ΔP_t) that powers a generator. There are three types of turbines: Non-reheat, reheat, and hydraulic turbines. This study investigated the presence of renewable energy in all three areas and its effect on frequency stability.

1) Non-Reheat Turbine

A non-reheat turbine is a first-order unit. The time delay, expressed in T_{ch} , occurs between switching the valve and generating the turbine torque. The transfer function for a non-reheat turbine is expressed as:

$$G_t(s) = \frac{P_t(s)}{P_g(s)} = \frac{1}{1+sT_{ch}} \quad (2)$$

2) Reheat Turbine

Reheat turbines are modeled as second-order units, as they have different stages due to high and low steam pressure. The transfer function can be expressed as:

$$G_t(s) = \frac{\Delta P_t(s)}{\Delta P_g(s)} = \frac{F_{hp}T_{rh}s+1}{T_{rh}s+1} \cdot \frac{1}{T_{ch}s+1} \quad (3)$$

where T_{rh} and F_{hp} are the low-pressure reheat time constant and the high-pressure stage, respectively.

3) Hydraulic Turbine

In a hydraulic turbine, the water pressure response is initially reversed to the gate position change and recovered after the transient response. Therefore, the transfer function of a hydraulic turbine can be described as follows:

$$G_t(s) = \frac{P_t(s)}{P_g(s)} = \frac{-T_ws+1}{0.5T_ws+1} \quad (4)$$

Regarding stability concerns, a temporary deflection compensator is required in the governor for the hydraulic turbine. The transfer function of the temporary slump compensation is given by:

$$G(s) = \frac{T_rs+1}{T_r \cdot \frac{R_t}{R_3} s+1} \quad (5)$$

where T_r , T_w , and R_t are the reset time of the hydraulic unit, water starting time, and the reset time, respectively.

B. Wind Turbine (RE sources)

A wind turbine [24] consists of a turbine-generator shaft mechanism, which is used to convert the rotor rotation into electrical energy. Equation (6) represents the mechanical output of the wind turbine:

$$P_{WT} = \frac{1}{2} \rho A C_p(\lambda, \beta) V_w^3 \quad (6)$$

where λ , ρ , V_w , and C_p are the tip speed ratio, air density factor (Kg/cu.m), wind speed (m/s), and power coefficient, respectively. The first-order transfer function model of a wind turbine is:

$$\frac{\Delta P_{wtg}(s)}{\Delta P_{wt}(s)} = \frac{1}{T_{wts}s+1} \quad (7)$$

where T_{wts} is the time constant of the wind turbine.

C. Solar Power (RE sources)

The transfer function model of solar power is [25]:

$$\frac{\Delta P_{spv}(s)}{\Delta P_{sp}(s)} = \frac{1}{T_{spv}s+1} \quad (8)$$

Figure 3 displays a block diagram of a three-area interconnected power system with RE sources and GDB along with GRC, established in the Matlab/Simulink platform. The three-area interconnected system was considered to be a typical power system model, as illustrated in Figure 2. In this system, there are nonlinear factors, such as GDB, GRC, and renewable energy sources affecting the frequency stability of the system. This is considered a complex system.

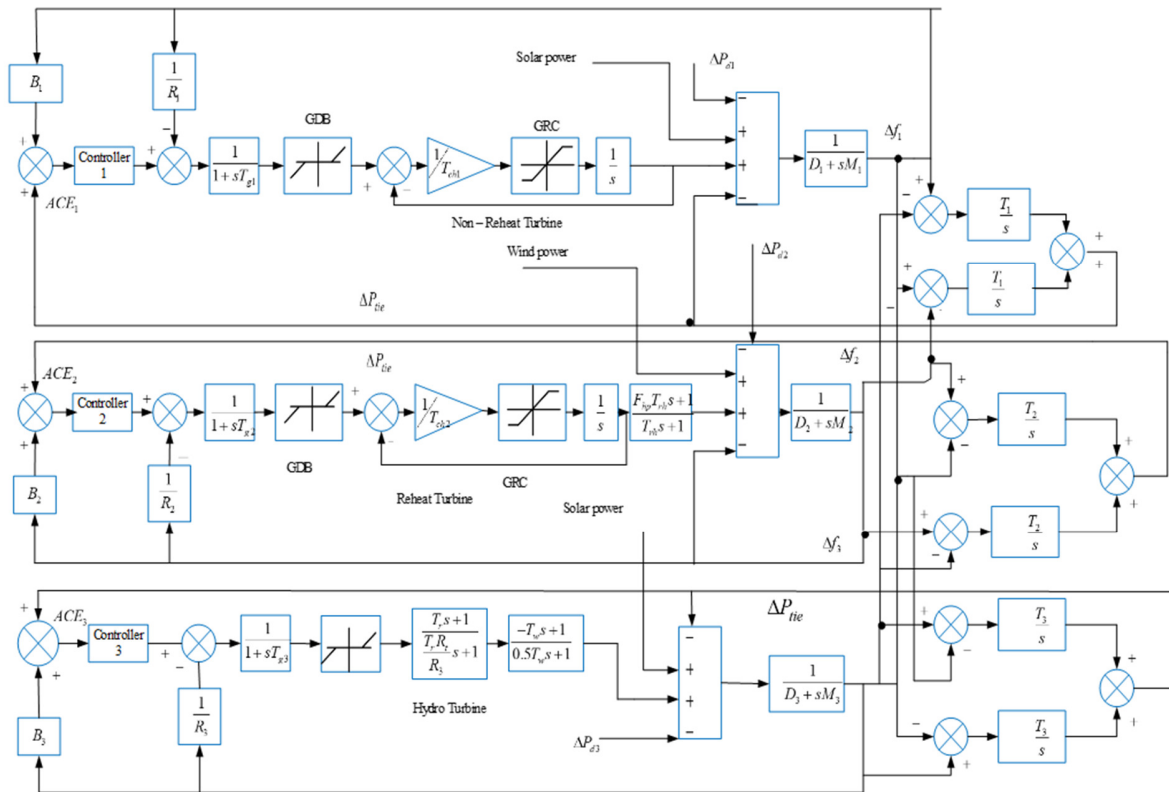


Fig. 3. A three-area interconnected power system model.

III. COMPARISON OF DIFFERENT CONTROLLERS IN VARIOUS LOAD CHANGE SCENARIOS

Figure 4 portrays the LFC structure of an interconnected power system utilizing different controllers. The main purpose of the secondary regulator is to restore the frequency offset and power flow deviation, exchanged between the systems, to zero by adjusting the Area Control Error (ACE) to zero. ACE is calculated according to:

$$ACE_i = \Delta P_{tie,i,j} + \beta_i \Delta f_i \quad (9)$$

with control inputs ACE and ΔACE . This study compared the Fuzzy-PSO-PID [26], Fuzzy-PID [27], Fuzzy-PI [28], PSO-PI, and FPID [29] controllers in five cases. Table I presents the parameters of the three-area interconnected power system. The Fuzzy-PSO-PID controller [26] was studied with the control diagram shown in Figure 5.

TABLE I. THE PARAMETERS OF THREE AREAS [27]

Area with non-reheat turbine	Value	Area with reheat turbine	Value	Area with hydraulic turbine	Value
M_1 (p.u.s)	10	M_2 (p.u.s)	10	T_w (s)	1.0
D_1 (p.u./Hz)	1.0	D_2 (p.u./Hz)	1.0	T_R (s)	5.0
T_{ch1} (s)	0.3	T_{ch2} (s)	0.3	F_{hp}	0.3
T_{e1} (s)	0.1	T_{e2} (s)	0.2	T_{rh} (s)	7.0
R_1 (Hz/p.u.)	0.05	R_2 (Hz/p.u.)	0.05	T_{spv} (s)	1.8
B_1 (p.u./Hz)	21.0	B_2 (p.u./Hz)	21.0	T_{wts} (s)	1.5
T_1 (p.u./rad.)	22.6	T_2 (p.u./rad.)	22.6		

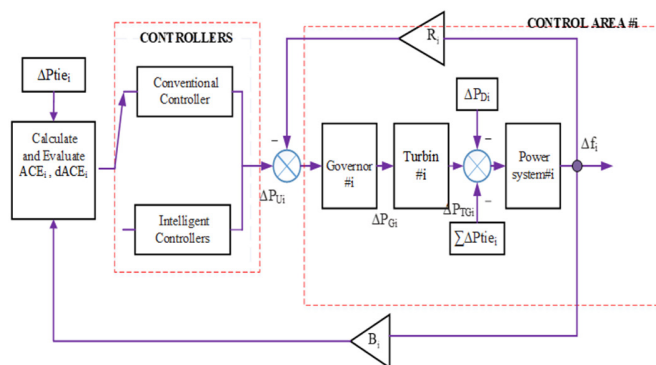


Fig. 4. The structure of the control area i using different controllers.

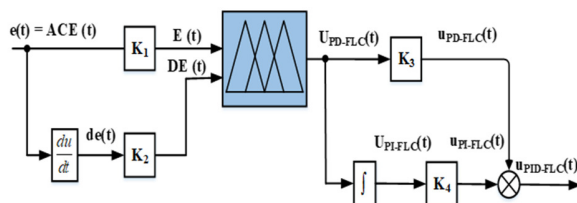


Fig. 5. The structure of the Fuzzy-PSO-PID controller.

The coefficients K_1 , K_2 , K_3 , and K_4 were optimized using the PSO method, as depicted in Table II.

TABLE II. OPTIMAL FUZZY-PID PARAMETERS USING PSO

Coefficients	K_1	K_2	K_3	K_4
Value	800.91	154.46	624.09	709.43

J is the objective function calculated using IATE to determine the optimal parameters of the Fuzzy-PSO-PID controller. The use of IATE to determine the objective function provides the best optimal parameters. The objective function is:

$$J = IATE = \int_0^{t_{sim}} \left(\frac{|\Delta f_1| + |\Delta f_2| + |\Delta f_3|}{\sum_{i,k=1}^3 |\Delta P_{tie,i,k}|} \right) dt \quad (10)$$

where t_{sim} is the simulation time. The Fuzzy-PID controller [27] employs fuzzy inference to determine three factors of the PID regulator as a master-slave mechanism. To compare the quality of frequency controllers for three areas, simulations of a three-area power system with the participation of renewable energy sources were carried out in five cases with varying load types.

A. Case 1

Figure 6 illustrates the simulation results of the case where:

$\Delta P_{d1} = \Delta P_{d2} = \Delta P_{d3} = 0.03$ (Types of step), and

$\Delta P_{spv} = \Delta P_{wts} = 0.03$ (Types of pulse).

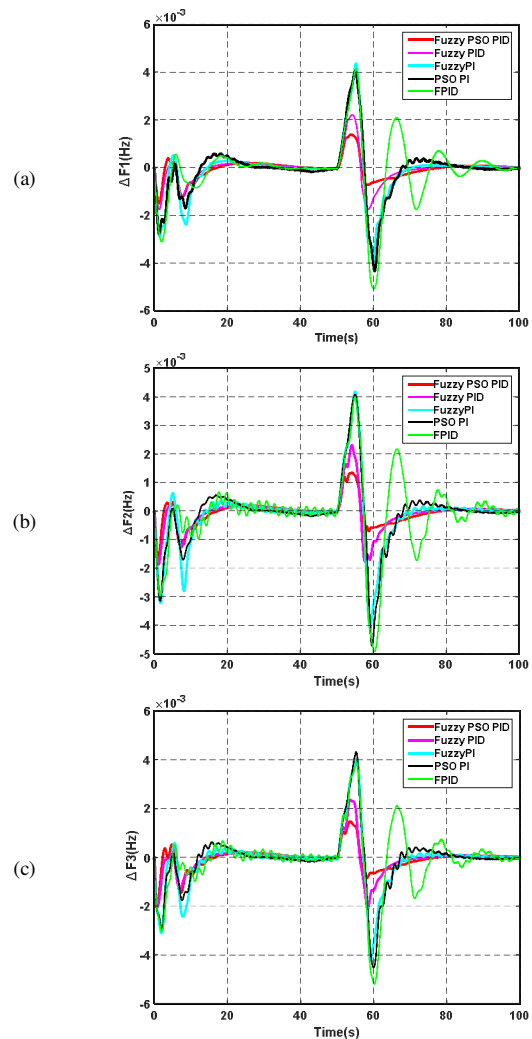


Fig. 6. Case 1: Dynamic frequency responses in three areas: (a) Δf_1 , (b) Δf_2 , (c) Δf_3 .

Table III displays the results of this case, demonstrating that Fuzzy-PSO-PID outperformed the other controllers.

TABLE III. COMPARISON OF CONTROLLERS

	Error of frequency (10^{-3} Hz)			Settling time (s)		
	$\Delta F1$	$\Delta F2$	$\Delta F3$	$\Delta F1$	$\Delta F2$	$\Delta F3$
Fuzzy- PSO-PID	1.6	1.1	1	2	2	2
Fuzzy-PID	2.01	2	2	5	5	6
Fuzzy-PI	4	4	4	10	15	17
PSO-PI	4.01	4	4	20	20	20
FPID	4.5	4.5	4.5	50	50	55

B. Case 2

Figure 7 shows the numerical simulation outcomes of the following case:

$$\Delta P_{d1} = \Delta P_{d2} = \Delta P_{d3} = 0.05 \text{ (Types of step),}$$

$$\Delta P_{spv} = \Delta P_{wts} = 0.03 \text{ (Types of pulse)}$$

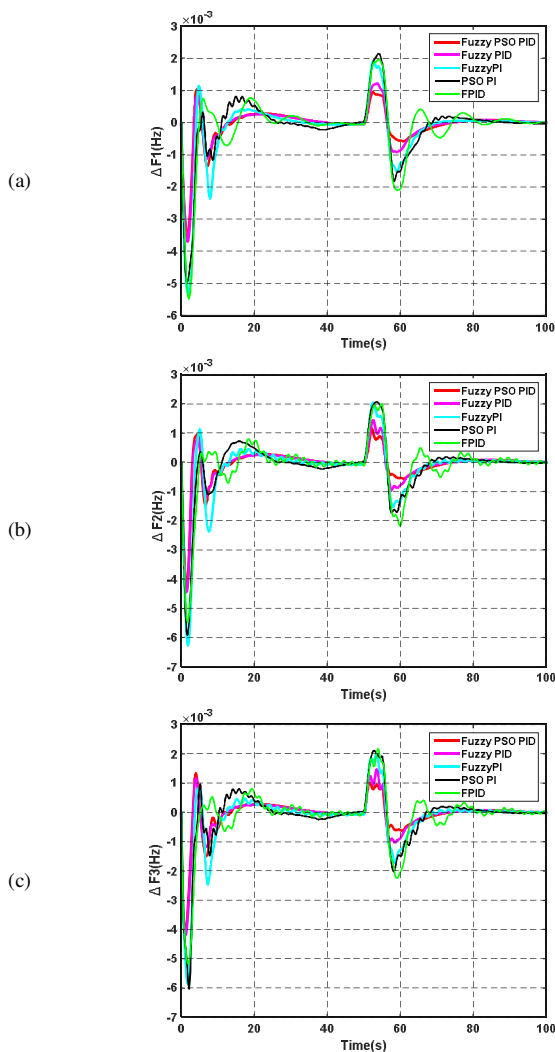


Fig. 7. Case 2: Dynamic frequency responses in three areas: (a) $\Delta F1$, (b) $\Delta F2$, (c) $\Delta F3$.

C. Case 3

Figure 8 describes the simulation results for the case of:

$$\Delta P_{d1} = \Delta P_{d2} = \Delta P_{d3} = 0.05 \text{ (Type of step), and}$$

$$\Delta P_{spv} = \Delta P_{wts} = 0.05 \text{ (Types of pulse).}$$

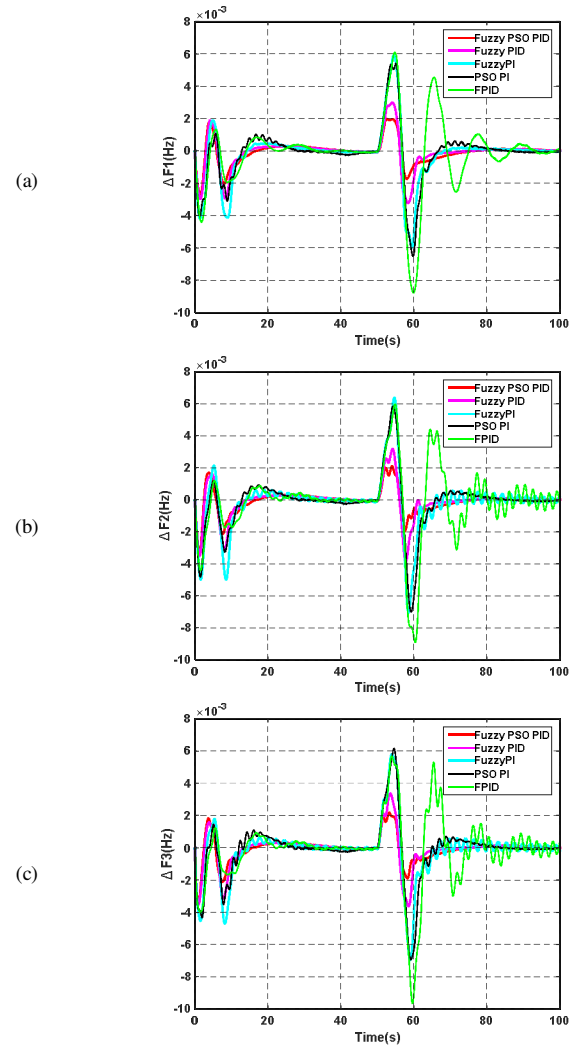


Fig. 8. Case 3: Dynamic frequency responses in three areas: (a) $\Delta F1$, (b) $\Delta F2$, (c) $\Delta F3$.

D. Case 4

Figure 9 portrays the plot curves of frequency responses in the case of:

$$\Delta P_{d1} = \Delta P_{d2} = \Delta P_{d3} = 0.05 \text{ (Type of random), and}$$

$$\Delta P_{spv} = \Delta P_{wts} = 0.05 \text{ (Type of pulse).}$$

E. Case 5

This case takes into account the time delay with $t = 2$ s, which greatly affects the quality of frequency control in each area. Figure 10 presents the simulation results with the following parameters:

$\Delta P_{d1} = \Delta P_{d2} = \Delta P_{d3} = 0.05$ (Type of random), and
 $\Delta P_{spv} = \Delta P_{wts} = 0.03$ (Type of pulse).

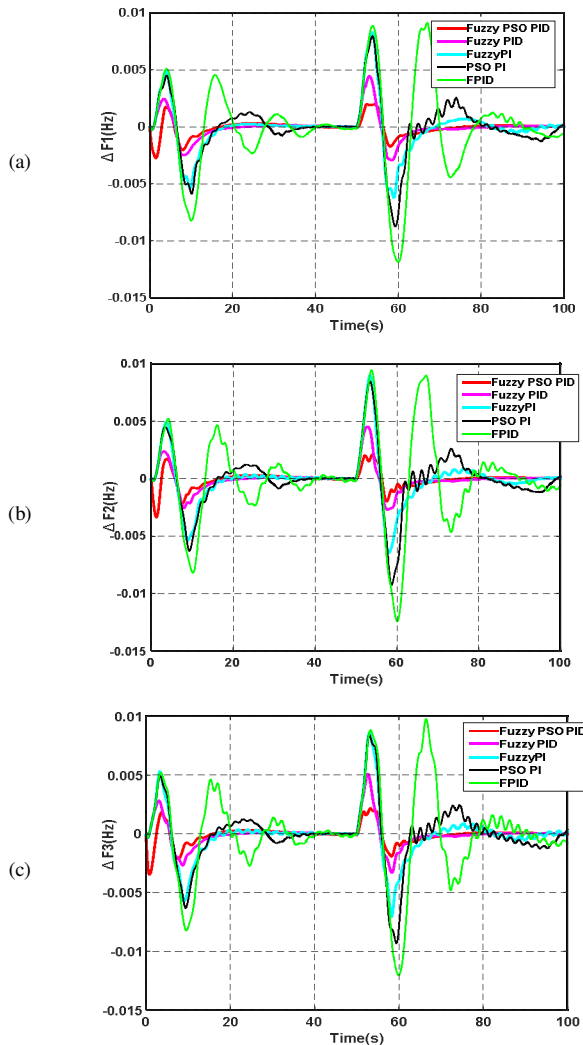


Fig. 9. Dynamic frequency responses in three areas in case 4: (a) $\Delta F1$, (b) $\Delta F2$, (c) $\Delta F3$.

The participation of renewable energy sources in an interconnected power system reduces its inertia. By observing the frequency responses of each region in each case, it is demonstrated that when using PID, PSO, and PID controllers, there is a strong system frequency deviation in each area with a large overshoot. When utilizing Fuzzy PI and Fuzzy PID controllers, the frequency error response had reduced oscillation but the overshoot was still large. The Fuzzy-PSO-PID controller did not cause overshoot and frequency deviation in stable areas. Figures 6-10 depict the frequency error responses in the three areas. When using the Fuzzy-PSO-PID controller, the frequency error and response time were the smallest. The FPID controller had the biggest bias and response time. This indicates that hybrid smart controllers give better responses than classic ones.

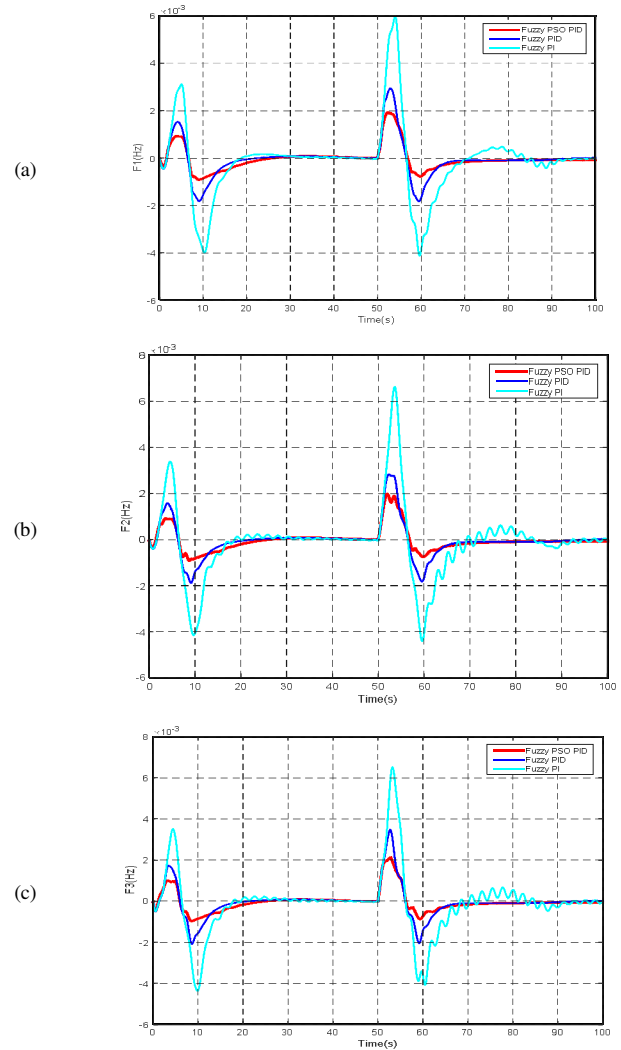


Fig. 10. Dynamic frequency responses in three areas in case 5: (a) $\Delta F1$, (b) $\Delta F2$, (c) $\Delta F3$.

An interconnected power system is a large system with many parameters that can be changed during operation. In addition, the system studied considered renewable energy to decrease its inertia. Renewable energy sources, such as solar and wind sources, are inherently variable and can experience sudden output changes. Unlike traditional power sources, like coal or gas plants, the renewable ones lack the rotational inertia provided by the spinning turbines. This inertia helps stabilize the grid against sudden fluctuations. When renewable sources dominate, the grid may experience reduced inertia, making it more susceptible to rapid frequency changes and requiring advanced control mechanisms to maintain stability. As a result, the power system faces challenges in maintaining frequency stability. As shown in Figures 6-10, simple controllers had worse performance. The results of the Fuzzy PID controller reveal that the Fuzzy logic-based controllers with fuzzy set selection generate input functions and fuzzy rules associated with the given PID controller and obtain better control quality than other controllers, such as FPID, Fuzzy PI, and PSO-PID.

However, Fuzzy PID was less efficient compared to the Fuzzy-PSO-PID controller. The reason is that due to fuzzy variable inputs, the rules are established based on the experience of the experts. With the Fuzzy-PSO-PID controller, the controller's parameters are designed according to the objective function of the optimization method, which gives more effective results.

IV. CONCLUSION AND FUTURE WORK

Load-frequency control in a power system is a serious problem in its operation and distribution. However, in a power system with nonlinear factors and renewable energy sources, it is difficult to stabilize its frequency. The comparative simulation results of different controllers in this study highlight that the hybrid intelligent controllers obtained better results than the classic ones. The future direction of this research is to study and design an efficient controller for a multi-area system with many transmitters in each area. In addition, the practical applications of the studied hybrid load-frequency controllers will be considered. In this scenario, a practical electrical power grid should be a suitable selection.

TABLE IV. ABBREVIATIONS

GDB	Governor Dead Band	FPID	Fractional PID
GRC	Generation Rate Constraint	ANFIS	Adaptive Neuro-Fuzzy Inference System
LFC	Load Frequency Control	DE-PSO	Differential Evolution and Particle Swarm Optimization
ACE	Area Control Error	BIA	Bat inspired algorithm
PSO	Particle Swarm Optimization	BBO	Biogeography-Based Optimization
GWO	Grey Wolf Optimizer	BFOA	Bacterial Foraging Optimization Algorithm
PI	Proportional Integral	PID	Proportional Integral Derivative

REFERENCES

- [1] M. H. Khooban and T. Niknam, "A new intelligent online fuzzy tuning approach for multi-area load frequency control: Self Adaptive Modified Bat Algorithm," *International Journal of Electrical Power & Energy Systems*, vol. 71, pp. 254–261, Oct. 2015, <https://doi.org/10.1016/j.ijepes.2015.03.017>.
- [2] A. Dwivedi, G. Ray, and A. K. Sharma, "Genetic Algorithm based Decentralized PI Type Controller: Load Frequency Control," *Journal of The Institution of Engineers (India): Series B*, vol. 97, no. 4, pp. 509–515, Dec. 2016, <https://doi.org/10.1007/s40031-015-0191-7>.
- [3] S. A. Azeer, R. Ramjug-Ballgobin, and S. Z. Sayed Hassen, "Intelligent Controllers for Load Frequency Control of Two-Area Power System," *IFAC-PapersOnLine*, vol. 50, no. 2, pp. 301–306, Dec. 2017, <https://doi.org/10.1016/j.ifacol.2017.12.062>.
- [4] S. Prakash and S. K. Sinha, "Neuro-Fuzzy Computational Technique to Control Load Frequency in Hydro-Thermal Interconnected Power System," *Journal of The Institution of Engineers (India): Series B*, vol. 96, no. 3, pp. 273–282, Sep. 2015, <https://doi.org/10.1007/s40031-014-0147-3>.
- [5] M. R. Sathya and M. Mohamed Thameem Ansari, "Design of biogeography optimization based dual mode gain scheduling of fractional order PI load frequency controllers for multi source interconnected power systems," *International Journal of Electrical Power & Energy Systems*, vol. 83, pp. 364–381, Dec. 2016, <https://doi.org/10.1016/j.ijepes.2016.04.006>.
- [6] M. Ma, C. Zhang, X. Liu, and H. Chen, "Distributed Model Predictive Load Frequency Control of the Multi-Area Power System After Deregulation," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 6, pp. 5129–5139, Jun. 2017, <https://doi.org/10.1109/TIE.2016.2613923>.
- [7] E. Rakhshani and J. Sadeh, "Practical viewpoints on load frequency control problem in a deregulated power system," *Energy Conversion and Management*, vol. 51, no. 6, pp. 1148–1156, Jun. 2010, <https://doi.org/10.1016/j.enconman.2009.12.024>.
- [8] B. Sonker, D. Kumar, and P. Samuel, "Dual loop IMC structure for load frequency control issue of multi-area multi-sources power systems," *International Journal of Electrical Power & Energy Systems*, vol. 112, pp. 476–494, Nov. 2019, <https://doi.org/10.1016/j.ijepes.2019.04.042>.
- [9] M. V. Srikanth and N. Yadaiah, "An AHP based optimized tuning of Modified Active Disturbance Rejection Control: An application to power system load frequency control problem," *ISA Transactions*, vol. 81, pp. 286–305, Oct. 2018, <https://doi.org/10.1016/j.isatra.2018.07.001>.
- [10] A. Delassi, S. Arif, and L. Mokrani, "Load frequency control problem in interconnected power systems using robust fractional PID controller," *Ain Shams Engineering Journal*, vol. 9, no. 1, pp. 77–88, Mar. 2018, <https://doi.org/10.1016/j.asej.2015.10.004>.
- [11] C. Chen, K. Zhang, K. Yuan, and X. Teng, "Tie-Line Bias Control Applicability to Load Frequency Control for Multi-Area Interconnected Power Systems of Complex Topology," *Energies*, vol. 10, no. 1, Jan. 2017, Art. no. 78, <https://doi.org/10.3390/en10010078>.
- [12] F. U. A. Ahammad and S. Mandal, "Robust load frequency control in multi-area power system: An LMI approach," in *2016 IEEE First International Conference on Control, Measurement and Instrumentation (CMI)*, Kolkata, India, Jan. 2016, pp. 136–140, <https://doi.org/10.1109/CMI.2016.7413726>.
- [13] S. Pande and R. Kansal, "Load Frequency Control of Multi Area System using Integral- Fuzzy Controller," *Surbhi Pande Int. Journal of Engineering Research and Applications*, vol. 5, no. 6, pp. 59–64, 2015.
- [14] S. Pahadasingh, C. Jena, C. K. Panigrahi, and B. P. Ganthia, "JAYA Algorithm-Optimized Load Frequency Control of a Four-Area Interconnected Power System Tuning Using PID Controller," *Engineering, Technology & Applied Science Research*, vol. 12, no. 3, pp. 8646–8651, Jun. 2022, <https://doi.org/10.48084/etasr.4891>.
- [15] D. V. Doan, K. Nguyen, and Q. V. Thai, "A Novel Fuzzy Logic Based Load Frequency Control for Multi-Area Interconnected Power Systems," *Engineering, Technology & Applied Science Research*, vol. 11, no. 4, pp. 7522–7529, Aug. 2021, <https://doi.org/10.48084/etasr.4320>.
- [16] M. A. Mohamed, A. A. Z. Diab, H. Rezk, and T. Jin, "A novel adaptive model predictive controller for load frequency control of power systems integrated with DFIG wind turbines," *Neural Computing and Applications*, vol. 32, no. 11, pp. 7171–7181, Jun. 2020, <https://doi.org/10.1007/s00521-019-04205-w>.
- [17] B. K. Sahu, S. Pati, and S. Panda, "Hybrid differential evolution particle swarm optimisation optimised fuzzy proportional–integral derivative controller for automatic generation control of interconnected power system," *IET Generation, Transmission & Distribution*, vol. 8, no. 11, pp. 1789–1800, 2014, <https://doi.org/10.1049/iet-gtd.2014.0097>.
- [18] N. K. Bahgaat, M. I. El-Sayed, M. A. M. Hassan, and F. A. Bendary, "Load Frequency Control in Power System via Improving PID Controller Based on Particle Swarm Optimization and ANFIS Techniques," in *Research Methods: Concepts, Methodologies, Tools, and Applications*, IGI Global, 2015, pp. 462–481.
- [19] T. Mahto, H. Malik, and V. Mukherjee, "Fractional Order Control and Simulation of Wind-Biomass Isolated Hybrid Power System Using Particle Swarm Optimization," in *Applications of Artificial Intelligence Techniques in Engineering*, Singapore, 2019, pp. 277–287, https://doi.org/10.1007/978-981-13-1819-1_27.
- [20] A. Y. Abdelaziz and E. S. Ali, "Cuckoo Search algorithm based load frequency controller design for nonlinear interconnected power system," *International Journal of Electrical Power & Energy Systems*, vol. 73, pp. 632–643, Dec. 2015, <https://doi.org/10.1016/j.ijepes.2015.05.050>.
- [21] S. Chaine, M. Tripathy, and D. Jain, "Non dominated Cuckoo search algorithm optimized controllers to improve the frequency regulation characteristics of wind thermal power system," *Engineering Science and Technology, an International Journal*, vol. 20, no. 3, pp. 1092–1105, Jun. 2017, <https://doi.org/10.1016/j.jestech.2017.05.005>.

- [22] M. Gheisarnejad, "An effective hybrid harmony search and cuckoo optimization algorithm based fuzzy PID controller for load frequency control," *Applied Soft Computing*, vol. 65, pp. 121–138, Apr. 2018, <https://doi.org/10.1016/j.asoc.2018.01.007>.
- [23] E. S. Ali and S. M. Abd-Elazim, "BFOA based design of PID controller for two area Load Frequency Control with nonlinearities," *International Journal of Electrical Power & Energy Systems*, vol. 51, pp. 224–231, Oct. 2013, <https://doi.org/10.1016/j.ijepes.2013.02.030>.
- [24] A. Kumar and S. N. V., "Teaching-Learning Optimization Based Adaptive Fuzzy Logic Controller for Frequency Control in an Autonomous Microgrid," *International Journal of Renewable Energy Research (IJRER)*, vol. 7, no. 4, pp. 1942–1949, Dec. 2017.
- [25] A. Annamraju and S. Nandiraju, "Robust Frequency Control in an Autonomous Microgrid: A Two-Stage Adaptive Fuzzy Approach," *Electric Power Components and Systems*, vol. 46, no. 1, pp. 83–94, Jan. 2018, <https://doi.org/10.1080/15325008.2018.1432723>.
- [26] D. V. Doan, K. Nguyen, and Q. V. Thai, "Load-Frequency Control of Three-Area Interconnected Power Systems with Renewable Energy Sources Using Novel PSO-PID-Like Fuzzy Logic Controllers," *Engineering, Technology & Applied Science Research*, vol. 12, no. 3, pp. 8597–8604, Jun. 2022, <https://doi.org/10.48084/etasr.4924>.
- [27] D. V. Doan, K. Nguyen, and Q. V. Thai, "On design of PSO-based Fuzzy-PI – like load-frequency control strategies for multi-area interconnected power systems," Ho Chi Minh City, Vietnam, Apr. 2022.
- [28] A. Kandhavel, "Load frequency control." <https://www.mathworks.com/matlabcentral/fileexchange/31514-load-frequency-control>.
- [29] S. A. Taher, M. Hajiakbari Fini, and S. Falahati Aliabadi, "Fractional order PID controller design for LFC in electric power systems using imperialist competitive algorithm," *Ain Shams Engineering Journal*, vol. 5, no. 1, pp. 121–135, Mar. 2014, <https://doi.org/10.1016/j.asej.2013.07.006>.

Machine Learning-based Predictive Maintenance for Fault Detection in Rotating Machinery: A Case Study

Ardalan F. Khalil

Department of Mechanical and Manufacturing Engineering, Technical College of Engineering, Sulaimani Polytechnic University, Kurdistan Region, Iraq
ardalan.fryad.k@spu.edu.iq

Sarkawt Rostam

Department of Mechanical and Manufacturing Engineering, Technical College of Engineering, Sulaimani Polytechnic University, Kurdistan Region, Iraq
sarkawt.rostam@spu.edu.iq (corresponding author)

Received: 25 December 2023 | Revised: 6 2024 | Accepted: 23 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6813>

ABSTRACT

In the realm of industrial production, condition monitoring plays a pivotal role in ensuring the reliability and longevity of rotating machinery. Since most of the production facilities rely heavily on vibration analysis, it has become the cornerstone of condition monitoring practices. However, manual analysis of vibration signals is a time-consuming and expertise-intensive task, often requiring specialized domain knowledge. The current research addresses the aforementioned challenges by proposing a novel semi-automated diagnostics system. The approach leverages historical vibration data in the form of Fast Fourier Transform (FFT) spectrums. The system extracts energy features from the frequency domain by dividing the frequency range into a predefined number of bins and summing the energy values within each bin. Subsequently, each datapoint is labeled based on the corresponding machine condition, enabling the system to learn diagnostic patterns by employing machine learning models. This approach facilitates efficient and accurate diagnostics with minimal manual intervention. The resulting dataset effectively represents and provides an interpretable result. Support Vector Machines (SVM), and ensemble algorithms are utilized to diagnose the faults instantaneously and with minimal error rates. The proposed system is capable of providing early warnings and thus prevents further deterioration and unplanned downtimes. Experimental validation using real-world data demonstrates the system's efficacy, achieving an accuracy of over 90%.

Keywords-condition monitoring; predictive maintenance; FFT; SVM; ensemble

I. INTRODUCTION

Industry 4.0 is characterized by the integration of advanced technologies, such as artificial intelligence, Machine Learning (ML), and Internet of Things (IoT) into industrial processes and maintenance. Maintenance technologies have evolved considerably to address the challenge of machine reliability. This integration allows for the collection and analysis of large amounts of data from machines, which can be used to optimize maintenance activities and improve the reliability and availability of machinery. Besides that, vibration signals and their analysis have become an integral part of every industrial plant worldwide, and ML techniques, either classical [1-3] or deep learning [4-6] models have been employed extensively for fault detection using these signals. The vibration signals from rotating equipment can be gathered easily during operation, reflecting their operating conditions in real-time, thus, current

trends in machinery fault diagnostics mainly rely on vibration signal analysis [7].

Most of the vibration signals are of non-linear nature. This characteristic of the machinery vibration signals is mainly due to the changing states of loading, interaction between parts, and loading variations. For this purpose, many works in the field of condition monitoring, justify the wide use of entropy as a measure for signal complexity and feature engineering. Therefore, the use of different entropy methods (i.e approximate, multiscale, and sample entropies) has been widely associated with vibration signal analysis and machinery fault diagnostics [8], even though many works in this field do not require the use of entropy for feature engineering. For instance, authors in [9] developed an automatic feature learning neural network eliminating the need for a conventional feature development step. This network achieved high accuracy.

Authors in [10] proposed a novel deep learning algorithm utilizing the power of Recurrent Neural Networks (RNNs) combining Bidirectional Long-Short-Term Memory (BiLSTM) and Support Vector Machines (SVMs). Additionally, authors in [11] utilized BiLSTM models with autoencoders. The proposed hybrid model for wind turbine gearbox fault detection and diagnosis exhibited higher accuracy than conventional models. SVM, as a versatile supervised learning algorithm excels in both classification and regression tasks [12-13] and is well-suited for classification problems involving small datasets [14]. In addition to linear classification, nonlinear classifications are also possible, by fine tuning the hyper parameters, and the kernel type [15-16]. Authors in [17] introduced a vibration signal dataset acquired from a laboratory testing rig. Three ML algorithms, namely k-Nearest Neighbors (kNN), SVM, and Gaussian Naive Bayes (GNB) achieved an overall accuracy of 99.75% on fivefold cross-validation using SVM. Authors in [18] developed an innovative fault diagnosis system for rolling bearings defects, powered by SVMs and Bayesian optimization. Discrete Fourier Transform (DFT) was used for feature extraction in both time and frequency domains. The work demonstrated a significant improvement from 85% to 100% accuracy compared with the base SVM model.

Further research explored other approaches to fault diagnosis using ML. Authors in [19] proposed a high-accurate early fault diagnosis method based on the Reinforcement Learning (RL) optimized SVM model. Authors in [20] developed a test rig to collect vibration signals under different bearing conditions, utilizing a Quadratic SVM model, achieving high accuracy in fault diagnosis. Authors in [21] utilized a combination of Principal Component Analysis (PCA) and SVM for fault diagnosis. The method demonstrated high accuracy in fault diagnosis under varying operating speeds. Authors in [22] proposed a fault identification method that combines variational mode decomposition, average refined composite multiscale dispersion entropy, and an SVM model optimized by Multi-strategy Enhanced Swarm Optimization achieving high classification accuracy. Authors in [23] described the development of a ML pipeline using SVMs for diagnosing bearing faults. The SVM classifier achieved an overall accuracy of 91%-99% and F1-score of 0.81-0.99. Other researchers utilized Wavelet Packet Transform (WPT), Empirical Mode Decomposition (EMD), and Variational Mode Decomposition (VMD) hybrid methods [24-26]. They were proven to be effective for fault diagnosis, achieving superior classification accuracies.

Continuing the above directions, ensemble learning has emerged as a prominent tool for classification problems, offering a robust and effective approach to enhancing predictive performance, and is poised to play an increasingly vital role in addressing real-world prediction tasks [27-29]. Authors in [30] presented an innovative diagnosis model using Complementary Ensemble Empirical Mode Decomposition (CEEMD) with SVM kernel to evaluate the health condition of bearings. This method has a high prediction accuracy and is easy to implement. Authors in [31] utilized the Multi-Scale Sample Entropy-based Energy Moment (M-SSampEn-EM) method and proposed an innovative approximate distance-based metric to optimize the feature extraction parameters.

Authors in [32] adopted an EMD method to extract features from denoised signals and classify them using multiple classifiers. The best results were achieved with a hybrid of time and spectral features using SVM with a Gaussian kernel.

To develop an interpretable dataset, for fault detection in rotating machinery, the first step is to extract the energy from each frequency band, through FFT transformation of the raw time waveform signal. The FFT has emerged as a widely used and indispensable tool in signal processing. By employing this transformation, signals can be effectively transformed to the frequency domain. Several algorithms have been developed to efficiently calculate the FFT, paving the way for the implementation of high-performance FFT processors. The Fourier transform essentially converts a signal's representation from the time domain to the frequency domain [33-35]:

$$X(f) = F\{X(t)\} = \int_{-\infty}^{\infty} X(t)e^{-j2\pi ft} dt \quad (1)$$

where $X(t)$ is the time domain signal, $X(f)$ is the FFT, and f is the frequency to be analyzed.

To measure the size or strength of a signal, signal energy as a concept in signal processing is used. If $(|x(t)|)$ is a continuous-time signal, its energy is given by (2) and if it is a discrete-time signal, it is given by (3) [36-37]:

$$E(x) = \int_{-\infty}^{\infty} |x(t)|^2 dt \quad (2)$$

$$E_x^{time} = \sum_{n=-\infty}^{\infty} |x[n]|^2 \quad (3)$$

Signal energy can also be computed in the frequency domain, using the Fourier transform of the signal. The total energy of the signal is equal to the integral of the spectral energy density over all frequencies. This is a consequence of Parseval's theorem [37], which states that:

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(f)|^2 df \quad (4)$$

where $(|X(f)|)$ is the Fourier transform of $(|x(t)|)$.

In order to calculate the cumulative energy of each individual bin, (5) [26] is used where $X(f)$ is the frequency domain signal, M denotes the number of frequency bins, each with width Δf , and the frequency range for the i^{th} bin is $[f_i, f_i + \Delta f]$:

$$\{E_1, E_2, \dots, E_M\} \quad (5)$$

To this end, the primary objective of this research work is to critically assess the practicality of incorporating ML methodologies as a diagnostic instrument within industrial environments. Also, its innovative aspect is underscored by the application of traditional ML algorithms to pre-existing historical data taken from a real-world industrial context as a case study, involving machinery that operates under severe conditions, exposed to both gradual and abrupt degradation of components. The process involves converting raw vibration waveform signal with FFT, then extracting the energies from the different frequency bands, creating a dataset that is fed into an ensemble model with SVC as the base estimator. The resulting model can classify various faults with high efficacy across various performance metrics including accuracy, precision, and F1-score.

II. METHODOLOGY

A. The Proposed Model

The proposed model for predictive maintenance comprises three key components: machine identification, data acquisition, and automated diagnosis. The system outputs one of the four predicted conditions: normal operation, unbalance, bearing defect, or a combined case of unbalance and bearing defect. This system is scalable to a wide range of similar machinery due to its reliance on dividing the FFT spectrum into frequency bins and measuring the cumulative energy in each bin. This approach offers interpretable results that are readily applicable in real-world scenarios. Figure 1 presents an overview of the proposed automated diagnosis system. The system consists of three primary components:

1. Rotating machinery: The selected rotating machinery in this study is an Induction Draft (ID) fan.
2. Vibration signal acquisition device and analysis system: This component collects the time-domain vibration signal from the rotating machinery and converts it into its FFT representation.
3. Automated diagnostic part: This component receives the FFT spectrum and extracts energy features by summing the energy values within pre-defined bins across the frequency domain. Each data point is labeled according to the corresponding machine condition.

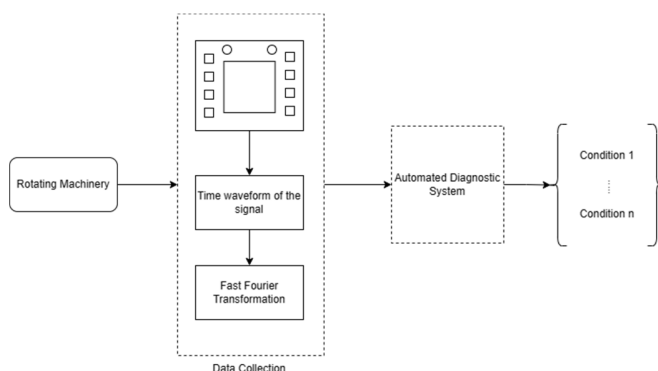


Fig. 1. Overview of the automated diagnosis system.

Subsequently, the extracted features are fed into an ML model with parameters optimized through cross-validation process. The model's output provides a textual classification, identifying one of the four predefined conditions. Figure 2 illustrates the internal workings of the automated diagnostic component. It depicts the flow of data through the component, from the input of the FFT spectrum to the final textual classification of the machine condition.

The combined use of vibration signal analysis and ML offers a promising approach for automated diagnostics of rotating machinery. This system provides potential advantages, such as improved accuracy, reduced reliance on human expertise, and increased efficiency in condition monitoring and fault detection.

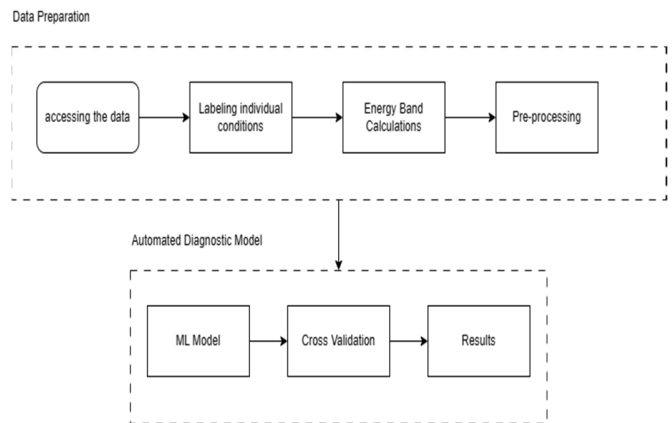


Fig. 2. Flow chart of the automated diagnostic system.

Rolling Contact Fatigue (RCF) [38-39] is a detrimental wear process that occurs in rolling bearings, characterized by the gradual deterioration of the bearing surface due to repeated rolling contact stresses. This progressive wear process can be broadly divided into five distinct stages, each marked by unique surface topology changes and accompanied by specific physical measurements. Elevated stresses in rolling contact arising from increased operating loads, faults like imbalance, misalignment, bent shaft, looseness, or distributed defects such as high surface roughness and waviness, contaminations, and inclusions, may induce topological alterations. The progression of RCF can be significantly accelerated by the presence of the above-mentioned faults within the rotating machinery. These faults introduce additional dynamic forces and stress concentrations, exacerbating the wear process and potentially shortening the bearing's lifespan as shown in Figure 3.

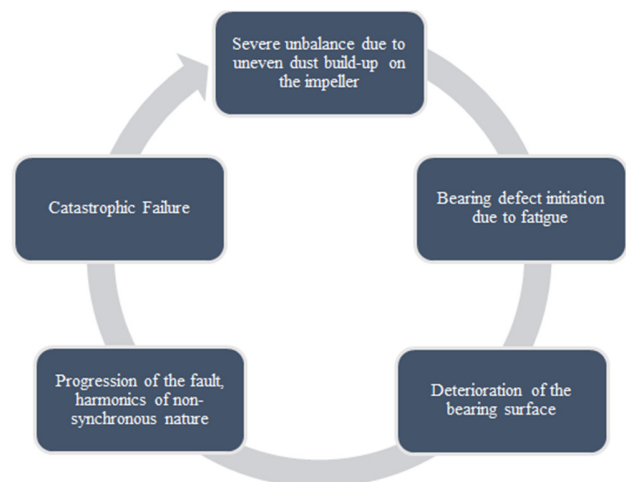


Fig. 3. Cycle of bearing failure pattern of the induction draft fan.

Rolling element bearings exhibit distinct vibrational signatures that serve as indicators of the degradation stage. These signatures arise from the interaction of rolling elements with fatigue-induced defects on bearing surfaces, generating periodic impulses known as "fundamental defect frequencies".

These frequencies are influenced by the bearing's geometrical configuration and the rotational speed of the shaft resulting in frequencies that lie within the high-frequency range of the vibration spectrum. These defect frequencies were categorized to: Ball Pass (BP) frequencies, Ball Pass Outer Race Frequency (BPFO), Ball Pass Inner Race Frequency (BPFI), Ball Spin Frequency (BSF), and Fundamental Train Frequency (FTF) as in (6-9), which are defined by:

$$BPFO = \frac{nf_r}{2} \left\{ 1 - \frac{d}{D} \cos\phi \right\} \quad (6)$$

$$BPFI = \frac{nf_r}{2} \left\{ 1 + \frac{d}{D} \cos\phi \right\} \quad (7)$$

$$FTF = \frac{f_r}{2} \left\{ 1 - \frac{d}{D} \cos\phi \right\} \quad (8)$$

$$BSF = \frac{D}{2d} \left\{ 1 - \left(\frac{d}{D} \cos\phi \right)^2 \right\} \quad (9)$$

B. Real World Dataset

Vibration signal data are acquired with an Emerson CSI 2140 data collector equipped with a 100 mV/g accelerometer. The data are collected from an ID centrifugal fan in the pyro-processing area of the local cement manufacturing facility at XYZ Company. The fan is driven by a 1600 kW medium voltage drive connected by a coupling to a shaft hanging an impeller centrally between two bearings. The rotating parts of the equipment are subjected to high-magnitude forces due to structural loads, multi-axis dynamic loads acting during operation, and rotational inertia. Rotating element bearings, including two spherical roller bearings of types SKF 22244, and SKF C2244, are mounted on the fan's shaft. The weight of the whole shaft along with the impeller exceeds 10,000 kg, exerting very high loads on these bearings, rendering them as the most critical components of the fan.

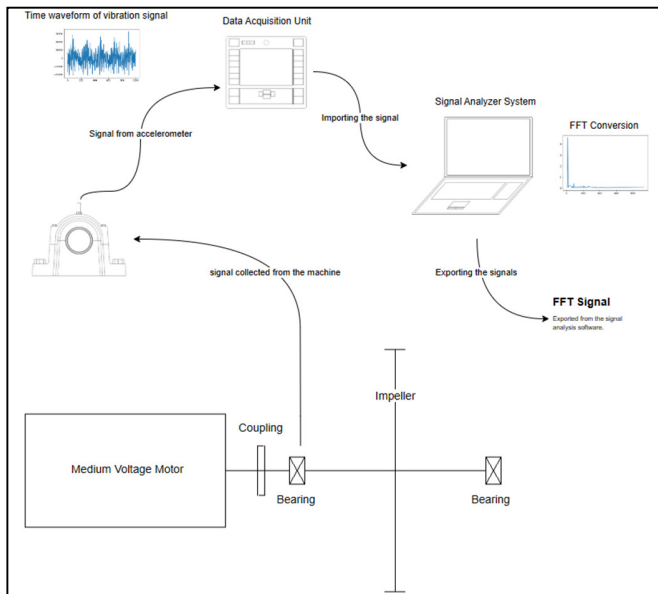


Fig. 4. The data acquisition process.

The signal is collected from three directions (horizontal, vertical, and axial) by placing a uniaxial accelerometer attached

to a data collection unit. The vibration signal is collected in the form of a time waveform, and then exported to the signal analyzer software. The signals are analyzed either by using the raw time waveform signals or by converting them digitally and displaying the signal in FFT form. This study exclusively employed the frequency-domain representation, extracting peak amplitudes and their corresponding frequencies as well as orders from the FFT spectra provided by the software. Figure 4 visually depicts the complete workflow of signal acquisition and subsequent data extraction.

III. RESULTS AND DISCUSSION

This section outlines the implementation process of the proposed model for early fault prediction, encompassing data labeling, dataset summarization, condition representation analysis, ML model evaluation, and accuracy assessment.

A. Labeling

Data labeling is a crucial step in the development of ML models, as it involves annotating raw data with meaningful labels that enable the model to learn patterns and make predictions. The historical dataset from 2016 to 2023 falls mainly into four major conditions, as shown in Figure 5.

B. Unbalanced Rotor

Unbalanced rotor is a prevalent type of machinery fault observed in rotating systems. The fault arises from an uneven distribution of mass within a rotating component about its axis of rotation. The resulting unbalanced force acts radially and increases with the square of the shaft's rotational speed [40]. The unbalanced force introduces additional fatigue stress on the bearings, potentially leading to premature failure of both bearings and rotating shafts as shown in Figure 6.

C. Results of Bearing Faults

In the case of the existence of bearing faults, the forcing frequencies show up as a multiple of the rotating speed in the frequency domain as per Table I (SKF- bearing properties).

TABLE I. FREQUENCIES FOR THE ID FAN BEARINGS (HZ)

Designation	Rotational frequency				Over-rolling frequency		
	Inner race	Outer race	Rolling element set and cage	Rolling element about its axis	Inner race	Rolling element set and cage	Outer race
C2244	1	0	0.418	2.978	9.308	6.692	5.955
22244 CCW33	0	0	0.431	3.484	10.817	8.183	6.968

Figure 7 displays an example of an outer race defect harmonics in the spectrum. The number of harmonics along with their corresponding amplitudes are indicative of the severity of the fault. The Figure indicates the existence of a defect on the outer race of the bearing. Since this work focuses on supervised ML, it is imperative to label every data point as per its corresponding condition by individually analyzing every data point acquired from the industrial facility, and combining them with the historical database on Computerized Maintenance Management Systems (CMMS). The aim of this procedure is to correlate the dates of repairs, and emergency repairs to the condition of the machine, to further verify the correctness of the labels used.

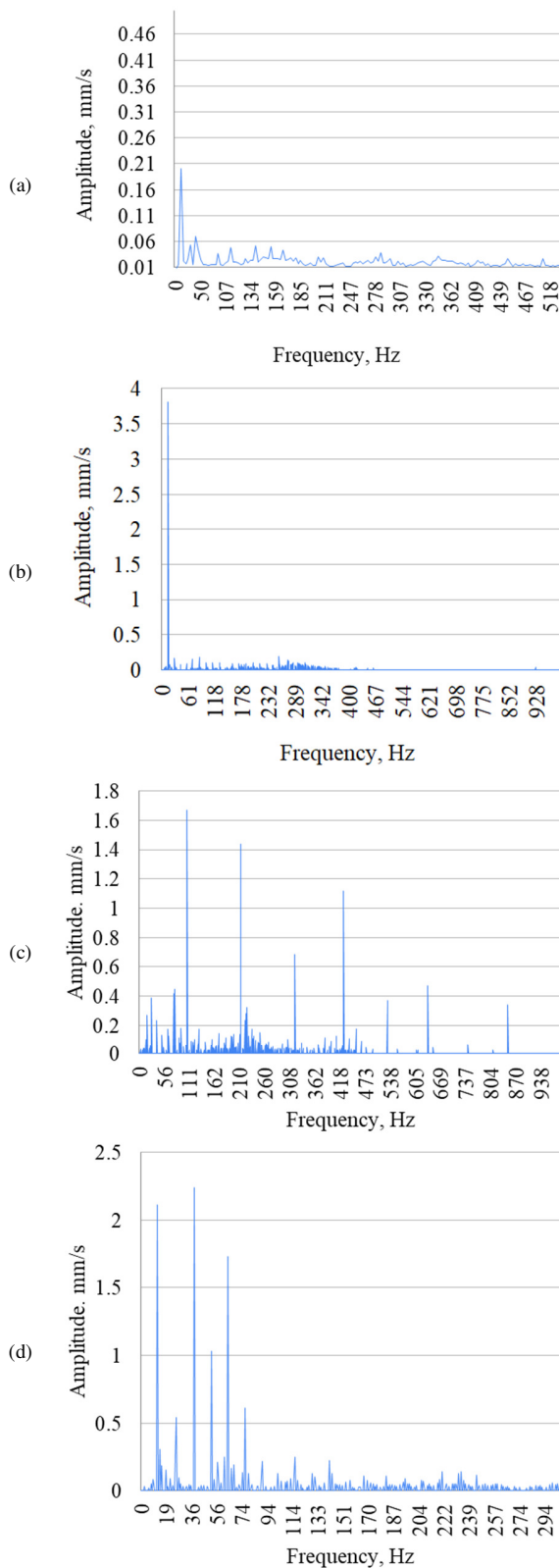


Fig. 5. Different machinery conditions. (a) Normal, (b) unbalanced, (c) bearing defect, (d) combined effect of unbalanced and bearing defect.

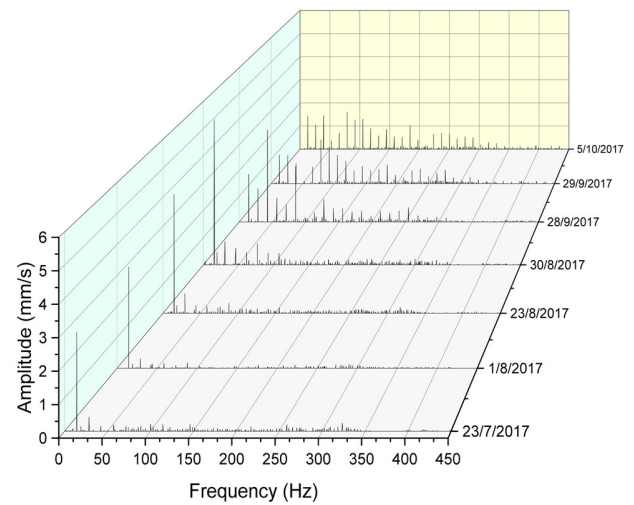


Fig. 6. Illustration of the bearing defect progression due to severe unbalance.

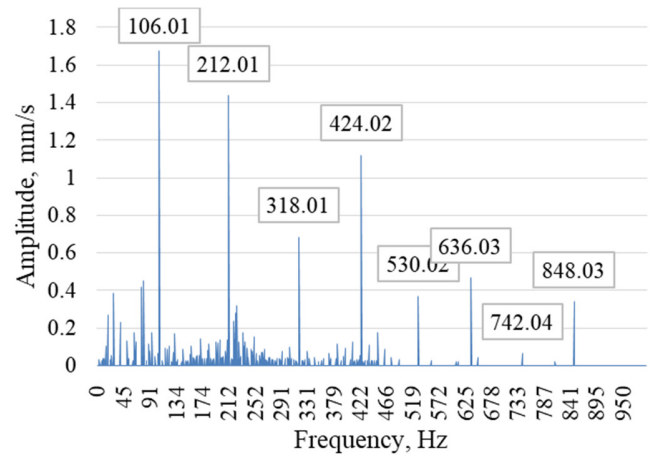


Fig. 7. FFT spectrum with 8x non-synchronous harmonics of BPFO of C2244 bearing.

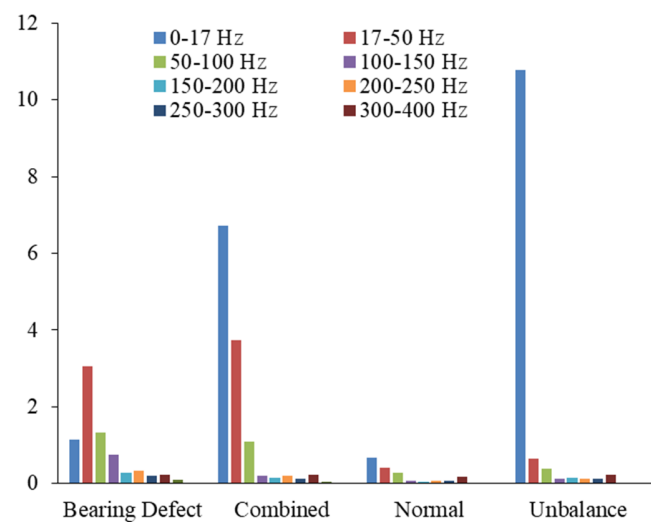


Fig. 8. Energy bands corresponding with the condition of the machinery.

The final dataset represents these cases by their energy bands. The division and width of each energy band are intrinsic to the type of fault it corresponds to. For example in pure unbalance case the band from 0-17 Hz shows the highest of the rest of the energy bands, thus indicating the unbalance case. Figure 8 summarizes energy bands of each case.

The resulting dataset labels are portrayed in Figure 9. Evidently, the dataset was unbalanced, it caused unpredictability to the model, and relies mainly on the randomness of the label distribution, to offset the effect of this unbalance. The dataset was resampled to 40 counts of each case (except the combined case), hence, overcoming the reliance on the randomness of the train/test sample distribution.

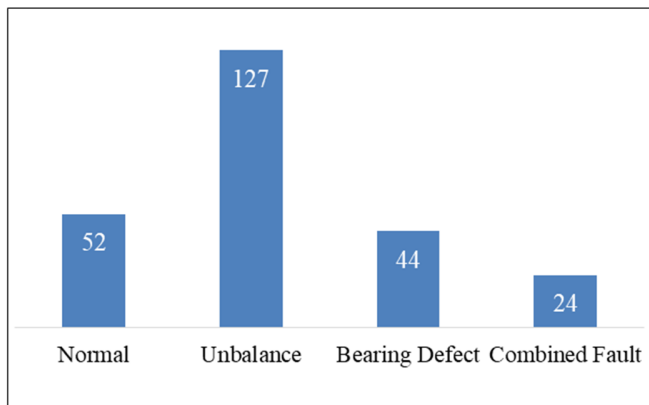


Fig. 9. The distribution of the labels in the dataset.

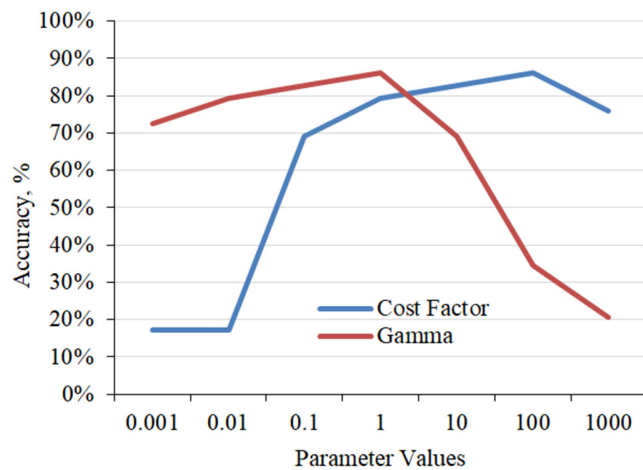


Fig. 10. SVC parameter values.

TABLE II. PERFORMANCE METRICS VALUES

Metric	SVC	Adaboost classifier	Bagging classifier	Multi-Layer Perception (MLP) classifier
Accuracy	86.21%	86.21%	89.66%	93.10%
Precision	89.22%	93.53%	92.67%	94.40%
F1 Score	86.90%	88.78%	90.43%	93.24%

Utilization of ensemble models has a very significant impact on improving the performance metrics of an SVM model significantly across all metrics. As illustrated in Table II,

the base model is SVC's accuracy after cross-validation of the parameters, which only reached 86.2% with parameter values $C=100$, $\text{Gamma}=0.1$, Kernel= Radial Base Function being chosen as optimal (Figure 10). The hyper-parameters for the models are shown in Table III. The SVC parameters, which are used as base estimator for the different ensemble models employed, were fixed across different ensemble algorithms. Figure 11 illustrates the results of experimenting with different numbers of estimators and their corresponding effect on the performance metrics. The findings show no significant difference when the number of estimators increased from 500 to 1000, justifying the choice of 500 estimators.

TABLE III. OPTIMIZED HYPER-PARAMETERS OF THE VARIOUS CLASSIFICATION MODELS

Model	Hyper parameters
SVC	(gamma=1, C=100, random_state = 42, kernel='rbf', probability=True)
Bagging	(base_estimator=svm, n_estimators=500, random_state=314)
Adaboost	(base_estimator=svm, n_estimators=500, learning_rate=0.0002, algorithm='SAMME.R')
MLP	(hidden_layer_sizes = (5000,), max_iter=3500, final_estimator =svm)

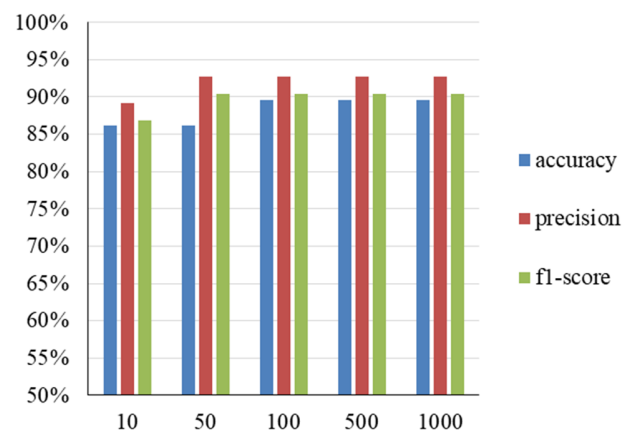


Fig. 11. Changes in performance metrics under varying number of estimators for Bagging ensemble classifier.

The same experiment was carried out on the MLP classifier. In that case, instead of the number of estimators, the experiment was conducted on varying the number of neurons in the ANN layer, and then examined the performance metrics for improvements. The results did not improve when the number of neurons increased from 500 to 5000 as depicted in Figure 12, and the number of iterations required for convergence decreased from 3500 for 10 and 50 neurons to 1500 iterations for 100 neurons and above, decreasing the computation time for the model to train.

The MLP classifier, an ensemble of SVC and ANN with two hidden layers of 500 neurons reached an accuracy of 93.1% and was most capable of recognizing the different conditions with ease and high precision, as shown in Figure 13. The numbers 0, 1, 2, and 3 denote the conditions normal, unbalanced, bearing defect, and combined, respectively.

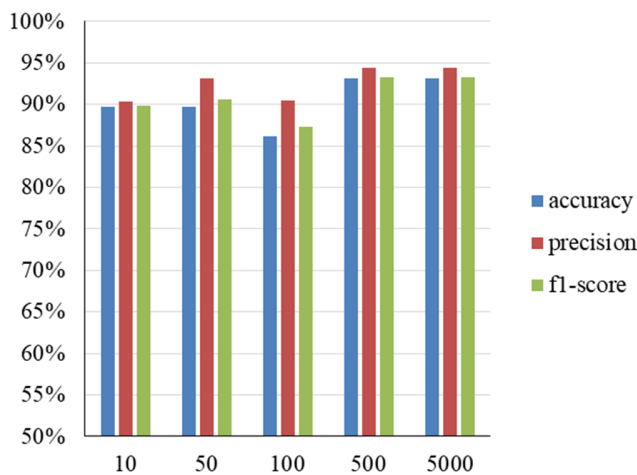


Fig. 12. Effect of the number of neurons on the performance metrics for MLP classifier.

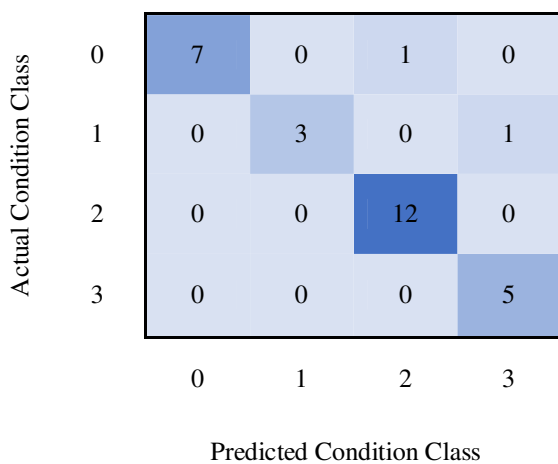


Fig. 13. Confusion matrix of stacking algorithm (SVC, ANN).

The first misclassified condition is a combined condition, whereas the model predicted an unbalance case, after examining the FFT spectrum, shown in Figure 14. However, instead of having multiple bearing defects harmonics, the existing harmonics were more sporadic and spaced out leading to misclassification.

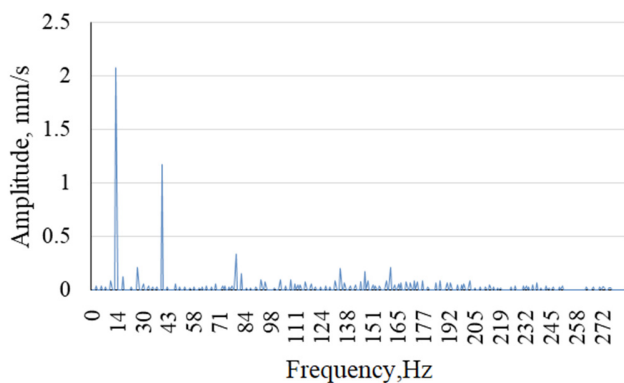


Fig. 14. FFT spectrum of the misclassified condition.

The second misclassified condition is an unbalance condition at 1x peak of 0.7 mm/s amplitude, coupled with high noise floor as displayed in Figure 15.

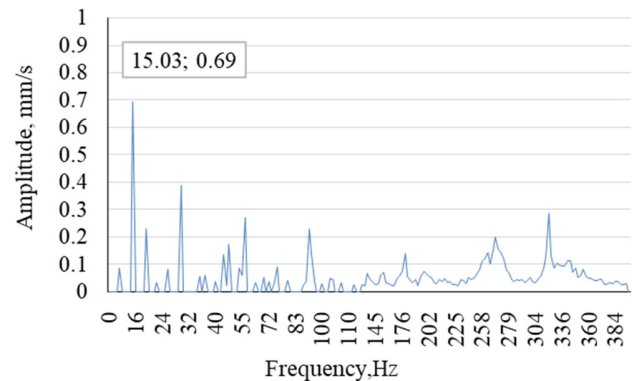


Fig. 15. Existence of 1x peak and high noise floor.

The above experiments can be seen as a feasibility study for integrating ML into an already functional system to improve quality and reduce the time taken for fault diagnosis. Thus, it is a stride towards automating fault detection and streamlining the usage of ML into local production facilities, by experimenting with the already existing data. Table IV portrays a comparison between the current work and other research works in this field.

TABLE IV. COMPARISON BETWEEN CURRENT AND RECENT WORKS

Ref.	Method	Accuracy
[17]	SVM (using test rig.)	99.97%
[22]	VMD-ARCMDELCPGWO-SVM (using test rig)	97.14%
[20]	SVM (using test rig)	97%
Current	SVM (using real-world data)	93.24%

The use of test rig data is widely employed in the field, (Table IV), and numerous datasets have been developed for this purpose, having a specified max frequency and sampling rates, as well as yielding less noisier data. The real-world dataset, however, faces several obstacles, including the non-public nature of the vibration signal data collection system, an insufficient quantity of data points, and non-standard data collection protocols (non-uniform maximum frequency limit and sampling rates). Despite those drawbacks, the findings are promising, indicating that the proposed methodology is a practicable and effective solution for automated machinery fault detection in industrial environments. Additionally, the model can effortlessly integrate with current systems without incurring additional financial costs. Besides, the findings demonstrate that the proposed model, employing a stacking ensemble of MLP classifier and SVC coupled with ANN, achieved an accuracy of 93% in identifying machinery faults, surpassing the baseline SVC model, which achieved an accuracy of 86%. It is evident that the findings are significant and showcase the potential of ML to automate machinery fault detection in cement plants.

The novelty of this work lies in its interpretability, and ease of use. As shown in the dataset, the energies from each band

can be associated with a specific case and can be used as a platform for further modifications to fit certain configurations of machinery as well as the fault types, based on the intervals of the frequency bands. This has the potential to revolutionize the way that cement plants (as an example) maintain their equipment, leading to broader adoption of ML techniques leveraging historical data to help improve production efficiency, reduce costs, and extend the equipment lifespan of the machinery.

IV. CONCLUSION

The current work assessed the feasibility of incorporating/integrating an automated fault diagnostic system upon the already existing manual condition monitoring layer, without additional costs. Once the failure pattern of the machine is understood, it is relatively easy to recognize when the fault is initiated and act ahead for a planned shutdown and replacement. In most cases, the bearings did not reach their normal end of life but were rather affected by the severely unbalanced rotor, exerting extreme loads on the bearings, leading to the initiation and development of cracks and spalling on the races of the bearing that eventually end with catastrophic failures.

The usage of machine learning algorithms was justified, as it can be scaled to encompass all similar equipment and be employed simultaneously across different devices at once, with minimal input from the domain expert. The dataset used contained varying conditions, however, there was a focus on the ones that contributed most to the failure of the equipment and excluded rare events. The most repeated conditions were normal, unbalanced, bearing defect, and a combination of the faulty states of unbalance and bearing defect. The performance across different traditional machine learning models shows a promising result, especially when a combination of more than one method was used, showing significant improvement from using standard SVC to a more accurate ensemble method yielding a change in accuracy from 86% to 93%.

As a future direction, the proposed model can be developed in several ways, including expanding the number of data points to improve the model's accuracy and robustness and developing an online real-time system.

REFERENCES

- [1] V. G. Salunkhe and R. G. Desavale, "An Intelligent Prediction for Detecting Bearing Vibration Characteristics Using a Machine Learning Model," *Journal of Nondestructive Evaluation, Diagnostics and Prognostics of Engineering Systems*, vol. 4, no. 3, Feb. 2021, Art. no. 031004, <https://doi.org/10.1115/1.4049938>.
- [2] D. Ganga and V. Ramachandran, "SVM Based Vibration Analysis for Effective Classification of Machine Conditions," in *International Congress and Workshop on Industrial AI*, Lulea, Sweden, Oct. 2021, pp. 415–423, https://doi.org/10.1007/978-3-030-93639-6_36.
- [3] J. Vives, "Vibration analysis for fault detection in wind turbines using machine learning techniques," *Advances in Computational Intelligence*, vol. 2, no. 1, Jan. 2022, Art. no. 15, <https://doi.org/10.1007/s43674-021-00029-1>.
- [4] Z. Chen, K. Gryllias, and W. Li, "Mechanical fault diagnosis using Convolutional Neural Networks and Extreme Learning Machine," *Mechanical Systems and Signal Processing*, vol. 133, Nov. 2019, Art. no. 106272, <https://doi.org/10.1016/j.ymssp.2019.106272>.
- [5] L. Eren, "Bearing Fault Detection by One-Dimensional Convolutional Neural Networks," *Mathematical Problems in Engineering*, vol. 2017, Jul. 2017, Art. no. e8617315, <https://doi.org/10.1155/2017/8617315>.
- [6] I. I. E. Amarouayache, M. N. Saadi, N. Guersi, and N. Boutasseta, "Bearing fault diagnostics using EEMD processing and convolutional neural network methods," *The International Journal of Advanced Manufacturing Technology*, vol. 107, no. 9, pp. 4077–4095, Apr. 2020, <https://doi.org/10.1007/s00170-020-05315-9>.
- [7] M. H. Mohd Ghazali and W. Rahiman, "Vibration Analysis for Machine Monitoring and Diagnosis: A Systematic Review," *Shock and Vibration*, vol. 2021, Sep. 2021, Art. no. e9469318, <https://doi.org/10.1155/2021/9469318>.
- [8] V. Vakharia, V. K. Gupta, and P. K. Kankar, "Ball Bearing Fault Diagnosis using Supervised and Unsupervised Machine Learning Methods," *The International Journal of Acoustics and Vibration*, vol. 20, no. 4, pp. 244–250, 2015, <https://doi.org/10.20855/ijav.2015.20.4387>.
- [9] X. Chen, B. Zhang, and D. Gao, "Bearing fault diagnosis base on multi-scale CNN and LSTM model," *Journal of Intelligent Manufacturing*, vol. 32, no. 4, pp. 971–987, Apr. 2021, <https://doi.org/10.1007/s10845-020-01600-2>.
- [10] Z. Qingbo, J. Han, C. Shi, and H. Gao, "Prediction of Bearing Vibration Fault State based on Fused Bi-LSTM and SVM," *Journal of Imaging Science and Technology*, vol. 67, no. 4, pp. 1–10, Jul. 2023, <https://doi.org/10.2352/J.ImagingSci.Technol.2023.67.4.040404>.
- [11] M. Sreenatha and P. B. Mallikarjuna, "A Fault Diagnosis Technique for Wind Turbine Gearbox: An Approach using Optimized BLSTM Neural Network with Undercomplete Autoencoder," *Engineering, Technology & Applied Science Research*, vol. 13, no. 1, pp. 10170–10174, Feb. 2023, <https://doi.org/10.48084/etasr.5595>.
- [12] S. Malek, C. Hui, N. Aziida, S. Cheen, S. Toh, and P. Milow, "Ecosystem Monitoring Through Predictive Modeling," in *Encyclopedia of Bioinformatics and Computational Biology*, S. Ranganathan, M. Gribskov, K. Nakai, and C. Schönbach, Eds. Oxford, UK: Academic Press, 2019, pp. 1–8.
- [13] R. Gholami and N. Fakhari, "Support Vector Machine: Principles, Parameters, and Applications," in *Handbook of Neural Computation*, P. Samui, S. Sekhar, and V. E. Balas, Eds. Cambridge, MA, USA: Academic Press, 2017, pp. 515–535.
- [14] Y. Sun, Y. Cao, G. Xie, and T. Wen, "Condition Monitoring for Railway Point Machines Based on Sound Analysis and Support Vector Machine," *Chinese Journal of Electronics*, vol. 29, no. 4, pp. 786–792, 2020, <https://doi.org/10.1049/cje.2020.06.007>.
- [15] S. K. Jalali, H. Ghandi, and M. Motamedi, "Intelligent Condition Monitoring of Ball Bearings Faults by Combination of Genetic Algorithm and Support Vector Machines," *Journal of Nondestructive Evaluation*, vol. 39, no. 1, Feb. 2020, Art. no. 25, <https://doi.org/10.1007/s10921-020-0665-7>.
- [16] P. Ilius, M. Almuhaiani, M. Javaid, and M. Abido, "A Machine Learning-Based Approach for Fault Detection in Power Systems," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11216–11221, Aug. 2023, <https://doi.org/10.48084/etasr.5995>.
- [17] B. T. Atmaja, H. Ihsannur, Suyanto, and D. Arifianto, "Lab-Scale Vibration Analysis Dataset and Baseline Methods for Machinery Fault Diagnosis with Machine Learning," *Journal of Vibration Engineering & Technologies*, May 2023, <https://doi.org/10.1007/s42417-023-00959-9>.
- [18] J. Zhou, M. Xiao, Y. Niu, and G. Ji, "Rolling Bearing Fault Diagnosis Based on WGWOA-VMD-SVM," *Sensors*, vol. 22, no. 16, Jan. 2022, Art. no. 6281, <https://doi.org/10.3390/s22166281>.
- [19] W. Zhao, Y. Lv, J. Liu, C. K. M. Lee, and L. Tu, "Early fault diagnosis based on reinforcement learning optimized-SVM model with vibration-monitored signals," *Quality Engineering*, vol. 35, no. 4, pp. 696–711, Oct. 2023, <https://doi.org/10.1080/08982112.2023.2193255>.
- [20] I. Lupea and M. Lupea, "Machine Learning Techniques for Multi-Fault Analysis and Detection on a Rotating Test Rig Using Vibration Signal," *Symmetry*, vol. 15, no. 1, Jan. 2023, Art. no. 86, <https://doi.org/10.3390/sym15010086>.
- [21] M. Pule, O. Matsebe, and R. Samikannu, "Application of PCA and SVM in Fault Detection and Diagnosis of Bearings with Varying Speed,"

- Mathematical Problems in Engineering*, vol. 2022, Apr. 2022, Art. no. e5266054, <https://doi.org/10.1155/2022/5266054>.
- [22] H. Shi, W. Fu, B. Li, K. Shao, and D. Yang, "Intelligent Fault Identification for Rolling Bearings Fusing Average Refined Composite Multiscale Dispersion Entropy-Assisted Feature Extraction and SVM with Multi-Strategy Enhanced Swarm Optimization," *Entropy*, vol. 23, no. 5, May 2021, Art. no. 527, <https://doi.org/10.3390/e23050527>.
- [23] D. Jallepalli and F. Davoudi Kakhki, "Data-Driven Fault Classification Using Support Vector Machines," in *International Conference on Intelligent Human Systems Integration*, Palermo, Italy, Feb. 2021, pp. 316–322, https://doi.org/10.1007/978-3-030-68017-6_47.
- [24] J. Guo, X. Liu, S. Li, and Z. Wang, "Bearing Intelligent Fault Diagnosis Based on Wavelet Transform and Convolutional Neural Network," *Shock and Vibration*, vol. 2020, Nov. 2020, Art. no. e6380486, <https://doi.org/10.1155/2020/6380486>.
- [25] L. Liu, L. Chen, Z. Wang, and D. Liu, "Early Fault Detection of Planetary Gearbox Based on Acoustic Emission and Improved Variational Mode Decomposition," *IEEE Sensors Journal*, vol. 21, no. 2, pp. 1735–1745, Jan. 2021, <https://doi.org/10.1109/JSEN.2020.3015884>.
- [26] P. Li, Y. Jiang, and J. Xiang, "Experimental Investigation for Fault Diagnosis Based on a Hybrid Approach Using Wavelet Packet and Support Vector Classification," *The Scientific World Journal*, vol. 2014, Feb. 2014, Art. no. e145807, <https://doi.org/10.1155/2014/145807>.
- [27] V. N. Gudivada, M. T. Irfan, E. Fathi, and D. L. Rao, "Cognitive Analytics: Going Beyond Big Data Analytics and Machine Learning," in *Handbook of Statistics*, V. N. Gudivada, V. V. Raghavan, V. Govindaraju, and C. R. Rao, Eds. New York, NY, USA: Elsevier, 2016, pp. 169–205.
- [28] S. Simske, "Introduction, overview, and applications," in *Meta-Analytics*, Amsterdam, Netherlands: Elsevier, 2019, pp. 1–98.
- [29] M. Machoke, J. Mbelwa, J. Agbinya, and A. E. Sam, "Performance Comparison of Ensemble Learning and Supervised Algorithms in Classifying Multi-label Network Traffic Flow," *Engineering, Technology & Applied Science Research*, vol. 12, no. 3, pp. 8667–8674, Jun. 2022, <https://doi.org/10.48084/etasr.4852>.
- [30] Y. Lu, R. Xie, and S. Y. Liang, "CEEMD-assisted kernel support vector machines for bearing diagnosis," *The International Journal of Advanced Manufacturing Technology*, vol. 106, no. 7, pp. 3063–3070, Feb. 2020, <https://doi.org/10.1007/s00170-019-04858-w>.
- [31] W. Jiao *et al.*, "Multi-Scale Sample Entropy-Based Energy Moment Features Applied to Fault Classification," *IEEE Access*, vol. 9, pp. 8444–8454, 2021, <https://doi.org/10.1109/ACCESS.2021.3049436>.
- [32] A. Kafeel *et al.*, "An Expert System for Rotating Machine Fault Detection Using Vibration Signal Analysis," *Sensors*, vol. 21, no. 22, Jan. 2021, Art. no. 7587, <https://doi.org/10.3390/s21227587>.
- [33] J. Zhang *et al.*, "Coupling a Fast Fourier Transformation With a Machine Learning Ensemble Model to Support Recommendations for Heart Disease Patients in a Telehealth Environment," *IEEE Access*, vol. 5, pp. 10674–10685, 2017, <https://doi.org/10.1109/ACCESS.2017.2706318>.
- [34] B. Popa, M. Roman, and R. L. Constantinescu, "Fast Fourier processing and real-time transformation system for a dynamic vibration signal," in *20th International Carpathian Control Conference*, Krakow-Wieliczka, Poland, Dec. 2019, pp. 1–6, <https://doi.org/10.1109/CarpathianCC.2019.8766039>.
- [35] H.-C. Lin, Y.-C. Ye, B.-J. Huang, and J.-L. Su, "Bearing vibration detection and analysis using enhanced fast Fourier transform algorithm," *Advances in Mechanical Engineering*, vol. 8, no. 10, Oct. 2016, Art. no. 1687814016675080, <https://doi.org/10.1177/1687814016675080>.
- [36] B. Boashash *et al.*, "Advanced time-frequency signal and system analysis," in *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, Amsterdam, Netherlands: Elsevier, 2016, pp. 141–236.
- [37] L. F. Chaparro and A. Akan, *Signals and Systems Using MATLAB*. Cambridge, MA, USA: Academic Press, 2018.
- [38] I. El-Thalji and E. Jantunen, "A descriptive model of wear evolution in rolling bearings," *Engineering Failure Analysis*, vol. 45, pp. 204–224, Oct. 2014, <https://doi.org/10.1016/j.engfailanal.2014.06.004>.
- [39] I. El-Thalji and E. Jantunen, "Dynamic modelling of wear evolution in rolling bearings," *Tribology International*, vol. 84, pp. 90–99, Apr. 2015, <https://doi.org/10.1016/j.triboint.2014.11.021>.
- [40] A. R. Mohanty, *Machinery Condition Monitoring: Principles and Practices*. Boca Raton, FL, USA: CRC Press, 2014.
- [41] Y. Bella, A. Oulmane, and M. Mostefai, "Industrial Bearing Fault Detection Using Time-Frequency Analysis," *Engineering, Technology & Applied Science Research*, vol. 8, no. 4, pp. 3294–3299, Aug. 2018, <https://doi.org/10.48084/etasr.2135>.

A Survey on the Latest Intrusion Detection Datasets for Software Defined Networking Environments

Harman Yousif Ibrahim Khalid

College of Science, University of Duhok, Kurdistan Region, Iraq
harman.khalid@uod.ac (corresponding author)

Najla Badie Ibrahim Aldabagh

College of Computer Science and Mathematics, University of Mosul, Iraq
najlabadie@uomosul.edu.iq

Received: 16 December 2023 | Revised: 2 January 2024 | Accepted: 10 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6756>

ABSTRACT

Software Defined Networking (SDN) threats make network components vulnerable to cyber-attacks, creating obstacles for new model development that necessitate innovative security countermeasures, like Intrusion Detection Systems (IDSs). The centralized SDN controller, which has global view and control over the whole network and the availability of processing and storing capabilities, makes the deployment of artificial intelligence-based IDS in controllers a hot topic in the research community to resolve security issues. In order to develop effective AI-based IDSs in an SDN environment, there must be a high-quality dataset for training the model to offer effective and accurate attack prediction. There are some intrusion detection datasets used by researchers, but those datasets are either outdated or incompatible with the SDN environment. In this survey, an overview of the published work was conducted using the InSDN dataset from 2020 to 2023. Also, research challenges and future work for further research on IDS issues when deployed in an SDN environment are discussed, particularly when employing machine learning and deep learning models. Moreover, possible solutions for each issue are provided to help the researchers carry out and develop new methods of secure SDN.

Keywords-*software defined networking; intrusion detection systems; network security; InSDN; datasets*

I. INTRODUCTION

SDN brings significant advantages to network security and has solved many security issues in conventional networks due to its promising features, such as centralized management, bird view, and statistics from forwarding devices to controller. These characteristics improve network security and make it easier to deploy threat detection systems via software applications that make use of open APIs. Despite the benefits offered by the SDN architecture compared to traditional networks, it is vulnerable to cyberattacks and faces new possible threats that do not occur in the traditional networks of today. SDN itself does not have security built-in [1, 2], and one major issue that may hinder the widespread use of SDN is the possibility of new assaults [3]. SDN characteristics, such as network programmability and centralization control introduce new fault and vulnerability, which open the doors for new threats that did not exist before [4, 5]. In [6], seven potential attack vectors against SDNs are listed. Three of these attacks are specific to SDN networks [4, 7–9]. Authors in [10] argue that SDN networks may be more susceptible to malicious traffic than traditional environments due to the decoupling of

the control plane and data plane. A security breach in conventional networks causes only minor damage to a small portion of the network, whereas an attack on the SDN controller might have catastrophic consequences for the entire network [7–11]. The attacks target different parts of SDN, as depicted in Figure 1. A countermeasure step to secure SDN is using Intrusion Detection Systems (IDSs) due to their ability to locate and identify malicious activity in the network by examining network traffic in real time [12]. IDS is a tool, which can be either in the form of software or hardware, employed in a network system to observe and assess the behavior of an individual computer or oversee and evaluate the entire network traffic with the aim of distinguishing between legitimate and malicious data flows. As shown in Figure 2, there are two types of IDSs, signature-based and anomaly-based. Signature-based systems are widely used commercially, and malicious traffic can be detected based on the predefined rule. The drawback of this type is that any change performed in the attack signature, even if it is very small, the system will not detect the attack making this type of system unable to detect zero-day attacks [13].

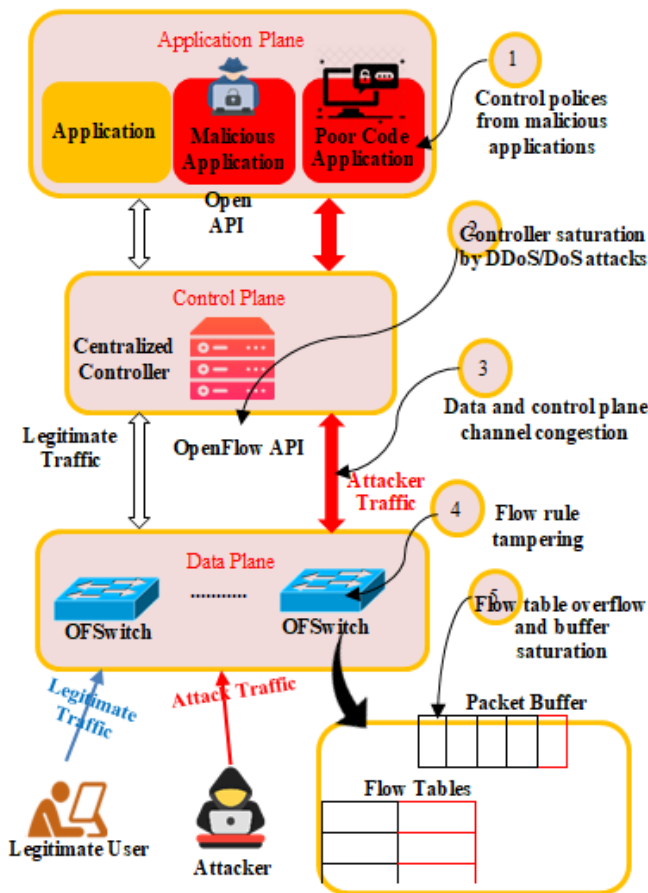


Fig. 1. A taxonomy of SDN security.

Anomaly-based IDSs have attracted significant interest from the academic community due to their capacity to detect new and previously unknown attacks by pinpointing any deviation from the typical traffic pattern. The source of data that is inspected by the IDSs can be the whole data packet, including packet headers and packet payload in packet-based IDS in contrast to the flow-based IDS, which only analyzes the basic information of the communication found in the packet header. Authors in [14] argued that packet content features are not directly accessible in OpenFlow protocol. The amount of data analyzed by flow-based IDSs is less. Therefore, such systems are faster and more computationally efficient than packet-based systems [15]. It is important to note that a flow-based approach is unable to detect attacks embedded in packet payloads. Therefore, to provide a high level of security, both approaches can be implemented together, forming a hybrid method [15]. Deploying IDS in an SDN controller, contrary to the distributed methodology used in traditional networks, not only reduces the expense of purchasing more detection equipment, but also improves detection efficiency due to the centralized view of the controller and the easily obtained statistical details for the IDS application [16]. Fortunately, the centralization of the SDN makes the training of a machine and Deep Learning (DL) based IDSs easier [17] and has become an active area of research [18]. With the use of training data, IDSs possess the ability to automatically learn and recognize patterns

from data [20] and the controller's visibility make the deployment and training of AI-based IDSs easier [17]. These IDSs have the ability to modify their behavior based on the changing traffic patterns in the network, making them better equipped to handle emerging security risks [18].

The quality of the detection mechanism depends on the quality of the dataset used for training [20]. One major issue, though, is the lack of high-quality datasets for network traffic and intrusion detection. In various fields, like language translation and computer vision, there are numerous high-publicly accessible quality datasets. The main cause of the intrusion detection domain's lack of publicly available datasets is related to privacy and legal concerns [12]. It is important to note that network data may include sensitive private customer information, and sharing such data with the public would be both unlawful and a breach of privacy. However, anomaly detection in SDN is predicted using public IDS datasets, which are gathered from conventional networks. It is argued [20], that dataset incompatibility with the SDN environment creates false predictions since the framework of SDN is different from the traditional network.

In this survey, an overview of the current approaches that apply DL and Machine Learning (ML) based IDSs in SDN environment is provided. Recent works that used InSDN [21] and the novel SDN-related dataset that was designed to capture unique character attacks in SDN were reviewed.

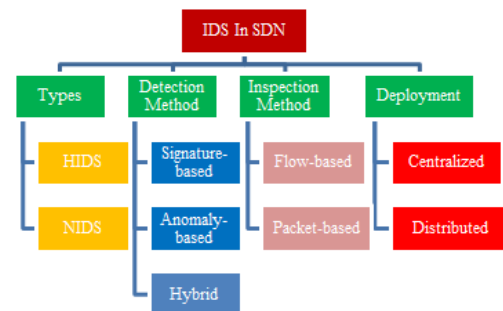


Fig. 2. Taxonomy of IDSs.

II. DATASETS

Using a high-quality dataset for training AI-based IDSs has a great impact on the accuracy of predictions. Only a few datasets related to intrusion detection are publicly available. The primary cause for the lack of publicly available datasets is related to privacy and legal concerns since network data can contain sensitive customer information [22]. However, some available datasets were used for intrusion detection in traditional networks, such as KDDcup99, NSL_KDD, ISCX2012, CICIDS 2017, and CICIDS. KDDcup99 and its revised version, NSL-KDD, are the most popular intrusion detection datasets and are considered benchmarking datasets. However, these datasets are unreliable and outdated, since they were released more than two decades ago [4, 9]. Authors in [16, 17] analyzed both datasets in the SDN context and found that only six (6) features are needed over the 41 available in the SDN environment. Some authors still use those datasets in their

work to predict the performance of new models or to compare them to other datasets, but it is not recommended to use them for real-time detection systems [16]. The ISCX2012 dataset contains only two types of attack: DoS and brute force and HTTP traffic is included as normal traffic, which is not the current standard on the Internet today. Therefore, the dataset is not suitable for modern evaluation. CICIDS 2017 is the updated version of ISCX 2012. It involves a total of 2,830,743 instances, with 19.7% of the total data representing attacks [20]. Those datasets are outdated, have a smaller number of attacks, or are incompatible with the SDN environment [17].

In 2020, a new open-source dataset known as InSDN was made publicly available to be used for the training and evaluation of IDSs within the SDN context [23]. The distribution of the 361,317 total samples in seven attack categories x in the InSDN dataset are shown in Table I. The normal class has 68,424 samples, while the attack class has 292,893 examples [20]. The dataset is split into three groups, including attacks traffic targeting the Metasploitable Server 2 virtual machine as well as the OpenvSwitch (OVS) VM [17]. The normal traffic encompassed many application services, entailing HTTPS, DNS, SSH, FTP, email, etc.

Authors in [7, 15] argue that most of the existing literature views the intrusion detection challenge of SDN as similar to that of a conventional network. Using outdated datasets to train SDN-based IDSs can lead to severe problems, as they are only capable of detecting attacks that exhibit identical behavior in both SDN and traditional networks. The IP-traditional network and SDN are significantly different in their operation. Additionally, the protocols utilized in SDN are not the same as those employed in conventional networks, like OpenFlow [7]. Similarly, new threats have surfaced as a result of the division between the data plane and the control plane. The SDN controller itself or the data communications may be the targets of these assaults.

TABLE I. INSDN DATASET ATTACK CATEGORIES AND DISTRIBUTION

Dataset	Normal traffic/ Attack traffic	Sample number	Total %
Normal	FTP, DNS, HTTPS	68424	68424 (19.9%)
Metasploitable 2	Attack Category		
	DDoS	73529	136743 (39.76%)
	Probe	61757	
	DoS	1145	
	Brute force attack	295	
	R2L	17	
OVS	DoS	52471	138772 (40.34%)
	DDoS	48413	
	Probe	36372	
	Brute force attack	1110	
	Web attack	192	
	Botnet	164	

One example of a novel assault in a SDN environment is a Distributed Denial of Service (DDoS) attack that specifically aims at the SDN controller [24]. The OpenFlow switch will forward the unmatched flow packets it receives to the SDN controller as a packet-in message for additional analysis and

handling [20]. An attacker can send various packets with a randomly spoofed destination address even if they do not have to spoof the source address, which is done in traditional networks. All traffic will be forwarded to the controller. The latter will overwhelm it and lead to controller failure, which eventually stops the entire network. Utilizing the InSDN dataset for assessing the effectiveness of anomaly detection models offers more precise outcomes, considering the types of assaults in SDN differ from those often observed in traditional networks. Thus, adopting such a dataset for model evaluation in SDN can serve as a reliable indicator of actual world circumstances. Furthermore, the InSDN dataset is free of any duplicate entries, preventing the learner model from showing a bias towards the most frequently occurring records [12].

III. CURRENT RESEARCH REVIEW

In this survey, existing work is deeply investigated using the InSDN dataset as depicted in Table II. Authors in [25] conducted an analysis of the InSDN dataset, considering all attacks, and presented attack specific feature selection to identify the features that have the greatest impact on anomaly detection to reduce the execution time of the model while maintaining high performance. They argue that because of overfitting and redundant features, a large dataset with many features will take longer for the detection model to execute and might not improve accuracy. To lessen the unbalanced set, they divided the dataset into six new ones, each of which contains normal traffic with a single attack. They performed multiple experiments using the SelectKBest feature selection algorithm to rank the top 10 features for each attack. They found that for all attacks, there were three common features in sequence: duration, Fwd IAT Tot, and Bwd IAT Tot. According to the experiment conducted, duration has the highest impact on all attacks in the InSDN dataset except for the DDoS attack, where Fwd IAT Tot was the most effective. This effectiveness may be attributed to the fact that the victim device is flooded with a huge number of requests as a consequence of the DDoS attack. From the result, it has been observed that both DDoS and probe attacks exhibit the same behavior in that the flow byte feature was dropped suddenly during the attack. The reason behind this behavior is that probe attacks usually scan the target system to discover some information, which results in a very low flow size measured in bytes per second. The normal flow bytes decreased from 0.3332 to zero for DDOS and 0.0001 for probes. However, their methodology of splitting into six datasets means that each attack with a normal sample produces overfitting when detecting attacks, such as botnets and web attacks due to the imbalanced dataset, where the number of samples in a normal traffic sample is much greater than those attacks. Moreover, their work only considered selecting important features for each attack and did not perform any detection classification.

In the same direction, to improve the performance of IDS to detect probe attacks more accurately, the Grey-Wolf Optimizer (GWO) algorithm for feature selection was implemented in [26]. The authors discussed the benefits of feature selection to the overall detection model. They highlighted that feature selection is essential in minimizing computation time, which will make the classifier have high accuracy with optimal

features selected and decrease the dataset size for testing and training. Moreover, for real-time detection, it is easier to extract fewer features, thus decreasing the detection time. They showed that by selecting a subset of 8 features using the LightGBM classifier, accuracy increased to 99.8%, while when using all features, it was 77.3%. Nevertheless, their topology was the same as that of the creator of the dataset [21]. Authors in [10] supported the fact that using many features could be useful in detection accuracy, but it could lead to issues, such as increased model complexity and training costs. Focus is given on various attacks and the Hierarchical Multi-Class (HMC) architecture is proposed to address the imbalance problem in the InSDN dataset and improve the performance of minority classes, like BFA, botnet, and web attacks. To boost the amount of minority class samples, they use SMOTE sampling. To detect abnormal traffic in SDN, they used many ML and DL models, namely K-Nearest Neighbors (KNN), Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), Adaboost, Bagging, Radial Basis Function Support Vector Machine (RBF-SVM), Linear Support Vector Machine (lin-SVM), Multilayer Perceptron (MLP), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU). Through the conducted experiments, they showed that DDoS, DoS, and probes had good identification performance compared to other attacks because of their majority in the InSDN dataset. At each stage, they used binary classification between the top major class, for instance, the normal traffic class, and the remaining classes in the dataset. In the next stage, after excluding the previous top class, the new top class, such as DDoS, is compared to the rest, and so on. The process of class excluding continues until all classes are filtered by binary classification. Using this method, they improved the detection accuracy of minority classes, like BFA, web attacks, and botnets. However, their method of hierarchical multi-class detection using binary classification at each stage leads to an increase in detection time and computation. Moreover, they implemented and verified their framework for DDoS attacks only.

Some works that have been published have used a hybrid method of ML with DL to improve performance. In [27], an attack detection and mitigation module was proposed that utilized a hybrid model of CNN and Extreme Learning Machine (CNN-ELM) to classify DDoS attacks in an SDN environment. Contrary to previous works, the authors developed a mitigation mechanism along with a detection module. Mitigation is done through IP traceback utilizing a blacklist, which records the abnormal traffic detected by the IDS. Their model detected DDoS attacks by using features extracted from the SDN environment, which were provided by packet-in messages toward the controller as well as the statistics messages provided by OpenFlow switches to the controller. A subset of 12 features from the InSDN dataset were mapped to the OpenFlow switch's extracted features. In addition, they considered four additional features, such as average speed flow, average duration, average packet size, and ratio asymmetric flow. Through the conducted experiments, they demonstrated that using a subset of 12 features not only increased accuracy, but also reduced test time. However, this

methodology creates overhead in the controller since every packet-in message, which will not be effective during a DDoS attack, should be checked. Moreover, there was no clear description of how features extracted from packet-in messages. Also, their methodology and the manually created four features were not verified. Similarly, authors in [28], proposed the Deep Convolutional Neural Network (DCNNs) to detect DDoS in SDN. They suggest similar detection and mitigation mechanisms but they used only the features provided by the flow table through OpenFlow statistics messages, and those messages were periodically sent to the controller for anomaly detection. They mentioned that only 12 features of InSDN were mapped to the extracted information from the OpenFlow switch. They argue that the existing solutions engaging a large number of features for ML or DL require more functions to extract those features, which create network congestion and latency. On the other hand, utilizing a small and limited number of features does not provide reliable attack detection. However, in practical implementation, they employed 78 features for training, not only 12, which makes it difficult to map the basic features provided by OpenFlow switches to this huge number. Furthermore, their methodology requires every packet-in message to be checked by the controller, as well as to periodically request statistics from the switch to create overhead in the controller. Similar to previous work, there was no clear description of how the features were extracted from packet-in, while, also, they were not verified.

Authors in [29] examined many DL models, such as the GRU, LSTM, and RNNs, in order to develop an IDS capable of detecting DDoS attacks. They selected 48 features from the InSDN dataset using the framework method described in [30] to acquire the specific features related to SDN. LSTM provided the best accuracy, but in terms of training time, RNN was the most optimal. Nonetheless, they did not consider the imbalanced dataset caused by minority classes when deploying multi-class detection to prevent overfitting. Additionally, there was no real-time classification; in their work, they only considered analyzing DL models. In the same direction, authors in [18] compared several ML models that have high performance with less execution time, such as DT, RF, and Adaboost, to determine the best candidate for the development of ML-based IDS. They revealed that using the SelectKBest function from scikit-learn, which selects relevant features, reduced the number of features from 83 to 7 without significantly degrading performance, while the execution time was reduced rapidly from 10 to 0.5 s. However, as in [29], the authors did consider an imbalanced dataset when using multi-class detection to prevent overfitting.

Authors in [7] propose a hybrid DL approach that combines CNN with RF, KNN, and SVM to classify network traffic. While the aforementioned ML techniques handled the categorization problem, authors employed CNN to extract more complex representations of the data attributes. Most research fails to take into account the impact of overfitting when putting models into practice, which results in poor detection of zero-day attacks. For this reason, a regularization method called SD-Reg was implemented to deal with the overfitting issue. The authors claimed that the imbalance in the InSDN dataset was caused by the insufficient amount of samples for U2R, Web,

and botnet assault. Some attackers always focus on those attacks since the detection models have false predictions about them, or other researchers may ignore them. Thus, in the training multi-classification process, they employed a combination of the oversampling (SMOTE) and undersampling approaches to eliminate randomly selected samples from the majority class and duplicate samples from the minority class. After performing several experiments, they showed that, using softmax with the SD-Reg regularization technique combined with CNN performed better than regularizations L1 and L2. In the next experiment, they replaced softmax with ML techniques, such as SVM, KNN, and RF to work with SD-Reg regularization and for 48 features, compared to single CNN models, CNN-SVM, CNN-KNN, and CNN-RF produced better results for binary and multi-classification. For detecting unknown attacks, for 9 attacks in the dataset, they conducted several experiments with one attack removed from the training set each time and then used for testing, and so on, for all attacks.

Authors in [20] presented a DL model based on LSTM and an autoencoder to detect DDoS attacks in SDN, with a limited number of features to create a lightweight approach to reduce the overhead of applying the detection model. They used Information Gain (IG) and RF algorithms for feature selection to analyze the most relevant feature to the DDoS attack. They utilized the same dataset, once with 48 features and once with only 10 features, and they showed that the accuracy does not decrease greatly. Their mechanisms include the flow collection and extraction module, which uses OpenFlow statistics messages to get necessary information from switch to controller periodically following a fixed time interval. The authors utilized CICFLOWMETER, which is a tool used for datasets, such as InSDN, CICIDS 2017, and CICIDS 2018, which extract 83 features from traffic flows. They argue that not every CICFLOWMETER feature could be extracted for usage in an SDN setting. Only through OpenFlow calls can the SDN controller obtain statistical data from OpenFlow switches, including flow duration, packet count, and byte count. Thus, they employ the methodology of [30] to identify subfeatures that may be readily retrieved, either directly from the SDN controller, or by computing flow statistics, like the standard deviation, mean, minimum, and maximum of flow features. They conducted experiments and feature selection on three different flow based datasets. It was observed that the subset of selected features in the InSDN dataset is different from those in CICIDS 2017 and CICID 2018, while many common features were found between CICID 2017 and CICIDS 2018 because they are from conventional environment. Traditional networks and SDN platforms are not comparable due to their distinct features and functionalities. Furthermore, the identification of features varies as well as their prominence inside each network. For instance, in a traditional network, flow duration refers to the length of time for which the connection between a source host and a destination host is active. In SDN, flow duration refers to the amount of time that a flow entry remains in the flow table of a switch. Hence, the "duration" attribute in SDN is closely associated with DDoS attacks, as these attacks involve the malicious flow remaining in the switch flow table for a long time. They validated their claim and demonstrated

how performance decreases significantly when training with one dataset and testing with another. This validation proves that other datasets that were collected in traditional networks need to be carefully deployed in the SDN network.

Authors in [31] developed an IDS using a hybrid model utilizing an LSTM and CNN combination to extract temporal and spatial information from input data. The accuracy was 96.32% for multi-classification in the InSDN dataset. To overcome the overfitting problem, they used two regularization methods: L2 Reg and dropout. They highlighted that to improve the performance of CNN and detect new intrusions, the overfitting must be reduced to increase accuracy. They used attack samples in testing that were different from those in training. Despite its high accuracy, the high false alarm percentage of the hybrid model could prevent its deployment in the production system. Moreover, the authors stated that their model failed to provide an acceptable result for detecting new attacks. Authors in [13] investigated a methodology to solve the problem of unlabeled and unbalanced dataset. For anomaly identification, they suggest a hybrid strategy based on an LSTM autoencoder and One-Class Support Vector Machine (OC-SVM). They used unsupervised training to solve the problem of an unbalanced dataset by training with a normal class. When there are anomalies, the model produces a significant error because it is unable to identify and rebuild anomaly instances. A threshold known as reconstruct error was employed to distinguish between normal and anomalous data. The shortcoming of OC-SVM, its low capability to work with high-dimension datasets, is solved by combining it with the LSTM autoencoder. The LSTM autoencoder model's data output is reduced to a smaller dimension and then trained using the OC-SVM algorithm to enhance the classification performance. The outcomes of the experiment show that the suggested model provides a higher detection rate. However, they used binary classification with approximately 57,000 normal samples and randomly selected 46,000 samples from all attacks. Combining all attacks in one category with a small number of samples does not provide accurate attack detection, since all attacks have different effective attributes. Moreover, they removed socket features, such as source IP, destination IP, flow ID, etc., only from the dataset to avoid the overfitting problem. Using a large number of features does not solve the issue; in fact, it increases the training and execution time, making the real time implementation impossible.

Authors in [11] introduced a lightweight supervised learning model to identify DDoS attacks targeting SDN controllers using only one feature of fluctuation of flow, which is the count of packet-in messages to the controller in a fixed time slice and for many consecutive times to avoid the behavior of a normal burst. They created their own dataset for the proposed system, but for testing and training their model, they used the InSDN dataset. The idea behind using only one feature is that it will be easier to obtain while it consumes less time and resources for training and real-time prediction. They implemented a multiple ML model with seven selected features of InSDN, which were flow-id, protocol, timestamp, flow-pkt/s, bwd-pkt/s, pkt-len-mean, init-bwd, and win-byts. The conducted experiment shows BT and KNN were the best in terms of accuracy, while in terms of accuracy and training time,

CPU utilization, and decision time, KNN was the most optimal. They tested their work using their own dataset and obtained an accuracy of 99.4% with BT using one feature. It is argued that employing many features will lead to either higher performance or overfitting for some models. However, they did not mention the methodology of feature selection and some of the selected features, such as flow-id and timestamp, which can affect the learning process during model training, leading to overfitting were irrelevant. What is more, continuous checking of the count of packet-in creates a load to the controller, and the methodology of time slices causes a delay in decision time. Ultimately, the technique loses potential because it assesses just one feature to train the intrusion system.

A deep evaluation and analysis was performed in [17], considering IDS implementation in the SDN network and relevant datasets. The author claims that there is a lack of research on IDS in SDN environments. Furthermore, most of the current publication views the IDS issue in SDN similar to that in conventional networks. Also, a significant number of them depend on datasets that were created for conventional networks. They presented novel evaluation work on the InSDN dataset, such as the classification of single/multiple attacks coming from different/same source of data, namely the OVS dataset and Metasploitable 2. Unlike other published research, this one considers training and testing using a single data source. In addition, the importance of the AUC metric for imbalanced dataset classification is highlighted. It was also highlighted that the InSDN dataset is a high quality dataset suitable for IDS in SDN and consists of identical attacks originating from several sources, distinguishing it from former datasets. Other researchers overlooked this particular point. The author conducted several experiments analyzing a new dataset; the same source of data was used for training and testing of single attack detection, and the results were quite satisfactory. However, when conducting the experiment again with different sources for testing and training for single attack detection, the results were good except for BFA and DoS, whose scores were lower in the case of metasploitable server 2 used for training and OVS for testing. It was mentioned that the reason is their limited number of metasploitable server, which affects training. The same problem arose for the probe when the training data source was OVS and testing was conducted on Metasploitable Server 2. Moreover, the experimental multi-classification detection results with the same source for testing and training were exceptional, but for different sources, there was degradation in the performance, due to the existence of heavily populated classes, such as DDoS. It is concluded that the existing technique must be shifted to address the issue of diverse data sources. U2R, botnet, and web attacks were excluded from the analysis due to the limited sample number.

In another direction, some researchers developed ensemble mechanisms to improve performance. Voting algorithms are characterized by low errors and overfitting. When combining multiple classifiers, they provide higher accuracy than a single classifier. By combining NB, KNN, DT, and ET in the V-NKDE ensemble classifier model, authors in [3] created a technique for DDoS detection and mitigation. The voting classifier's concept is to combine many ML algorithms and predict the class label using either an average prediction

probability or a majority vote. Furthermore, their systems consist of a collaborative module that notifies other controllers about the attacks, a classifier module for detection, and a mitigation module for blocking attacker ports. The reason behind the collaborative module is that after blocking attacks, there is a vast number of malicious flows in flow tables. These flows are useless, but they waste the memory of the switch. Collaborative modules clear flow tables from such flows. They implement a data collector unit in the controller, which periodically receives OpenFlow statistics messages from OpenFlow switches to obtain details about flow in the flow table. For attack detection, five tuple features in SDN have been considered. To evaluate their model, they used 48 features of InSDN for multi-class detection, with 99.84% accuracy. The experiment showed that the real-time traffic classification accuracy was 99.1% with 0.002 False Positive Rate (FPR).

IV. DISCUSSION

In this survey, a deep analysis was conducted for research employing the InSDN dataset in the SDN environment. After the InSDN dataset was published in 2020, many researchers utilized this dataset during training and testing their models. Among those works, 16 papers were selected. Most of the existing works using InSDN reduced the number of features as follows: 7 of the 16 works used less than 20 features, 6 of them used 48 features [30], and two others used variations in the number of features. The motive behind employing less features was to eliminate overfitting and provide a lightweight intrusion detection system. IG and SelectKBest were the most used feature selection algorithms. Due to the importance of the SDN controller and the fact that it is an attractive target for DDoS attacks, as depicted in Figure 3, seven among the 16 existing solutions performed binary classification and four of them were specific for DDoS detection, one for probe attack detection, and the remaining for classification between normal and abnormal traffic. Various classifiers have been used in the literature and it has been shown that RF was the most optimal. Moreover, LSTM and CNN were often utilized in hybrid classifiers, in combination with CNN achieving good results in image processing due to its power of learning from spatial features, while the power of LSTM resides in its ability to learn from temporal correlations of network traffic that generate times series data [32]. Additionally, LSTM performs better with large datasets [33].

It is clearly illustrated in Figure 4, that among 16 works, four only analyzed the dataset and did not perform any mechanism for attack detection. The remaining 12 works implemented IDS, 5 of them extracted features from OpenFlow statistics messages, only 3 of them used packet-in messages for feature extraction, whereas the remaining did not provide any details about the feature extraction methods used.

Real-time detection is considered an important metric for evaluating the intrusion detection mechanism in the real world and online. However, among those works that deployed intrusion mechanisms, only three of them provided real-time detection. It is clear that real-time detection is not a straightforward process due to the difficulty of getting the required features in real-time and mapping them to dataset features for feeding the classifier model.

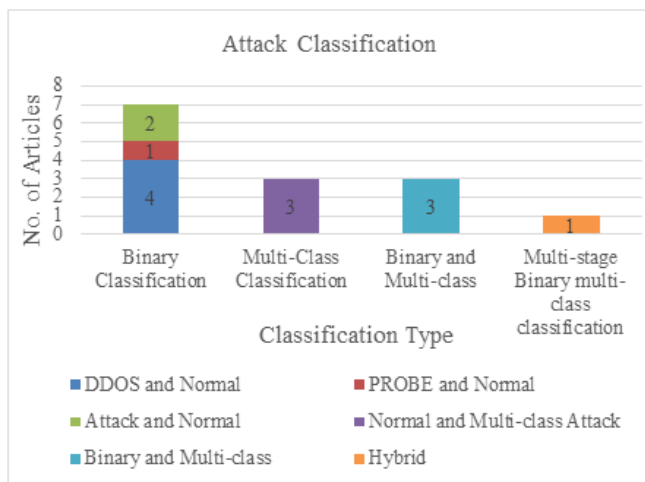


Fig. 3. Binary and multiclass attack distribution in the literature.

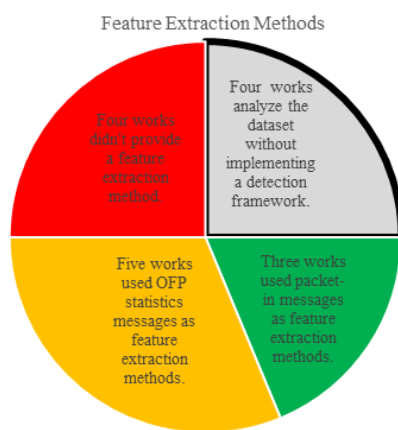


Fig. 4. Feature extraction methods in SDN used in the literature.

The existing models for intrusion detection have several drawbacks. One of them is overfitting, when the model's accuracy is very high during model training, but its performance decreases significantly when tested. This usually happens when we use high-dimensional features. Increasing the number of data samples is an optimal solution, but this method is expensive and constrained by the availability of network data, especially for network traffic. Among the reviewed works, 9 of 16 reduced the feature dimensions using a feature selection algorithm to reduce the effect of overfitting. Removing socket information features, such as timestamp, source IP, destination IP, flow ID, etc. to avoid the overfitting problem as in [12] is not enough. Some other works [4, 27, 30] apply regularization techniques to prevent overfitting. Moreover, some authors used a voting algorithm that was characterized by low error and low overfitting [3]. Another problem is that class imbalance occurs when datasets result from some minority attacks, such as U2R, botnet, and web attacks. Some works did not consider solving this issue because their models performed binary classification between normal and abnormal or normal and attack traffic, namely DDoS and probes [22-24]. Since these are considered majority classes, they did not affect the performance. The effect of class

imbalance becomes unavoidable when performing multi-attack classification. Some studies simply neglect minority class attacks, while others use oversampling and undersampling techniques to resolve this issue [10]. Another solution is to use multi-class hierarchical binary classification [10, 15].

Most researchers only focus on providing detection, neglecting mitigation. In the reviewed literature investigated, only four studies provide mitigation. It is difficult and costly to provide a successful protection mechanism, which is why a mitigation mechanism is a favorable option. However, the current work implementing mitigation uses a blocking port mechanism.

V. OPEN ISSUES, CHALLENGES, AND FUTURE RESEARCH DIRECTIONS

In the previous sections, existing solutions for deploying AI-based IDS in an SDN environment were reviewed using the recent novel SDN based dataset. However, in the literature, several issues and challenges were found. In this section, the constraints of the existing methods are identified and potential research concepts to address these limitations are proposed. Additionally, certain future research concerns and challenges are highlighted.

A. Sampling Time Interval for IDS Traffic Monitoring

The IDS must be active all the time to recognize malicious traffic; therefore, IDSs are required to continuously check the traffic and extract features to feed to the detection module. A widespread method for extracting features for AI-based intrusion detection in SDN involves periodically sending OpenFlow statistics messages from the data plane to the controller and extracting the necessary features from the response message. The definition of the time interval to collect flow entries is of great importance. Long time intervals create a delay in the detection of attacks and a reduction in the time available for possible mitigation, allowing an excellent opportunity for the attacker to damage the network. Moreover, it would impose an enormous load on the controller and switch due to the necessity of processing a large number of flows. Conversely, if it is extremely short, the controller will repeatedly engage the detection module, resulting in higher computational costs, increased resource usage by the controller, and increased communication between the controller and switch, leading to bandwidth consumption. The impact of this issue might be insignificant for a network of limited size. However, with larger networks, the problem becomes more severe. There is a need for a mechanism that selects an optimal time interval or a method of invoking the detection mechanism when needed.

B. Accuracy in Real-Time Detection

Previous studies on deploying intrusion IDS in SDN have primarily concentrated on attaining high accuracy through the utilization of novel ML and DL algorithms, or by employing feature selection methods during offline training. There was no clear implementation of real-time detection, and the accuracy of the literature did not reflect reality, as it did not run in real-time. Therefore, detecting attacks in real-time needs to be considered by the research community.

TABLE II. VARIOUS RESEARCH WORKS USED INSDN DATASET

Reference	[25]	[26]	[10]	[27]	[28]	[29]	[18]	[11]
Year	2021	2022	2022	2022	2023	2021	2023	2022
Controller used	N/A	ONOS	N/A	Ryu	Ryu	N/A	N/A	N/A
Simulation environment	N/A	Mininet	Estinet	Mininet	Mininet	N/A	N/A	N/A
IDS	x	x	x	✓	✓	✓	✓	✓
Dataset	InSDN	InSDN	InSDN	InSDN	InSDN	InSDN	InSDN	InSDN and self-generated
No. of features selected	10	8	20	12	78	48	7	7
Binary / Multi-class	---	Binary	Both	Binary	Binary	Multi-class	Multi-class	Binary
Attack types	---	Probe	All	DDOS	DDOS	All	All	DDOS
Classifier used	Only used feature selection algorithms	LightGBM	Multi-stage Binary Classification	CNN-ELM	DCNN	LSTM, RNN, GRU	DT, RF, Adaboost	Many Classifier. The best were BT and KNN
Feature selection method	SelectKBest	GWO	IG	Manually selected by author + packet-in	Manually selected by author + packet-in	Followed [30]	SelectKBest	N/A
Feature extraction method	Not required	N/A	N/A	OFP_stats message + manually create new feature + packet-in	OFP_stats message + packet-in	N/A	N/A	Count packet-in for self-generated Dataset. Not given for InSDN
Topology	X	✓	✓	X	X	X	X	✓
Framework	X	X	✓	✓	✓	X	X	✓
Real-time detection (deployment)	X	X	✓	X	X	X	X	✓
Accuracy	---	99.80%	96%-99%	99.86%	99.90%	92%	99.80%	Self-generated Dataset: 99.4%. InSDN: 97% - 99%
Evaluation metrics	---	Accuracy, Recall, Precision, F1-score	F1-score	Accuracy, Recall, Precision, F1-score, Test time, Confusion matrix	Accuracy, Recall, Precision, F1-score, Loss rate, Confusion matrix	Accuracy, Recall, Precision, F1-score, Training time, AUC	Accuracy, Recall, Precision, F1-score, Execution time	Accuracy, Recall, Precision, F1-score, Training time, Decision time, CPU utilization
Provide mitigation	X	X	X	✓	✓	X	X	X
Overfitting consideration	Not needed since few features were selected.				N/A	N/A	Not needed since few features were selected.	
Imbalance consideration	N/A	No need to consider because the binary classification between normal and probes	Used Multi-Stage Binary and perform SMOTE for botnet and U2R	No need to consider because the binary classification between normal and DDOS	No need to consider because the binary classification between normal and DDOS	N/A	N/A	No need to consider because the binary classification between normal and DDOS
Remarks	They only analyze the dataset to select relevant features. They did not consider the class imbalance between normal and other minority class attacks, such as botnets and web attacks.	Topology of [21] was considered.	Real-time implementation and verification are performed only for DDOS. Their method will lead to an increase in detection time and computation due to multi-class binary classification.	Four manually created features were not verified. There was no clear methodology for detecting DDOS through packet-in messages.	Overfitting due to the huge number of features. No clear methodology for detecting DDOS through packet-in messages. The mechanism was not verified.	They did not consider imbalanced dataset and overfitting. No real-time classification.	Only analyzed the dataset and did not provide details about IDS implementation. Did not consider an imbalanced dataset.	Did not mention the feature selection methodology and some of the selected features were irrelevant. Continuous checking of the count of packet-ins creates a load on the controller. Time-slice methodology delays decisions.

Reference	[17]	[3]		[7]		[20]		[31]	[34]	[12]	[35]
Year	2021	2021		2021		2022		2021	2023	2020	2023
Controller used	N/A	Ryu		N/A		Ryu		N/A	N/A	N/A	N/A
Simulation environment	N/A	Mininet		N/A		N/A		N/A	N/A	N/A	N/A
IDS	x	X		X		✓		✓	✓	✓	✓
Dataset	InSDN	InSDN		InSDN		InSDN		InSDN	InSDN	InSDN	InSDN
No. of features selected	N/A	48	5	48	9	48	10	48	48	78	9
Binary / Multi-class	Both	Multi-class	Binary	Both		Binary		Multi-class	Binary then multi-class	Binary	Binary
Attack types	All	All	DDOS	All		DDOS		All	All	All	All
Classifier used	XGBoost	V-NKDE		CNN-RF, CNN-SVM, CNN-KNN		LSTM		CNN-LSTM	LSTM	LSTM-autoencoder + OC-SVM	RF, XGboost
Feature selection method	N/A	Follows [30]	Manually selected	Follow [30]	PCA for 9 features selected	Follows [30]	IG and RF for 10 features selected	Follow [30]		Not used	Follows [7]
Feature extraction method	Not required	N/A	OFF_stats message	Not required		OFF_stats message		N/A	N/A	N/A	OFF_stats message
topology	X	✓		X		X		X	X	X	X
Framework	X	✓		X		✓		X	X	X	X
Real-time detection (deployment)	X	✓		X		X		X	X	X	X
accuracy	N/A	99.84%	99.10%	97% - 99%		99.95%		96.32%	99.50%	90.50%	99.69%
Evaluation metrics	AUC, confusion matrix	Accuracy, Recall, Precision, F1- score, TPR, FPR		Accuracy, Recall, Precision, F1- score, AUC, Confusion matrix		Accuracy, Recall, Precision, F1- score, Execution time, Throughput, Latency		Accuracy, Recall, Precision, F1 score, AUC, Confusion matrix	Accuracy, Detection rate, Precision, F1- score, AUC	Accuracy, Precision, Recall, F1- measure	Accuracy, Precision, Recall, F1- score, Confusion matrix
Provide mitigation	X	✓		X		✓		X	X	X	X
Overfitting consideration	N/A	Voting algorithms are characterized by low error and low overfitting.		Not needed since few features were selected.				Regularization techniques to solve the overfitting problem.		Only removed socket information.	Not needed since only a few features were selected.
Imbalance consideration	N/A	N/A		✓		No need to consider due to the binary classification between normal and DDOS		N/A	N/A	N/A	N/A
Remarks	Only analyzed the dataset and did not perform any detection. Did not mention the number of features selected and using all features without removing features and socket information such as IP, port, and MAC, lead to overfitting. They did not consider an imbalanced dataset.	They did not mention how they trained their model using five tuple features.		Only analyzed the dataset.		Requires periodically getting statistics from the switch to the controller, which creates a load on the controller.		Using a huge number of features makes the feature extraction process very difficult and the detection time very long.	Using a huge number of features makes the detection time very long, and feature extraction creates overhead in an IoT environment that is characterized by limited resource devices. No real-time implementation	The number of randomly selected samples of all attacks in one category does not provide accurate detection since the attacks have different effective attributes. Used a huge number of features, which created overfitting. Using high-dimensional features increases training and execution time, making real time implementation impossible.	Did not consider an imbalanced class. Methodology of checking OpenFlow statistics message according to limited number of packets in flow instead of fix time window create overhead in the controller.

C. Controller Resource Consumption by IDS

The controller is considered an optimal location for IDS deployment to get the benefit of SDN features as well as exploit the power of the controller [36]. Almost all works implemented their IDS in SDN controllers, but less attention was given to controller resources consumed by IDS. There must be a deep investigation of the controller when network size increases. Integrating IDS as a separate platform and linking it to a controller helps alleviate the stress on the

controller [37]. Another solution is to implement a distributed controller or dedicate a specific controller for implementing IDS to migrate excessive traffic for checking to other controllers for processing. However, those methods may increase the time delay for detecting attacks.

D. Consideration of Other Attacks

The nature of the centralized controller of SDN, which is considered the network brain, makes it an attractive target for

DDoS attacks. As a result, many studies of intrusion detection in SDN environments were concerned with DDoS only. The detection procedure needs to be flexible enough to accommodate the additional attack types that the literature has overlooked, such as probe attacks, which have a different methodology of work in SDN from traditional networks. Furthermore, there is a lack of studies considering DDoS attacks targeting OpenFlow switches. Due to the limited memory size of flow tables, attackers can send a burst of forged source packets. The switch will then forward those packets to the controller for decision. After the decision is made by the controller, the rules for those packets are specified and forwarded to the switch. The flow tables in the OpenFlow switch will be unable to store all the fake flow rules. Consequently, the flow table will soon fill up, and the transmission of legitimate traffic will stop.

E. Attack Early Detection

Many research works have developed different high accuracy models, including hybrid, ensemble, etc. These complex models require additional time to detect attacks [38]. Quick identification of an intrusion is crucial, as it enables the initiation of mitigation actions at an earlier stage.

F. Other Methods for Attack Mitigation

The existing solution for attack mitigation in the literature is blocking the attacker port, ignoring the impact the attack caused, and still exist in the network. Performance can be harmed by blocking assaults without taking into account the malicious flow entries that are stored in the switch flow table. When attacks start, there are some useless flow rules installed based on the attacker's behavior, which consume switch resources until they are removed. In addition to the blocking mechanism, there must be other methods used to store the attacker's behavior, which might be helpful in the future for analysis or reference. In this direction, instead of blocking the attacker port, redirecting the malicious traffic to honeypot or mirroring the flow to a deep packet inspector to further analyze the flow is a probable suitable action.

G. Socket Information and Overfitting

Using socket information, such as IP, port, MAC, etc. as direct features [10], leads to overfitting, which has an effect on model prediction. However, it is worth noting that it is a good future direction for researchers employing entropy to measure the distribution of those values for model training or creating other features, namely speed of source IP, standard deviation of flow packets, and standard deviation of flow bytes.

VI. CONCLUSION

SDN features, like flexibility, programmable networks, and dynamic management, successfully resolve the drawbacks of the former network. However, security issues arise because SDN lacks built-in security, and their architecture produces additional security concerns due to the decoupling of data and controller planes. Therefore, implementing intrusion detection in SDN has become a hot topic for the research community to consider in SDN security issues. In this survey, an emphasis was placed on the SDN architecture as a suitable platform for deploying AI-based IDSs to monitor networks and detect

threats. Since the efficiency of AI-based intrusion detection depends on the quality of the dataset, a review of InSDN, a new SDN dataset that was collected in an SDN environment, was conducted. Moreover, various research works that used this dataset for intrusion detection development were outlined while their strengths and weaknesses were highlighted. Related research challenges and issues were briefly analyzed and examined. In addition, hypotheses for solving some of those open challenges are provided. There is a strong belief that this survey will help and guide the researchers who aim to develop AI-based IDS solutions in the SDN context.

REFERENCES

- [1] L. Kou, S. Ding, T. Wu, W. Dong, and Y. Yin, "An Intrusion Detection Model for Drone Communication Network in SDN Environment," *Drones*, vol. 6, no. 11, Nov. 2022, Art. no. 342, <https://doi.org/10.3390/drones6110342>.
- [2] H. Y. I. Khalid, P. M. Ismael, and A. B. Al-Khalil, "Efficient Mechanism for Securing Software Defined Network against Arp Spoofing Attack," *The Journal of Duhok University*, vol. 22, no. 1, pp. 124–131, Nov. 2019, <https://doi.org/10.26682/sjuod.2019.22.1.14>.
- [3] O. E. Tayfour and M. N. Marsono, "Collaborative detection and mitigation of DDoS in software-defined networks," *The Journal of Supercomputing*, vol. 77, no. 11, pp. 13166–13190, Nov. 2021, <https://doi.org/10.1007/s11227-021-03782-9>.
- [4] T. A. Tang, D. McLernon, L. Mhamdi, S. A. R. Zaidi, and M. Ghogho, "Intrusion Detection in SDN-Based Networks: Deep Recurrent Neural Network Approach," in *Deep Learning Applications for Cyber Security*, M. Alazab and M. Tang, Eds. New York, NY, USA: Springer, 2019, pp. 175–195.
- [5] H. Y. Ibrahim, P. M. Ismael, A. A. Albabawat, and A. B. Al-Khalil, "A Secure Mechanism to Prevent ARP Spoofing and ARP Broadcasting in SDN," in *International Conference on Computer Science and Software Engineering*, Duhok, Iraq, Apr. 2020, pp. 13–19, <https://doi.org/10.1109/CSASE48920.2020.9142092>.
- [6] D. Kreutz, F. M. V. Ramos, P. E. Verissimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-Defined Networking: A Comprehensive Survey," *Proceedings of the IEEE*, vol. 103, no. 1, pp. 14–76, Jan. 2015, <https://doi.org/10.1109/JPROC.2014.2371999>.
- [7] M. S. ElSayed, N.-A. Le-Khac, M. A. Albahar, and A. Jurcut, "A novel hybrid model for intrusion detection systems in SDNs based on CNN and a new regularization technique," *Journal of Network and Computer Applications*, vol. 191, Oct. 2021, Art. no. 103160, <https://doi.org/10.1016/j.jnca.2021.103160>.
- [8] G. Logeswari, S. Bose, and T. Anitha, "An Intrusion Detection System for SDN Using Machine Learning," *Intelligent Automation & Soft Computing*, vol. 35, no. 1, pp. 867–880, 2023, <https://doi.org/10.32604/iasc.2023.026769>.
- [9] T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi, and M. Ghogho, "Deep learning approach for Network Intrusion Detection in Software Defined Networking," in *International Conference on Wireless Networks and Mobile Communications*, Fez, Morocco, Oct. 2016, pp. 258–263, <https://doi.org/10.1109/WINCOM.2016.7777224>.
- [10] H.-M. Chuang, F. Liu, and C.-H. Tsai, "Early Detection of Abnormal Attacks in Software-Defined Networking Using Machine Learning Approaches," *Symmetry*, vol. 14, no. 6, Jun. 2022, Art. no. 1178, <https://doi.org/10.3390/sym14061178>.
- [11] S. Wang *et al.*, "Detecting flooding DDoS attacks in software defined networks using supervised learning techniques," *Engineering Science and Technology, an International Journal*, vol. 35, Nov. 2022, Art. no. 101176, <https://doi.org/10.1016/j.jestch.2022.101176>.
- [12] M. Said Elsayed, N.-A. Le-Khac, S. Dev, and A. D. Jurcut, "Network Anomaly Detection Using LSTM Based Autoencoder," in *16th ACM Symposium on QoS and Security for Wireless and Mobile Networks*, Alicante, Spain, Nov. 2020, pp. 37–45, <https://doi.org/10.1145/3416013.3426457>.

- [13] N. A. Alsharif, S. Mishra, and M. Alshehri, "IDS in IoT using Machine Learning and Blockchain," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11197–11203, Aug. 2023, <https://doi.org/10.48084/etasr.5992>.
- [14] A. D. Althobiti, R. M. Almohayawi, and O. O. Bamsag, "Machine Learning approach to Secure Software Defined Network: Machine Learning and Artificial Intelligence," in *4th International Conference on Future Networks and Distributed Systems*, Saint Petersburg, Russian, Nov. 2020, pp. 1–8, <https://doi.org/10.1145/3440749.3442597>.
- [15] M. Latah and L. Toker, "An efficient flow-based multi-level hybrid intrusion detection system for software-defined networks," *CCF Transactions on Networking*, vol. 3, no. 3, pp. 261–271, Dec. 2020, <https://doi.org/10.1007/s42045-020-00040-z>.
- [16] E. M. Zeleke, H. M. Melaku, and F. G. Mengistu, "Efficient Intrusion Detection System for SDN Orchestrated Internet of Things," *Journal of Computer Networks and Communications*, vol. 2021, Nov. 2021, Art. no. e5593214, <https://doi.org/10.1155/2021/5593214>.
- [17] Q.-V. Dang, "Intrusion Detection in Software-Defined Networks," in *Future Data and Security Engineering*, Nov. 2021, pp. 356–371, https://doi.org/10.1007/978-3-030-91387-8_23.
- [18] A. Mzibri, R. Benaini, and M. B. Mamoun, "Case Study on the Performance of ML-Based Network Intrusion Detection Systems in SDN," in *International Conference on Networked Systems*, Benguerir, Morocco, Dec. 2023, pp. 90–95, https://doi.org/10.1007/978-3-031-37765-5_7.
- [19] S. Singh and S. Banerjee, "Machine Learning Mechanisms for Network Anomaly Detection System: A Review," in *International Conference on Communication and Signal Processing*, Chennai, India, Jul. 2020, pp. 976–980, <https://doi.org/10.1109/ICCSP48568.2020.9182197>.
- [20] M. S. E. Sayed, N.-A. Le-Khac, M. A. Azer, and A. D. Jurcut, "A Flow-Based Anomaly Detection Approach With Feature Selection Method Against DDoS Attacks in SDNs," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 4, pp. 1862–1880, Sep. 2022, <https://doi.org/10.1109/TCCN.2022.3186331>.
- [21] M. S. Elsayed, N.-A. Le-Khac, and A. D. Jurcut, "InSDN: A Novel SDN Intrusion Dataset," *IEEE Access*, vol. 8, pp. 165263–165284, 2020, <https://doi.org/10.1109/ACCESS.2020.3022633>.
- [22] M. Ring, S. Wunderlich, D. Scheuring, D. Landes, and A. Hotho, "A survey of network-based intrusion detection data sets," *Computers & Security*, vol. 86, pp. 147–167, Sep. 2019, <https://doi.org/10.1016/j.cose.2019.06.005>.
- [23] "Index of /datasets/SDN," <https://aseados.ucd.ie/datasets/SDN/>.
- [24] M. H. H. Khairi, S. H. S. Ariffin, N. M. A. Latiff, A. S. Abdullah, and M. K. Hassan, "A Review of Anomaly Detection Techniques and Distributed Denial of Service (DDoS) on Software Defined Network (SDN)," *Engineering, Technology & Applied Science Research*, vol. 8, no. 2, pp. 2724–2730, Apr. 2018, <https://doi.org/10.48084/etasr.1840>.
- [25] N. Abbas, Y. Nasser, M. Shehab, and S. Sharafeddine, "Attack-Specific Feature Selection for Anomaly Detection in Software-Defined Networks," in *3rd IEEE Middle East and North Africa COMMUNICATIONS Conference*, Agadir, Morocco, Dec. 2021, pp. 142–146, <https://doi.org/10.1109/MENACOMM50742.2021.9678279>.
- [26] A. Almazyad, L. Halman, and A. Alsaheed, "Probe Attack Detection Using an Improved Intrusion Detection System," *Computers, Materials & Continua*, vol. 74, no. 3, pp. 4769–4784, 2023, <https://doi.org/10.32604/cmc.2023.033382>.
- [27] J. Wang and L. Wang, "SDN-Defend: A Lightweight Online Attack Detection and Mitigation System for DDoS Attacks in SDN," *Sensors*, vol. 22, no. 21, Jan. 2022, Art. no. 8287, <https://doi.org/10.3390/s22218287>.
- [28] V. Hnamte and J. Hussain, "An efficient DDoS attack detection mechanism in SDN environment," *International Journal of Information Technology*, vol. 15, no. 5, pp. 2623–2636, Jun. 2023, <https://doi.org/10.1007/s41870-023-01332-5>.
- [29] A. S. Alshra'a, A. Farhat, and J. Seitz, "Deep Learning Algorithms for Detecting Denial of Service Attacks in Software-Defined Networks," *Procedia Computer Science*, vol. 191, pp. 254–263, Jan. 2021, <https://doi.org/10.1016/j.procs.2021.07.032>.
- [30] P. Krishnan, S. Duttgupta, and K. Achuthan, "VARMAN: Multi-plane security framework for software defined networks," *Computer Communications*, vol. 148, pp. 215–239, Dec. 2019, <https://doi.org/10.1016/j.comcom.2019.09.014>.
- [31] M. Abdallah, N. An Le Khac, H. Jahromi, and A. Delia Jurcut, "A Hybrid CNN-LSTM Based Approach for Anomaly Detection Systems in SDNs," in *16th International Conference on Availability, Reliability and Security*, Vienna, Austria, Aug. 2021, pp. 1–7, <https://doi.org/10.1145/3465481.3469190>.
- [32] O. M. Ahmed, L. M. Haji, A. M. Ahmed, and N. M. Salih, "Bitcoin Price Prediction using the Hybrid Convolutional Recurrent Model Architecture," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11735–11738, Oct. 2023, <https://doi.org/10.48084/etasr.6223>.
- [33] R. Alsulami, B. Alqarni, R. Alshomrani, F. Mashat, and T. Gazdar, "IoT Protocol-Enabled IDS based on Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12373–12380, Dec. 2023, <https://doi.org/10.48084/etasr.6421>.
- [34] R. A. Elsayed, R. A. Hamada, M. I. Abdalla, and S. A. Elsaid, "Securing IoT and SDN systems using deep-learning based automatic intrusion detection," *Ain Shams Engineering Journal*, vol. 14, no. 10, Oct. 2023, Art. no. 102211, <https://doi.org/10.1016/j.asej.2023.102211>.
- [35] M. S. Towhid and N. Shahriar, "Early Detection of Intrusion in SDN," in *IEEE/IFIP Network Operations and Management Symposium*, Miami, FL, USA, Dec. 2023, pp. 1–6, <https://doi.org/10.1109/NOMS56928.2023.10154272>.
- [36] A. Abubakar and B. Pranggono, "Machine learning based intrusion detection system for software defined networks," in *Seventh International Conference on Emerging Security Technologies*, Canterbury, UK, Sep. 2017, pp. 138–143, <https://doi.org/10.1109/EST.2017.8090413>.
- [37] S. Kumar et al., "DDoS Detection in SDN using Machine Learning Techniques," *Computers, Materials & Continua*, vol. 71, no. 1, pp. 771–789, 2022, <https://doi.org/10.32604/cmc.2022.021669>.
- [38] A. O. Alzahrani and M. J. F. Alenazi, "Designing a Network Intrusion Detection System Based on Machine Learning for Software Defined Networks," *Future Internet*, vol. 13, no. 5, May 2021, Art. no. 111, <https://doi.org/10.3390/fi13050111>.

A System Dynamics Approach to Feedback Processes in Project Scheduling

Babatunde Omoniyi Odedairo

Department of Industrial Engineering, College of Engineering in Al-kharj, Prince Sattam Abdulaziz University, Saudi Arabia | Department of Industrial and Production Engineering, University of Ibadan, Nigeria

b.odedairo@psau.edu.sa (corresponding author)

Ali Alarjani

Department of Industrial Engineering, College of Engineering in Al-kharj, Prince Sattam Abdulaziz University, Saudi Arabia

a.alarjani@psau.edu.sa

Received: 24 November 2023 | Revised: 23 December 2023 and 3 January 2024 | Accepted: 3 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6666>

ABSTRACT

Projects, as catalysts for proactive transformation, offer a temporary and adaptable framework that effectively handles complexities (or uncertainties) within a competitive corporate landscape. Hence, the use of an effective project management framework, such as Dynamic Project Scheduling (DPS), is a method to handle intricacies in order to accomplish organizational objectives. DPS refers to a triangle interaction involving baseline scheduling, schedule risk analysis, and project control while supporting schedule adjustment in response to changes and uncertainties. However, there is a lack of information regarding studies that have investigated the feedback mechanisms among DPS components. This study was designed to examine the counterintuitive relationships between these components using system dynamics. The quantities within the DPS system were identified and defined. A causal loop diagram was used to illustrate the interactions among these quantities. Subsequently, a Stock and Flow Diagram (SFD) was created to identify the inputs, states, and flow mechanisms within the DPS. Using the SFD, a system dynamics expression was generated which was then employed to compute the rate of change of the Budgeted Cost of Work Remaining (BCWR) for two projects at different time intervals. The results properly indicated the period of idleness during project execution. The use of BCWR rather than schedule variance provides a more effective visual representation for evaluating performance and tracking progress. The BCWR and planned value exhibit contrasting trends, highlighting the importance of earned value quantities in project control. The use of system dynamics in project management can enhance the planning and scheduling phase, allow project managers to monitor pertinent performance measures, and optimize project outcomes through informed decisions.

Keywords-*project scheduling; system dynamics; dynamic project scheduling; project performance measures; budgeted cost of work remaining*

I. INTRODUCTION

Nowadays, technological advancements, innovative business strategies, collaborations, and worldwide integration have contributed to the increasing complexity of undertaken projects. Apart from traditional projects like construction and engineering, significant interest in a management-by-projects strategy for business operations is due to the advent of new business endeavors (or applications). These new endeavors, such as the development of a new product, research and development, the acquisition of new certifications, operational process enhancement, novel information technologies deployment, system improvement, etc. [1-4] are called modern projects. The criteria for determining the success of a project may vary across different stakeholders [5]. A project is deemed

successful if it is finished within the designated timeframe and financial constraints, while meeting the client's expectations in terms of stated standards [4, 6]. The issues that result in project failure frequently arise from the intricacy, unpredictability, and ever-changing characteristics of project execution. Some projects may incur cost overruns between 100 and 200% and can be delayed beyond the anticipated market delivery date [7]. According to prior research, 70% of projects were unsuccessful because they lacked a clearly defined objective, whereas 55% of project managers identified budget overruns as a contributing factor to failure [8].

Complexity is considered a characteristic of a program, a project, or its environment that is difficult to manage due to human or system behavior or ambiguity [9]. On system

behavior, a project can be viewed as a system or sub-system depending on the project structure. Invariably, the interactions between project components may result in complex situations. These interrelationships include dependencies between the project and the environment, interactions among project activities, and the resulting system dynamics. The effect of dynamism is the counterintuitive relationships and behavior among components in a project system [10]. Therefore, to navigate through the expected complex situation, a good understanding of the components and their resulting connections will enhance project performance. Also, to avoid negative consequences, a rigorous qualitative description of the processes, boundaries, and strategies of the project management process can be achieved using system dynamics.

System Dynamics (SD) has been applied in various fields including social sciences, economics, and strategic management [11-15]. SD is a widely employed methodology for modeling and simulating complex systems, which helps in making informed decisions and gaining a deeper comprehension of these systems [14, 16-18]. Project management has greatly benefited from the application of SD [19, 20]. Authors in [21] examined the interrelationships and linkages among the initial stages of construction projects utilising Stock and Flow Diagrams (SFDs) and Causal Loop Diagrams (CLDs). Authors in [22] determined that SD has the capacity to detect and establish links between risk elements, as evidenced by their study on the influence of risk analysis in construction projects. However, the methods of structural decomposition and Delphi risk analysis did not possess this potential. Authors in [23] investigated the root cause of rework in construction projects executed in southwest Nigeria, and proposed strategic interventions to improve project performance. However, the study did not consider modelling approach, which provides a deeper comprehension of the challenges related to project activity rework. Authors in [24] created a model that integrated SD and earned value management to track changes in the distribution of staff histograms and problematic project behaviour in real-world scenarios. Similarly, in [18] it was concluded that a SD model possesses the capability to comprehensively assess the impact of risk factors in building projects.

Scheduling is a vital aspect of project management that involves assigning resources to tasks in order to ensure efficient processing of the project [25-26]. Scheduling is an optimization problem with the primary goal of optimizing one or more criteria. The significance of scheduling lies in its ability to effectively handle interdependencies, foster team synergy, provide a clear project implementation plan, and monitor progress toward achieving project goals [25, 27-29]. However, researchers have acknowledged the constraints of conventional scheduling techniques and, as a result, have implemented dynamic scheduling [29-30]. Dynamic Project Scheduling (DPS) takes into consideration the effects of several elements, such as modifications in project requirements, availability of resources, and external influences on the project schedule [29-31]. Furthermore, project data may be consistently reviewed, and if needed, schedule adjustments can be implemented to guarantee the accomplishment of project objectives [32]. The DPS refers to a triangle interaction involving Baseline

Scheduling (BS), Schedule Risk Analysis (SRA), and Project Control (PC) [33]. The interactions among these components may have complex impacts, despite their apparent simplicity [10, 26, 33-34]. Therefore, a project manager must have a good understanding of the interconnectivity and interrelationship among the DPS components and their associated metrics. Numerous businesses and project managers have adopted the utilization of performance metrics to monitor project advancement for efficient decision-making.

Performance measurements encompass metrics that facilitate decision-making and evaluate the efficient utilization of resources in accomplishing strategic objectives [35]. Nevertheless, relying on a single metric across several scenarios may prove to be inefficient. For instance, while a metric can reveal the impact of the critical path on the total time, it may offer limited insights into its influence on cost, quality, and other factors crucial for project success. Therefore, in the DPS system, it is possible to explore holistically the counterintuitive relationships among the metrics. However, information is sparse on studies that explored the feedback mechanisms among DPS components. Hence, in this research, the counterintuitive relationships among the components of DPS will be investigated using SD.

II. METHODOLOGY

In this section, DPS components were defined, and a CLD was used to aggregate the interrelationships between the elements. Thereafter, the mathematical relationship between these elements was developed using the SFD. The applicability and effectiveness of the mathematical equations were validated with two project datasets. The datasets were obtained from the operation research and scheduling group dynamic scheduling library [36].

A. Identification of Quantities in DPS

Table I presents the quantities in the DPS system. A full definition of these quantities is beyond the scope of the current research.

TABLE I. QUANTITIES IN DPS

	BS	SRA	PC
1	Activity id	Criticality index	Planned value
2	Activity name	Cruciality index	Earned value
3	Actual duration	Significance index	Actual cost
4	Precedence relations	Schedule sensitivity index	Actual start
5	Successors		Schedule variance
6	Baseline start		Cost variance
7	Baseline finish		Budget at completion
8	Activity cost		Tine at completion
9	Total project cost		Performance factor
10			Percentage of project completed
11			Schedule performance index
12			Cost performance index

The quantities presented in Table I were further categorized into four types, namely rates, state, inputs, and auxiliaries, as shown in Table II.

TABLE II. FLOW DIAGRAM NOTATIONS

	Quantity type	Quantity	Notation
1	Rates	Planned value rate	PV_{rate}
		Earned value rate	EV_{rate}
2	State	Budgeted cost of work remaining	BCWR
		Planned value	PV
3	Inputs	Performance factor	PF
		Actual cost	AC
		Percentage of project completed	%PC
		Actual duration	AD
		Planned duration	PD
		Budget at completion	BAC
4	Auxiliaries	Earned value	EV
		Schedule variance	SV
		Cost variance	CV
		Schedule performance index	SPI
		Cost performance index	CPI
		Estimated time at completion	EAC_{time}
		Estimated cost at completion	EAC_{cost}

B. Causal Loop Diagram (CLD) for DPS

The relationship between the quantities defined in Table I is described by the CLD in Figure 1. In a CLD, the relationships between the factors that influence the cost and duration of a project activity are described.

C. Flow Diagram for DPS

In Table II, the BCWR is defined as a state. The value of BCWR can be determined from the difference in the rate of inflow and outflow represented by Planned Value rate (PV_{rate}) and Earned Value rate (EV_{rate}), respectively. The PV_{rate} is estimated from the value of budget at completion and planned duration. The Schedule Variance (SV) is the difference between the earned value and the planned value at time t expressed in monetary value. Similarly, the EV_{rate} can be calculated using the budgeted cost of work performed at time t. The budgeted cost of work performed, also known as the Earned Value (EV) is determined using the values of Budget At Completion (BAC) and percentage project completion. The SFD for DPS is presented in Figure 2.

D. Formulation of System Dynamics Equations for DPS

The SD model for DPS was developed based on the following assumptions.

1. At the start of the tracking process, the BCWR is equal to the total budgeted cost of work scheduled.
2. The planning phase precedes project tracking and monitoring.
3. The planned duration and total budgeted cost of the project are constant.

4. Project activity rework is not permitted.
5. Resource cost is assumed to be a part of the budgeted cost for each activity.
6. The baseline scheduling and schedule risk analysis information is assumed to be available and accurate.

E. Rate Equations

Equation (1) is the Rate Of Change (ROC) or the budgeted cost of work remaining.

$$ROC = \text{inflow} - \text{outflow} = \frac{BCWR}{dt} = PV_{rate} - EV_{rate} \quad (1)$$

Equation (2) gives the PV_{rate} , from assumption 3, with BAC and PD being constant:

$$PV_{rate} = \frac{BAC}{PD} \quad (2)$$

Equation (3) gives the EV_{rate} :

$$EV_{rate} = \frac{EV}{AD} \quad (3)$$

Equation (4) can be derived by inserting (2) and (3) into (1):

$$ROC = K - \frac{EV}{AD} \quad (4)$$

By differentiating (4), change in ROC can be expressed as a function of time (t) as presented in (5):

$$\Delta ROC(t) = -\Delta EV(t) \quad (5)$$

F. Auxilliary Equations

The auxiliary equations are stated from (6)-(10).

Earned value:

$$EV = \%PC * BAC \quad (6)$$

Schedule variance:

$$SV = EV - PV \quad (7)$$

Cost variance:

$$CV = EV - AC \quad (8)$$

Estimated time at completion:

$$EAC_{time} = PD - \frac{SV}{PV_{rate}} \quad (9)$$

where $\frac{SV}{PV_{rate}}$ is also known as the time variance.

Estimated cost at completion:

$$EAC_{cost} = AC + \frac{BAC - EV}{PF} \quad (10)$$

G. State Equations

From the SFD presented in Figure 2, the BCWR is a stock that is subject to change over time. Therefore, the value of BCWR can be estimated using Euler solution method [10]. The state variable (S) can be replaced by BCWR at different time intervals. If i represents intervals of time, then:

$$S(t_{i+1}) = S(t_i) + \Delta ROC(t_i) \quad (11)$$

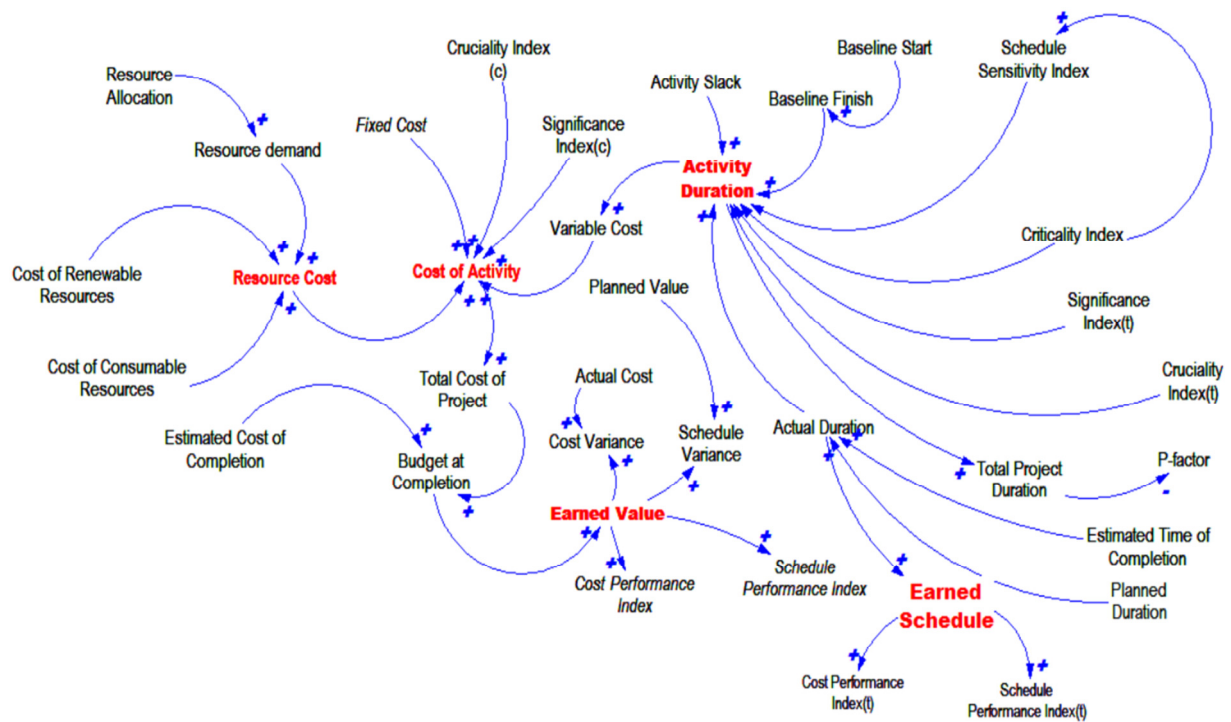


Fig. 1. CLD showing the relationship between the quantities of DPS.

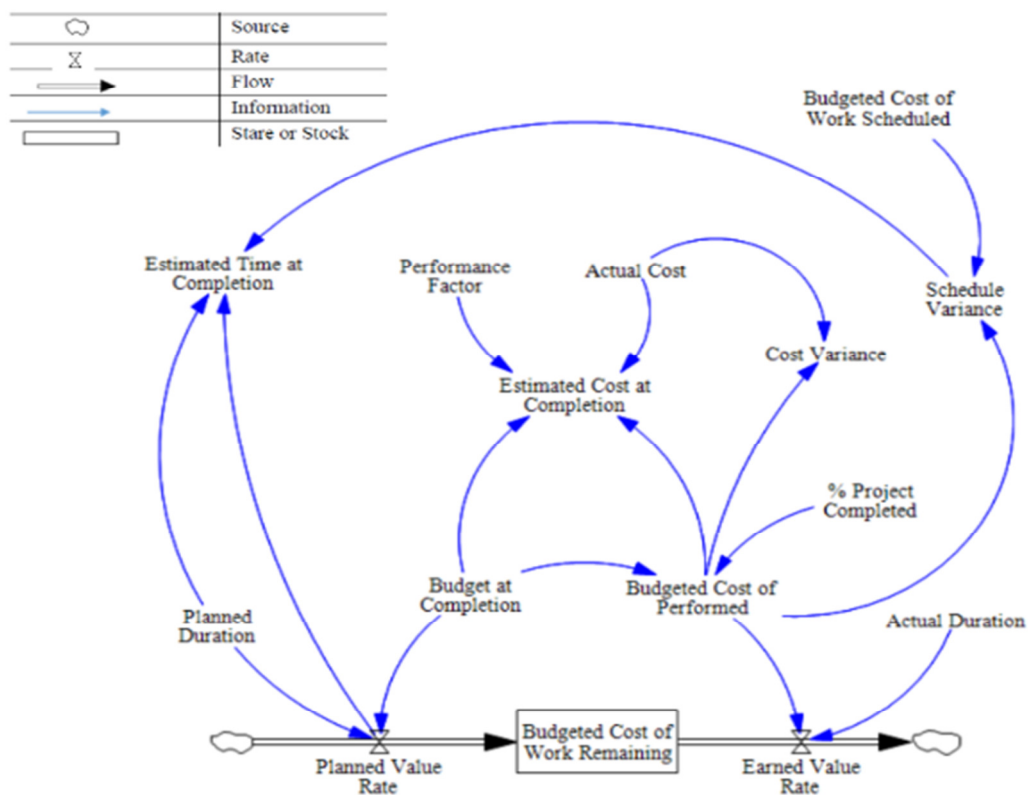


Fig. 2. The SFD for DPS

The state variable in (11) is replaced with BCWR as shown in (12):

$$BCWR(t_{i+1}) = BCWR(t_i) + \Delta ROC(t_i) \quad (12)$$

Equation (5) can be inserted in (12). Thus, we get:

$$BCWR(t_{i+1}) = BCWR(t_i) - \Delta EV(t) \quad (13)$$

Equation (13) can be used to simulate the BCWR.

H. Model Application

From the dynamic scheduling library (DSLIB), two datasets regarding projects, namely (i) Christmas Market Project (CMP), and (ii) Tournament Infrastructure Project (TIP) were obtained to test the validity and applicability of the developed model. These projects were selected based on the detailed tracking information available. From the datasets, relevant project metrics were extracted such as planned duration, budget at completion, actual duration, percentage of project completed, planned value, actual cost, cost variance, SV, schedule performance index, and cost performance index.

III. RESULTS AND DISCUSSION

To verify the applicability of the model, several additional metrics were estimated to enhance the analysis. The estimated metrics include: (i) planned value rate, (ii) earned value rate, (iii) estimated cost of completion, and (iv) estimated time of completion. The BCWR at each project tracking point is the output of the model.

A. Christmas Market Project (CMP)

The Actual Duration (AD) and the simulated budgeted cost of work for CMP are 185 days and €58900, respectively. There is a five-day interval between tracking updates. The tracking pattern allows for a comprehensive understanding of the project's status and aids in the evaluation of the amount of work remaining in terms of monetary value using (13).

1) Relationship between SV and ROC

The relationship between the SV and the ROC of the budgeted cost of work shown in Figure 3 is essential for performance evaluation and project progress tracking.

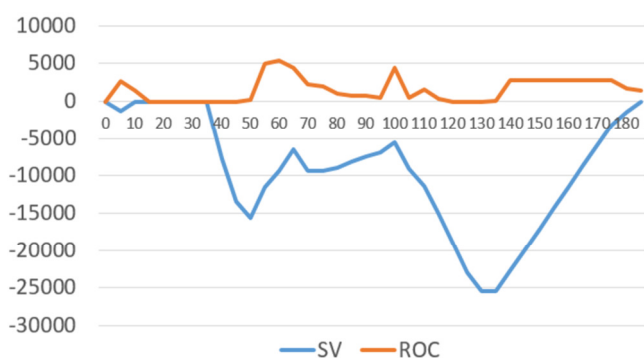


Fig. 3. Relationship between SV and ROC for CMP.

The SV considers the difference between the planned and actual progress, but it ignores periods of project stagnation. On the other hand, ROC is a helpful metric to determine whether

project work is finished. Project modifications are clearly shown with ROC, giving a clearer picture of whether work is completed or whether there are delays and interruptions. A period of zero activity can be seen in Table III from AD = 15 to AD = 45 and from AD = 120 to AD = 130. From a tracking perspective, Figure 3 illustrates the dynamic nature of ROC, there was no project activity between Project Tracking ID (PTID) 10 to 35, and PTID 120 to 135 as indicated by the ROC.

2) Relationship between BCWR and PV

An important aspect of good budget management is the alignment with planned values, as demonstrated by the relationship between BCWR and PV which have opposing trajectories and intersect at project tracking ID 18 as shown in Figure 4. This is equivalent to AD of 90 days. The intersection between BCWR and PV data indicated the point at which the planned value of the remaining work aligned with the budgeted cost. With the opposite trajectories and eventual convergence between budgeted costs and planned values, ongoing monitoring and analysis are necessary to detect and address potential budgetary concerns. By closely tracking these metrics, project managers can ensure financial control and make informed decisions to optimize project outcomes.

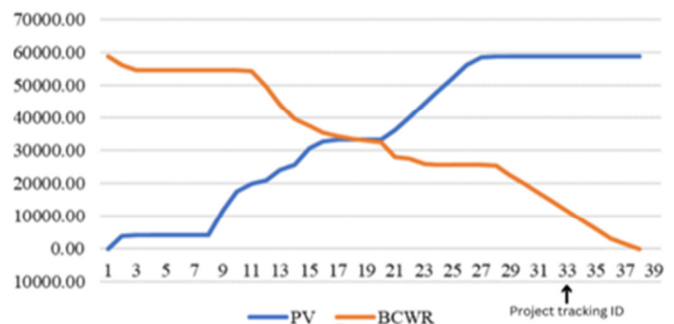


Fig. 4. Relationship between BCWR and PV for CMP.

B. Tournament Infrastructure Project (TIP)

The planned duration of the project is 27 days while the actual duration is 23 days. The budget at completion was estimated to be € 126,955.30 while the actual cost at the end of the project summed up as € 124,502.30.

1) Relationship between SV and ROC

In Figure 5, the maximum value for ROC was obtained at AD = 20. The rate of change between AD=5 and AD=20 can be compared to a batch-tub scenario. This implies that between AD=5 and AD=9, there was a decline in the BCWR, which remained unchanged between AD=9 and AD=17, and increased between AD=17 and AD= 20. However, the SV remains unchanged between AD=9 and AD=13. In reality, an SV with zero value denotes that the project is on time (i.e. the earned value equals the planned value). This explains why the value of BCWR was zero between AD=9 and AD=13. Also, this could be explained from a steady natural rhythm of work maintained among team members [37].

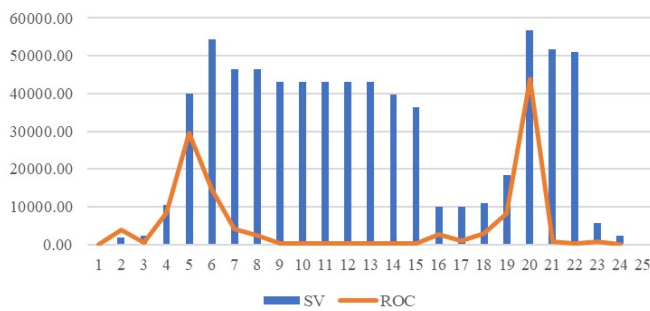


Fig. 5. Relationship between SV and ROC for TIP.

2) Relationship between BCWR and PV

In Figure 6, BCWR and PV follow opposite trajectories. Hence the need to monitor this relationship throughout the project lifecycle. Any significant deviation or consistent disparities between these metrics could signify potential budgetary issues or challenges in meeting planned targets.

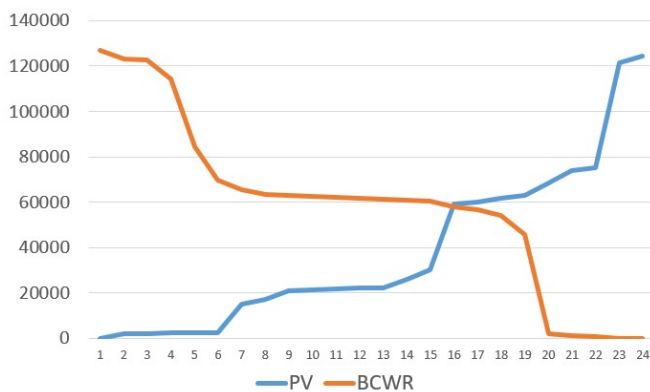


Fig. 6. Relationship between BCWR and PV for TIP.

C. Practical Implications of SD in DPS

Within the project performance framework, baseline scheduling, schedule risk analysis, and project control are metrics available for monitoring the progress of a project. In this study, from the results obtained using the developed SD model, the assumed counterintuitive relationships among DPS components and their implications for project management practices were confirmed. Also, the herculean computational process associated with DPS was reduced. In addition, the project tracking system provided real-time data on project scheduling, activity sensitivity, and actual time/cost performance, facilitating a more comprehensive understanding of the project and enabling informed decision-making. Nevertheless, it is implicitly assumed that all other factors influencing the effective completion of a project are under control.

D. Discussion

Scheduling project activities is a complex task and a key component of operational-level decision-making. While dynamic project scheduling establishes an inherent relationship among its components, a mathematical framework that aligns the interdependencies between the components is sparse. To

achieve this, a quantitative structure was created utilizing SD to model the flow of inputs and outputs between different components of DPS. The project tracking opportunity provided by this method allows for the identification of actions that may not be deemed significant in the baseline schedule, but are nonetheless essential for the successful completion of the project. For example, activities which may not be critical (from baseline scheduling), could be crucial to the completion of the project as identified from the BCWR and SV indices. Therefore, our research differs from past studies by examining all three components, whereas earlier research only focused on the use of system dynamics on individual elements of DPS.

IV. CONCLUSION AND FUTURE WORK

In this research work, system dynamics was used to explore the intricate interconnections among the various components of dynamic project scheduling. Through the development of a causal loop diagram, the underlying relationships between these quantities were revealed. Using a stock and flow model, a mathematical expression to calculate the Budgeted Cost of Work Remaining (BCWR) at each tracking period was derived. As highlighted in the literature, BCWR is relevant in the context of forecasting time and accuracy. The BCWR is a valuable indicator to monitor the completion of project activity and it serves as a tool to identify idle period. The modeling technique demonstrated how baseline scheduling and project control impacted the scope. However, quantities associated with schedule risk analysis focused on project activities rather than project scope.

This study examined the connections between the components of dynamic project scheduling at the macro level, with less focus on the relationships between activities. In future works, the system dynamics equations can be extended to capture activity level interactions, and the risk factors associated with project activities through schedule risk analysis. Furthermore, it is possible to model various scenarios of project activity rework and analyze how they impact DPS quantities and project outcomes.

REFERENCES

- [1] A. Jaafari, "Project Management in the Age of Complexity and Change," *Project Management Journal*, vol. 34, no. 4, pp. 47–57, Dec. 2003, <https://doi.org/10.1177/875697280303400407>.
- [2] G. Winch, M.-C. Meunier, J. Head, and K. Russ, "Projects as the content and process of change: The case of the health and safety laboratory," *International Journal of Project Management*, vol. 30, no. 2, pp. 141–152, Feb. 2012, <https://doi.org/10.1016/j.ijproman.2011.06.005>.
- [3] N. G. Hall, "Project management: Recent developments and research opportunities," *Journal of Systems Science and Systems Engineering*, vol. 21, no. 2, pp. 129–143, Jun. 2012, <https://doi.org/10.1007/s11518-012-5190-5>.
- [4] B. O. Odedairo, "Development of scheduling heuristic for the resource constrained project management problem with idleness cost," Ph.D. dissertation, University of Ibadan, Ibadan, Nigeria, 2016.
- [5] P. W. G. Morris, J. K. Pinto, J. Söderlund, P. W. G. Morris, J. K. Pinto, and J. Söderlund, Eds., *The Oxford Handbook of Project Management*. in Oxford Handbooks. Oxford, UK: Oxford University Press, 2011.
- [6] A. Shtub, J. F. Bard, and S. Globerson, *Project Management: Engineering, Technology, and Implementation*. Hoboken, NJ, USA: Prentice Hall, 1994.
- [7] J. D. Sterman, *Business Dynamics: Systems Thinking and Modeling for a Complex World*. New York, NY, USA: McGraw-Hill Education, 2000.

- [8] "Project Failure Statistics: The Shocking Truth," www.projectmanagementworks.co.uk, Oct. 14, 2021. <https://www.projectmanagementworks.co.uk/project-failure-statistics/>.
- [9] PMI, *Navigating Complexity: A Practice Guide*. Newtown Square, PA, USA: Project Management Institute, 2014.
- [10] B. O. Odedairo, E. H. Alaba, and I. Edem, "A System Dynamics Model To determine the value of inventory holding cost," *Journal of Engineering Studies and Research*, vol. 26, no. 3, pp. 112–123, Jul. 2020, <https://doi.org/10.29081/jesr.v26i3.213>.
- [11] L. A. Malczynski, "Best practices for system dynamics model design and construction with powersim studio.," Sandia National Laboratories, Albuquerque, NM, USA, SAND2011-4108, Jun. 2011, <https://doi.org/10.2172/1029812>.
- [12] P. S. Szwed, *Expert Judgment in Project Management: Narrowing the Theory-Practice Gap*. Newtown Square, PA, USA: Project Management Institute, 2016.
- [13] F. Cosenz and G. Noto, "Applying System Dynamics Modelling to Strategic Management: A Literature Review," *Systems Research and Behavioral Science*, vol. 33, no. 6, pp. 703–741, 2016, <https://doi.org/10.1002/sres.2386>.
- [14] D. J. Currie, C. Smith, and P. Jagals, "The application of system dynamics modelling to environmental health decision-making and policy - a scoping review," *BMC Public Health*, vol. 18, no. 1, Mar. 2018, Art. no. 402, <https://doi.org/10.1186/s12889-018-5318-8>.
- [15] F. Redivo, "Introduction to System Dynamics Modeling," *System Dynamics Society*, Oct. 21, 2021. <https://systemdynamics.org/introduction-to-system-dynamics-modeling/>.
- [16] Y. Wang, Y. Li, and P. Guo, "Modelling Construction Project Management Based on System Dynamics," *Metallurgical and Mining Industry*, no. 9, pp. 1056–1061, 2015.
- [17] M. Kamali Saraji and A. Morovati Sharifabadi, "Application of System Dynamics in Forecasting: A Systematic Review," *International Journal of Management, Accounting and Economics*, vol. 4, no. 12, pp. 1192–1205, Dec. 2017.
- [18] O. L. Adepiton, "Risk Assessment of Nigerian Road Construction Projects: A System Dynamics Approach," Ibadan, Nigeria, 2021.
- [19] J. M. Lyneis and D. N. Ford, "System dynamics applied to project management: a survey, assessment, and directions for future research," *System Dynamics Review*, vol. 23, no. 2–3, pp. 157–189, 2007, <https://doi.org/10.1002/sdr.377>.
- [20] T. Sirovs, "System Dynamic Theoretical Framework for Construction Management: A Case of Baltic States," *Baltic Journal of Real Estate Economics and Construction Management*, vol. 10, no. 1, pp. 113–128, Jan. 2022, <https://doi.org/10.2478/bjreecm-2022-0008>.
- [21] A. M. Zainab and J. K. Cheng, "Using System Dynamics Simulation To Understand The Feedback Process In A Construction Project," *International Journal of Industrial Management*, vol. 3, 2017.
- [22] J. Wan and Y. Liu, "A System Dynamics Model for Risk Analysis during Project Construction Process," *Open Journal of Social Sciences*, vol. 2, no. 6, Jun. 2014, Art. no. 46686, <https://doi.org/10.4236/jss.2014.26052>.
- [23] O. A. Aiyetan and D. Das, "Using system dynamics modelling principles to resolve problems of rework in construction projects in Nigeria," *Journal of Construction Project Management and Innovation*, vol. 5, no. 2, pp. 1266–1295, Dec. 2015, <https://doi.org/10.10520/EJC183321>.
- [24] J. M. Nevison and K. J. Chichakly, "Latent Errors and Visible Earned Value: How the Evolutionary Model Integrates Earned Value Metrics with Project System Dynamics," *Systems*, vol. 9, no. 4, Dec. 2021, Art. no. 88, <https://doi.org/10.3390/systems9040088>.
- [25] B. O. Odedairo and V. Oladokun, "Relevance and Applicability of Multi-objective Resource Constrained Project Scheduling Problem: Review Article," *Engineering, Technology & Applied Science Research*, vol. 1, no. 6, pp. 144–150, Dec. 2011, <https://doi.org/10.48084/etasr.53>.
- [26] M. Vanhoucke, *Integrated Project Management Sourcebook*. New York, NY, USA: Springer, 2016.
- [27] S. Sohu, A. H. Abdullah, S. Nagapan, T. A. Rind, and A. A. Jhatial, "Controlling Measures for Cost Overrun Causes in Highway Projects of Sindh Province," *Engineering, Technology & Applied Science Research*, vol. 9, no. 3, pp. 4276–4280, Jun. 2019, <https://doi.org/10.48084/etasr.2749>.
- [28] W. Herroelen, E. Demeulemeester, and B. De Reyck, "A Classification Scheme for Project Scheduling," in *Project Scheduling*, vol. 14, J. Weglarz, Ed. Boston, MA, USA: Springer, 1999, pp. 1–26.
- [29] M. Padalkar and S. Gopinath, "Six decades of project management research: Thematic trends and future opportunities," *International Journal of Project Management*, vol. 34, no. 7, pp. 1305–1321, Oct. 2016, <https://doi.org/10.1016/j.ijproman.2016.06.006>.
- [30] M. Vanhoucke, "On the use of schedule risk analysis for project management," *Journal of Modern Project Management*, vol. 2, no. 3, pp. 108–117, 2015.
- [31] E. Uytewaala, "A schedule as a valid and dynamic model," presented at the PMI® Global Congress 2005—North America, Toronto, ON, Canada, 2005.
- [32] M. Vanhoucke, "An Overview of Recent Research Results and Future Research Avenues Using Simulation Studies in Project Management," *International Scholarly Research Notices*, vol. 2013, Oct. 2013, Art. no. e513549, <https://doi.org/10.1155/2013/513549>.
- [33] J. Song, A. Martens, and M. Vanhoucke, "Using Earned Value Management and Schedule Risk Analysis with resource constraints for project control," *European Journal of Operational Research*, vol. 297, no. 2, pp. 451–466, Mar. 2022, <https://doi.org/10.1016/j.ejor.2021.05.036>.
- [34] R. K. Chakraborty, H. F. Rahman, and M. J. Ryan, "Efficient priority rules for project scheduling under dynamic environments: A heuristic approach," *Computers & Industrial Engineering*, vol. 140, Feb. 2020, Art. no. 106287, <https://doi.org/10.1016/j.cie.2020.106287>.
- [35] B. O. Odedairo and N. Nwabukei, "Framework for Operational Performance Measurements in Small and Medium Scale Industries Using Discrete Event Simulation Approach," *Engineering, Technology & Applied Science Research*, vol. 8, no. 4, pp. 3103–3107, Aug. 2018, <https://doi.org/10.48084/etasr.2106>.
- [36] "Research | Operations Research & Scheduling Research Group." <https://www.projectmanagement.ugent.be/research>.
- [37] B. O. Odedairo, "Managing Spare Parts Inventory by Incorporating Holding Costs and Storage Constraints," *Journal of Engineering, Project, and Production Management*, vol. 11, no. 2, pp. 139–144, May 2021, <https://doi.org/10.2478/jepm-2021-0014>.

Deflection and Elastic Modulus Assessment of Subgrade in Flexible Pavement mixed with Waste Tire Scrap Material

Sujoy Sarkar

Department of Civil Engineering, Jadavpur University, India
sujoyrnd82@gmail.com (corresponding author)

Sumit Kumar Biswas

Department of Civil Engineering, Jadavpur University, India
thesissumit@gmail.com

Saibal Chakraborty

Department of Civil Engineering, Jnan Chandra Ghosh Polytechnic, India
sahilju06@yahoo.co.in

Received: 2 January 2024 | Revised: 19 January 2024 and 23 January 2024 | Accepted: 24 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6858>

ABSTRACT

This study aims to assess the deflection and elastic modulus (E_s) of subgrade in flexible pavements, focusing on a comparative analysis between pavements with clayey soil subgrade and subgrade modified with tire scrap. The research utilized Falling Weight Deflectometer (FWD) for measuring subgrade deflection, essential in evaluating pavement performance. The FWD applied a dynamic load to the pavement, with deflection measurements processed using the KGP-BACK software to calculate the E_s of the pavement subgrade. This approach included assessing the Lower Layer Index (LLI) and E_s of the subgrade. Findings revealed a notable reduction of 37.5% in deflection and 2.68 times increase in E_s for the tire scrap modified subgrade pavement compared to the standard clayey soil subgrade pavement. These results demonstrate significant enhancements in pavement structure, underlining the potential of recycled materials in sustainable civil engineering practices.

Keywords-deflection; elastic modulus; falling weight deflectometer; subgrade

I. INTRODUCTION

The rapid growth in global vehicle numbers has resulted in a parallel increase in waste tire and tube production. This trend is anticipated to approximately yield 2 billion scrap vehicles by 2030, posing significant environmental and waste management challenges [1]. Annually, one billion tires reach the end of their lifespan, with only about half undergoing recycling while the rest end up in landfills [2], where large volumes of scrap tires are accumulated, emphasizing the environmental risks if not managed properly [3]. Often, these tires are disposed of uncontrolledly, exacerbating the rapid depletion of waste disposal sites and leading to severe environmental issues. The use of scrap tires has been increasingly observed in various civil engineering applications. In this regard, authors in [4] investigated the cumulative effect of crumb rubber and steel fiber on the flexural toughness of concrete. The study explores the use of these materials as potential enhancements for concrete properties, particularly in terms of toughness and ductility, which are important for resisting impacts or blast

loads. The research delves into the behavior of concrete beams and slabs when combined with steel fiber and crumb rubber, offering insights into the potential of these materials in creating more resilient and sustainable construction materials. However, emerging research in geotechnical engineering provides many possible benefits for repurposing waste tires. These recycled materials boast high tensile strength, durability, toughness, and resistance to aging, making them a viable solution for environmental concerns [5].

Authors in [6] conducted a study on the mechanical properties of waste tire rubber powder for use in civil engineering. They performed physical, chemical, and direct shear tests on different rubber powder sizes, establishing empirical relationships between cohesion, friction angle, and particle size. It was found that a cubic regression model explains these relationships better than linear or quadratic models, offering new perspectives on recycling waste tire materials in construction and engineering. Authors in [7] examine the impact of varying scrap tire rubber contents (10%, 20%, 25%, 50%) on Fergoug sediment and Tizi Tuff soil. Tests

considered grain size, Atterberg limits, shear strength, compaction, and CBR. Cohesion, and CBR values were found to decrease with increasing rubber content, while compression and recompression indices increased. Notably, a 75% Tizi Tuff and 25% rubber mix showed contrasting CBR results at different water contents, highlighting the potential of scrap tire, as a reinforcement material in dredged soil, with careful consideration of its impact on compressibility.

Within pavement engineering, flexible pavements are particularly beneficial due to their ability to incrementally strengthen in response to growing traffic loads. Pavements are vital for efficient transportation of passengers, freight, and other community services. A flexible pavement, as a load-bearing structure, comprises layers of various granular materials over a soil subgrade. The durability of these pavements depends on several factors, including the strength of the subgrade soil, material quality, layer thickness, environmental conditions, and traffic characteristics. Ensuring the structural integrity and load-bearing capacity of the pavement subgrade is crucial for distributing loads effectively, mitigating strain on the pavement layers, and potentially extending the lifespan of the pavement. In alignment with IRC115:2014 [8], the Falling Weight Deflectometer (FWD) is a pivotal tool in determining the subgrade deflection and elastic modulus of pavements. This study focuses on comparing the deflection and elastic modulus of flexible pavements with normal clayey soil subgrade and scrap tire mix clayey soil subgrade. Data were collected from previous studies conducted at the Soil Mechanics and Foundation Engineering Division of Jadavpur University, Kolkata, West Bengal, India. This investigation aims to compare existing pavement with scrap tire modified subgrade pavement or modified pavement. The primary objective is to conduct a comparative analysis between the subgrade deflection and elastic modulus of the existing and modified pavement using the FWD system. Recent studies in geotechnical and transportation engineering have emphasized the significance of FWD in assessing soil and pavement conditions.

Further works on advanced pavement engineering with multi-directional FWD testing on concrete plates, highlighting asymmetry in structural behavior have been conducted. FWD tests at plate centers, measuring vertical deflections in eight directions were carried out. Significant asymmetries were found in a 22-year-old plate, while a new plate showed double-symmetric behavior. The study applied Kirchhoff-Love plate theory and optimized the uniform modulus of subgrade reaction, introducing an auxiliary surface load for more accurate results. Inertia forces were also considered, affecting the effective modulus of subgrade reaction by less than 3.5%. This research enhances pavement engineering by offering a detailed approach to assess structural integrity and asymmetry of concrete plates [9]. The FWD is a widely used non-destructive tool for pavement assessment, valued for its reliability, rapid operation, and user-friendliness. It employs back-calculation methods to compute layer moduli, highlighting the necessity of correction factors for reliable layer modulus determination. Furthermore, the statement references the development of low-cost, indigenous FWD models, particularly emphasizing their potential in pavement

engineering, with a special focus on applications in countries like India [10]. This indicates a growing interest in adapting advanced technologies to suit local economic and infrastructural contexts. Authors in [11] focus on using FWD data to assess pavement structural health. Their study covers 97 pavement sections in the South-Central United States, using 3D-Move software for simulating FWD deflection bowls. The research introduces the normalized Comprehensive Area Ratio (CAR) for evaluating pavement structures. It reveals that 3D-Move simulations correlate highly with the actual FWD deflections. The most significant contribution of this study is that it includes a new classification scale for pavement conditions, accounting for different drop loads, which aids in effective pavement maintenance and rehabilitation decisions. In [12], attention is paid to the development of intelligent pavement performance models to enhance the efficiency of highway maintenance and repair. The research highlights the importance of these models in managing pavement maintenance and rehabilitation, considering various factors, such as traffic, environmental, and climatic conditions. The study involves the use of FWD tests to analyze and understand patterns of deterioration in flexible pavements. For assessing the structural conditions of pavement, Deflection Basin Parameters (DBPs) from FWD data are utilized as an efficient alternative. Use of DBPs through finite-element modeling and field analyses, offers a comprehensive view of pavement conditions for rehabilitation decisions [13]. In the technical assessment of flexible pavements, especially those constructed on tropical soils, the bonding state of the subgrade is a critical aspect that can be effectively examined using FWD data. In this context, the deflection measurements obtained from FWD are pivotal, serving as key indicators of the subgrade condition. This emphasis on deflection data is crucial for the early detection of subgrade issues, which is fundamental in maintaining the overall integrity and longevity of the pavement structure [14]. Authors in [15] investigated the use of Reclaimed Asphalt Pavement (RAP) for stabilizing unbound layers in road structures, focusing on four sections of State Main Road A7 from Riga to the Lithuanian border. Utilizing a blend of laboratory and field assessments, including FWD data, the research back-calculated the equivalent modulus of elasticity (E_{eq}) of stabilized RAP. Findings revealed E_{eq} values of 370 MPa at the surface, decreasing to 100 MPa at 60 cm depth, with 67 FWD measurements confirming the effectiveness of cement-stabilized RAP. However, significant variability was noted, with some values reaching up to 4000 MPa, indicating anisotropy. The study concluded that cement-stabilized RAP is technically, economically, and environmentally viable for road construction, but emphasized the need for improved design and construction specifications due to the variability of the results, highlighting the importance of more detailed investigations into pavement structures with stabilized road bases. In [16], an extensive evaluation was conducted on a 20 km segment of the Barnala-Mansa State Highway. The primary tool for this assessment was the FWD, which was employed to gauge the pavement conditions both prior to and following the application of an overlay. The focal point was the calculation of critical parameters, notably the Surface Curvature Index (SCI) and Middle Layer Index (MLI). These indices provide valuable insights into the condition of

various pavement layers. Collectively, such studies are instrumental in enhancing the field of pavement engineering. They introduce innovative methodologies that are crucial for the assessment, design, and maintenance of pavement performance. Authors in [17] demonstrated through large-scale models and field tests that geocells reduce surface deflections and vertical pressure on the subgrade. These tests also examined the effect of aspect ratio, indicating improved performance with increased height to diameter ratio.

In the current study, a comprehensive methodology to examine the application of scrap tires in deflection and elastic modulus of pavement subgrades was used. The primary methods and procedures include data collection and experimental studies. Laboratory soaked CBR and thickness data of the pavements were collected from the Soil Mechanics Research Division of the Civil Engineering Department, Jadavpur University, Kolkata. FWD studies on the existing pavement and scrap tire-modified subgrade pavement were conducted to assess the deflection of the subgrade.

II. DATA COLLECTION AND SITE SELECTION

A. Background

The necessary data for analyzing FWD results were collected from prior research conducted by the Soil Mechanics and Foundation Engineering Division of the Jadavpur University. This research involved a detailed examination of a specific roadway segment under Public Works Department (PWD) in West Bengal. The roadway, stretching from Jibantala Bazar to Taldi Bazar near Canning (District-South 24 Parganas, West Bengal, India), initiates at 0.00 km near Jibantala crossing the market (coordinates: Latitude 22°20'37.7" N, Longitude 88°36'29.6" E) and concluding at Taldi Bazar near the railway station (Latitude 22°25'11.8" N, Longitude 88°39'44.4" E), covering a total distance of 12.45 km. This road belongs to the PWD of the Government of West Bengal. During the research, soil samples, termed as road chainage, were collected at various points along the length of the road. These samples, which are brownish grey silty clay, were taken to the laboratory for testing to determine their soaked California Bearing Ratio (CBR) and other metrics (Table I). The design CBR was obtained from Table I by following Clause 6.2.2 of IRC:37-2018 [18]. The results yielded a design CBR value of 3.36 for this particular road section. Additionally, soil samples were specifically selected from road subgrade locations where the soaked CBR closely aligned with the design CBR value. The innovative aspect of the study involved experimenting with various sizes of scrap tire pieces, ranging from 10 mm × 10 mm to 30 mm × 30 mm, mixed with the collected soil in proportions varying from 5% to 30%. The optimum improvement in CBR value was noted with 15mm × 15mm tire scraps at 10% weight of the soil, achieving a CBR of 8.90. Based on these laboratory findings, a 30 m long and 5.5 m wide flexible pavement section was constructed by following the design methodology of IRC 37:2018. This construction was situated 20 m away from the existing pavement and utilized the optimally determined tire scrap mix combined with original soil collected from the selected areas. The purpose of this construction was to replicate the laboratory-obtained CBR

values of the tire mix soil under field conditions, thereby validating the laboratory results.

TABLE I. LABORATORY TEST RESULTS ON ROAD SOIL WITH RESPECT TO CHAINAGE

Chainage (m)	Description of soil	Modified proctor		Soaked CBR (%)
		MDD (gm/cc)	OMC (%)	
100	Grey clayey silt	1.760	15.18	3.80
1000	Grey silty clay/ clayey silt	1.751	15.25	3.75
2000	Brownish grey silty clay	1.741	16.80	3.62
3000	Grey silty clay	1.724	17.06	3.39
4000	Grey silty clay	1.720	17.42	3.15
5000	Grey silty clay	1.710	18.14	3.05
6000	Brownish grey silty clay	1.730	16.50	3.30
7000	Grey silty clay/ clayey silt	1.728	17.12	3.45
8000	Brownish grey silty clay	1.750	15.60	3.75
9000	Grey silty clay	1.740	15.86	3.57
10000	Brownish grey silty clay	1.740	16.20	3.55
11000	Grey silty clay	1.730	17.16	3.41
12000	Grey silty clay	1.730	16.90	3.28
12450	Greyish brown silty clay	1.730	17.05	3.40

B. Site Selection

The focus was on the 3.00 km to 3.03 km stretch of Jibantala-Taldi Road, chosen for its smooth surface and uniform cross-section, which is ideal for FWD tests. Data from Jadavpur University supported this site selection. Both the modified and existing pavements, each 30.00 m in length, were divided into four equal segments of 10 m for testing, allowing for direct performance comparison. This methodological approach ensured a precise evaluation of the impact of tire scrap on pavement quality. Table II shows the different chainage points under study.

TABLE II. TEST POINTS AND CHAINAGE

Pavement type	Selected chainage for FWD test (m)	FWD test points			
		1 st point	2 nd point	3 rd point	4 th point
Existing pavement	3.00×10 ³ to 3.03×10 ³	3.00×10 ³ m	3.01×10 ³ m	3.02×10 ³ m	3.03×10 ³ m
Modified pavement	0.00 to 30.00	0.00 m	10.00 m	20.00 m	30.00 m

C. CBR Test results

For further processing in FWD and to account for the worst-case scenario, the minimum CBR values obtained from laboratory tests were considered. It is worth noting that laboratory CBR values were used in FWD analysis for the Barnala - Mansa Section of SH13 in the district of Barnala, Punjab, India, which spans a length of 20 km [16].

D. Pavement Details

Table III shows different layer pavement thickness values, for existing and modified subgrade acquired from Soil Mechanics Research Division of Jadavpur University. IITPAVE software was used for obtaining pavement thickness. Both pavements have a subgrade depth of 0.5 m.

TABLE III. DIFFERENT LAYER PAVEMENT THICKNESS FOR NORMAL AND TYRE SCRAP MIXED SOIL

Category	Layers	Pavement thickness	
		Existing road subgrade	Scrap tire modified subgrade
Bituminous Layer	Bituminous Concrete (BC)	40 mm	30 mm
	Dense Bituminous Macadam (DBM)	80 mm	50 mm
Granular layer	Wet Mix Macadam (WMM)	250 mm	250 mm
	Granular Sub Base (GSB)	200 mm	150 mm
Total thickness		570 mm	480 mm
Difference in thickness		90 mm	

III. EXPERIMENTAL STUDY

A. Testing Procedure and Methodology

The FWD is key Non-Destructive Testing (NDT) equipment for evaluating pavement strength, capable of calculating the elastic modulus of individual layers [19-20]. In the current study, the primary objective is to conduct a comparative analysis between the subgrade deflection and elastic modulus of existing and modified pavement. For the purpose of conducting the FWD survey, a loading force within the range of 0-100 kN is utilized. This range allows the FWD to efficiently simulate various types of vehicles loads on the pavement surface. In FWD study, the setup includes deflection sensors or geophones which are strategically placed at specific distances from the center of the loading plate. According to IRC 115:2014, the distances are - 0 mm (D0), 300 mm (D1), 600 mm (D2), 900 mm (D3), 1200 mm (D4), 1500 mm (D5), and 1800 mm (D6). These sensors have been used to measure the surface deflection resulting from dropped weights, such as 40kN (0.56 MPa contact stress) over a contact area of 150 mm radius. The loading time of the FWD typically ranges between 25 to 30 ms. A typical FWD schematic representation is illustrated in Figure 1.

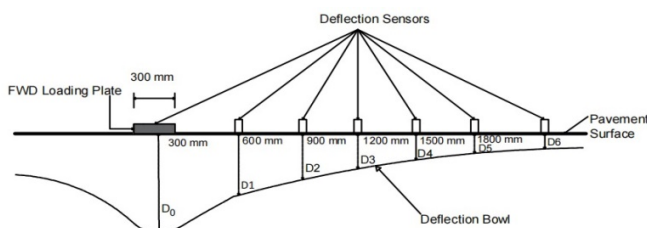


Fig. 1. Typical representation of FWD operation.

B. Testing Frequency

In the present work, FWD is applied to measure the subgrade deflection of the pavements, in line with the procedures outlined in Section 3 of IRC 115: 2014 [8]. This analysis involves testing at various locations as described in Table I. For both pavements, the intermediate distance for testing is 10 m. The referenced studies [21-22] conducted a technical analysis focusing on the deflection bowls captured during FWD testing across various typical South African pavement structures. These structures included granular, bituminous, and cemented base pavements. The FWD testing

was characterized by applying a load of 40 kN or exerting a contact pressure of 565.9 kPa. To ensure comprehensive coverage and detailed mapping of pavement conditions, high-density FWD surveys were strategically performed at intervals ranging from 5 to 10 m. This approach was meticulously designed to encompass both the outer and inner wheel tracks, covering the slow, fast, and shoulder lanes of these roads. Such a methodical and detailed approach in the FWD survey provided an in-depth understanding of the pavement behavior under load, which is essential for evaluating the structural integrity and serviceability of these roads.

C. FWD Test Results for Existing and Modified Subgrades

The deflection data from four points as specified in Table I, were gathered specifically for pavement performance analysis. These data points are presented in Tables IV and V.

TABLE IV. SUMMARY OF AVERAGE DEFLECTION (EXISTING SUBGRADE)

Chainage (m)	Distance from load center (mm)						
	0	300	600	900	1200	1500	1800
	Deflection (mm)						
	D0	D1	D2	D3	D4	D5	D6
3.00×10^3	0.519	0.322	0.197	0.099	0.063	0.048	0.022
3.01×10^3	0.509	0.324	0.217	0.109	0.065	0.047	0.037
3.02×10^3	0.529	0.340	0.236	0.093	0.063	0.047	0.038
3.03×10^3	0.518	0.342	0.146	0.096	0.062	0.047	0.036
Average deflection	0.519	0.332	0.199	0.099	0.063	0.047	0.033

TABLE V. SUMMARY OF AVERAGE DEFLECTION (FOR MODIFIED SUBGRADE)

Chainage (m)	Distance from load center (mm)						
	0	300	600	900	1200	1500	1800
	Deflection (mm)						
	D0	D1	D2	D3	D4	D5	D6
0.00	0.412	0.172	0.078	0.051	0.029	0.019	0.013
10.00	0.422	0.165	0.080	0.049	0.027	0.021	0.018
20.00	0.387	0.150	0.075	0.044	0.038	0.025	0.015
30.00	0.393	0.158	0.074	0.051	0.025	0.022	0.012
Average deflection	0.404	0.161	0.077	0.049	0.030	0.022	0.015

The study compares the two pavements by dividing each into four equal segments and by establishing specific Reference Change (RC) points for further analysis. Both pavements are 30 m long, but with different chainages. To simplify deflection data representation, the chainages are categorized as RC1 (0.00 m for modified and 3×10^3 m for existing pavement), RC 2 (10 m for modified and 3.01×10^3 m for existing pavement), RC 3 (20 m for modified and 3.02×10^3 m for existing pavement), and RC 4 (30 m for modified and 3.03×10^3 m for existing pavement). Figure 2 portrays the deflection data collected at these intervals. In this study, the primary focus is dedicated to the analysis and comparison of subgrade deflection and elastic modulus. To competently characterize the subgrade condition and gauge its performance, deflections have been measured at two key distances: 1200 mm (referred to as D1200) and 1500 mm (referred to as D1500). The difference between these two deflections is referred to as the Lower Layer Index (LLI), which is a deflection bowl parameter derived from the results of deflection tests [23, 24].

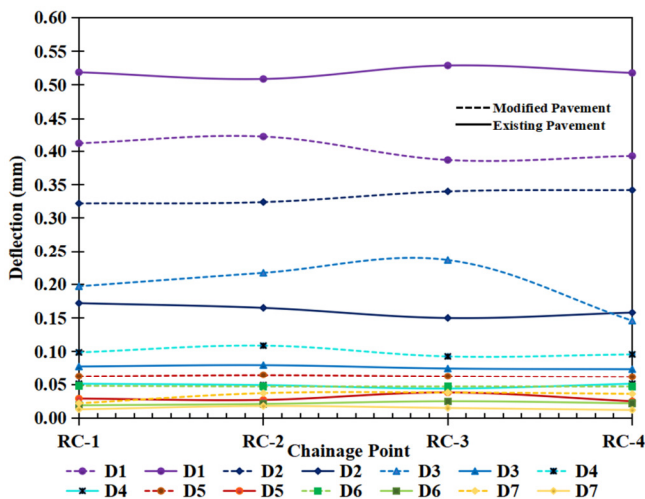


Fig. 2. Graphical presentation of the deflection of both pavements.

Table VI provides a summary of the average deflection for D1200 and D1500 along with the LLI for both types of pavements, offering insights into the performance and condition of subgrade.

TABLE VI. LLI OF THE SUBGRADE LAYER

Pavement type	Chainage (m)	Distance from load center (mm)		LLI (mm)
		1200	1500	(D4-D5)
		Deflection(mm)		
		D4	D5	
Existing subgrade	3.00×10 ³	0.063	0.048	0.015
	3.01×10 ³	0.065	0.047	0.018
	3.02×10 ³	0.063	0.047	0.016
	3.03×10 ³	0.062	0.047	0.015
Average LLI for existing pavement subgrade (LLI _{eps}) in mm				0.016
Scrap tyre modified subgrade pavement	0.00	0.029	0.019	0.010
	10.00	0.027	0.021	0.006
	20.00	0.038	0.025	0.013
	30.00	0.025	0.022	0.003
Average LLI for modified pavement subgrade (LLI _{msc}) in mm				0.008

D. Back Calculation of Layer Modulus for Both Pavements

The FWD was employed to apply a dynamic load to the existing pavement, and the response to this load was measured. The obtained deflection values are then utilized in the KGP-BACK software to calculate the elastic modulus of the modelled pavement layers, following the guidelines of IRC: 115-2014 [8]. The layer modulus was back-calculated with KGP-BACK program. The pavements were modelled as three-layer systems with bituminous layer, granular layer, and subgrade. The KGP-BACK program, a specific version of the BACKGA program developed by the transportation engineering section at IIT Kharagpur, India is a vital tool for the back-analysis procedure used to calculate the elastic modulus of pavement surfaces. This calculation relies on deflection measurements obtained from the FWD. The purpose of this procedure is to assess the structural condition of in-service pavements by determining the in situ elastic modulus of different pavement layers. Utilizing normalized data and the

additional input parameters specified in Table VII, the KGP-BACK software was employed to derive the pavement layer modulus. The sample input and output of the KGP-BACK for the existing pavement are illustrated in Figures 3 and 4. Using the inputs given in Table VII, the back calculated modulus of each layer is calculated and presented in Table VIII.

TABLE VII. INPUT PARAMETERS FOR KGP-BACK ANALYSIS

Parameters	Values	
	Existing pavement	Scrap tire modified subgrade pavement
Single wheel load (N)	40000	
Contact pressure (MPa)	0.56 [8]	
Number of deflections	7	7
Radial distance between each geophone (mm)	0, 300, 600, 900, 1200, 1500, 1800	
Design CBR (%)	3.36	8.90
Measured deflections (mm)	As per Table II	As per Table III
Pavement layer thickness (mm)	570	480
Poisson's ratio values	0.5, 0.4, 0.4 (bituminous layer, granular layers, subgrade as per [8])	

```

!!!!!! PRINT INPUT DATA !!!!!
!!!!!! PL. SEE THE MANUAL SUPPLIED FOR HELP !!!!!

TYPE PEAK FWD LOAD (N), CONTACT PRESSURE (MPa)
Standard Values are 40000 0.56
40000 .56

HOW MANY DEFLECTIONS WERE MEASURED (4 TO 10)?
7

PRINT RAD.DISTANCES (mm) WHERE DEFLECT. WERE MEASURED
eg: 0, 300, 600, 900, 1200, 1500 is a Typical
Configuration for six Geophones
0 300 600 900 1200 1500 1800

PRINT MEASURED DEFLECTIONS IN mm.
.5195 .322 .1968 .099 .0633 .048 .022
GIVE THE PAVEMENT RELATED INPUTS (3-LAYER SYSTEM)
TYPE EACH LAYER THICKNESS(mm). START FROM TOP
120 450

TYPE POISSON RATIO OF EACH LAYER. START FROM TOP
Suggested values are 0.5 0.4 0.4
.5 .4 .4

INPUT RANGE (lower and upper) FOR EACH LAYER MODULUS
Please note that Backcalculation Results will depend
on the selection of appropriate Ranges. The selection
of Ranges has to be made judiciously on the basis of
of the Pavement Condition

PRINT LOWER AND UPPER BOUND MODULI (MPa) LAYERS
PL. See the Manual supplied for guidance

750 3000
100 500
16.8 67.2

```

Fig. 3. Input window of KGP-BACK for the existing pavement.

```

#####
# !!! THANKS FOR USING KGPBACK !!! #
# THE RESULTS ARE GIVEN BELOW #
#####

#####
# INPUT DATA #
#####
No. of Layers = 3
FWD Load (N) = 40000.00
Contact Pressure (MPa) = .56
No. of Deflection points = 7
Deflections measured using FWD (mm) = .51950 .32200 .19680 .09900 .06330 .04800 .02200
Radial distances from centre of load (mm) = .0 300.0 600.0 900.0 1200.0 1500.0 1800.0
Layer thickness (mm) = 120.00 450.00
Poisson ratio values = .50 .40 .40
Layer Modulus (MPa) Ranges Selected :-
(a) Bituminous Surfacing = 750.0 3000.0
(b) Granular Base = 100.0 500.0
(c) Subgrade = 16.8 67.2

#####
# OUTPUT DATA #
#####

Backcalculated Layer Moduli are:
Surface (MPa) = 2465.5
Base (MPa) = 103.9
Subgrade (MPa) = 67.2

```

Fig. 4. Output window of KGP-BACK for the existing pavement.

TABLE VIII. BACK CALCULATED LAYER MODULUS VALUES

Pavement type	Back calculated Moduli (MPa)			
Existing pavement	Chainage (m)	Bituminous	Granular	Subgrade
	3.00×10^3	2296.20	102.30	67.20
	3.01×10^3	2975.80	107.80	67.20
	3.02×10^3	2300.60	111.30	67.20
Modified pavement	3.03×10^3	2925.20	189.90	67.20
	0.00	2997.80	176.60	178.00
	10.00	2918.60	189.90	178.00
	20.00	2958.20	232.20	178.00
Modified pavement	30.00	2815.20	186.80	178.00

E. Determination of Corrected Back Calculated Moduli

The back-calculated modulus of the granular, and subgrade layers obtained from software analysis were adjusted using a correction factor for seasonal variation, implemented for the granular and subgrade layers, in accordance with clause 6.5.1 of IRC:115-2014 [8]. Table IX displays the calculation of correction factors and the resulting corrected back-calculated modulus for the granular and subgrade layers, specifically accounting for seasonal variations. Figure 5 shows the corrected back calculated modulus chart for both pavements.

TABLE IX. CORRECTED BACK CALCULATED MODULI FOR GRANULAR ($E_{\text{gran_win}}$) AND SUBGRADE ($E_{\text{sub_win}}$) LAYER OF PAVEMENT

Pavement type	Chainage (m)	Back calculated $E_{\text{sub_win}}$ (MPa)	Back calculated $E_{\text{gran_win}}$ (MPa)	Corrected back calculated modulus for subgrade	Corrected back calculated modulus for granular layer
				$E_{\text{sub_mon}} = 3.351 \times (E_{\text{sub_win}})^{0.7688 - 28.9}$ (MPa)	$E_{\text{gran_mon}} = 3.351 \times (E_{\text{gran_win}})^{0.624 - 113.857}$ (MPa)
Existing pavement	3.00×10^3	67.2	102.30	56.22	75.60
	3.01×10^3	67.2	107.80	56.22	81.89
	3.02×10^3	67.2	111.30	56.22	85.83
	3.03×10^3	67.2	189.90	56.22	164.85
	Average back calculated modulus			56.22	102.04
Modified pavement	0.00	178	176.60	151.11	152.50
	10.00	178	189.90	151.11	164.85
	20.00	178	232.20	151.11	202.11
	30.00	178	186.80	151.11	162.00
	Average back calculated modulus			151.11	170.365

Figure 5 clearly illustrates that in the case of the scrap tire-modified subgrade pavement, there was an increase in the back calculated modulus for each component of the pavement. These back calculated modulus values play a crucial role in the analysis of the in-service pavement and the assessment of its structural condition, as outlined in Clause 6.3.1 of [8].

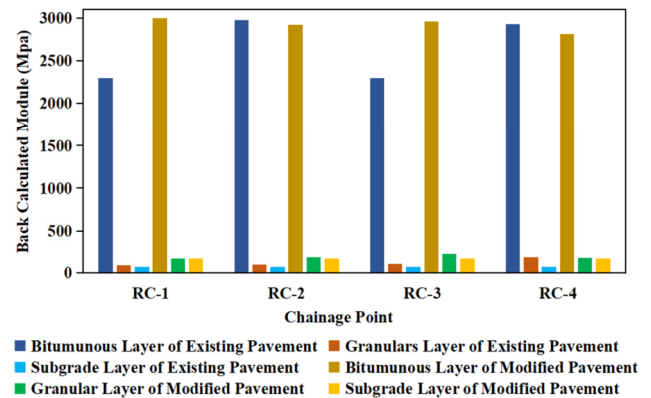


Fig. 5. Corrected back calculated moduli chart for both pavements.

IV. DISCUSSION ON FWD OBTAINED DEFLECTION AND ELASTIC MODULUS

A. Discussion on Subgrade Deflection

In this study, LLI serves for the characterization of the subgrade condition and was proved valuable in predicting the performance and assessing the overall condition [24-26]. To calculate the LLI, the average deflection values of D1200 and D1500 for both pavements were considered according to Table VI. The resulting LLI values are:

- LLI for existing pavement subgrade: $LLI_{\text{eps}} = 0.016$ mm (from Table IX).
- LLI for modified pavement subgrade: $LLI_{\text{mps}} = 0.008$ mm (from Table VIII).

This means that the decrease of LLI for modified pavement with respect to that of the existing pavement becomes:

$$[(0.016 - 0.008) \times 100 / 0.016] \% = 50 \%$$

Thus, the obtained data suggest that the improvement, in the form of decrease of LLI, is 50%. Figure 6 depicts the deflection variation of both pavements.

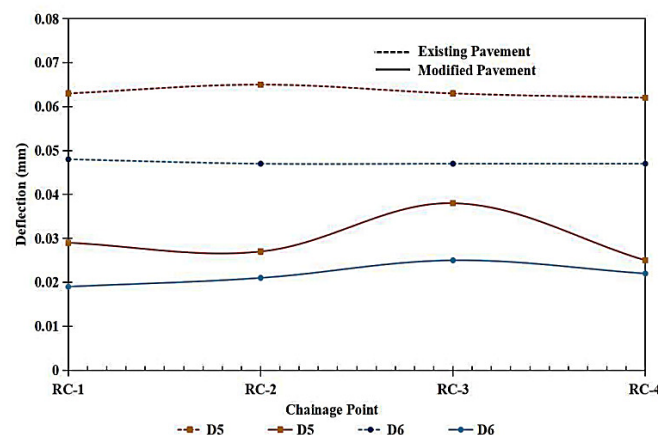


Fig. 6. Deflection variation in subgrade for both pavements.

LLI provides a quantitative measure of the subgrade's ability to distribute loads, while effectively characterizing the stiffness and load-bearing capacity of the subgrade. The LLI values are indicative of the structural integrity of the subgrade [23]. This implies that the LLI is exceptionally capable of identifying possible structural issues in the subgrade. A lower LLI value suggests a stiffer subgrade that is better at distributing loads, thus implying a potentially longer lifespan and reduced maintenance needs [26]. The LLI of the existing pavement subgrade indicates a relatively less stiff subgrade. This could translate to a higher likelihood of deformations under load, leading to possible issues like rutting or cracking in the overlying pavement layers. The LLI of the modified pavement subgrade suggests a considerable improvement in subgrade stiffness. This could be a result of modifications like the incorporation of materials (scrap tires) that enhance the load-bearing capacity. A stiffer subgrade as indicated by this lower LLI value could lead to better load distribution, reduced strain on the pavement layers, and probably a longer lifespan.

B. Discussion on the Elastic Modulus (E_s) of the Subgrade

This study provides crucial insights into the comparative performance of the existing pavement and the scrap tire modified subgrade pavement. Utilizing FWD for deflection measurements and the KGP-BACK software for back-calculating moduli values, the analysis aligns with the standards set forth [8]. For the existing pavement, the elastic modulus (E_{eps}) is measured at 56.22 MPa. This value falls within the typical range (20 to 100 MPa) for conventional pavement structures, indicating a standard level of stiffness. Such a modulus level suggests that the pavement is likely to perform adequately under normal traffic conditions [16]. However, this also implies potential limitations in its load-bearing capacity, possibly making it more susceptible to wear and degradation over time. In contrast, the modified pavement, characterized by an elastic modulus of 151.09 MPa, exhibits a markedly 2.68 times higher stiffness level with respect to the existing pavement, due to the integration of scrap tire rubber. This substantial increase in the modulus points to an enhanced load-bearing capacity and overall structural integrity. Consequently, pavements with such modifications are expected to offer improved durability, resist deformation more effectively, and potentially enjoy a longer service life [24]. The observation that the scrap tire modified pavement has a significantly higher modulus than the existing pavement underscores the effectiveness of using recycled materials in enhancing pavement performance.

V. CONCLUSIONS

The use of scrap tire material in pavement subgrades presents a promising method for enhancing pavement performance. Not only does this approach address environmental concerns related to tire waste, but also contributes to the development of more durable and sustainable road infrastructures. The following conclusions can be drawn from the research findings:

- The lower LLI value of the modified pavement subgrade suggests a considerable improvement in subgrade stiffness, possibly due to the incorporation of materials like scrap tire

rubber that enhance load-bearing capacity. This improvement in subgrade stiffness could result in better load distribution, reduced strain on pavement layers, and a possibly longer lifespan.

- The scrap tire modified pavement has 2.68 times higher modulus than the existing pavement, underscoring the effectiveness of utilizing recycled materials for enhancing pavement performance. This suggests that incorporating scrap tire rubber in pavement construction can lead to structural improvements, offering a sustainable and beneficial approach to ameliorate the durability and performance of road surfaces.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Soil Mechanics and Foundation Engineering Division of Jadavpur University, Kolkata, India, for providing essential laboratory equipment, technical support, and valuable data from their prior research. Additionally, the authors are thankful for the opportunity to conduct the FWD test on existing pavements, facilitated by the Public Works Department (PWD) of the Government of West Bengal, India. Their contributions have been instrumental to the success of this project.

REFERENCES

- [1] M. Sommer, J. Dargay, and D. Gately, "Vehicle Ownership and Income Growth, Worldwide: 1960-2030," *The Energy Journal*, vol. 28, pp. 143–170, Oct. 2007, <https://doi.org/10.2307/41323125>.
- [2] S. A. Tahami, A. F. Mirhosseini, S. Dessouky, H. Mork, and A. Kavussi, "The use of high content of fine crumb rubber in asphalt mixes using dry process," *Construction and Building Materials*, vol. 222, pp. 643–653, Oct. 2019, <https://doi.org/10.1016/j.conbuildmat.2019.06.180>.
- [3] F. D. B. de Sousa, "Devulcanization of Elastomers and Applications," in *Elastomers*, N. Cankaya, Ed. IntechOpen, 2017, <https://doi.org/10.5772/intechopen.68585>.
- [4] B. H. A. Bakar, A. T. Noaman, and H. M. Akil, "Cumulative Effect of Crumb Rubber and Steel Fiber on the Flexural Toughness of Concrete," *Engineering, Technology & Applied Science Research*, vol. 7, no. 1, pp. 1345–1352, Feb. 2017, <https://doi.org/10.48084/etasr.854>.
- [5] M. S. Mashiri, J. S. Vinod, M. N. Sheikh, and H.-H. Tsang, "Shear strength and dilatancy behaviour of sand–tyre chip mixtures," *Soils and Foundations*, vol. 55, no. 3, pp. 517–528, Jun. 2015, <https://doi.org/10.1016/j.sandf.2015.04.004>.
- [6] M. Bekhiti, H. Trouzine, and A. Asroun, "Properties of Waste Tire Rubber Powder," *Engineering, Technology & Applied Science Research*, vol. 4, no. 4, pp. 669–672, Aug. 2014, <https://doi.org/10.48084/etasr.439>.
- [7] H. Sellaf, H. Trouzine, M. Hamhami, and A. Asroun, "Geotechnical Properties of Rubber Tires and Sediments Mixtures," *Engineering, Technology & Applied Science Research*, vol. 4, no. 2, pp. 618–624, Apr. 2014, <https://doi.org/10.48084/etasr.424>.
- [8] *IRC 115: Guidelines for Structural Evaluation and Strengthening of Flexible Road Pavements Using Falling Weight Deflectometer (FWD) Technique*. New Delhi, India: Indian Roads Congress, 2014.
- [9] R. Díaz Flores, M. Aminbaghai, L. Eberhardsteiner, R. Blab, M. Buchta, and B. L. A. Pichler, "Multi-directional Falling Weight Deflectometer (FWD) testing and quantification of the effective modulus of subgrade reaction for concrete roads," *International Journal of Pavement Engineering*, vol. 24, no. 1, Dec. 2023, Art. no. 2006651, <https://doi.org/10.1080/10298436.2021.2006651>.
- [10] P. Makwana, "Assessment of Flexible Pavement Using the Falling Weight Deflectometer: A Comprehensive Review," *International Research Journal of Engineering and Technology*, vol. 10, no. 10, pp. 140–147, Oct. 2023.

- [11] N. R. Bastola, M. I. Souliman, S. Dessouky, and R. Daoud, "Structural health assessment of pavement sections in the southern central United States using FWD parameters," *Frontiers in Built Environment*, vol. 8, 2022, <https://doi.org/10.3389/fbuil.2022.1026469>.
- [12] M. Alam, M. Nazim, and S. K. Singh, "Deterioration Pattern of Flexible Pavement with the Help of Falling Weight Deflectometer," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 8, pp. 737–744, Jun. 2020, <https://doi.org/10.35940/ijitee.H6755.069820>.
- [13] Md. F. Rabbi and D. Mishra, "Using FWD deflection basin parameters for network-level assessment of flexible pavements," *International Journal of Pavement Engineering*, vol. 22, no. 2, pp. 147–161, Jan. 2021, <https://doi.org/10.1080/10298436.2019.1580366>.
- [14] M. Razali, N. A. C. Mahmood, K. A. Hashim, S. Mansor, and N. I. Zainuddin, "The falling weight deflectometer (FWD) for characterization bonding state of subgrade," in *Advances in Civil Engineering and Science Technology*, Penang, Malaysia, Oct. 2018, vol. 2020, Art. no. 020022, <https://doi.org/10.1063/1.5062648>.
- [15] J. Qiu, M. Huurman, M. H. T. Frunt, and R. Gelpke, "Towards high quality re-use of porous asphalt," in *Bearing Capacity of Roads, Railways and Airfields*, 1st ed., A. Loisos, I. Al-Qadi, and T. Scarpas, Eds. London, UK: CRC Press, 2017, <https://doi.org/10.1201/9781315100333-167>.
- [16] U. Solanki, P. Gundaliya, and M. Barasara, "Structural evaluation of flexible pavement using Falling Weight Deflectometer," in *Multi-disciplinary Sustainable Engineering: Current and Future Trends*, P. N. Tekwani, M. Bhavsar, and B. A. Modi, Eds. London, UK: CRC Press, 2016, pp. 141–146.
- [17] A. Emersleben and N. Meyer, "Bearing Capacity Improvement of Gravel Base Layers in Road Constructions using Geocells," in *Proceedings of the 12th International Conference of International Association for Computer Methods and Advances in Geomechanics (IACMAG'08)*, Goa, India, Oct. 2008, pp. 3538–3545.
- [18] *IRC 37-2018: Guidelines for the Design of Flexible Pavements*. New Delhi, India: Indian Roads Congress, 2018.
- [19] L. F. Walubita *et al.*, "Texas Flexible Pavements and Overlays: Year 1 Report - Test Sections, Data Collection, Analyses, and Data Storage System," Texas Transportation Institute, The University of Texas at El Paso, El Paso, TX, USA, Technical Report: November 2010–September 2011 FHWA/TX-12/0-6658-1, Jun. 2012.
- [20] U. J. Solanki, P. J. Gundalia, and M. K. Barasara, "A Review on Structural Evaluation of Flexible Pavements using Falling Weight Deflectometer," *Trends in Transport Engineering and Applications*, vol. 2, no. 1, pp. 1–10, Jan. 2014.
- [21] J. H. Maree and R. J. L. Bellekens, *The effect of asphalt overlays on the resilient deflection bowl response of typical pavement structures*. Pretoria, South Africa: Dept. of Transport, Chief Directorate National Roads, 1991.
- [22] J. H. Maree and F. Jooste, "Structural classification of pavements through the use of IDM deflection basin parameters," RDAC Report PR, 91, 325, 1999.
- [23] E. Horak, "Benchmarking the structural condition of flexible pavements with deflection bowl parameters," *Journal of the South African Institution of Civil Engineers*, vol. 50, no. 2, pp. 2–9, Jun. 2008.
- [24] O. Talvik and A. Aavik, "Use of FWD Deflection Basin Parameters (SCI, BDI, BCI) for Pavement Condition Assessment," *The Baltic Journal of Road and Bridge Engineering*, vol. 4, no. 4, pp. 196–202, Dec. 2009, <https://doi.org/10.3846/1822-427X.2009.4.196-202>.
- [25] U. J. Solanki, "A Study on FWD Deflection Bowl Parameters for Structural Evaluation of Flexible Pavement," *Trends in Transport Engineering and Applications*, vol. 4, no. 2, pp. 1–7, Jul. 2019, <https://doi.org/10.3759/tea.v4i2.2873>.
- [26] L. Fuentes, K. Taborda, X. Hu, E. Horak, T. Bai, and L. F. Walubita, "A probabilistic approach to detect structural problems in flexible pavement sections at network level assessment," *International Journal of Pavement Engineering*, vol. 23, no. 6, pp. 1867–1880, May 2022, <https://doi.org/10.1080/10298436.2020.1828586>.

The Efficiency of Surface Impedance Technique in the Transverse Wave Approach for the EM-Modeling of Fractal-Like Tree Structure used in 5G Applications

Mohamed Ayari

Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia | Syscom Laboratory, National Engineering School of Tunis, University of Tunis El-Manar, Tunisia
mohamed.ayari@nbu.edu.sa (corresponding author)

Saleh M. Altowaijri

Department of Information Systems, Faculty of Computing and Information Technology, Northern Border, University, Saudi Arabia
saleh.altowaijri@nbu.edu.sa

Received: 30 December 2023 | Revised: 13 January 2024 and 25 January 2024 | Accepted: 30 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6744>

ABSTRACT

Fractal antenna technology is a promising approach for 5G applications because its complex nature offers optimization potential in terms of time and space trade-offs. However, the computational effort required to analyze such antennas is significant. This paper investigates the Advanced Transverse Wave Approach (ATWA), which utilizes the surface impedance technique to improve simulation efficiency. This study introduces and analyzes a fractal-like 5G tree structure, displaying improved computational accuracy and efficacy, as well as peak memory utilization compared to current works. The proposed approach demonstrates significant effectiveness in enhancing the performance of complex fractal antennas for 5G technology and shows promise for integration with cloud, fog, and edge computing environments. This integration could potentially optimize data processing and network efficiency in these advanced computing landscapes.

Keywords-5G fractal-like tree; advanced transverse wave approach; surface impedance technique; computational efficiency; cloud computing; fog computing; edge computing

I. INTRODUCTION

In the dynamic landscape of wireless communications, the perpetual quest for innovation drives researchers and engineers to continually improve the efficiency of antennas. Fractal antennas are a breakthrough that has captured considerable attention. Drawing inspiration from the captivating realm of fractal geometry, these antennas have ushered in a revolution in wireless communications. Fractal antennas emerge from the fusion of fractal geometry with antenna design. The concept revolves around creating antenna structures that replicate themselves on diminishing scales, reflecting the self-replicating nature of fractal patterns [1-3]. This innovative approach enables fractal antennas to exhibit extraordinary characteristics, including compact size, broadband frequency coverage, and enhanced performance. Fractal antennas offer several key advantages. First, their compact size stands out in contrast to traditional antennas that require specific lengths for resonance. This compactness is particularly valuable in devices with

limited space, such as smartphones and laptops [4]. Second, their broadband capability allows them to cover a wide range of frequencies simultaneously, which is essential in modern communication systems using multiple bands, such as Wi-Fi, GPS, and cellular networks [5]. Third, fractal antennas are designed for efficient operation in multiple frequency bands, demonstrating crucial versatility in supporting various wireless communication standards [6]. Lastly, their intricate and self-similar structures contribute to enhanced radiation patterns and reduced interference, leading to superior signal quality and reception [7].

Fractal antennas find diverse applications in different industries. They are widely used in wireless communications, ensuring efficient data transmission. In the field of radio astronomy, fractal antennas serve as valuable tools capable of capturing signals from celestial sources by covering multiple frequencies [8]. In military and defense, these antennas are employed in communication systems and radar applications, offering adaptability to different frequencies in a compact form

factor. The medical field benefits from fractal antennas, as they play crucial roles in wireless telemetry devices, remote monitoring systems, and even medical imaging equipment [9]. In addition, the automotive industry has adopted fractal antennas for applications, like GPS, satellite radio, and keyless entry systems [10]. Recent advances in fractal antenna research showcase the potential for transformative solutions in the ever-progressing field of wireless communication. Notable studies focus on optimizing fractal-based antennas for 5G networks, addressing challenges related to accommodating multiple frequency bands and the need for compact and efficient designs [11]. Miniaturization efforts, particularly for portable and IoT devices, have been explored, revealing that fractal geometries allow the creation of compact antennas without compromising performance [12]. The versatility of fractal antennas in covering a wide range of frequencies was emphasized in multiband and wideband operation, addressing the demands of various wireless communication standards [13]. In the realm of satellite communication, fractal antennas contribute to enhancing the efficiency and reliability of satellite links, supporting global connectivity [14]. Furthermore, the integration of 3D printing technology into fractal antenna design has been investigated, demonstrating cost-effective and customizable solutions for various wireless communication applications [15]. These recent studies underscore the evolving landscape of fractal antennas, affirming their potential to address contemporary challenges and shape the future of wireless communications.

Simulating and analyzing the performance of fractal antennas is a vital step in their design and optimization, with computational time standing out as a crucial aspect impacting the efficiency of the design process. The complexity of fractal geometry, especially with intricate self-replicating patterns, increases computational demands, particularly for designs such as the Koch or Sierpinski curves [16]. The wide frequency range and bandwidth coverage of fractal antennas require comprehensive simulations, which affect computational time, especially for wideband scenarios or multiband operations. The choice of simulation software and algorithms, along with considerations for mesh density and accuracy, further influences computational time. Hardware resources, including high-performance computing clusters and GPUs, play a significant role in reducing simulation time for complex fractal antenna designs. Specific simulation goals, iterative optimization, and sensitivity analysis add layers to computational time considerations, emphasizing the need for a balanced approach between result accuracy and available resources to harness the potential of fractal antennas for wireless communications. Numerical electromagnetic (EM) methods play a pivotal role in the realm of 5G planar structures, serving as the foundational framework. These methods encompass various techniques, among which are the Method Of Moments (MOM) [17] with its alternative Fast Multipole Method (FMM) and the extensive Partial Element Equivalent Circuit (PEEC) method specifically designed to tackle integral equations related to EM waves. In addition, methods, such as the Finite Element Method (FEM), the Finite Difference Time Domain (FDTD) method, and the Discontinuous Galerkin Method (DGM) [18-21] serve as

differential equation solvers. These numerical approaches require a lot of memory and processing power, combined with the requirement of high accuracy in the functions and design. This is particularly important for performing EM-modeling of planar structures used in wireless systems.

This study investigates the Advanced Transverse Wave Approach (ATWA) [22-24], which stands out from previous numerical EM methods due to its notable advantages in terms of speed, compactness, and memory efficiency, introducing the surface impedance technique to tailor ATWA for fractal 5G antenna simulations and enhance its adaptability. Thorough tests were performed to assess the efficiency and stability of the adapted ATWA approach within this specific domain. Drawing inspiration from the aforementioned studies, a Fractal-Like Tree Structure was used as a suitable prototype, to validate the proposed approach in the context of 5G applications. This prototype provides a foundation for investigating more intricate antennas, contributing to the advancements in ongoing wireless technology.

II. THEORETICAL FRAMEWORK OF ATWA

This section addresses challenges in dealing with EM radiation and scattering by acknowledging the difficulty of obtaining analytic solutions for integral equations. As a solution, computational techniques are employed, leading to a concise exploration of the mathematical foundation and theoretical framework of the ATWA. ATWA, a numerical EM method, is characterized by its efficiency and strengths. The method, specifically designed for full and millimeter-wave analysis of planar structures, stands out for its avoidance of matrix inversion, freedom from constraints on component forms, prevention of numerical instabilities linked to large matrices, and assured convergence irrespective of planar structure interfaces. The fundamental integral relation connecting the electric field and current density is expressed as:

$$E(p, q) = \int_S G(p, q, p', q') J(p', q') dp' dq' \quad (1)$$

where $G(p, q, p', q')$ is the dyadic Green's function. Derived from the wave concept, the equation linking transverse electric field E_T to transverse magnetic field H_T is:

$$\begin{bmatrix} \vec{A}_r \\ \vec{B}_r \end{bmatrix} = \hat{K} \begin{bmatrix} \vec{E}_T^r \\ \vec{H}_T^r \times \vec{n}_r \end{bmatrix} = \hat{K} \begin{bmatrix} \vec{E}_T^r \\ \vec{J}_T^r \end{bmatrix} \quad (2)$$

\hat{K} guarantees the transition from integral EM field to algebraic EM waves.

$$\hat{K} = \frac{1}{2} \begin{bmatrix} Z_{0r}^{-\frac{1}{2}} & Z_{0r}^{\frac{1}{2}} \\ Z_{0r}^{-\frac{1}{2}} & -Z_{0r}^{\frac{1}{2}} \end{bmatrix} \quad (3)$$

The wave impedance Z_{0r} for homogeneous isotropic region r is:

$$Z_{0r} = \sqrt{\frac{\mu_0}{\epsilon_0 \epsilon_{r_r}}} \quad (4)$$

where μ_0 is vacuum permeability, ϵ_0 is vacuum permittivity, and ϵ_{r_r} is the dielectric constant from region r . The incident (A) and reflected (B) waves are related by:

$$\begin{cases} A = \hat{F}B & (\text{modal domain}) \\ B = \hat{S}A + B_{(0)} & (\text{spatial domain}) \end{cases} \quad (6)$$

where $B_{(0)}$ denotes the overall excitation wave at the source.

The shift from spatial to modal spaces entails implementing the two-Dimensional Fast Fourier Transform (2D-FFT) or the two-Dimensional Non-Uniform Fast Fourier Transform (2D-NUFFT) at the discontinuity interface for accelerated iterative processes. Transforming from spectral to modal spaces includes specific rotation as defined in [22, 25]. This iterative process continues until system convergence. The 2D-TWA is expedited by an anisotropic mesh technique, ensuring both memory and computational complexity for the Forward and Backward of the ATWA process at $O(N \log N)$, N being the meshing density, which is faster than conventional computational EM methods such as MOM or FEM.

A. Surface Impedance Technique

Considering (1) and (2), the Green operator, in the case of homogeneous media, can be viewed as an impedance operator. The complex surface impedance establishes a connection between the electric field and the current density within any subdomain delineating the discontinuity plane. The behavior of the surface impedance subdomain, denoted by Z_S , resembles that of a metallic domain when Z_S is zero and functions akin to a dielectric domain when Z_S approaches infinity. Within this surface impedance subdomain, the continuity and boundary conditions of transverse EM fields in LIM media necessitate the equality of electric fields on both sides of the interface, linking the transverse electric field to the total tangential current density in the following manner:

$$\begin{cases} E_T = E_T^{r=1} = E_T^{r=2} \\ J_{T_{total}} = J_T^{r=1} + J_T^{r=2} \\ E_T = Z_S \cdot J_{T_{total}} \end{cases} \quad (7)$$

In terms of waves:

$$\begin{cases} Z_{01}^2 \times (A_1 + B_1) = Z_{02}^2 \times (A_2 + B_2) \\ \left(Z_{01}^2 \times (A_1 - B_1) + Z_{02}^2 \times (A_2 - B_2) \right) \cdot Z_S = Z_{01}^2 \times (A_1 + B_1) \end{cases} \quad (8)$$

From (6), the expression of $S_{\Omega_{SUB}}$ on the surface impedance subdomain $\Omega_{SUB} = I_{surf}$ becomes:

$$S_{\Omega_{SUB} = I_{surf}} = \begin{bmatrix} \frac{Z_S \times (1 - \rho^2) - Z_{01}}{Z_S \times (1 + \rho^2) + Z_{01}} & \frac{2\rho \times Z_S}{Z_S \times (1 + \rho^2) + Z_{01}} \\ \frac{2\rho \times Z_S}{Z_S \times (1 + \rho^2) + Z_{01}} & \frac{Z_S \times (\rho^2 - 1) - Z_{01}}{Z_S \times (1 + \rho^2) + Z_{01}} \end{bmatrix} \quad (9)$$

where:

$$\rho = \sqrt{\left(\frac{\mu_{r1} \epsilon_{r2}}{\mu_{r2} \epsilon_{r1}} \right)^{\frac{1}{2}}} \quad (10)$$

III. FRACTAL-LIKE TREE STRUCTURE FOR 5G

The fractal-like tree structure for 5G is a specialized antenna design crafted to meet the specific requirements of 5G wireless communication systems. Drawing inspiration from fractal geometry, it integrates self-replicating patterns to

optimize antenna performance. With a tree-like configuration, it offers distinct advantages, including compactness, multiband capability, and efficient operation across a wide frequency spectrum. Tailored to address the challenges of modern wireless networks, this antenna holds significant promise in enhancing data rates, expanding network coverage, and overall efficiency within the rapidly evolving landscape of 5G technology. This antenna was chosen to validate the proposed approach and demonstrate its effectiveness. The evolution of fractal-like trees is achieved through recursive processes. Every terminal segment endeavors to spawn fresh branches extending forward, at a 45° angle to the left, or a 45° angle to the right. The integration of a segment into the structure is contingent upon its noncollision with any existing or prospective segments. Different colors serve to delineate different stages in the generation process, as shown in Figure 1.

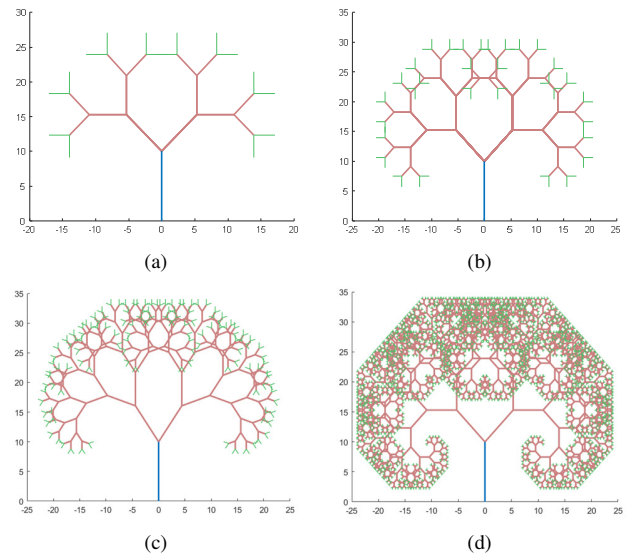


Fig. 1. Fractal-like antennas with scale factor 0.75 and angular vector $\pi/4$ at different iterations: (a) $n = 4$, (b) $n = 6$, (c) $n = 8$, (d) $n = 12$.

IV. SIMULATION RESULTS AND DISCUSSION

Two methods were used to examine and enhance the fractal-like tree structure of the antenna across varying iterations ($n = 4, 6, 8, 12$) within the domain of 5G technology: the direct ATWA-based solver and the ATWA method incorporating the Surface Impedance Technique (SIT). In the initial phase, the computation of the complex value of surface impedance (Z_S) was executed for each pixel using (3) and [27]. The mean average square of Z_S was subsequently employed to characterize the equivalent fractal zone on each pixel. As an illustrative example, for the antenna-like tree with $n = 8$ iterations, the computed Z_S manifested as $0.737i$.

A discernible correlation emerged when comparing the direct and SIT methods. The reflection coefficient (S_{11}) demonstrated multifrequency bands, notably at $[21, 23.1]$ and $[50.5, 55.8]$ GHz for the direct method, and $[24.9, 28]$ and $[50.8, 51.5]$ GHz for the SIT technique. The resonance frequencies were closely aligned at 22 and 52.2 GHz for the direct method and 26 and 51 GHz for the SIT approach. The

corresponding return losses were determined at -14.3 and -38.7 dB for the direct method, and -14.83 and -12.2 dB for the SIT approach. The relative error between the two methods was approximately 1.7%, a deemed acceptable margin considering the computational and memory demands intrinsic to fractal systems. The occurrence of double resonances was ascribed to the inherent fractal nature of the antenna, a validation supported by simulation results based on 2D-ATWA. The Voltage Standing Wave Ratio (VSWR) at 1.027 dB within the 51 GHz band showcased favorable impedance characteristics, particularly pertinent to 5G wireless systems. An in-depth review of the simulation results, encompassing the impedance evolution across the [1, 60] GHz frequency range, as well as azimuthal, elevational, and 3D patterns at 51 GHz, accentuated the validation of the antenna through the chosen approach. When combined, the simulation results robustly confirm the efficacy of the SIT technique in the EM analysis of fractal-like tree structures within the domain of 5G technology.

TABLE I. FRACTAL-LIKE TREE DESIGN AND MODELING PARAMETERS

Design Parameters	Description
Length	7.3185 mm
Height	0.12227 mm
Number of Stages (Iteration)	4
Ground plane length	8.1297 mm
Ground plane width	8.945 mm
Feed diameter	0.16795 mm
Substrate (dielectric)	RO4725JXR $\epsilon_{r1} = 1$ $\epsilon_{r2} = 2.55$ Loss tangent= 0.0022 Thickness= 0.12227 mm
Meshing	512x512
Modeling Parameters	Description
Nature of box	Periodic walls
Type of polarization	Bilateral in x-direction
Number of iterations	$N_{iter} = 200$
Value of surface impedance	$Z_s = 0.737i$
Waveband	$F_{min} = 1$ GHz
	$F_{max} = 60$ GHz
	$Step_{Frq} = 0.1$ GHz

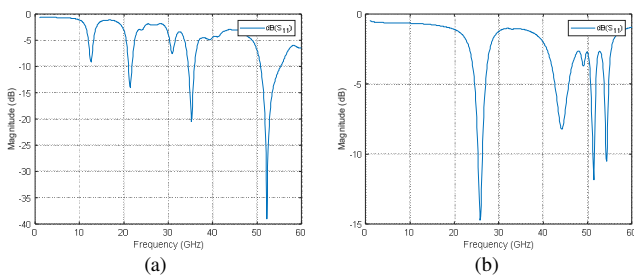


Fig. 2. Evolution of insertion parameter S_{11} (dB) with a comparison between (a) direct method and (b) SIT.

Table II compares the performance of the selected antenna with two benchmark antennas, focusing on the multiband functionality, which is essential for Ultra-Wideband (UBW) and 5G networks. This comparison goes beyond basic efficiency metrics, showcasing the proposed antenna's superior bandwidth and gain. The results highlight the significant

advantage of the proposed antenna in handling the diverse frequency demands of modern UBW and 5G applications, demonstrating its potential as a robust solution in the evolving landscape of wireless communication. Furthermore, considering the effectiveness of both the SIT and the AMT techniques [28] introduced by the ATWA approach, a promising avenue is created for exploring compact and responsive fractal structures applicable to emerging 5G wireless technologies.

TABLE II. COMPARISON BETWEEN THE PROPOSED FRACTAL-LIKE TREE STRUCTURE AND REFERENCE ANTENNAS

Antenna/Parameter	Proposed antenna	[29]	[30]
Type	fractal-like tree	fractal-like tree	fractal-like tree
Feeding Technique	CPW	CPW	CPW
Application	UWB, 5G	UWB, 5G	WLAN, Wi-Fi
Angular Vector	$\pi/4$	$\pi/2$	$\pi/3$
Scale Factor	0.75	0.7	0.7
Iteration level	4	2	2
Substrate	RT/duroid 5880	RT/duroid 5880	FR4_epoxy
Size (mm)	13.3x7.43x0.58	90x110x16	13x8x1.58
Simulator	Proposed simulator	HFSS	HFSS
Numerical EM method	A-TWA	FEM	FEM
Mode of Operation	multiband	multiband	multiband
Impedance bandwidth	$S_{11} \leq -10$ dB	$S_{11} \leq -10$ dB	$S_{11} \leq -10$ dB
Ranges (GHz)	(21–23.1), (34–36.2), (50.5–55.8)	(1.6–7.5), (8–10), (10.5–12.2), (12.7–16.1)	(1.12–2.3), (2.4–3.6)
VSWR	< 2	< 2	< 2

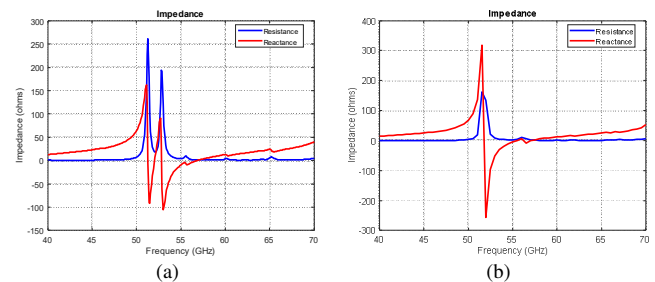


Fig. 3. Evolution of the impedance as a function of frequency with a comparison between (a) direct method and (b) SIT.

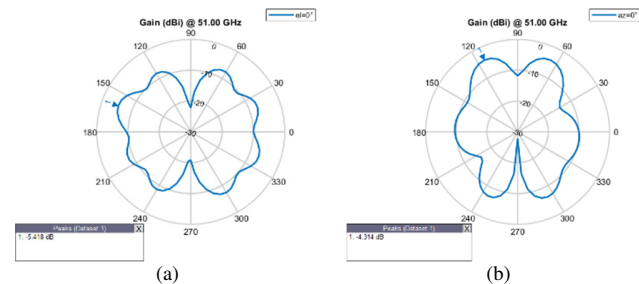


Fig. 4. Azimuth pattern at 51GHz with a comparison between (a) direct method and (b) SIT.

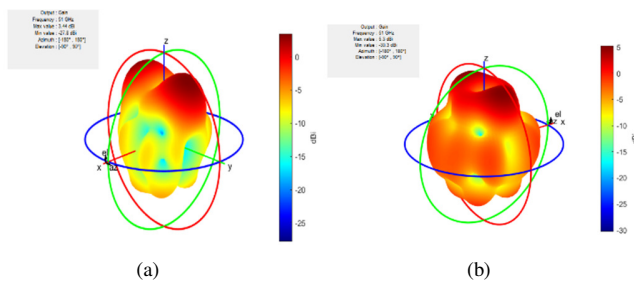


Fig. 5. 3D-pattern at 51GHz with a comparison between (a) direct method and (b) SIT.

V. CONCLUSIONS

This study presented a succinct theoretical framework and mathematical underpinnings for the proposed 2D-ATWA, incorporating the SIT for comprehensive full and mm-wave applications. Design and analysis of a 5G fractal-like tree antenna were executed precisely, and rigorous validation, evaluation, and comparison of simulation results were performed against both the direct and SIT-based ATWA approaches. The results emphasized the EM stability and the efficiency of the SIT technique when applied to the analysis of fractal-like tree structures within the 5G technology landscape. The integration of enhanced ATWA with an anisotropic mesh technique holds promise for providing effective solutions for stabilizing and selecting optimal parameters, especially in the intricate domain of 5G fractal antennas. The improved efficacy of the ATWA method in modeling complex structures may be crucial to developing antennas for cloud, fog, and edge computing environments. These computing paradigms often require advanced wireless communication technologies, such as 5G, to ensure high-speed and reliable connectivity. The fractal-like tree structures modeled in this study can support dense, distributed networks characteristic of cloud, fog, and edge computing, offering optimized coverage and bandwidth capabilities. Moreover, the emphasis of ATWA on computational efficiency aligns well with the resource-constrained environments of fog and edge computing, where processing power and memory are at a premium. This fact will be investigated in future work. Furthermore, in future work, the goal is to use cloud computing resources to model larger problem sizes. This enables the exploration of more complex fractal-like structures and their applications in advanced 5G networks. Utilizing cloud computing's vast processing and storage capabilities will allow for more detailed and comprehensive simulations, further enhancing the practicality and applicability of the ATWA method in real-world scenarios.

ACKNOWLEDGMENT

The authors gratefully acknowledge the approval and the support of this research study by the grant no.-CSCR-2022-11-1616 from the Deanship of Scientific Research at Northern Border University, Arar, K.S.A.

REFERENCES

- [1] B. B. Mandelbrot, *The Fractal Geometry of Nature*. San Francisco, CA, USA: Times Books, 1982.
- [2] M. F. Barnsley, R. L. Devaney, B. B. Mandelbrot, H. O. Peitgen, D. Saupe, and R. F. Voss, *The Science of Fractal Images*. New York, NY, USA: Springer, 1988.
- [3] A. Jamil, K. Rafique, M. D. Idrees, A. S. Rajput, A. Abdullah, and A. S. Saand, "Design and Geometric Transformations of Koch Curve Monopole Antennas," *Engineering, Technology & Applied Science Research*, vol. 12, no. 2, pp. 8452–8457, Apr. 2022, <https://doi.org/10.48084/etasr.4767>.
- [4] C. B. Nsir, J. M. Ribero, C. Boussetta, and A. Gharsallah, "Design of a 1x2 CPW Fractal Antenna Array on Plexiglas Substrate with Defected Ground Plane for Telecommunication Applications," *Engineering, Technology & Applied Science Research*, vol. 11, no. 6, pp. 7897–7903, Dec. 2021, <https://doi.org/10.48084/etasr.4558>.
- [5] S. Chouhan, D. K. Panda, V. S. Kushwah, and S. Singhal, "Spider-shaped fractal MIMO antenna for WLAN/WiMAX/Wi-Fi/Bluetooth/C-band applications," *AEU - International Journal of Electronics and Communications*, vol. 110, Oct. 2019, Art. no. 152871, <https://doi.org/10.1016/j.aeue.2019.152871>.
- [6] A. T. Abed *et al.*, "Challenges and limits of fractal and slot antennas for WLAN, LTE, ISM, and 5G communication: a review paper," *Annals of Telecommunications*, vol. 76, no. 9, pp. 547–557, Oct. 2021, <https://doi.org/10.1007/s12243-020-00828-6>.
- [7] K. V. Ajetao and A. P. Dhande, "Phi Shape UWB Antenna with Band Notch Characteristics," *Engineering, Technology & Applied Science Research*, vol. 8, no. 4, pp. 3121–3125, Aug. 2018, <https://doi.org/10.48084/etasr.2013>.
- [8] A. Kaur, R. Khanna, and M. V. Kartikeyan, "A stacked sierpinski gasket fractal antenna with a defected ground structure for UWB/WLAN/RADIO astronomy/STM Link applications," *Microwave and Optical Technology Letters*, vol. 57, no. 12, pp. 2786–2792, 2015, <https://doi.org/10.1002/mop.29442>.
- [9] A. Martínez-Lozano *et al.*, "UWB-Printed Rectangular-Based Monopole Antenna for Biological Tissue Analysis," *Electronics*, vol. 10, no. 3, Jan. 2021, Art. no. 304, <https://doi.org/10.3390/electronics10030304>.
- [10] F. Ez-Zaki, H. Belahrach, and A. Ghammaz, "Broadband microstrip antennas with Cantor set fractal slots for vehicular communications," *International Journal of Microwave and Wireless Technologies*, vol. 13, no. 3, pp. 295–308, Apr. 2021, <https://doi.org/10.1017/S1759078720000719>.
- [11] A. Rahim and P. K. Malik, "Analysis and design of fractal antenna for efficient communication network in vehicular model," *Sustainable Computing: Informatics and Systems*, vol. 31, Sep. 2021, Art. no. 100586, <https://doi.org/10.1016/j.suscom.2021.100586>.
- [12] P. K. Goswami and G. Goswami, "Trident Shape Ultra-Large Band Fractal Slot EBG Antenna for Multipurpose IoT Applications," *Progress In Electromagnetics Research C*, vol. 96, pp. 73–85, 2019, <https://doi.org/10.2528/PIERC19073002>.
- [13] M. S. Fouad, A. E. Farahat, K. F. A. Hussein, A. H. A. M. Shaaan, and M. F. Ahmed, "Super-wideband Fractal Antenna for Future Generations of Wireless Communication," *Progress In Electromagnetics Research C*, vol. 136, pp. 137–149, 2023, <https://doi.org/10.2528/PIERC23042507>.
- [14] A. Kumar and A. P. Singh, "Design of micro-machined modified Sierpinski gasket fractal antenna for satellite communications," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 29, no. 8, 2019, Art. no. e21786, <https://doi.org/10.1002/mmce.21786>.
- [15] A. E. Ahmed and W. A. E. Ali, "Analysis and Design of 3-D Printed Fractal Triangular Antenna for Wireless Communications Applications," in *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, Istanbul, Turkey, Jun. 2020, pp. 1–5, <https://doi.org/10.1109/ICECCE49384.2020.9179333>.
- [16] I. S. Bangi and J. S. Sivia, "Moore, Minkowski and Koch Curves Based Hybrid Fractal Antenna for Multiband Applications," *Wireless Personal Communications*, vol. 108, no. 4, pp. 2435–2448, Oct. 2019, <https://doi.org/10.1007/s11277-019-06531-7>.
- [17] M. Ayari, Y. E. Touati, and S. Altowajiri, "Method of Moments versus Advanced Transverse Wave Approach for EM Validation of Complex Microwave and RF Applications," *Journal of Electromagnetic*

- Engineering and Science*, vol. 20, no. 1, pp. 31–38, Jan. 2020, <https://doi.org/10.26866/jees.2020.20.1.31>.
- [18] H. Temimi, "Error analysis of a novel discontinuous Galerkin method for the two-dimensional Poisson's equation," *Applied Numerical Mathematics*, vol. 189, pp. 130–150, Jul. 2023, <https://doi.org/10.1016/j.apnum.2023.04.005>.
- [19] M. Baccouch and H. Temimi, "A high-order space–time ultra-weak discontinuous Galerkin method for the second-order wave equation in one space dimension," *Journal of Computational and Applied Mathematics*, vol. 389, Jun. 2021, Art. no. 113331, <https://doi.org/10.1016/j.cam.2020.113331>.
- [20] M. Baccouch, H. Temimi, and M. Ben-Romdhane, "A discontinuous Galerkin method for systems of stochastic differential equations with applications to population biology, finance, and physics," *Journal of Computational and Applied Mathematics*, vol. 388, May 2021, Art. no. 113297, <https://doi.org/10.1016/j.cam.2020.113297>.
- [21] H. Temimi, S. Adjerid, and M. Ayari, "Implementation of the discontinuous Galerkin method on a multi-story seismically excited building model," *Engineering Letters*, vol. 18, no. 1, Feb. 2010.
- [22] M. Ayari, "On the Efficiency of the Advanced TWA Approach to the 60-GHz Microstrip Antenna Analysis for 5G Wireless Communication Systems," *Engineering, Technology & Applied Science Research*, vol. 13, no. 1, pp. 10151–10157, Feb. 2023, <https://doi.org/10.48084/etasr.5593>.
- [23] M. Ayari, T. Aguil, H. Temimi, and H. Baudrand, "An extended version of transverse wave approach (TWA) for full-wave investigation of planar structures," *Journal of Microwaves and Optoelectronics*, vol. 7, no. 2, pp. 123–138, Dec. 2008.
- [24] M. Ayari, T. Aguil, and H. Baudrand, "More efficiency of transverse wave approach (TWA) by applying anisotropic mesh technique (AMT) for full-wave analysis of microwave planar structures," *Progress In Electromagnetics Research B*, no. 14, pp. 383–405, 2009, <https://doi.org/10.2528/PIERB09022001>.
- [25] M. Ayari, Y. E. Touati, and S. Altowaijri, "Advanced Transverse Wave Approach for MM-Wave Analysis of Planar Antennas applied in 5G-Technology," *International Journal of Computer Science and Network Security*, vol. 22, no. 1, pp. 295–299, Jan. 2022, <https://doi.org/10.22937/IJCSNS.2022.22.1.41>.
- [26] A. K. Gupta, H. C. Mohanta, P. S. R. Chowdary, M. V. Krishna, and H. G. Mohamed, "Design and Analysis of Fractal-Shaped High-Impedance Surface Unit Cell Characteristics," *Fractal and Fractional*, vol. 7, no. 6, Jun. 2023, Art. no. 472, <https://doi.org/10.3390/fractalfract7060472>.
- [27] M. Ayari, T. Aguil, H. Temimi, and H. Baudrand, "On the Use of Anisotropic Mesh Technique (AMT) in Transverse Wave Approach (TWA) for the EM Analysis of Microwave Circuits in Wireless Applications," in *2009 Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing*, Brisbane, QLD, Australia, Jul. 2009, pp. 246–251, <https://doi.org/10.1109/UIC-ATC.2009.8>.
- [28] S. E. El-Khamy, A. Zaki, S. Hamdy, and A. El-Khouly, "A new fractal-like tree structure of circular patch antennas for UWB and 5G multi-band applications," *Microwave and Optical Technology Letters*, vol. 59, no. 9, pp. 2168–2174, 2017, <https://doi.org/10.1002/mop.30707>.
- [29] V. R. Anitha, M. Cho, and J. Shim, "Design of two by three element fractal tree antenna array for WLAN and WiFi applications," *Journal of Communications Technology and Electronics*, vol. 62, no. 1, pp. 61–65, Jan. 2017, <https://doi.org/10.1134/S1064226917010016>.

AUTHORS PROFILE

Mohamed Ayari received Dipl., M.S., and Ph.D. degrees in telecommunications in 2003, 2004, and 2009, respectively, from the National Engineering School of Tunis (ENIT), Tunisia, in collaboration with the National Polytechnic Institute of Toulouse, France, and Virginia Tech, USA. He has been teaching at several universities since 2003. He has been a permanent research member at the 6'COM laboratory at ENIT since 2003. In 2005 he joined RCEM-Inc. in Toulouse, France. Since 2010, he has been a tenure-track associate professor at the National Engineering School of Carthage (ENICAR), Carthage University, Tunisia. In 2015, he joined the IT Department of the Faculty of

Computing and Information Technology, Northern Border University (NBU), Kingdom of Saudi Arabia. Currently, he is an associate professor. His current research interests are electromagnetic (EM) fields, numerical EM methods, and computer-aided design of microwave circuits and antennas. His research interests also include information security, image processing, and wireless applications.

Saleh M. Altowaijri obtained his Ph.D. from Swansea University in the area of cloud computing. Currently he is an associate professor at the Department of Information Systems, Faculty of Computing and Information Technology, Northern Border University, Rafha, Kingdom of Saudi Arabia. He has over 15 years of research experience and has published several book chapters, conference, and journal papers. He is a reviewer of several international conferences and journals. His research interests include grid, cloud, and fog computing, database management systems, data mining, information systems, information technology risk management, IoT, and emerging ICT systems in the healthcare and transportation sectors.

Predicting the Number of Software Faults using Deep Learning

Wahaj Alkaberi

Computer Science and Artificial Intelligence Department, College of Computer Science and Engineering,
University of Jeddah, Saudi Arabia
2100295@uj.edu.sa (corresponding author)

Fatmah Assiri

Software Engineering Department, College of Computer Science and Engineering, University of Jeddah,
Saudi Arabia
fyassiri@uj.edu.sa

Received: 22 December 2023 | Revised: 20 January 2024 | Accepted: 28 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6798>

ABSTRACT

The software testing phase requires considerable time, effort, and cost, particularly when there are many faults. Thus, developers focus on the evolution of Software Fault Prediction (SFP) to predict faulty units in advance, therefore, improving software quality significantly. Forecasting the number of faults in software units can efficiently direct software testing efforts. Previous studies have employed several machine learning models to determine whether a software unit is faulty. In this study, a new, simple deep neural network approach that can adapt to the type of input data was designed, utilizing Convolutional Neural Networks (CNNs) and Multi-Layer Perceptron (MLP), to predict the number of software faults. Twelve open-source software project datasets from the PROMISE repository were used for testing and validation. As data imbalance can negatively impact prediction accuracy, the new version of synthetic minority over-sampling technique (SMOTEND) was used to resolve data imbalance. In experimental results, a lower error rate was obtained for MLP, compared to CNN, reaching 0.195, indicating the accuracy of this prediction model. The proposed approach proved to be effective when compared with two of the best machine learning models in the field of prediction. The code will be available on GitHub.

Keywords-deep learning; MLP; CNN; software testing; prediction; fault; class imbalance

I. INTRODUCTION

Quality attributes are among the most important desired attributes in developed software, and faults have a significant negative impact on software quality. The cost of software projects rises as they progress towards the deployment phase, whereas the presence of faults entails additional costs that can affect the allocated budget and actual time [1]. The field of Software Fault Prediction (SFP) is not limited to software engineering only, it has also contributed to the medical [2-3] and industrial fields [4]. Recently, SFP, a method of designing models to classify software units as faulty [5], has gained considerable attention [6]. The SFP predicts defective software modules/units before the testing phase [7]. Various supervised Machine Learning (ML) models, such as Decision Tree (DT), Neural Networks (NNs), and Support Vector Machines (SVMs) [8-10], as well as unsupervised algorithms, namely K-means++ and QuadTree K-means (QDK) [10], have been used with varying degrees of success. Also, ensemble learning models have been developed and utilized for this purpose [11]. Some studies have applied Principal Component Analysis (PCA) and feature selection techniques to improve the prediction results

[12, 13]. Furthermore, a few researchers have turned to Deep Learning (DL) as it produces a higher accuracy rate for SFP [7, 14-17]. The existing studies focus on classifying software units as faulty or not. However, predicting the number of faults is more useful as it helps developers focus more on software units with a large number of faults, thereby reducing the testing effort [7-14]. Software developers aim to find the best solutions to predict the number of faults in early phases. Many models and techniques have been tested. With the emergence of DL models many experiments have proven the ability of DL to predict software faults.

Given the advantages that Convolutional Neural Networks (CNN) have in terms of their ability to predict and adapt to structure, quality, and number of data, this feature is considered important in the field of fault prediction, because the data that we obtain for prediction vary due to the existing differences in programs and features [18-20]. Furthermore, the source code usually consists of millions of lines and hundreds of files [6], which is a huge amount of data that can be better utilized with DL techniques [21]. CNNs along with MLP models have been effectively applied in the area of prediction [7]. In this study, DL models will be developed using CNNs and Multi-Layer

Perceptron (MLP) to predict the number of faults. The model design of [22] was followed, with modifications producing better results. To further improve the prediction results, SMOTEND technique was applied to solve data imbalance, which introduced majority bias and considered noise. The main contribution of the current study lies in the investigation of the effectiveness of the proposed CNN and MLP design to the field of SFP. Additionally, the acquired results were compared with those of well-known ML models to showcase the effectiveness of DL. As far as is known, only the studies [22-34] have considered DL algorithms to predict the number of faults at the class level.

II. RELATED WORK

Initially, SFP was approached as a classification problem, aiming to predict whether software units were faulty. Researchers focused on ML models in earlier studies. Authors in [9] summarized the related articles published between January 1991 and October 2013 and identified the best techniques and measures for SFP models [9]. The results showed that Random Forest (RF) was the best-used model with 80% accuracy. Authors in [8] compared the performances of three of the most popular supervised ML techniques: MLP, Bayesian network, and Naïve Bayes (NB). The findings indicated that NB achieved the highest accuracy (97%).

In addition to supervised ML techniques, unsupervised ML algorithms have been also used. The author in [23] conducted a comparative study on clustering algorithms, specifically K-means and their variants (K-means++, QDK, and Fuzzy C-means (FCM)). NASA datasets with 29 static code attributes were used, with the QDK algorithm exhibiting the best performance. Research continues to improve classification accuracy. Authors in [10] predicted software faults using ML and source code metrics. They also investigated the effect of the feature selection technique on the prediction performance. In a study conducted employing a NASA project, RF achieved the highest accuracy of 93.7%, while correlation-based feature subset selection (CFS) led to a slight increase in performance, reaching 93.84%. Authors in [24] studied the effect of feature-selection techniques utilizing five classification algorithms: MLP, SVM, k-Nearest Neighbors (kNN), NB, and Logistic Regression (LR), with a telecommunications software system [24]. LR exhibited the best performance, whereas the wrapper-based subset selection technique outperformed the other techniques. Authors in [6] performed a new experiment by combining well-known classification algorithms with PCA, which reduced dataset dimensionality. The datasets were taken from Kaggle using the WEKA simulation tool and the experiment achieved an accuracy of 98.70% implementing the SVM model.

Considering improvements in prediction models, in [25], 28 datasets from the PROMISE repository were utilized to investigate the performance of seven ensemble techniques using three classification algorithms. A total of 532 models were built to demonstrate the effectiveness of ensemble techniques in predicting faults. While most existing research has focused on classifying software units as faulty (buggy) or non-faulty (clear), only a few studies have focused on predicting the number of faults [11, 14, 26, 27]. The DTR model aimed to

predict the number of faults, including intra- and inter-release predictions, along with several software project datasets [14]. DTR yielded good accuracy, with the intra-release prediction showing better results. Additionally, GP and NNs were used to predict the number of faults employing ten software projects from the PROMISE repository. GP achieved better results for large datasets [27]. The quality of the classification model is highly dependent on the quality of the data [28]. Authors in [5] focused on two data-related problems: class overlap and data imbalance. These problems were solved utilizing neighborhood cleaning and random oversampling. After data processing, four models were trained: kNN, NB, DT, and LR. The Synthetic Minority Oversampling Technique (SMOTE) and an ensemble classifier were used whereas NASA datasets were considered and the proposed approach provided improved results. While ML techniques are widely employed for software unit classification, the research community is still experimenting to find better models, with a few researchers utilizing DL [7, 15, 16, 29]. Authors in [29] used DL algorithms, including ANNs, CNNs, Self-Organizing Maps (SOMs), Learning Vector Quantization-3 (LVQ3), and multipass-LVQ (multiLVQ). CNN achieved the best performance, reaching 99.28%. Additionally, dimensionality reduction techniques have also been applied to improve DL algorithms. Authors in [13] applied PCA and Kernel Principal Component Analysis (KPCA) using DT and ANNs, with the ANN algorithm combined with the KPCA yielded the best results. Authors in [22] designed a CNN to predict the number of software faults, utilizing SMOTEND to overcome the problem of imbalanced data. In the most recent considered studies [30, 31], other DL techniques were implemented to predict software faults, namely Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and Radial Basis Function Network (RBFN) and to classify the modules into faulty or non-faulty. LSTM and BiLSTM gave the better performance with 93.53% and 93.75% accuracy, but RBFN was the fastest [30]. Utilizing DL models based on Recurrent Neural Networks (RNNs), the accuracy reached 95% [31].

III. DATASET

The considered dataset consisted of 12 publicly available software projects obtained from the PROMISE repository, which was used for fault prediction studies. These projects include Apache Ant, Camel, Xerces, Xalan, Ivy, Poi, Log4j, Velocity, Lucene, and Synapse, JEdit, and PROP. Table I summarizes the dataset. In this study, the term class is utilized to refer to each module. The complete details of the data can be found in [32]. Each project consisted of the project name, version number, class name, and 20 object-oriented metrics as a set of features [26], as shown in Table II. In total, 83,113 records and 24 columns, including the target column were obtained. A change in the target column (bug) name is referred to as a fault. This study focused on the class level, and the dataset consisted of the number of faults in each class for each project. Figure 1 shows the total number of classes related to the number of faults. For example, there were 70,000 classes without any faults. The number of software faults is centered around zero, indicating an imbalance in the dataset (the number of classes with no faults is greater than the number of classes having at least one fault).

TABLE I. DATASET

Project	Version	Number of classes	Number of faulty classes	Project	Version	Number of classes	Number of faulty classes	Project	Version	Number of classes	Number of faulty classes
Ant	1.3	125	20	Lucene	2.0	195	91	Jedit	3.2	272	90
	1.4	178	40		2.2	246	144		4.0	306	75
	1.5	293	32		2.4	340	203		4.1	312	79
	1.6	351	92		1.5	237	141		4.2	367	48
	1.7	745	166	Poi	2.0	314	37		4.3	492	11
Camel	1.2	609	217		3.0	442	281		3.2	272	90
	1.4	873	145		1.4	196	147	Prop	1	15134	2439
	1.6	966	189	Velocity	1.5	214	142		2	17065	2158
Xalan	2.4	724	111		1.6	229	78		3	7949	1048
	2.5	804	388		1.0	157	16		4	6981	756
Xerces	1.3	453	69	Synapse	1.1	222	60		5	6649	1178
	1.4	588	437		1.2	256	86		6	631	64
Ivy	1.1	111	63		1.0	135	34				
	1.4	241	16	Log4j	1.1	109	37				
	2.0	352	40		1.2	205	189				

TABLE II. OBJECT-ORIENTED SOFTWARE METRICS

Abbreviation	Name	Abbreviation	Name	Abbreviation	Name
WMC	Weighted methods per class	LOC	Lines of code	CE	Efferent couplings
DIT	Depth of inheritance tree	DAM	Data access metric	LCOM3	Lack of cohesion in methods
NOC	Number of children	MOA	Measure of aggregation	IC	Inheritance coupling
CBO	Coupling between object classes	MFA	Measure of functional abstraction	AMC	Average method complexity
RFC	Response for a class	CAM	Cohesion among methods of class	Max_cc	Max McCabe's cyclomatic complexity
LCOM	Lack of cohesion in methods	CBM	Coupling between methods	Avg_cc	Avg McCabe's cyclomatic complexity
CA	Afferent couplings	NPM	Number of public methods	FAULT	Number of faults

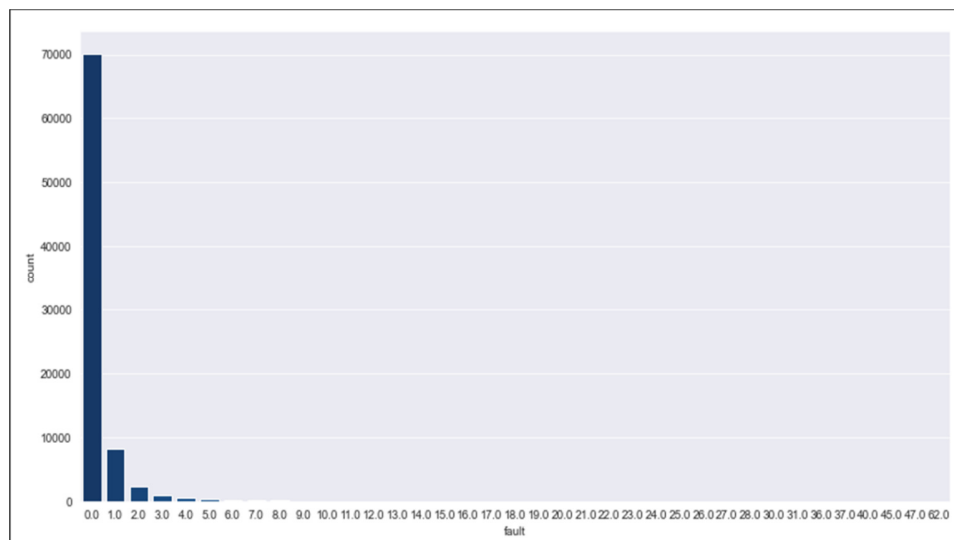


Fig. 1. Number of classes vs number of faults.

A. Data Analysis

During the analysis, it was evident that approximately 37,123 duplicate records were present. Consequently, the data were analyzed to determine the underlying relationships between the features and the dependent variable (fault). First, bivariate analysis was applied to identify possible relationships or distinct patterns between the two features. A commonly used technique for this purpose is the correlation matrix, which

effectively detects the linear relationship (correlation) between two continuous features. Correlation helps identifying important features with respect to the dependent variable and checks for multiple linear relationships among the features. Figure 2 depicts the only the strongly correlated feature. However, there are only a few strong positive correlations. There was a strong positive correlation between the number of WMC and NPM, and the total number of methods for RFC and LCOM indicates that as the number of the former increased, the number of the

latter also increased within each class. From the first relationship it can be concluded that most of the methods in projects are of a public type. Additionally, there was a strong positive correlation among DIT, IC, and MFA, indicating that most of the methods were inherited. Additionally, RFC exhibited a positive correlation with LOC. There are a few

linear relationships between the dependent variable (fault) and the independent variables (features). It was found out that RFC, LOC, and WMC have the strongest linear relationships with the number of faults. These features are important for predicting the number of faults.

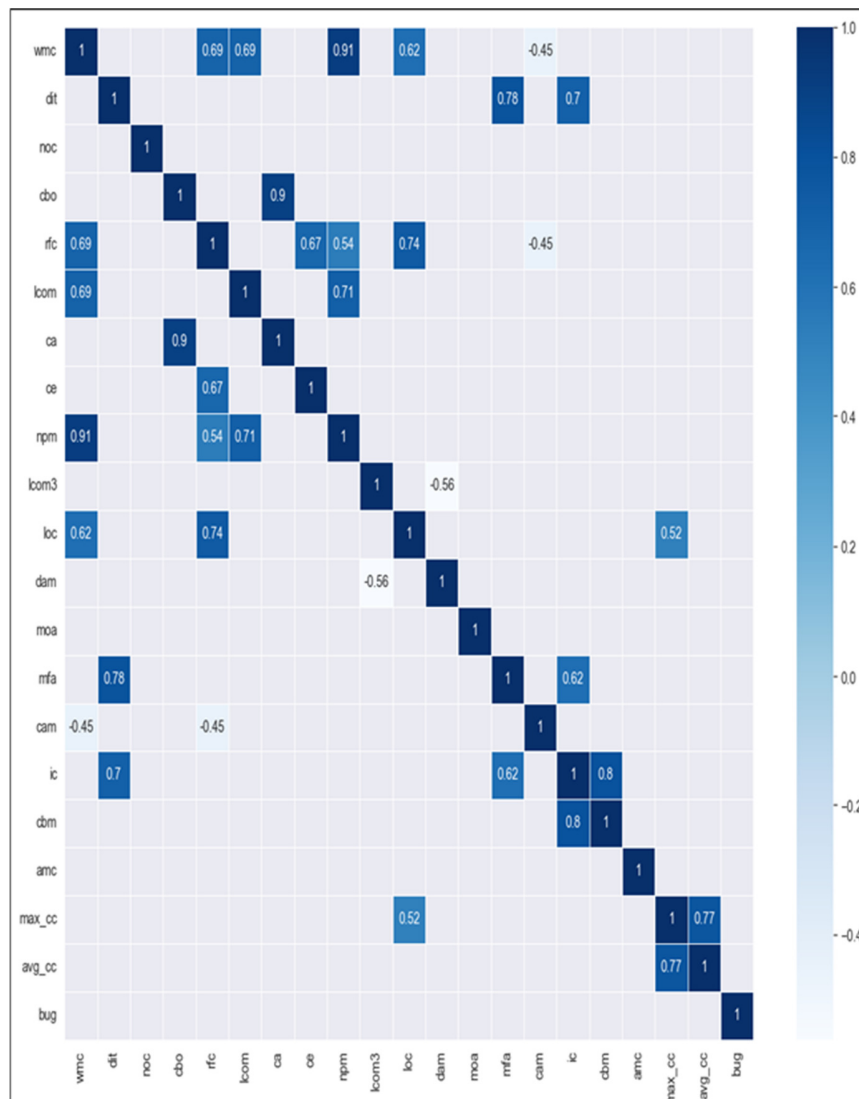


Fig. 2. Correlation matrix, only strongly correlated features.

B. Data Pre-Processing

Data are combined into one file to make them suitable for DL models. In this phase, duplicate records were deleted, resulting in a data size of 45,988. Subsequently, columns that do not affect the target column (fault) were removed, namely the program name (Name), version number (Version), and class name (Name), reducing the total number of columns to 21. The number of faulty records, 11,122, was way less than that of non-faulty records (34,867). Faulty records represented only 24%, while non-faulty records comprised 75% of the total data size. The SMOTEND method was applied to address this problem [11], which increases the number of minority records,

specifically the number of faulty records in this case. Thus, the dataset size was 161,086 records after deleting the duplicate and missing records, which led to a 78% increase in classes containing faults. The distribution of the dependent variable (fault) in the original data following the application of the SMOTEND method is displayed in Figure 3. The distribution was improved slightly by increasing the number of faulty records shown, which balanced the data (Figure 4). Subsequently, natural logarithmic transformation and standardization were performed. A natural logarithm transformation was adopted to transform a highly skewed dataset distribution into a less skewed one [33].

A log transformation on the dependent variable (fault) was applied to improve the distribution of the dependent variable further, as portrayed in Figure 5. Standardization was then performed to enhance the feature distribution by eliminating the bias using (1):

$$z = \frac{(x-\mu)}{\sigma} \quad (1)$$

where x expresses a specific element (feature), μ is the mean of the feature, and σ is the standard deviation of the feature.

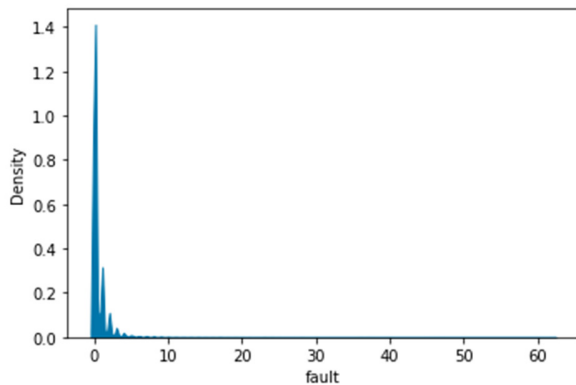


Fig. 3. Distribution of dependent variable, original data.

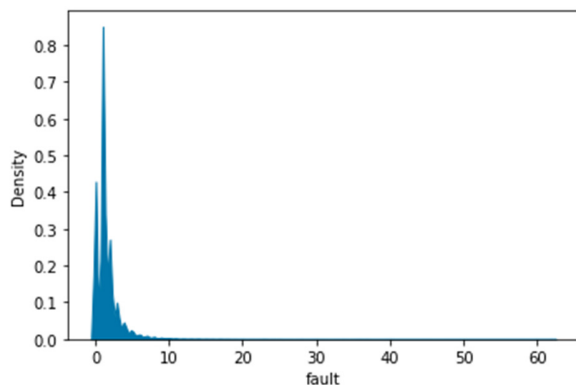


Fig. 4. Distribution of the dependent variable, after applying SMOTEND.

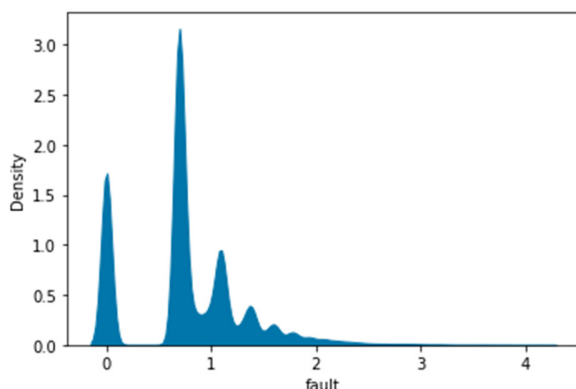


Fig. 5. Distribution of data with a log transformation.

IV. RESEARCH METHODOLOGY

Figure 6 presents the block diagram of the followed process. After collecting and pre-processing the data, the latter were divided into training and testing sets to train the model (70% training and 30% testing). Two DL models, CNN and MLP, were designed to predict faulty units. The results were evaluated using two performance measures, Mean Squared Error (MSE) and Kendall, and the performances of the two models were compared. Finally, the obtained results were compared with the two best ML fault predicting models, which are DTR and SVR.

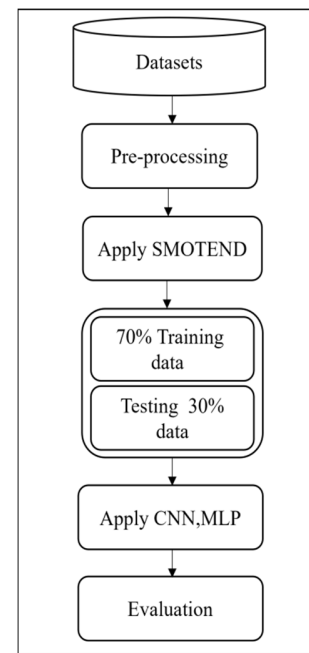


Fig. 6. The overall process.

A. Prediction Model

DL, a subset of ML, is based on ANNs [34]. The structure of an ANN is composed of multiple inputs, outputs, and hidden layers, which makes the learning structure deep. Each layer consists of neurons that process input data and pass the information to the next layer for a specific predictive task. The final output depends on the input data, the choice of activation function, and weight parameters [34]. CNN and MLP are alternative Deep Neural Network (DNN) architectures [34]. To improve prediction accuracy, CNNs were chosen to perform a new experiment based on the architecture developed in [22]. However, MLP has been effectively applied in the area of prediction [7]. Also, it has not been used for predicting the number of software faults.

1) Convolutional Neural Networks

A CNN is a DL technique that has attracted considerable attention for handling high-dimensional data. It engages a mathematical operation called convolution and incorporates a special type of NN consisting of feature extractors that determine weights during the training process [29]. The basic structure of a CNN is composed of three layers: a convolutional layer, a pooling layer, and a fully connected layer. In the

traditional approach, all the features are inserted into one CNN layer and then into the next layer, unlike the way we used.

In this paper, the CNN structure was applied, as described in [22], and subsequent modifications were made to improve the design. The difference was spotted in the division of the features. That is, they were divided according to dimensions, and encapsulation metrics, coupling metrics, abstraction metrics, cohesion metrics, while complexity metrics were considered. There was an effort for the same division to be utilized, but certain groups had a higher number of features than others. For example, the encapsulation metric group contained only one feature, namely the Data Access Metric (DAM), whereas the complexity metric group contained seven features. The resulting model did not learn and did not give satisfactory results.

As a next step, the features were divided randomly into five groups, all groups containing four features equally. Group one

contained (WMC, DIT, NOC, CBO), group two (RFC, LCOM, CA, CE), group three (NPM, LCOM3, LOC, DAM), group four (MOA, MFA, CAM, IC), and group five (CBM, AMC, MAX_CC, AVG_CC). Each feature group was fed into a single CNN layer. The output of the CNN input layer becomes the input to a maxpooling layer. The maxpooling layer was added to merge similar features and reduce dimensionality [34]. Subsequently, dropout layers were added to reduce the possibility of overfitting [7]. Finally, flatten layers are added to convert the resulting matrix into an one-dimensional vector. These layers were repeated five times consecutively, which was equal to the number of feature groups, with a different dataset being entered each time. The flattened layers generated five outputs, which were then merged using the merge layer and fed into the first fully connected layer. The final layer is the second fully connected layer, which produced the predicted number of faults. Figure 7 illustrates the design structure of the CNN model.

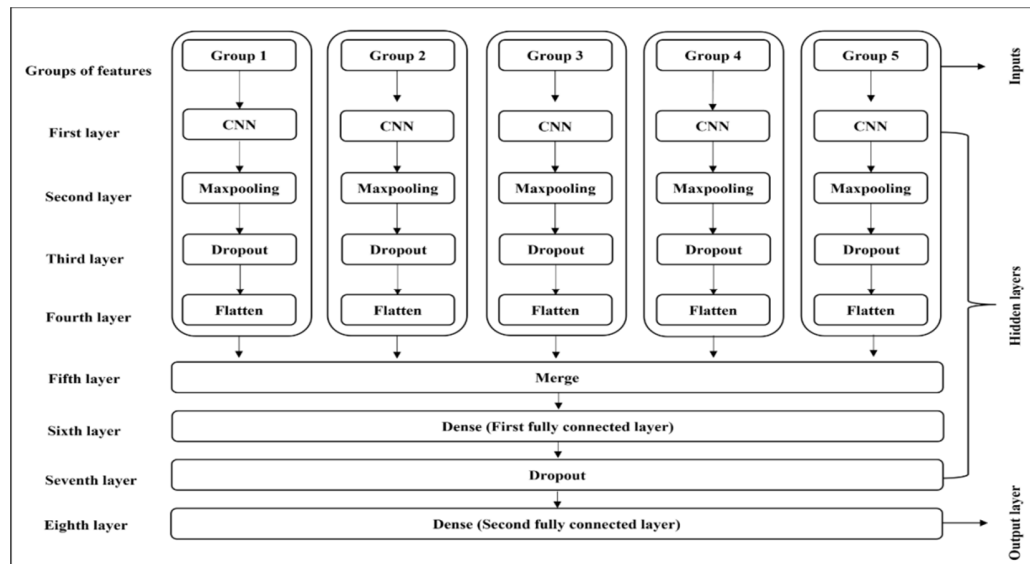


Fig. 7. The structure of the CNN model.

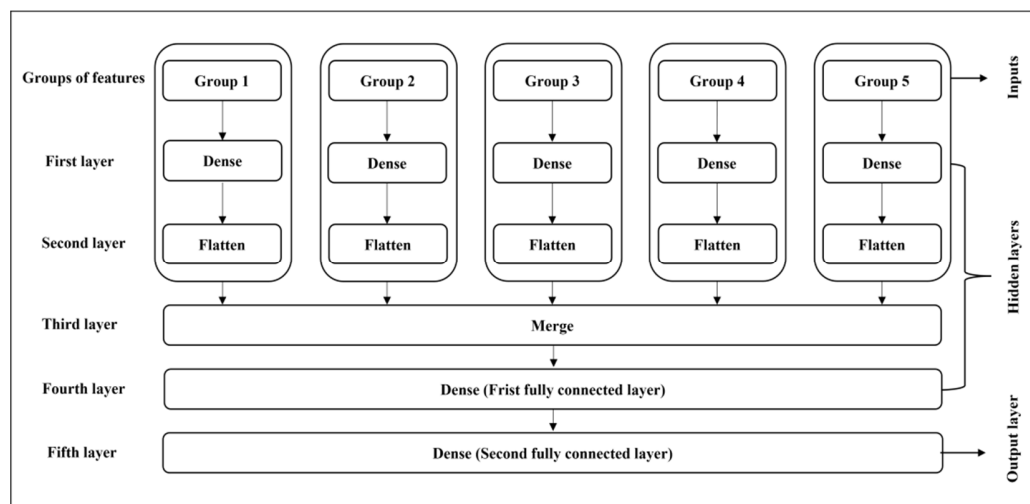


Fig. 8. The structure of the MLP model.

2) Multilayer Perceptron

MLPs are multi-hidden layer NNs composed of a minimum of three layers: an input layer, one or more hidden layers, and an output layer. The input features are weighted and simultaneously fed into the first hidden layer. This layer generates inputs for the next hidden layer, and this process continues until the last layer. The weighted output from the last hidden layer is fed into the output layer, enabling the network to make predictions [7]. In terms of implementation, the deep MLP is the simplest DL model [34]. As a preliminary experiment, this model was built similarly to the previous CNN model. The features were divided into five groups. A dense layer was inserted into each group, followed by a flat layer. Thus, five outputs were obtained. Following that, the merge layer collected the outputs and fed them into the first and second fully connected layers. Figure 8 shows the structure of the MLP model.

V. MODEL EVALUATION

Two different performance metrics were used to evaluate the prediction models: Kendall and MSE. Kendall provides the best performance measure for predicting the number of faults compared to other metrics [11]. On the other hand, the MSE is widely employed and is considered a measure of the quality of a prediction model, with a preference for a smaller value [33]. The Kendall rank correlation coefficient, known as Kendall, is a non-parametric statistic used to test the similarities in data ordering. It has been also used to measure the ordinal association between two measured quantities [11]. In this experiment, the ordinal correlation between the actual and predicted numbers of faults was determined. A higher Kendall coefficient value shows better performance. The Kendall coefficient is given by:

$$\tau = \frac{\# \text{concordant pairs} - \# \text{discordant pairs}}{n(n-1)/2} \quad (2)$$

MSE measures the squared difference between the predicted and the actual values [33]. The formula for calculating MSE is given by:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (|y_i - \hat{y}|^2) \quad (3)$$

where \hat{y} is the predicted value of y .

VI. RESULTS AND DISCUSSION

A. Experimental Settings

The models were developed in Google Colab using Python 3. An Asus PC with Windows operating system was utilized. The total number of records employed in the experiments was 161,086. Two experiments were conducted. First, the prediction models were evaluated before and after using the SMOTEND method. During model training, the epoch and batch size parameters were modified regularly to achieve the best results. The batch size varied between 10 and 20 and the number of epochs between 10 and 200, with the improvement stopping at 100. The CNN and MLP model parameters were defined as: activation = ReLU and kernel_initializer = gloriot_uniform.

B. Experimental Results

The results of Experiments 1 and 2, evaluating both models before and after applying the SMOTEND method, are depicted in Table III. When the data were imbalanced, the MLP performed poorly compared to CNN in both the training and testing phases. However, the results demonstrated further improvement when SMOTEND was used to resolve the imbalance issue, as in the second experiment. In the testing set, the MLP outperformed the CNN, with an enhanced Kendall's value of 0.416 and MSE of 0.195 in the testing phase when utilizing MLP with balanced data, and this proves the effectiveness of MLP in predicting the number of faults. The results were not in accordance with those in [22], where CNN performed well. To ensure the integrity of the model's progression, we need to monitor the validation loss value. This value is adopted to check the performance after each complete epoch and determine if the model requires adjustments. The training and validation losses indicate how effectively the model was trained. The MSE was selected as the loss function due to its compatibility with the data type.

TABLE III. EXPERIMENTS TO TRAIN AND TEST THE IMPACT OF BALANCE AND IMBALANCE DATA WITH CNN AND MLP

Experiment		Experiment 1 (without SMOTEND)		Experiment 2 (with SMOTEND)	
MODEL	METRIC	Kendall	MSE	Kendall	MSE
CNN	Train	0.19	1.776	0.361	0.222
	Test	0.162	1.316	0.363	0.218
MLP	Train	0.186	1.887	0.444	0.185
	Test	0.183	1.73	0.416	0.195

Figure 9 presents the training and validation losses for both models before processing the unbalanced data. It is noticed that the MLP did not learn well with the current data, thus, the SMOTEND method was applied to solve this issue. Figure 10 shows the optimized training and validation losses for both CNN and MLP. The best performing models are displayed in Figure 10(b), where a clear reduction in loss can be seen. In Figure 10(a), the number of batch size and epochs changed many times and no improvement was noticed.

The difference between CNN and MLP is that CNN contains layers that filter features, unlike MLP, which deals with the data directly without any intervention or filtering. Therefore, it was concluded that the data pre-processing is good enough. MLP performed better, contrary to the conclusions reached in [16-29]. The reason is the type of data used (numeric data) and the non-linearity of the problem.

While DL models employ big data to solve complex problems, they have been effective in addressing SFP problems for better predictions. To verify the effectiveness of the proposed DL models for SFP, two state-of-the-art ML models, SVR and DTR, were applied with the results illustrated in Table IV. These models were utilized due to their accuracy in predicting the number of faults in software modules [33]. The obtained results were compared with those of the implemented DL algorithms. The ML models were applied in the same two experiments using the same datasets and features.

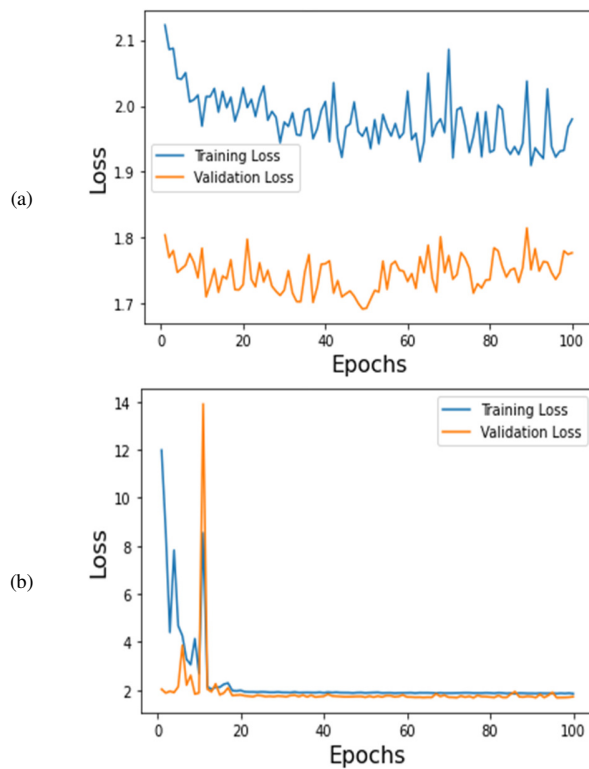


Fig. 9. (a) The CNN and (b) MLP without SMOTEND.

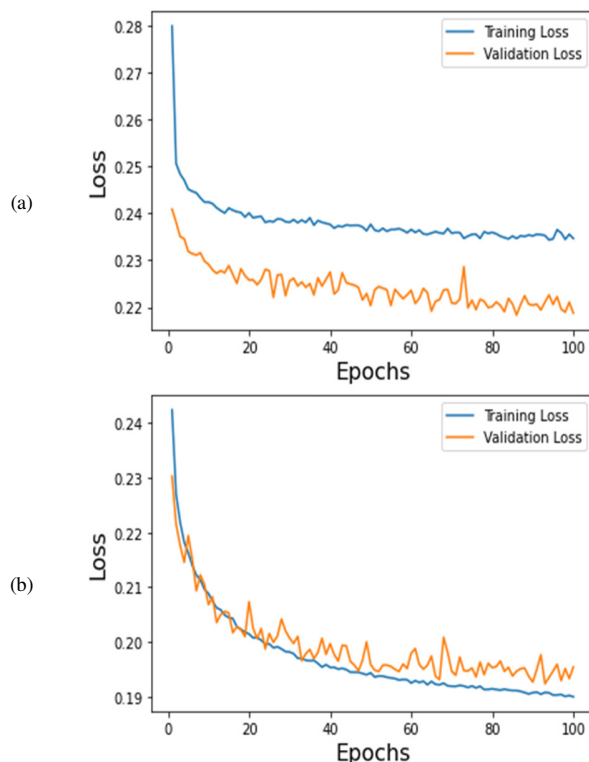


Fig. 10. (a) The CNN and (b) MLP with SMOTEND.

As shown in Table IV, Experiment 1 demonstrated underfitting, and the model failed to learn with the original data. However, the improvement was noticeable as the data were refined and cleaned. The DTR model provided better results than the SVR model, which required a longer time to run and deliver results. Additionally, the SVR did not present the outcomes of Experiment 1 in all training and test groups. Even though DTR gave better results in Experiment 2, with Kendall's coefficient of 0.486 and MSE of 0.170, the model did not train well.

TABLE IV. PERFORMANCE OF THE ML MODELS IN THE TRAINING AND TEST SETS AND THE SAME EXPERIMENTS

Experiment		Experiment 1 (without SMOTEND)		Experiment 2 (with SMOTEND)	
MODEL	METRIC	Kendall	MSE	Kendall	MSE
DTR	Train	0.307	1.692	0.532	0.146
	Test	0.219	1.929	0.486	0.17
SVR	Train	-	-	0.279	0.255
	Test	-	-	0.276	0.257

Figure 11 depicts the training of the DTR model with balanced data. The red color represents the actual data, while the blue color represents the predicted data. Therefore, it cannot be said that DTR gave better performance than MLP, because even though the MLP results were slightly lower, the model learned well.

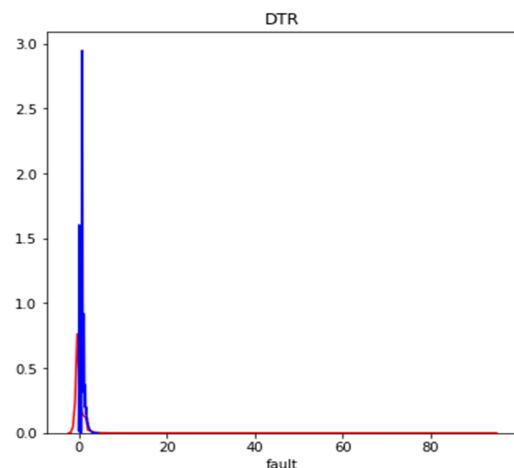


Fig. 11. DTR with SMOTEND.

VII. THREATS TO VALIDITY

This section discusses the external validity of our experiments. External validity relates to the generality of the results. To mitigate this threat, we used 12 publicly available datasets for training and evaluating the proposed model. However, all projects involved in our experiments were written in Java, raising the potential threat of generalizing the results to other programming languages. Furthermore, our chosen datasets posed other threats to external validity, as they consisted of 20 object-oriented metrics. However, other metrics yielded different results.

VIII. CONCLUSION AND FUTURE WORK

SFP has attracted significant research attention due to its ability to decrease the testing cost. While most existing studies focused on the classification of software units as faulty, predicting the number of faults in each software unit is also beneficial. In this study, two deep learning models, CNN and MLP, were designed to predict the number of faults in each software unit. Their performance was evaluated over 12 software project datasets. Furthermore, SMOTEND was applied to solve the data imbalance issue and improve prediction performance, which adversely affects result accuracy.

Based on the findings of this research, the MLP showed better results, with a Kendall value of 0.416 and MSE equal to 0.195, achieving a lower error rate than CNN. In addition, using SMOTEND improved the findings significantly. MLP was better in terms of the results and model performance. The main reason for this is that the problem is non-linear and the data type is numeric, making MLP suitable for facing this issue. Future work plans involve finding methods for obtaining data on fault types and their criticalities to prioritize faults. In addition, software semantics can be utilized to predict faults. Thus, this approach can be extended to techniques that consider code semantics while predicting faulty units.

REFERENCES

- [1] A. Kumar and A. Bansal, "Software Fault Proneness Prediction Using Genetic Based Machine Learning Techniques," in *4th International Conference on Internet of Things: Smart Innovation and Usages*, Ghaziabad, India, Apr. 2019, pp. 1–5, <https://doi.org/10.1109/IoT-SIU.2019.8777494>.
- [2] K. Punitha and B. Latha, "Validation of Medical Imaging Software Using Metaheuristic Knowledge Discovery," *Journal of Medical Imaging and Health Informatics*, vol. 6, no. 8, pp. 1966–1971, Dec. 2016, <https://doi.org/10.1166/jmihi.2016.1958>.
- [3] C. Shyamala and S. A. Sahaaya Arul Mary, "Defect Prediction in Medical Software Using Hybrid Genetic Optimized Support Vector Machines," *Journal of Medical Imaging and Health Informatics*, vol. 6, no. 7, pp. 1600–1604, Nov. 2016, <https://doi.org/10.1166/jmihi.2016.1857>.
- [4] J. Hryszko and L. Madeyski, "Cost Effectiveness of Software Defect Prediction in an Industrial Project," *Foundations of Computing and Decision Sciences*, vol. 43, no. 1, pp. 7–35, Mar. 2018.
- [5] E. Elahi, A. Ayub, and I. Hussain, "Two staged data preprocessing ensemble model for software fault prediction," in *International Bhurban Conference on Applied Sciences and Technologies*, Islamabad, Pakistan, Jan. 2021, pp. 506–511, <https://doi.org/10.1109/IBCAST51254.2021.9393182>.
- [6] C. L. Prabha and N. Shivakumar, "Software Defect Prediction Using Machine Learning Techniques," in *4th International Conference on Trends in Electronics and Informatics*, Tirunelveli, India, Jun. 2020, pp. 728–733, <https://doi.org/10.1109/ICOEI48184.2020.9142909>.
- [7] O. A. Qasem and M. Akour, "Software Fault Prediction Using Deep Learning Algorithms," *International Journal of Open Source Software and Processes*, vol. 10, no. 4, pp. 1–19, Oct. 2019, <https://doi.org/10.4018/IJOSSP.2019100101>.
- [8] S. Dhankhar, H. Rastogi, and M. Kakkar, "Software fault prediction performance in software engineering," in *2nd International Conference on Computing for Sustainable Global Development*, New Delhi, India, Mar. 2015, pp. 228–232.
- [9] R. Malhotra, "A systematic review of machine learning techniques for software fault prediction," *Applied Soft Computing*, vol. 27, pp. 504–518, Feb. 2015, <https://doi.org/10.1016/j.asoc.2014.11.023>.
- [10] G. P. Bhandari and R. Gupta, "Machine learning based software fault prediction utilizing source code metrics," in *3rd International Conference on Computing, Communication and Security*, Kathmandu, Nepal, Oct. 2018, pp. 40–45, <https://doi.org/10.1109/CCCS.2018.8586805>.
- [11] X. Chen, D. Zhang, Y. Zhao, Z. Cui, and C. Ni, "Software defect number prediction: Unsupervised vs supervised methods," *Information and Software Technology*, vol. 106, pp. 161–181, Feb. 2019, <https://doi.org/10.1016/j.infsof.2018.10.003>.
- [12] C. Pan, M. Lu, and B. Xu, "An Empirical Study on Software Defect Prediction Using CodeBERT Model," *Applied Sciences*, vol. 11, no. 11, Jan. 2021, Art. no. 4793, <https://doi.org/10.3390/app11114793>.
- [13] M. Massoudi, N. K. Jain, and P. Bansal, "Software Defect Prediction using Dimensionality Reduction and Deep Learning," in *Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks*, Tirunelveli, India, Feb. 2021, pp. 884–893, <https://doi.org/10.1109/ICICV50876.2021.9388622>.
- [14] S. S. Rathore and S. Kumar, "A Decision Tree Regression based Approach for the Number of Software Faults Prediction," *ACM SIGSOFT Software Engineering Notes*, vol. 41, no. 1, pp. 1–6, Oct. 2016, <https://doi.org/10.1145/2853073.2853083>.
- [15] H. Alsawalqah, H. Faris, I. Aljarah, L. Alnemer, and N. Alhindawi, "Hybrid SMOTE-Ensemble Approach for Software Defect Prediction," in *6th Computer Science On-line Conference*, Prague, Czech Republic, Apr. 2017, pp. 355–366, https://doi.org/10.1007/978-3-319-57141-6_39.
- [16] K. Wongpheng and P. Visutsak, "Software Defect Prediction using Convolutional Neural Network," in *35th International Technical Conference on Circuits/Systems, Computers and Communications*, Nagoya, Japan, Jul. 2020, pp. 240–243.
- [17] A. Hasanpour, P. Farzi, A. Tehrani, and R. Akbari, "Software Defect Prediction Based On Deep Learning Models: Performance Study," arXiv, Apr. 02, 2020, <https://doi.org/10.48550/arXiv.2004.02589>.
- [18] T. Liu, S. Fang, Y. Zhao, P. Wang, and J. Zhang, "Implementation of Training Convolutional Neural Networks," arXiv, Jun. 03, 2015, <https://doi.org/10.48550/arXiv.1506.01195>.
- [19] L. B. Salah and F. Fourati, "Systems Modeling Using Deep Elman Neural Network," *Engineering, Technology & Applied Science Research*, vol. 9, no. 2, pp. 3881–3886, Apr. 2019, <https://doi.org/10.48084/etasr.2455>.
- [20] S. Sahel, M. Alsahafi, M. Alghamdi, and T. Alsubait, "Logo Detection Using Deep Learning with Pretrained CNN Models," *Engineering, Technology & Applied Science Research*, vol. 11, no. 1, pp. 6724–6729, Feb. 2021, <https://doi.org/10.48084/etasr.3919>.
- [21] E. E. Miandoab and F. S. Gharehchopogh, "A Novel Hybrid Algorithm for Software Cost Estimation Based on Cuckoo Optimization and K-Nearest Neighbors Algorithms," *Engineering, Technology & Applied Science Research*, vol. 6, no. 3, pp. 1018–1022, Jun. 2016, <https://doi.org/10.48084/etasr.701>.
- [22] L. Qiao, G. Li, D. Yu, and H. Liu, "Deep Feature Learning to Quantitative Prediction of Software Defects," in *45th Annual Computers, Software, and Applications Conference*, Madrid, Spain, Jul. 2021, pp. 1401–1402, <https://doi.org/10.1109/COMPSAC51774.2021.00204>.
- [23] R. Jothi, "A Comparative Study of Unsupervised Learning Algorithms for Software Fault Prediction," in *Second International Conference on Intelligent Computing and Control Systems*, Madurai, India, Jun. 2018, pp. 741–745, <https://doi.org/10.1109/ICCONS.2018.8663154>.
- [24] H. Wang and T. M. Khoshgoftaar, "A Study on Software Metric Selection for Software Fault Prediction," in *18th IEEE International Conference On Machine Learning And Applications*, Boca Raton, FL, USA, Dec. 2019, pp. 1045–1050, <https://doi.org/10.1109/ICMLA.2019.00176>.
- [25] S. S. Rathore and S. Kumar, "An empirical study of ensemble techniques for software fault prediction," *Applied Intelligence*, vol. 51, no. 6, pp. 3615–3644, Jun. 2021, <https://doi.org/10.1007/s10489-020-01935-6>.
- [26] S. S. Rathore and S. Kumar, "An empirical study of some software fault prediction techniques for the number of faults prediction," *Soft Computing*, vol. 21, no. 24, pp. 7417–7434, Dec. 2017, <https://doi.org/10.1007/s00500-016-2284-x>.
- [27] S. S. Rathore and S. Kuamr, "Comparative analysis of neural network and genetic programming for number of software faults prediction," in *National Conference on Recent Advances in Electronics & Computer*

- Engineering, Roorkee, India, Feb. 2015, pp. 328–332, <https://doi.org/10.1109/RAECE.2015.7510216>.
- [28] A. Agrawal and T. Menzies, "Is 'better data' better than 'better data miners'? on the benefits of tuning SMOTE for defect prediction," in *40th International Conference on Software Engineering*, Gothenburg, Sweden, Jun. 2018, pp. 1050–1061, <https://doi.org/10.1145/3180155.3180197>.
- [29] G. P. Bhandari and R. Gupta, "Measuring the Fault Predictability of Software using Deep Learning Techniques with Software Metrics," in *5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering*, Gorakhpur, India, Nov. 2018, pp. 1–6, <https://doi.org/10.1109/UPCON.2018.8597154>.
- [30] I. Batool and T. A. Khan, "Software fault prediction using deep learning techniques," *Software Quality Journal*, vol. 31, no. 4, pp. 1241–1280, Dec. 2023, <https://doi.org/10.1007/s11219-023-09642-4>.
- [31] E. Borandag, "Software Fault Prediction Using an RNN-Based Deep Learning Approach and Ensemble Machine Learning Techniques," *Applied Sciences*, vol. 13, no. 3, Jan. 2023, Art. no. 1639, <https://doi.org/10.3390/app13031639>.
- [32] M. Jureczko, "Significance of Different Software Metrics in Defect Prediction," *Software Engineering: An International Journal*, vol. 1, no. 1, pp. 86–95, 2011.
- [33] L. Qiao, X. Li, Q. Umer, and P. Guo, "Deep learning based software defect prediction," *Neurocomputing*, vol. 385, pp. 100–110, Apr. 2020, <https://doi.org/10.1016/j.neucom.2019.11.067>.
- J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, no. 1, Mar. 2019, Art. no. 27, <https://doi.org/10.1186/s40537-019-0192-5>.

Sensorless Maximum Power Point Control for Single-stage Grid Connected PV Systems

Mokhtar Abbassi

LISI Laboratory, University of Carthage Tunis, Tunisia
mok98474304@gmail.com

Abdelkarim Aouiti

LISI Laboratory University of Carthage Tunis, Tunisia
abdelkarim.aouiti.ensit@gmail.com

Faouzi Bacha

Electrical Engineering department, University of Tunis, Tunisia
faouzi.bacha@esstt.rnu.tn (corresponding author)

Received: 18 November 2023 | Revised: 3 January 2024 and 8 January 2024 | Accepted: 9 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6642>

ABSTRACT

In this paper, a novel approach for implementing the maximum power point that could be generated from a photovoltaic (PV) panel while eliminating the need for current sensors through the application of the Hill Climbing algorithm is proposed. The active power generated by the PV panel is injected into the grid via a three-phase inverter using voltage-oriented voltage control with Spatial Vector Modulation (SVM). The developed strategy ensures minimal ripples for both active and reactive power and produces a sinusoidal alternating current waveform, even under varying lighting conditions. A comprehensive description of the adopted control strategy is provided and validated through numerical simulations conducted in MATLAB/Simulink environment. Furthermore, the performance of the proposed method is assessed by analyzing the simulation results. In an attempt to validate the effectiveness of the proposed approach, an implementation of the inverter control was conducted with the DSpace 1104 board, and the results underscored the feasibility and effectiveness of the employed approach for grid-connected PV systems.

Keywords-MPPT; photovoltaic; VOC; SVM

I. INTRODUCTION

During the recent decades, significant attention and focus have been directed towards the operation of three-phase bidirectional converters employing Pulse-Width Modulation (PWM). These converters have undergone rapid evolution, and have become attractive solutions for various industrial applications. Their key benefits include bidirectional power flow capabilities, decoupled power control, and the ability to maintain low level harmonic current distortions [1]. Numerous efforts have been undertaken and various control strategies have been employed to achieve low current distortion and stable switching frequencies [1]. Two predominant converter control strategies have emerged, namely the Direct Power Control (DPC) [2-4] and the Voltage-Oriented Control (VOC) [1]. DPC topologies often result in significant switching frequency variation with active and reactive power changes. In an effort to enhance these strategies, Model Predictive Control (MPC) [5] has been employed, effectively reducing harmonic distortions in grid-side currents. However, MPC demands a relatively high sampling frequency to achieve optimal performance, incurring substantial computational costs and,

consequently, requiring expensive DSP or FPGA hardware. These drawbacks are evident in discrete control approaches such as DPC for converter control [5]. The control method relies on a predefined switching table, where selection is made to minimize errors between the output and its reference. This minimization process aims to reduce discrepancies between the control references and the actual instantaneous active and reactive power values in the case of DPC. One of the most mature and widely recognized control strategies for PWM converters is VOC [6-8], where electrical grid currents are decomposed into direct and quadrature components in the dq plane. According to [9], the VOC strategy ensures swift transient response and high static performance through internal current control loops.

Solar energy, owing to its relatively low maintenance and installation costs, has emerged as a promising resource in energy systems [10, 11], and rapidly became a significant component of the energy balance. This technology harnesses the sun as a primary energy source, meaning that the energy supplied by PVs depends on the irradiance and ambient temperature. The PV current-voltage characteristics are

typically represented by nonlinear equations that lack analytical solutions. These numerous climate-related variations necessitate the use of algorithms to track the maximum power generated by the PV generators. The two most commonly used configurations in a grid-connected PV system are the single-stage and the two-stage configurations. The presence of multiple energy conversion stages impacts the overall efficiency, reliability, and cost of the system [12]. Single-stage energy conversion offers several advantages, including basic configuration, high overall efficiency, minimized cost, and compact weight. However, a control method must be developed in extent to ensure that maximum available energy can be efficiently extracted and transmitted from the PV generator to the grid.

To harness the maximum power from PV panels, robust and fast Maximum Power Point Tracking (MPPT) techniques have become of great importance in the operation of all PV systems. Many approaches and techniques have been proposed to track the Maximum Power Point (MPP). Conventional strategies, such as the Perturb and Observe (P&O) algorithm, Incremental Conductance (IC), Hill Climbing (HC) algorithm, and advanced methods like Particle Swarm Optimization (PSO), adaptive control, fuzzy logic controllers, and hybrid methods, combining classical and intelligent algorithms, have been explored. Authors in [13] introduced a P&O method for a converter connected to a PV system, whereas in [14], an IC controller for a two-stage grid-connected PV system was developed. These methods are straightforward and easy to implement. However, the obtained results showed that these approaches exhibit oscillations, which diminish the efficiency and stability of the system under steady-state conditions, in addition to requiring the use of two measuring devices, thereby increasing the costs of the proposed controllers. Methods based on fuzzy logic, sliding mode approaches, predictive techniques, adaptive strategies, and other techniques have garnered the attention of numerous researchers. In [15], the authors developed a neural network algorithm for the Single-Stage Grid-Connected PV (SSGC-PV) system. The functions are heavily depended on experimentation and subsequently exert a significant influence on control performance. Adaptive control [16] and backstepping and terminal sliding mode [10] have also been employed for MPPT control in a PV systems. While these approaches demonstrate improved performance in various weather conditions, their complexity, slow dynamics, and lower efficiency in poorly defined conditions are major challenges. In [17], a predictive method was proposed for MPPT tracking, revealing high efficiency under various disturbance conditions. However, its fundamental issue lies in the computational load, resulting in significantly higher computation requirements and longer processing times.

Optimized MPPT techniques based on metaheuristic algorithms such as PSO [18] and an Adaptive Neuro Fuzzy Inference System (ANFIS) [19] for PV system control have also been introduced. Simulations and experimental outcomes have demonstrated extremely rapid MPPT for sudden changes in irradiance, with minimized steady-state oscillations. However, these methods require highly powerful processors to perform the computations.

This paper introduces a novel sensorless MPPT control strategy based on the line current in the rotating frame. Voltage-oriented control has been achieved through a fixed switching frequency and it has been combined with a space vector modulation. The performance of the proposed VOC-SVM control has been tested under varying irradiance conditions through simulation and experimental tests.

II. PROPOSED SYSTEM CONTROL

A. Inverter Control

The VOC control strategy involves two coordinate system transformations. The first transformation is performed in the fixed (α, β) axis, and the second is carried out in the rotating (d,q) system. The goal of these transformations is to align the current vectors with the voltage vectors. This method provides the decoupling of the current vector components, ensuring a network current with no Total Harmonic Distortion (THD) and the injection of reactive power into the network through the i_q current component. The VOC control comprises two control loops, with one loop focused on the current (internal loop), while the second loop is applied to voltage (external loop). A limitation of this approach is the variable-frequency switching of IGBTs. To overcome this limitation, the application of SVM is used to establish a constant switching frequency [20]. In the subsequent sections of this paper, it is assumed that the filter's resistance is negligible. In the case of a balanced three-phase system without a neutral conductor, the equation governing the voltages and currents of the controlled converter are expressed by:

$$\begin{bmatrix} L \frac{di_a}{dt} \\ L \frac{di_b}{dt} \\ L \frac{di_c}{dt} \end{bmatrix} = \begin{bmatrix} -R_g & 0 & 0 \\ 0 & -R_g & 0 \\ 0 & 0 & -R_g \end{bmatrix} \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} + \begin{bmatrix} v_{ga} - v_a \\ v_{gb} - v_b \\ v_{gc} - v_c \end{bmatrix} \quad (1)$$

The three-phase quantities in a balanced system in abc coordinates, defined as $u_{a,b,c}$, can also be transformed into the stationary $\alpha\beta$ coordinates using the transformation matrix provided by (2)

$$\begin{bmatrix} y_\alpha \\ y_\beta \end{bmatrix} = \frac{2}{3} \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} y_a \\ y_b \\ y_c \end{bmatrix} \quad (2)$$

The transformation of (2) from $\alpha\beta$ to dq is given by:

$$\begin{bmatrix} y_d \\ y_q \end{bmatrix} = \begin{bmatrix} \cos \omega t & \sin \omega t \\ -\sin \omega t & \cos \omega t \end{bmatrix} \begin{bmatrix} y_\alpha \\ y_\beta \end{bmatrix} \quad (3)$$

In Figure 1, the voltage-oriented control diagram based on virtual flux is depicted, illustrating the coordinate system transformations ($\alpha\beta$ to dq) and the application of SVM to maintain a constant switching frequency.

The transformation of voltage equations into the rotating (dq) Park coordinate system leads to the following system of equations:

$$\begin{bmatrix} c_{gd} \\ c_{gq} \end{bmatrix} = \begin{cases} L \frac{di_d}{dt} + \omega L i_q + v_{cd} \\ L \frac{di_q}{dt} + \omega L i_d + v_{cq} \end{cases} \quad (4)$$

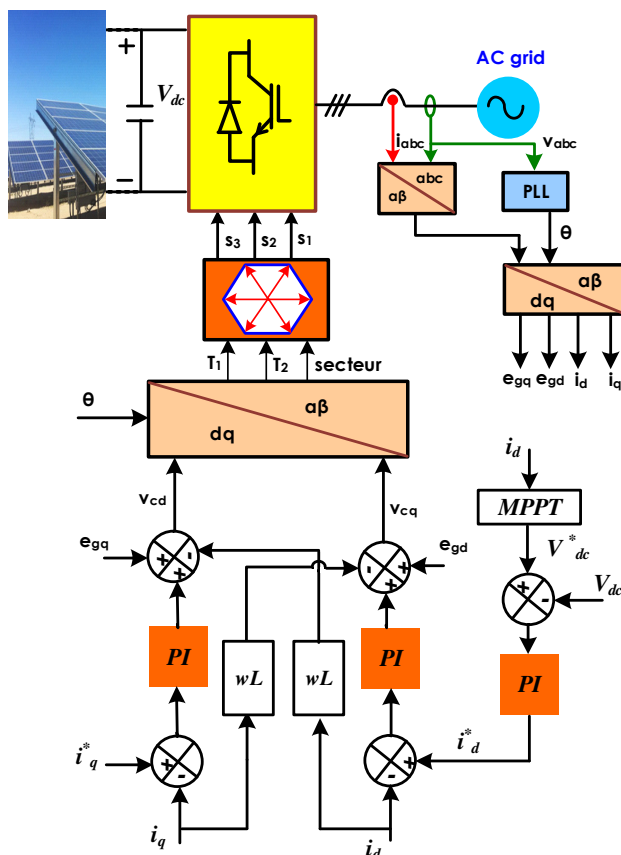
$$\begin{bmatrix} p_{gd} \\ q_{gq} \end{bmatrix} = \begin{cases} \frac{3}{2}(v_{gd}i_{gd} - v_{gq}i_{gq}) \\ \frac{3}{2}(v_{gd}i_{gq} - v_{gq}i_{gd}) \end{cases} \quad (5)$$


Fig. 1. Voltage-oriented control diagram based on virtual flux.

B. MPPT Control

The Hill Climbing algorithm is employed to adjust the i_d current, shifting the operating point along the PV generator's characteristic curve, within a predefined maximum limit. The search process persists until the maximum current point is reached or closely approached. This approach has gained popularity among manufacturers for its simplicity and avoidance of complex calculations. Additionally, it enables the switch of the measurement sensor from the DC current to the line current in the rotating frame (i_d), resulting in benefits such as simplicity, speed, reliability, and cost reduction. Figure 2 displays the proposed MPPT algorithm control.

III. DISCUSSION OF EXPERIMENTAL AND SIMULATION RESULTS

Experimental validation of the novel approach and the performance of the proposed control were tested utilizing the

DSpace 1104 board. Simulation analysis was conducted in MATLAB/Simulink for the single-stage grid-connected PV generator under various irradiance variations. The simulation block diagram of the control block is portrayed in Figure 3. It is mainly composed of the MPPT algorithm block that guarantees maximum power tracking and generates the corresponding reference voltage from the measured current i_d , from the external control to provide the reference current in the d-axis, and from two internal current control blocks to have the inverter control voltages V_d and V_q . Fixed frequency control is carried out by the SVM block. The currents and voltages in the dq-axis and the locking loop are obtained by transformation blocks from the voltages and currents measured at the electrical grid. Tables I and II provide the simulation and practical validation parameters.

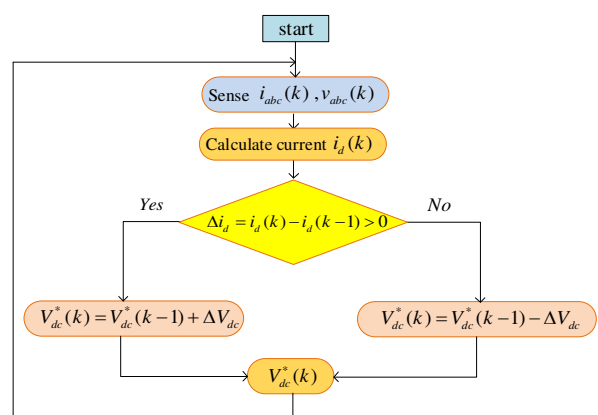


Fig. 2. The proposed MPPT algorithm.

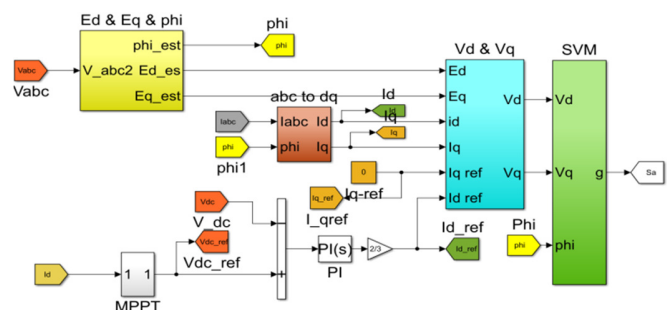


Fig. 3. Simulink model of the proposed control approach.

TABLE I. SIMULATION AND PRACTICAL VALIDATION PARAMETERS

Designation	Simulation	Experimentation
RMS Line to Line voltage	220 V - 50 Hz	220 V - 50 Hz
Line inductors L	12 mH	12 mH
DC bus Capacitor: C	1100 μ F	1100 μ F
K_p (outer loop)	0.2	0.2
K_i (outer loop)	10 s^{-1}	10 s^{-1}
K_{pi} (inner loop)	0.2	0.2
K_{ii} (inner loop)	10 s^{-1}	10 s^{-1}
K_{pi} (inner loop)	0.2	0.2
K_{ii} (inner loop)	10 s^{-1}	10 s^{-1}
Sampling time T_s / Ode	10 μ s (Fixed step)	70 μ s (Fixed step)
DC bus voltage	325 V	325 V
Switching frequency	5 KHz	5 KHz

TABLE II. PV PANEL PARAMETERS

Designations	Values
Module type	ENN solar Energy EST-110
Maximum power generated by PV	109.2 W
PV voltage V_{oc}	138 V

The initial scenario aims to evaluate the performance of the MPPT algorithm control strategy within a single-stage grid-connected PV system. This section encompasses a comprehensive array of practical and numerical simulation results for the monitoring and tracking of the MPP of the PV generator and the power injected into the grid. Figure 4 shows the irradiance scenario, while Figure 5 depicts the results.

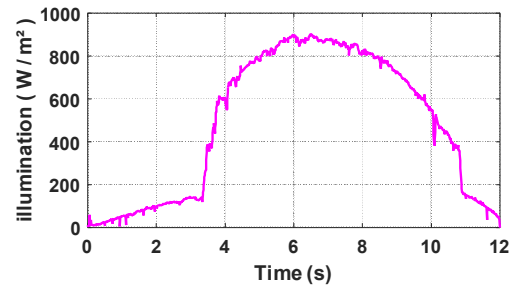


Fig. 4. Irradiance scenario.

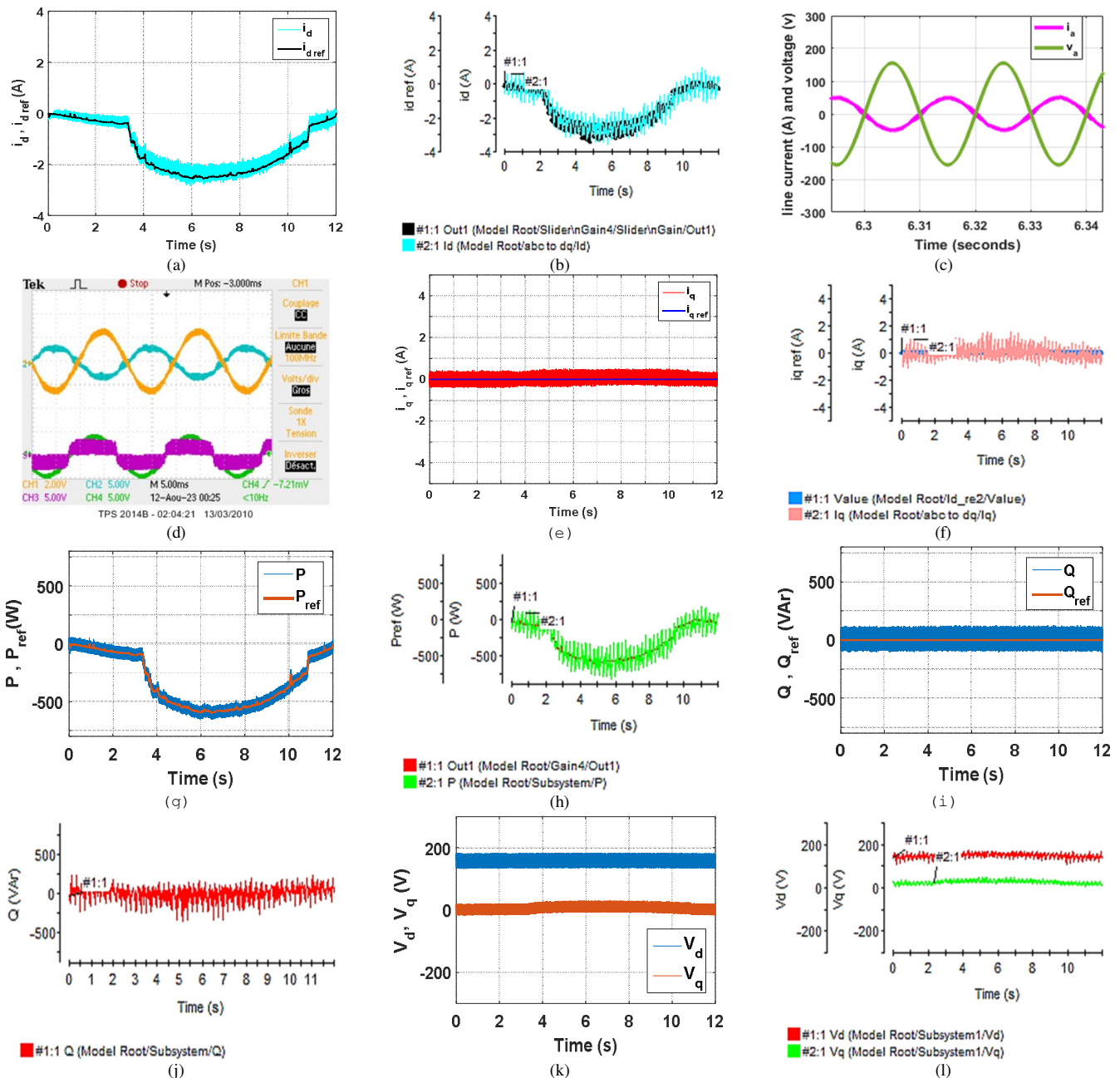


Fig. 5. Simulation and experimental results.

Figures 5 (a) and (b) showcase the inner current loop performance for the d-axis current. The i_d current accurately follows its reference, which is generated by the PV panels, whether in the simulation results or in the practical validation. In Figures 5 (c) and (d), it can be observed that the current injected into the electrical grid exhibits a sinusoidal form with minimal distortion. The internal controller is responsible for maintaining the q-axis current at an average value of zero during variations in the irradiance, as depicted in the simulation and experimental result illustrated in Figures 5 (e) and (f). Figures 5 (g) and (h) depict the variations in the active power injected into the grid for simulation and experimental tests, respectively. It can be seen that the injected power closely tracks its reference value with negligible error across the entire range of irradiance variations. This strongly implies that the Hill Climbing algorithm is very effective and underlines the sensorless MPPT control system's rapid response. The reactive power response can be evaluated through Figures 5 (i) and (j), where the defined values of reactive power remain close to their reference. This implies that the PV system generates an electrical energy with an important power factor. Figures 5 (k) and (l) evidently showcase that the SVM control parameters are pristine. In fact, V_q maintains a null value and exhibits no harmonics, resulting in a notably efficient control of the system. Table III offers a comparison between the method proposed in [10] and the approach employed in this paper.

TABLE III. COMPARISON TABLE

Evaluated parameters	Proposed method	[10]
Convergence ability to the MPP	Medium	High
Efficiency of MPPT (%)	95 to 99.50	99.33
Control variable	Voltage	Duty cycle
Required sensors	Estimated	1 current sensor 1 Voltage sensor
Control strategy	Sampling method	Intelligent control based on FLC
Algorithm complexity	Low	High
Overall system performance (%)	96	93

IV. CONCLUSION

This paper describes a voltage-oriented control of a three-phase grid-connected photovoltaic inverter, based on a sensorless maximum power point tracking strategy. The main objective of the control system is to obtain the maximum power from the photovoltaic panel and to regulate the bus voltage to the desired reference value, employing a MPPT algorithm, while guaranteeing sinusoidal current injection into the network for active and reactive power injection scenarios. The newly introduced strategies were synthesized and evaluated via the Matlab environment and the DSpace 1104 platform. Moreover, the implemented control system exhibits excellent performance in stable and transient scenarios at low cost. In conclusion, the proposed control method aims to decrease the number of sensors, improve system efficiency, reduce cost, and lighten the weight of the system while minimizing CPU-level computation.

REFERENCES

[1] J. Y. Lee, R. Verayiah, K. H. Ong, A. Ramasamy, and M. B. Marsadek, "Voltage Oriented Control and Direct Power Control Strategies in

solving under and overvoltage conditions for heavy load application on Malaysian Distribution Representative Network," *Electrical Engineering*, vol. 103, no. 3, pp. 1597–1612, Jun. 2021, <https://doi.org/10.1007/s00202-020-01143-y>.

- [2] B. Bouaziz and F. Bacha, "Direct power control of grid-connected converters using sliding mode controller," in *2013 International Conference on Electrical Engineering and Software Applications*, Hammamet, Tunisia, Mar. 2013, pp. 1–6, <https://doi.org/10.1109/ICEESA.2013.6578497>.
- [3] K. Sayahi, A. Kadri, F. Bacha, and H. Marzougui, "Implementation of a D-STATCOM control strategy based on direct power control method for grid connected wind turbine," *International Journal of Electrical Power & Energy Systems*, vol. 121, Oct. 2020, Art. no. 106105, <https://doi.org/10.1016/j.ijepes.2020.106105>.
- [4] F. Tlili, A. Kadri, and F. Bacha, "Advanced control strategy for bidirectional three phase AC/DC converter," *Electric Power Systems Research*, vol. 179, Feb. 2020, Art. no. 106078, <https://doi.org/10.1016/j.epsr.2019.106078>.
- [5] M. Easley, M. B. Shadmand, and H. Abu-Rub, "Hierarchical Model Predictive Control of Grid-Connected Cascaded Multilevel Inverter," *Journal of Emerging and Selected Topics in Power Electronics*, vol. 9, no. 3, pp. 3137–3149, Jun. 2021, <https://doi.org/10.1109/JESTPE.2020.3015128>.
- [6] A. Soyed, A. Kadri, O. Hasnaoui, and F. Bacha, "Direct Power and Voltage Oriented Control Strategies of Grid-Connected Wind Energy Conversion System Based on Permanent Magnet Synchronous Generator," *Cybernetics and Systems*, vol. 53, no. 1, pp. 103–125, Jan. 2022, <https://doi.org/10.1080/01969722.2021.2008684>.
- [7] M. Baazouzi, K. Sayahi, and F. Bacha, "Control strategy used for AC fault ride-through in VSC-HVDC transmission systems," *IEEE Transactions of the Institute of Measurement and Control*, Jun. 2023, Art. no. 01423312231180546, <https://doi.org/10.1177/01423312231180546>.
- [8] M. N. Tasnim *et al.*, "Voltage-Oriented Control-Based Three-Phase, Three-Leg Bidirectional AC–DC Converter with Improved Power Quality for Microgrids," *Energies*, vol. 16, no. 17, Jan. 2023, Art. no. 6188, <https://doi.org/10.3390/en16176188>.
- [9] R. Kadri, J. -P. Gaubert, and G. Champenois, "An Improved Maximum Power Point Tracking for Photovoltaic Grid-Connected Inverter Based on Voltage-Oriented Control," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 1, pp. 66–75, Jan. 2011, <https://doi.org/10.1109/TIE.2010.2044733>.
- [10] F. Z. Kebbab, L. Sabah, and H. Nouri, "A Comparative Analysis of MPPT Techniques for Grid Connected PVs," *Engineering, Technology & Applied Science Research*, vol. 12, no. 2, pp. 8228–8235, Apr. 2022, <https://doi.org/10.48084/etasr.4704>.
- [11] K. Behih and H. Attoui, "Backstepping Terminal Sliding Mode MPPT Controller for Photovoltaic Systems," *Engineering, Technology & Applied Science Research*, vol. 11, no. 2, pp. 7060–7067, Apr. 2021, <https://doi.org/10.48084/etasr.4101>.
- [12] W. Libo, Z. Zhengming, and L. Jianzheng, "A Single-Stage Three-Phase Grid-Connected Photovoltaic System With Modified MPPT Method and Reactive Power Compensation," *IEEE Transactions on Energy Conversion*, vol. 22, no. 4, pp. 881–886, Sep. 2007, <https://doi.org/10.1109/TEC.2007.895461>.
- [13] A. A. H. AlZubaidi, L. A. Khaliq, H. S. Hamad, W. K. Al-Azzawi, M. S. Jabbar, and T. A. Shihab, "MPPT implementation and simulation using developed P&O algorithm for photovoltaic system concerning efficiency," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 5, pp. 2460–2470, Oct. 2022, <https://doi.org/10.11591/eei.v11i5.3949>.
- [14] R. I. Putri, S. Wibowo, and M. Rifa'i, "Maximum Power Point Tracking for Photovoltaic Using Incremental Conductance Method," *Energy Procedia*, vol. 68, pp. 22–30, Apr. 2015, <https://doi.org/10.1016/j.egypro.2015.03.228>.
- [15] L. Bouselham, M. Hajji, B. Hajji, and H. Bouali, "A New MPPT-based ANN for Photovoltaic System under Partial Shading Conditions," *Energy Procedia*, vol. 111, pp. 924–933, Mar. 2017, <https://doi.org/10.1016/j.egypro.2017.03.255>.

- [16] S. K. Kollimalla and M. K. Mishra, "A Novel Adaptive P&O MPPT Algorithm Considering Sudden Changes in the Irradiance," *IEEE Transactions on Energy Conversion*, vol. 29, no. 3, pp. 602-610, Sep. 2014, <https://doi.org/10.1109/TEC.2014.2320930>.
- [17] A. Lashab, D. Sera, J. M. Guerrero, L. Mathe and A. Bouzid, "Discrete Model-Predictive-Control-Based Maximum Power Point Tracking for PV Systems: Overview and Evaluation," *IEEE Transactions on Power Electronics*, vol. 33, no. 8, pp. 7273-7287, Aug. 2018, <https://doi.org/10.1109/TPEL.2017.2764321>.
- [18] V. V. Prabhakaran and A. Singh, "Enhancing Power Quality in PV-SOFC Microgrids Using Improved Particle Swarm Optimization," *Engineering, Technology & Applied Science Research*, vol. 9, no. 5, pp. 4616-4622, Oct. 2019, <https://doi.org/10.48084/etasr.2963>.
- [19] A. Baraeen, M. Kassas, M. S. Alam, and M. A. Abido, "Hybrid Neural Network and Adaptive Terminal Sliding Mode MPPT Controller for Partially Shaded Standalone PV Systems," *Arabian Journal for Science and Engineering*, vol. 48, no. 11, pp. 15527-15539, Nov. 2023, <https://doi.org/10.1007/s13369-023-08179-9>.
- [20] S. Ouchen, M. Benbouzid, F. Blaabjerg, A. Betka, and H. Steinhart, "Direct Power Control of Shunt Active Power Filter Using Space Vector Modulation Based on Supertwisting Sliding Mode Control," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 9, no. 3, pp. 3243-3253, Jun. 2021, <https://doi.org/10.1109/JESTPE.2020.3007900>.

Residual Attention Augmentation Graph Neural Network for Improved Node Classification

Muhammad Affan Abbas

Department of Electrical and Information Engineering, Control Science and Engineering, Tianjin University, China
affan_abbas8@yahoo.com

Waqar Ali

Department of Environmental Sciences, Informatics, and Statistics, Ca' Foscari University of Venice, Italy
waqar.ali@unive.it

Florentin Smarandache

Mathematics, Physics, and Natural Science Division, University of New Mexico, USA
smarand@unm.edu

Sultan S. Alshamrani

Department of Information Technology, College of Computer and Information Technology, Taif University, Saudi Arabia
susamash@tu.edu.sa

Muhammad Ahsan Raza

Department of Information Sciences, University of Education Lahore, Multan Campus, Pakistan
ahsan.raza@ue.edu.pk

Abdullah Alshehri

Department of Information Technology, Faculty of Computing and Information, Al-Baha University, Saudi Arabia
aashehri@bu.edu.sa

Mubashir Ali

Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China
mubashir@siat.ca.cn (corresponding author)

Received: 31 December 2023 | Revised: 23 January 2024 | Accepted: 24 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6844>

ABSTRACT

Graph Neural Networks (GNNs) have emerged as a powerful tool for node representation learning within graph structures. However, designing a robust GNN architecture for node classification remains a challenge. This study introduces an efficient and straightforward Residual Attention Augmentation GNN (RAA-GNN) model, which incorporates an attention mechanism with skip connections to discerningly weigh node features and overcome the over-smoothing problem of GNNs. Additionally, a novel MixUp data augmentation method was developed to improve model training. The proposed approach was rigorously evaluated on various node classification benchmarks, encompassing both social and citation networks. The proposed method outperformed state-of-the-art techniques by achieving up to 1% accuracy improvement. Furthermore, when applied to the novel Twitch social network dataset, the proposed model yielded remarkably promising results. These findings provide valuable insights for researchers and practitioners working with graph-structured data.

Keywords-graph neural networks; node classification; over-smoothing; citation and social networks; mixup data augmentation

I. INTRODUCTION

Graphs are used in many different disciplines, such as social networks, biological systems, and recommendation engines, to depict intricate relationships and structures. To properly interpret and utilize graph data, it is critical to learn meaningful node representations within these complex network topologies. In light of this, Graph Neural Networks (GNNs) have become a powerful paradigm that presents a hopeful resolution to this problem [1]. GNNs facilitate efficient node classification, graph classification, and link prediction, among other tasks, by encoding both the local and global graph structure [2]. However, creating effective GNN architectures that meet the unique requirements of node classification is still a challenging issue. Numerous GNN variations have been proposed, each with a unique architectural design and components, creating a vast array of alternatives [3]. The ongoing research for the most effective and efficient GNN architectures that can function effectively on a range of real-world graph data is highlighted by this diversity.

This study aims to offer a thorough and workable approach for enhancing GNN performance in the context of node classification in light of these difficulties. The proposed method aims to improve the capabilities of GNN designs while making them simpler, drawing inspiration from recent developments in the field. This approach combines graph convolutional layers with fully connected layers in a simplified architectural layout. It uses Attention Mechanism (AM) [4] and Data Augmentation (DA) strategies, namely MixUp [5], to further enhance the performance of GNNs in node classification. Strategically incorporated into the GNN architecture, these strategies help improve the generalization of the GNN model. Skip Connections (SCs) are used to reduce the accuracy loss caused by over-smoothing. Extensive tests were performed on various node classification tasks to thoroughly evaluate the performance of the proposed strategies on known benchmarks, such as social networks [6] and citation networks [7]. Concisely, this study:

- Developed a Residual Attention Augmentation Graph Neural Network (RAA-GNN) to enhance the evaluation of the node classification task.
- Developed a novel DA method, called MixUp DA, which combines labels and node attributes to produce synthetic data points and improve the model's ability to classify nodes. Additionally, well-designed skip connections and an effective multi-head attention technique were introduced to improve information aggregation and over-smoothing issues, which together improve GNN performance for node classification.
- Evaluated the proposed method on the Twitch social network dataset, and the results showed up to a 1% gain in accuracy, providing further insights for graph-structured data applications.

II. RELATED WORKS

Several studies have investigated SCs, AMs, and DA in the context of graph-based machine learning. Although DA has been beneficial in enhancing model performance in several fields, its implementation in graph-based machine learning has encountered difficulties. Conventional augmentation methods for graph data, including noise addition or perturbing node properties [8], frequently fail because they break the natural graph structure [9]. Furthermore, the addition of synthetic noise can impede the learning and generalization of the model. The proposed MixUp DA strategy [10] provides a logical method to enhance graph data by seamlessly combining node attributes and labels.

Attention mechanisms have revolutionized information aggregation in GNNs by allowing nodes to choose to attend to the relevant neighbors [11]. However, problems with scalability and processing complexity may make them less successful. Current methods are frequently computationally intensive, and therefore, they are unfeasible for large-scale graphs. Currently, the SuperHyperGraph presents the most general form of graph [11]. These issues are addressed and make it easier to apply attention methods to larger graphs by introducing a multi-head AM that strikes a compromise between expressive capacity and computational efficiency. SCs are important in deep learning architectures because they facilitate the transfer of information between layers [12]. Applying SCs in GNNs has proven difficult, despite their usefulness. Their poor integration can cause over-smoothing, reducing classification accuracy by making nodes indistinguishable through excessive information exchange. This study introduces SCs into the GNN design to mitigate the effects of over-smoothing [13], resulting in improved performance without sacrificing accuracy. Consequently, SCs, AM, and DA [14-15] have all been crucial in the advancement of graph-based machine learning. This study addresses these systems' drawbacks by providing a computationally efficient multi-head AM, a more principled approach to DA, and a method for preventing over-smoothing using SCs. Together, these developments enhance node classification in GNNs and enable a greater variety of complicated, real-world graph data to be used in GNNs.

III. METHODOLOGY

Figure 1 shows the architecture of the proposed RAA-GNN model. In the first step, the MixUp augmentation strategy employs a feature-label augmentation method to increase the robustness of the training dataset. Then, the AM is used, which permits the adaptive weighting of pertinent neighbors, boosting the model's capacity to identify significant local structures and raising classification accuracy all around. Following this, SCs are used to solve the over-smoothing problem of GNNs for node classification.

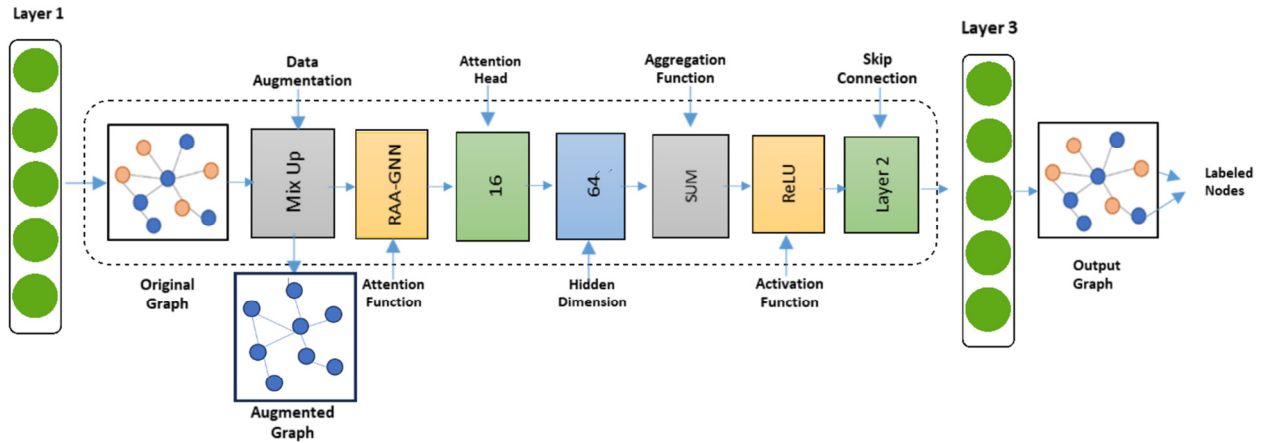


Fig. 1. The RAA-GNN architecture integrates SCs, AMs, and DA for improved node classification.

A. Node Augmentation MixUp Method

GNNs can be designed with MixUp augmentation as a practical and efficient approach to improve node classification performance. The proposed method is based on meticulous preprocessing of input data represented by X , which guarantees consistency and standardization. MixUp augmentation serves as a dynamic catalyst by carefully combining the training dataset to add robustness and diversity. The amount of augmentation is dynamically influenced by the mixing parameter λ . This coefficient highlights the controlled variability included during training and is randomly generated from a beta distribution with parameters:

$$X_{mix} = \lambda X + (1 - \lambda)X_{shuffled} \quad (1)$$

MixUp works in unison with the larger GNN design, where SCs and attention processes are essential building blocks to improve the model's comprehension of complex graph structures. The following mathematical formulas capture MixUp's effect on the data and explain how original and shuffled features and labels are well combined, which adds to the model's flexibility:

Mixed Data =

$$\lambda \cdot \text{Batch Data} + (1 - \lambda) \cdot \text{Batch Data}[\text{Indices}] \quad (2)$$

Mixed Labels =

$$\lambda \cdot \text{Batch Labels} + (1 - \lambda) \cdot \text{Batch Labels}[\text{Indices}] \quad (3)$$

Integrating MixUp into the GNN model promotes a more robust and flexible learning process for node categorization tasks. The proposed GNN architecture is at the forefront of node classification research because of this deliberate augmentation, which also strengthens the model's ability to generalize across a variety of graph configurations and enrich the training dataset.

B. Attention Mechanism (AM) for Node Classification

RAA-GNN is used to represent the attention function, which is essential to the model's ability to concentrate on the most pertinent data inside the graph structure. With sixteen attention heads, the model captures many structural details and complex relationships, leading to a thorough comprehension.

The ReLU activation function is used to combine the contributions of each attention head, represented as $\text{RAA-GNN}_i(X, A)$, to get the final attention scores. Each attention head makes a unique contribution to the overall AM. The adjacency matrix is represented by A . Intricate graph patterns are captured by 64 hidden dimensions, which balance model expressiveness and efficiency. The aggregation function combines data from nearby nodes by applying the summation:

$$H_{agg} = \sum_N H_N \quad (4)$$

where N is for the neighbor. By introducing non-linearity, ReLU activation improves the model's capacity to learn intricate relationships:

$$H_{activated} = \max(0, H_{agg}) \quad (5)$$

C. Skip Connections (SCs)

The SCs, denoted by H_{skip} , enable the smooth transfer of data between the model's layers. These SCs serve to bridge the gap between subsequent layers by integrating the activated features ($H_{activated}$) with the features of the preceding layer ($H_{previous}$), therefore facilitating the transfer and retention of crucial information. By ensuring that important information from previous layers is merged, this additive process improves the model's ability to represent both local and global interdependence.

$$H_{skip} = H_a + H_p \quad (6)$$

where a is the activated and p is the previous. The model's three layers improve node classification by capturing hierarchical representations:

$$H_{output} = \text{GNN}(H_{skip}^{(3)}) \quad (7)$$

The output graph, which displays node classifications based on learned features, is generated by the last layer:

$$Y_{pred} = \text{Softmax}(H_{output}) \quad (8)$$

This architecture provides a basis for strong node classification in a variety of datasets by utilizing cutting-edge methods to address graph-based learning difficulties.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Extensive simulations were carried out to evaluate the performance of the proposed GNN with the novel features of DA, AM, and SCs. The findings demonstrate the complex relationships between these elements and the accuracy of node classification on a variety of datasets. The results in Table I shed important light on the relative importance of the various parts of the proposed GNN model. These features are crucial for accurately capturing complex relationships in graphs, as demonstrated by the model's strong performance across a range of datasets. The results advance the knowledge of GNN architectures and provide useful advice for creating powerful models that are suited to particular uses.

The experimental results demonstrate that the proposed GNN architecture changes increase the model's test set accuracy up to 1%. Adam optimizer, with a learning rate of 0.01, DA, AM, SC, and three GCN layers with the ReLU activation function were utilized in the top-performing design. Table II shows the reported mean classification accuracy for

the fully supervised node classification task for various graph neural network models. The bold numbers represent the best results while the second bests are underlined. The results for GCN, Mix-Hop, and GraphSAGE were obtained from [16]. The results for GCNII, NodeAug, FSGNN, GPRGNN, and GEOM-GCN were taken from [17, 18]. Traditionally, GNN models such as GCN and GAT have more efficiency on homophily datasets, although they give poor results on datasets with heterophily. Advanced models such as WRGAT, and GPRGNN function are reasonably superior on datasets with both homophily and heterophily. The proposed model performs significantly better on heterophily datasets, particularly with a notable boost on the CiteSeer and Chameleon datasets. Improvements were also noted for the datasets from Actor, Texas, and Cornell. The proposed model achieves consistent and comparable performance to state-of-the-art methods on homophily datasets. It also performed exceptionally well in the evaluation of the new Twitch social network dataset for node classification, demonstrating its flexibility to various graph architectures [19].

TABLE I. NODE CLASSIFICATION ACCURACY (%) FOR THE PROPOSED MODEL

	Cora	CiteSeer	PubMed	Chameleon	Wisconsin	Texas	Cornell	Squirrel	Actor
Proposed	88.94	78.32	89.14	79.31	86.14	86.21	86.57	72.22	36.71
Without DA	85.33	73.52	87.37	78.30	84.90	85.12	86.21	71.39	34.99
Without AM	86.22	75.59	87.13	79.01	84.90	84.22	84.43	71.35	33.78
Without SC	85.59	77.23	86.97	78.30	83.09	84.32	85.55	71.87	32.34

TABLE II. NODE CLASSIFICATION ACCURACY (%) FOR DIFFERENT MODELS ON VARIOUS DATASETS

Model	Cora	CiteSeer	PubMed	Chameleon	Wisconsin	Texas	Cornell	Squirrel	Actor	Mean Acc.
GCN	87.28	76.68	87.38	59.82	59.80	59.46	57.03	36.89	30.26	61.62
GraphSAGE	86.90	76.04	88.45	58.73	81.18	82.43	75.95	41.61	34.23	69.50
MixHop	87.61	76.26	85.31	60.50	75.88	77.84	73.51	43.80	32.22	68.10
GEOM-GCN	85.27	77.99	90.05	60.90	64.12	67.57	60.81	38.14	31.63	64.05
GCNII	88.01	77.13	90.30	62.48	81.57	77.84	76.49	N/A	N/A	–
NodeAug	86.20	75.40	82.1	N/A	N/A	N/A	N/A	–	–	–
GPRGNN	88.49	77.08	88.99	66.47	85.88	86.49	81.89	49.03	36.04	73.37
FSGNN	87.61	77.17	89.70	78.93	87.25	85.90	86.23	73.32	34.89	77.88
RAA – GNN	88.94	78.32	89.14	79.31	86.14	<u>86.21</u>	86.57	<u>72.22</u>	36.71	78.17

TABLE III. RESULTS ON A NEW TWITCH SOCIAL NETWORKS DATASET

Dataset	Train Loss	Train Accuracy	Validation Loss	Test Accuracy
Twitch Social Networks	0.2463	0.8989	0.9186	0.9046

V. CONCLUSION

This study presents the RAA-GNN model for node classification that incorporates SC, AM, and DA, showing that these elements can work together to improve its discriminative ability. While SCs handle over-smoothing issues, the AMs specifically allow the model to perform better on graphs for node classification. DA is an essential component that adds variation to the training dataset and promotes robustness against overfitting. The experimental study highlighted each component's independent effectiveness, as well as their combined impact on overall performance. The proposed model demonstrated its adaptability by consistently outperforming

state-of-the-art approaches in node classification across multiple datasets. In summary, this study extends GNN architectures and sheds light on the complex interactions between SC, DA, and AM. It also sets a new Sota node classification in graph structure learning through the first attempt to integrate SC, AM, and DA into RAA-GNN, thus advancing our understanding of GNNs.

ACKNOWLEDGMENT

The authors would like to acknowledge the Deanship of Scientific Research, Taif University for helping with this work.

REFERENCES

- [1] S. Khoshraftar and A. An, "A Survey on Graph Representation Learning Methods," *ACM Transactions on Intelligent Systems and Technology*, vol. 15, no. 1, Jan. 2024, Art. no. 19, <https://doi.org/10.1145/3633518>.
- [2] P. W. Battaglia *et al.*, "Relational inductive biases, deep learning, and graph networks." *arXiv*, Oct. 17, 2018, <https://doi.org/10.48550/arXiv.1806.01261>.

- [3] J. Zhou *et al.*, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, Jan. 2020, <https://doi.org/10.1016/j.aiopen.2021.01.001>.
- [4] A. Elhassouny and F. Smarandache, "Trends in deep convolutional neural Networks architectures: a review," in *2019 International Conference of Computer Science and Renewable Energies (ICCSRE)*, Agadir, Morocco, Jul. 2019, <https://doi.org/10.1109/ICCSRE.2019.8807741>.
- [5] H. Guo, Y. Mao, and R. Zhang, "MixUp as Locally Linear Out-of-Manifold Regularization," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 3714–3722, Jul. 2019, <https://doi.org/10.1609/aaai.v33i01.33013714>.
- [6] P. H. C. Avelar, A. R. Tavares, M. Gori, and L. C. Lamb, "Discrete and Continuous Deep Residual Learning Over Graphs." *arXiv*, Nov. 26, 2019, <https://doi.org/10.48550/arXiv.1911.09554>.
- [7] R. Liao, Z. Zhao, R. Urtasun, and R. S. Zemel, "LanczosNet: Multi-Scale Deep Graph Convolutional Networks." *arXiv*, Oct. 23, 2019, <https://doi.org/10.48550/arXiv.1901.01484>.
- [8] A. Deptuła, "Application of the Dependency Graph Method in the Analysis of Automatic Transmission Gearboxes," *Engineering, Technology & Applied Science Research*, vol. 11, no. 2, pp. 7033–7040, Apr. 2021, <https://doi.org/10.48084/etasr.4098>.
- [9] K. Ding, Z. Xu, H. Tong, and H. Liu, "Data Augmentation for Deep Graph Learning: A Survey," *ACM SIGKDD Explorations Newsletter*, vol. 24, no. 2, pp. 61–77, Sep. 2022, <https://doi.org/10.1145/3575637.3575646>.
- [10] X. Han, Z. Jiang, N. Liu, and X. Hu, "G-Mixup: Graph Data Augmentation for Graph Classification," in *Proceedings of the 39th International Conference on Machine Learning*, Jun. 2022, pp. 8230–8248.
- [11] F. Smarandache, "Extension of HyperGraph to n-SuperHyperGraph and to Plithogenic n-SuperHyperGraph, and Extension of HyperAlgebra to n-ary (Classical-/Neutro-/Anti-)HyperAlgebra," *Neutrosophic Sets and Systems*, vol. 33, pp. 289–295, Feb. 2020, <https://doi.org/10.5281/zenodo.3783103>.
- [12] K. Xu, M. Zhang, S. Jegelka, and K. Kawaguchi, "Optimization of Graph Neural Networks: Implicit Acceleration by Skip Connections and More Depth," in *Proceedings of the 38th International Conference on Machine Learning*, Jul. 2021, pp. 11592–11602.
- [13] T. K. Rusch, M. M. Bronstein, and S. Mishra, "A Survey on Oversmoothing in Graph Neural Networks." *arXiv*, Mar. 20, 2023, <https://doi.org/10.48550/arXiv.2303.10993>.
- [14] D. D. Van, "Application of Advanced Deep Convolutional Neural Networks for the Recognition of Road Surface Anomalies," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10765–10768, Jun. 2023, <https://doi.org/10.48084/etasr.5890>.
- [15] H. Sasaki, S. Yamamoto, A. Agchbayar, and N. Nkhbayasgalan, "Extracting Problem Linkages to Improve Knowledge Exchange between Science and Technology Domains using an Attention-based Language Model," *Engineering, Technology & Applied Science Research*, vol. 10, no. 4, pp. 5903–5913, Aug. 2020, <https://doi.org/10.48084/etasr.3598>.
- [16] J. Zhu, Y. Yan, L. Zhao, M. Heimann, L. Akoglu, and D. Koutra, "Beyond Homophily in Graph Neural Networks: Current Limitations and Effective Designs," in *Advances in Neural Information Processing Systems*, 2020, vol. 33, pp. 7793–7804.
- [17] M. Chen, Z. Wei, Z. Huang, B. Ding, and Y. Li, "Simple and Deep Graph Convolutional Networks," in *Proceedings of the 37th International Conference on Machine Learning*, Nov. 2020, pp. 1725–1735.
- [18] S. K. Maurya, X. Liu, and T. Murata, "Simplifying approach to node classification in Graph Neural Networks," *Journal of Computational Science*, vol. 62, Jul. 2022, Art. no. 101695, <https://doi.org/10.1016/j.jocs.2022.101695>.
- [19] B. Rozemberczki and R. Sarkar, "Twitch Gamers: a Dataset for Evaluating Proximity Preserving and Structural Role-based Node Embeddings." *arXiv*, Feb. 16, 2021, <https://doi.org/10.48550/arXiv.2101.03091>.

Production of Thermoplastic Composites reinforced with Posidonia Oceanica Fibers

Faouzi Slimani

LMPE/LR18ES01, National Higher Engineering School of Tunis, University of Tunis, Tunisia
slimani.fauzi@yahoo.fr (corresponding author)

Ines Ghanmi

Laboratory/LR18ES45-Mathematical Physics, Quantum Modeling and Mechanical Design, Nabeul Preparatory Engineering Institute, Carthage University, Tunisia
ines11ghanmi@gmail.com

Samir Ghanmi

Laboratory/LR18ES45-Mathematical Physics, Quantum Modeling and Mechanical Design, Nabeul Preparatory Engineering Institute, Carthage University, Tunisia
samir.ghanmi@ipein.rnu.tn

Mohamed Guedri

Laboratory/LR18ES45-Mathematical Physics, Quantum Modeling and Mechanical Design, Nabeul Preparatory Engineering Institute, Carthage University, Tunisia | National Higher Engineering School of Tunis, University of Tunis, Tunisia
mohamed.guedri@isetn.rnu.tn

Received: 6 December 2023 | Revised: 19 January 2024 | Accepted: 28 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6712>

ABSTRACT

This study investigates the development and characterization of a new biocomposite and biodegradable material based on natural fibers. This new biocomposite is composed of commercially available biodegradable polylactic acid (PLA) as a matrix and Posidonia Oceanica (PO) fibers collected from the coasts of Tunisia as reinforcement. This new material is produced by heating and pressing the two components in a special device. The use of PO, or sea balls, will allow exploiting one of the marine residues abundant on Tunisian beaches, instead of exploited industrially, and to preserve the beaches from debris given the impact of tourist activity in the Tunisian economy. The PLA/PO coupling allowed obtaining a biocomposite with promising mechanical properties. The improvement in maximum stress and strain after the addition of PO is one of the highlights of the results of this work.

Keywords: biocomposites; Posidonia oceanica; natural fibers; characterization

I. INTRODUCTION

Natural fiber-reinforced biocomposites using biodegradable polymers as a matrix are currently experiencing a major boom in their applications and implementation techniques. They are attracting attention in applications that require both rigidity and lightness, such as transportation, marine, aerospace, sports and leisure, heavy industry, and civil engineering. The fabrication of these biocomposites requires the correct selection of fibers and matrix to obtain acceptable mechanical properties and sufficient adhesion between the fibers and the matrix. Several researchers have conducted research on cellulose-based fibers such as kenaf [1], flax [2], hemp [3], banana [4], or natural fibers such as fibers from date palm waste [5]. These fibers are

used with thermoplastics or thermosets as a matrix. The most commonly used thermoplastics are polypropylene (PP) [6], polylactic acid (PLA) [7], polyethylene, and polyvinyl chloride (PVC). On the other hand, phenolic resins, epoxy resins, and polyester resins are the most commonly used thermosetting matrices [8]. This study represents a contribution to the development and characterization of new biocomposite and biodegradable materials based on Posidonia Oceanica (PO) fibers as reinforcement and PLA as matrix. PO is a coastal marine plant with elongated leaves and a stem with roots known as rhizomes. The rhizome ends in a bundle of elongated leaves [9, 10]. The plant has a leaf or bulb shape (Figure 1). Balls, also known as aegagropiles or sea balls, are formed from fibrous accumulations of leaves that have fallen around pieces

of rhizomes by the movement of the sea and accumulate in large quantities on the coast every summer. The choice of sea ball fibers is justified for two reasons: They are abundant on beaches and they are biodegradable. Similarly, the choice of PLA is justified by two reasons. It is present in the market at a reasonable price and it is also biodegradable. This biocomposites material is produced by hot pressing. For this reason, we designed and manufactured a digitally controlled thermomechanical system consisting of a die and punch that can be mounted on a conventional tension compressor.



Fig. 1. Posidonia balls harvested from the Tunisia sea.

A biocomposite was created through a four-element, two-modality experimental design. These factors mainly depend on the manufacturing conditions of the biocomposite, such as the proportion of fibers used, the heating temperature of PLA, the retention time of the punch within the matrix during hot compaction, and the compressive force required between the die and the punch. A series of tensile tests were carried out on a standardized bio-composite [14] cut by a laser machine.

II. MATERIALS AND METHODS

To develop the biocomposite with PLA matrix and PO reinforcement, we designed and manufactured a thermomechanical device that can develop it by thermal compression (Figure 2). This device basically consists of:

- Steel dice and punch.
- Two heating resistors, one in the die and another on the punch.
- Two thermocouples, one on the die and another on the punch.
- A digital controller to adjust the set temperature as needed.

The die holding the layer of the biocomposite (PLA/PO) is fixed by the lower jaw of the traction machine, and the punch is fixed by the upper jaw (Figure 2). Figure 3 shows the die, punch, bottom, and top PLA layers, and the PO layer. The two holes represent the location of the heating resistor and the thermocouple connected to the temperature controller. The die punch assembly attaches to the HM-S 200KN 04M4210 series tensile testing machine (Figure 2).

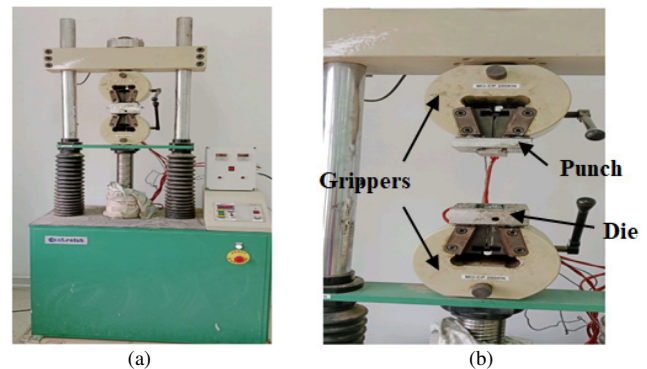


Fig. 2. Experimental set-up for the elaboration of the new material: (a) PLA/PO compression phase, (b) die/punch after the disassembly of the PLA/PO plate.

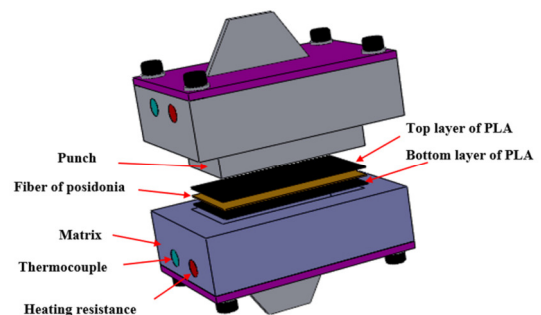


Fig. 3. 3D compound of the die and punch of the new experimental device.

Our novel biocomposite characterization process is divided into three major steps:

1) Step 1

- Preparation of the PO fibers (washing and fiber disentangling) according to the dimensions of the desired biocomposite panel (Figure 4).
- PLA preparation: PLA was manually cut into 0.5-1 cm pieces.
- Panels from the new biocomposite were obtained by hot pressing.

2) Step 2

- Laser cutting of PLA/PO biocomposite samples was conducted according to the ISO 527-2 standard (Figure 5).
- A series of tensile tests were performed.

3) Step 3

The results are analyzed (determination of the mechanical properties of biocomposites).

III. EXPERIMENTS

To develop the new PLA/PO material, we used the experimental setup described in Figures 2 and 3. Four design parameters were taken into account for the development of the new material: the heating temperature T (°C), the percentage of reinforcement (% of PO), the pressure holding time t (min) of

the bio-composite and the pressure force F (10^3 daN). These parameters are at two levels (min and max) as shown in Table I. Eight specimens of the bio-composite were made according to Table II and Figure 6.

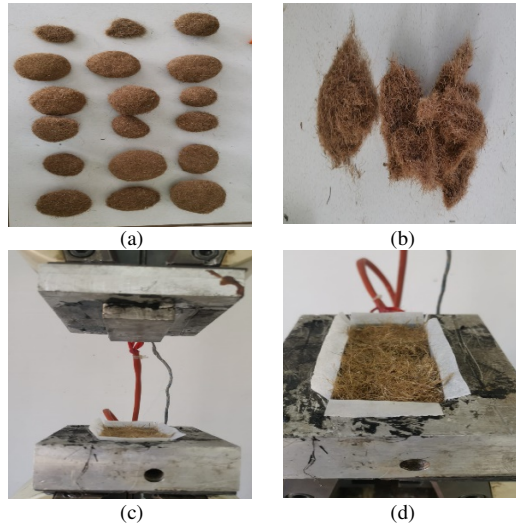


Fig. 4. Preparation of PO reinforcement: (a) PO before treatment, (b) fiber detanglement, (c) and (d) preparation of the PO layer in the matrix.

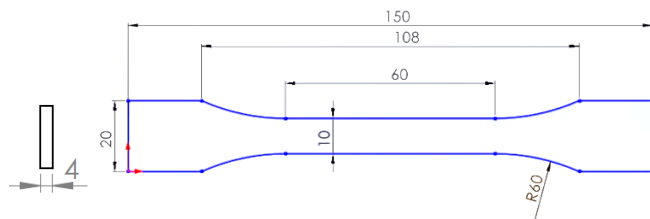


Fig. 5. Tensile test specimens in accordance with ISO 527-2.

TABLE I. DESIGN PARAMETERS CONSIDERED FOR THE DEVELOPMENT OF THE NEW BIO-COMPOSITE

Independent variable	Factor levels	
PLA heating temperature ($^{\circ}\text{C}$)	160	170
% of fiber in bio-composite (% PO/PLA)	5	10
Punch holding time in the die (min)	10	20
Pressure force between die and punch (10^3 daN)	10	14



Fig. 6. Example of a PLA/PO plate.

IV. RESULTS AND DISCUSSION

In order to characterize the new PLA/ PO bio-composite, a series of tensile tests were carried out on specimens in accordance with ISO 527 [14] (Figure 5) from the 8 levels obtained by hot pressing using the thermomechanical system parameters shown in Table II. Each level was laser-cut into 3 specimens.

TABLE II. EXPERIMENTAL PARAMETERS

Specimen N°	T ($^{\circ}\text{C}$)	PO (%)	t (min)	F (10^3 daN)
1	160	5	20	10
2	160	5	20	14
3	170	5	20	10
4	170	5	20	14
5	160	10	10	10
6	160	10	10	14
7	170	10	10	10
8	170	10	10	14

These tests resulted in at least one fracture for each tested specimen. The stress-strain curves are plotted in Figures 7 and 8, showing the behavior of the bio-composite containing 5% (PLA/PO-5) and 10% (PLA/PO-10) fibers, respectively. The tensile behavior of all specimens is similar and can be broken down into three main phases:

The first phase corresponds to a linear increase in the stress applied with the strain in the strain zone $[0.05\% - 0.25\%]$ in accordance with ISO 527-4 [15]. This phase is the elastic zone of the stress-strain curve, where the modulus of elasticity can be determined from Hook's law by [1]:

$$E = \frac{\Delta\sigma}{\Delta\varepsilon} \quad (1)$$

where E , $\Delta\sigma$, and $\Delta\varepsilon$ are respectively the modulus of elasticity (GPa), the difference in the applied tensile stress between two strain points, and the difference between the two strain points.

In the second phase, the beginning of this part is inclined, becoming almost linear up to the maximum stress value, except for a few peaks in this linear zone due to fiber breaks.

The third and final phase, in which stress is completely released, is the phase of total rupture of PLA and fibers in the bio-composite.

Table III and Figure 9 summarize the tensile test results for two types of the developed bio-composite, PLA/PO -5, PLA/PO-10, and PLA alone. They represent the mean values and standard deviations of the mechanical properties (ρ , σ_{Max} , ε , and E) of the two tensile-stressed bio-composites compared to those of PLA alone.

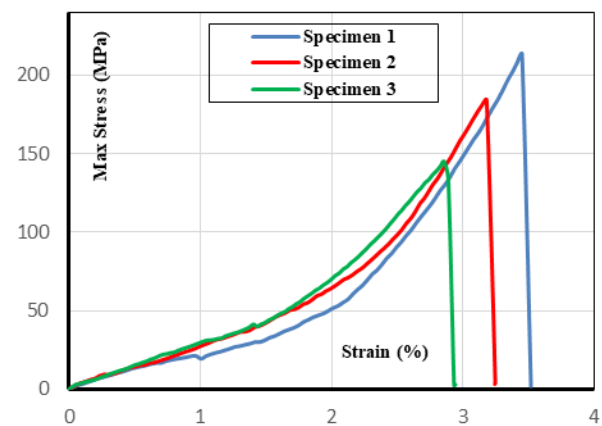


Fig. 7. Tensile curves for samples of 5% PO (level 2).

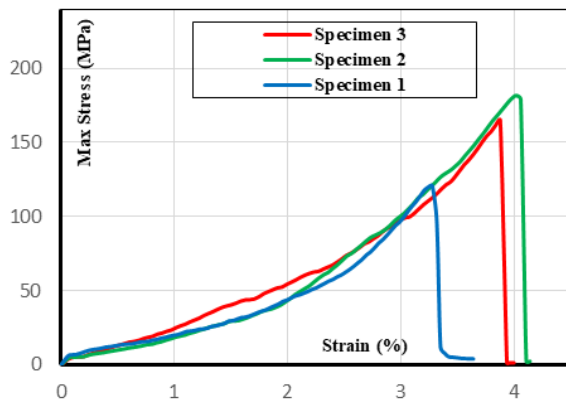


Fig. 8. Tensile curves for samples of 10% PO (level 6).

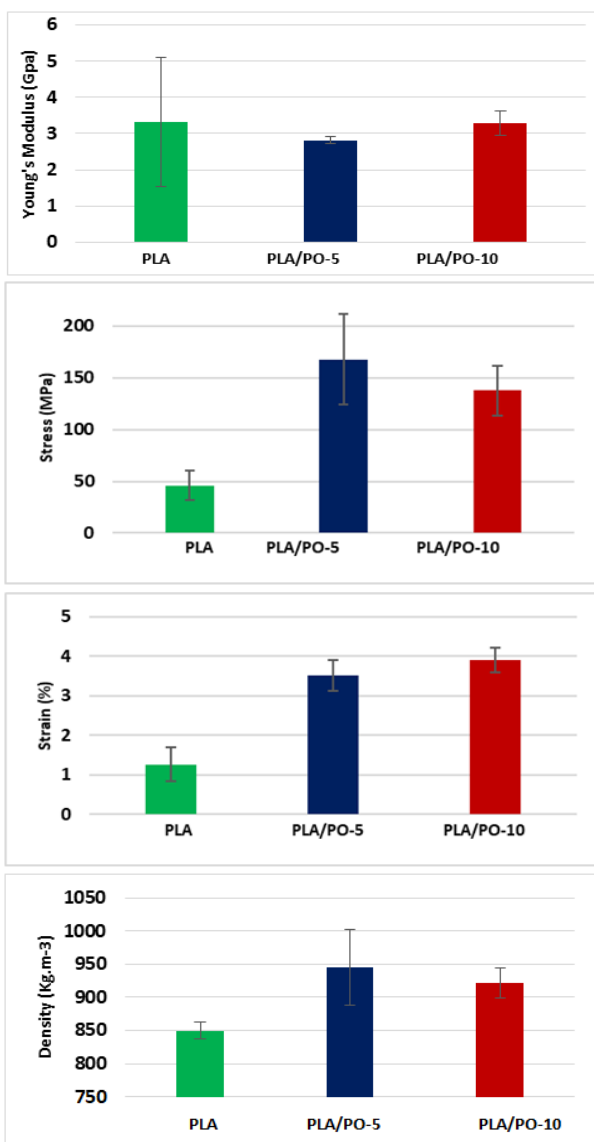


Fig. 9. Variation in mechanical properties of bio-composite as a function of PO fiber content.

There was a significant improvement in the maximum stress, strain, and density of PLA/PO-5 and PLA/PO-10 biocomposites in comparison with PLA alone, at the cost of degradation in the Young's modulus of PLA/PO-5. Indeed, the increases in σ_{Max} and ϵ for PLA/PO-5 are 264.94% and 177.78%, respectively, while for PLA/PO-10 the increases are 199.44% and 209.52%, respectively. The degradation of Young's modulus for PLA/PO-5 bio-composite is 15.06%, while it is not significant for PLA/PO-10. We also note that an increase in the fiber content of 5% allowed an improvement of the Young modulus of 14.28% at the cost of 17.94% degradation of the maximum stress.

TABLE III. AVERAGE PROPERTIES OF BIO-COMPOSITES PLA/PO

Material	ρ (Kg/m ³)	σ_{Max} (MPa)	ϵ (%)	E (GPa)
PLA	850 \pm 12	46 \pm 14	1.26 \pm 0.43	3.32 \pm 1.77
PLA-PO-5	945 \pm 57	167.87 \pm 43.46	3.50 \pm 0.39	2.82 \pm 0.1
PLA-PO-10	922 \pm 23	137.74 \pm 24.32	3.9 \pm 0.30	3.29 \pm 0.34

V. CONCLUSION

The main objective of this work was the development and characterization of a new biocomposite composed of commercial PLA as matrix and PO fibers as reinforcement. Posidonia balls are in abundance in Tunisia beaches and are considered an annoyance for seaside tourism. In the first step, we designed and manufactured a digitally controlled thermomechanical device for the development of the new biocomposite by hot compression.

As a second step, and to characterize our new biocomposite, a series of tensile tests on standardized specimens from 8 PLA/PO specimens was carried out. These specimens were manufactured taking into account four experimental parameters: the heating temperature, the percentage of reinforcement, the time of holding the biocomposite under pressure t , and the pressure stress.

The results showed a clear increase in the maximum stress and strain of PLA/PO-10 and PLA/PO-5 compared to plain PLA, at the cost of degradation of the Young modulus of PLA/PO-5. In addition, an increase in fiber content of 5% allowed an improvement of the Young modulus of 14.28% at the cost of the degradation of the maximum stress of 17.94%. The study of the impact of humidity and its effect on the mechanical characteristics of the biocomposite, the study of the formability of this new material, and the improvement of its biodegradability could be new areas for improvement in this work.

REFERENCES

- [1] M. S. Huda, L. T. Drzal, A. K. Mohanty, and M. Misra, "Effect of fiber surface-treatments on the properties of laminated biocomposites from poly(lactic acid) (PLA) and kenaf fibers," *Composites Science and Technology*, vol. 68, no. 2, pp. 424–432, Feb. 2008, <https://doi.org/10.1016/j.compscitech.2007.06.022>.
- [2] C. Poilâne, Z. E. Cherif, F. Richard, A. Vivet, B. Ben Doudou, and J. Chen, "Polymer reinforced by flax fibres as a viscoelastoplastic material," *Composite Structures*, vol. 112, pp. 100–112, Jun. 2014, <https://doi.org/10.1016/j.compstruct.2014.01.043>.

- [3] B. Pietruszka, M. Gołębiewski, and P. Lisowski, "Characterization of Hemp-Lime Bio-Composite," *IOP Conference Series: Earth and Environmental Science*, vol. 290, no. 1, Mar. 2019, Art. no. 012027, <https://doi.org/10.1088/1755-1315/290/1/012027>.
- [4] P. Sivaranjana and V. Arumugaprabu, "A brief review on mechanical and thermal properties of banana fiber based hybrid composites," *SN Applied Sciences*, vol. 3, no. 2, Jan. 2021, Art. no. 176, <https://doi.org/10.1007/s42452-021-04216-0>.
- [5] K. Almi, A. Benchabane, S. Lakel, and A. Kriker, "Potential utilization of date palm wood as composite reinforcement," *Journal of Reinforced Plastics and Composites*, vol. 34, no. 15, pp. 1231–1240, Aug. 2015, <https://doi.org/10.1177/0731684415588356>.
- [6] J.-H. Lee *et al.*, "Mechanical Properties of Biocomposites Using Polypropylene and Sesame Oil Cake," *Polymers*, vol. 13, no. 10, Jan. 2021, Art. no. 1602, <https://doi.org/10.3390/polym13101602>.
- [7] M. Sh. Al-Otaibi, O. Y. Alothman, M. M. Alrashed, A. Anis, J. Naveen, and M. Jawaid, "Characterization of Date Palm Fiber-Reinforced Different Polypropylene Matrices," *Polymers*, vol. 12, no. 3, Mar. 2020, Art. no. 597, <https://doi.org/10.3390/polym12030597>.
- [8] A. K. Trivedi, M. K. Gupta, and H. Singh, "PLA based biocomposites for sustainable products: A review," *Advanced Industrial and Engineering Polymer Research*, vol. 6, no. 4, pp. 382–395, Oct. 2023, <https://doi.org/10.1016/j.aiepr.2023.02.002>.
- [9] C. Vasapollo, "Spatio-temporal variability of plant features and motile invertebrates in Posidonia oceanica seagrass meadows," Ph.D. dissertation, The Open University, Milton Keynes, UK, 2010.
- [10] "Biology of Posidonia," *Mediterranean Posidonia Network*. <https://medposidonianetwork.com/biology/>.
- [11] A. N. Hassooni and S. R. A. Zaidee, "Behavior and Strength of Composite Columns under the Impact of Uniaxial Compression Loading," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 8843–8849, Aug. 2022, <https://doi.org/10.48084/etasr.4753>.
- [12] S. H. Soomro *et al.*, "Synthesis and Characterization of Poly(styrene)-block-Poly(acrylic acid) and Organoclay Based Hybrid Composite Materials," *Engineering, Technology & Applied Science Research*, vol. 8, no. 5, pp. 3411–3415, Oct. 2018, <https://doi.org/10.48084/etasr.2273>.
- [13] T. Nehari, A. Ziadi, D. Ouinas, and B. Boutabout, "Numerical Study of the Effect of the Penetration of a Crack in the Matrix of a Composite," *Engineering, Technology & Applied Science Research*, vol. 4, no. 3, pp. 649–655, Jun. 2014, <https://doi.org/10.48084/etasr.447>.
- [14] ISO/TC 61/SC 2 Technical Committee, *ISO 527-2:2012. Plastics. Determination of tensile properties. Part 2: Test conditions for moulding and extrusion plastics*. ISO, 2012.
- [15] ISO/TC 61/SC 13 Technical Committee, *ISO 527-4:2021. Plastics. Determination of tensile properties. Part 4: Test conditions for isotropic and orthotropic fibre-reinforced plastic composites*. ISO, 2021.

Study of the Injection of Secondary Air into the Intake Manifold of the Gas Turbine to Avoid the Compressor Surging Phenomenon

George Iulian Balan

Defence and Security Systems Engineering, Faculty of Aircraft and Military Vehicles, Ferdinand I Military Technical Academy, Romania
george333@gmail.com

Dragos Gabriel Zisopol

Mechanical Engineering Department, Petroleum-Gas University Ploiești, Romania
zisopold@upg-ploiesti.ro (corresponding author)

Amado Stefan

Department of Integrated Aviation Systems and Mechanics, Faculty of Aircraft and Military Vehicles, Ferdinand I Military Technical Academy, Romania
amado.stefan@mta.ro

Vasile Nastasescu

Department of Integrated Aviation Systems and Mechanics, Faculty of Aircraft and Military Vehicles, Ferdinand I Military Technical Academy, Romania
vasile.nastasescu@mta.ro

Lucian Grigore

Center of Excellence in Robotics and Autonomous Systems - CERAS, Ferdinand I Military Technical Academy, Romania
lucian.grigore64@gmail.com

Received: 18 January 2024 | Revised: 28 January 2024 | Accepted: 31 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6927>

ABSTRACT

This paper presents part of the research on avoiding or reducing the surging effects that appear in the axial compressor intake manifold of a gas turbine. This research has led to an original solution validated by numerical simulations and experimental investigations. The increased amount of air suddenly required in the transient regime of the gas turbine is introduced into the intake manifold through slits arranged perpendicular to the direction of flow, on an aerodynamic profile at a certain angle to it and a certain distance from the minimum transversal section. The slits are arranged on the opposite sides of the gallery and connect with a transverse channel of the airfoil, in which there is air under pressure, from which the introduction of additional air is ordered. The numerical and experimental results extended to the influence of many geometric and mechanical parameters, proving that the proposed solution is as effective as possible compared to the classic ejector solution.

Keywords-gas turbine; secondary air injection; intake ducting; flow simulation

I. INTRODUCTION

The increasing use of gas turbines in vehicles and ships makes scientific research on their operation at the highest possible parameters topical. This paper presents the results of previous research on air intake, with the avoidance or reduction

of the compressor surge phenomenon specific to transient regimes. The air intake in a gas turbine is a continuous process at high and variable flow rates to ensure the amount required not only for fuel combustion but also for cooling the engine components. A maximum of 10% of this air quantity can be used for auxiliary processes of the gas turbine [1-2]. This study

adopted an original solution of additional air injection, which, through numerical simulations and experimental investigations, proved to be superior to the classic solution that uses the ejection phenomenon.

Depending on the field of power plant use, the air intake ducting system is designed to ensure the amount of filtered, dehumidified, and guided air to the intake nozzle of the gas turbine. Since the supply air in gas turbines is compressed by passing successively through the compressor rotors, the main requirements regarding the quality of the air and the gas-dynamic conditions can be obtained as follows: limiting the admission of impurities [3], limiting the admission of water, except in the case where a controlled flow of water vapor is ensured to improve the efficiency of the engine or compressor washing solutions, avoiding the admission of objects between the compressor blades, avoiding the accumulation of ice in the supply path or on the air guide struts [4], and configure the gas-dynamic route to reduce losses but also to avoid the occurrence of surging conditions during transient operating regimes [5-6].

In [2, 7], a study of the airflow through the intake manifold of the ST40 engine was presented, highlighting the distribution of velocities and pressures using a CFD analysis. The results had a high degree of generality, but numerical simulation and experimental research were performed for the intake manifold of the ST40M-type gas turbine, as shown in Figure 1. In addition to the intake plenum configuration, the air guide vanes are used in the first stage of the compressor. Guide vanes can be of fixed type if the gas turbine engine works mainly at a predetermined regime, or of orientable type to achieve air guidance according to the load of the power plant. Since the amount of air required to supply gas turbines is considerably greater than the air required for diesel engine operation, intake paths with complicated configurations can generate significant gas-dynamic resistances. In [8], the pressure variation in the outlet section of the gas turbine intake manifold was analyzed according to the pressure variations produced by the wind.

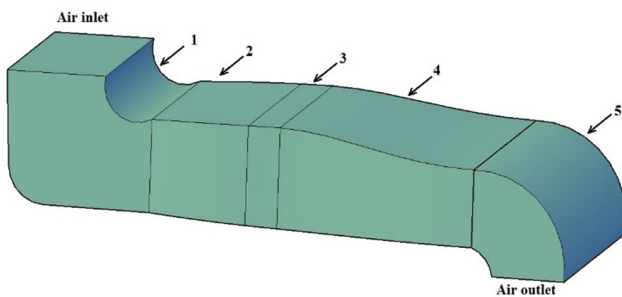


Fig. 1. Air intake manifold sector of the gas turbine. 1 - 90 degree transition volute, 2 - intake manifold TR1, 3 - elastic couple and anti-stone sieve, 4 - intake manifold TR2, 5 - 90 degree transition volute.

For marine power plants, air intake paths for gas turbines must satisfy all the listed requirements to ensure the operation of the ship. In [9] the design and verification of the intake manifold of the gas turbine engine was addressed, where the manifold configurations were analyzed at the inlet to the turbine compressor.

II. RESEARCH OBJECTIVES

Laboratory investigation of the phenomena specific to the transient regime of air supply was performed by injecting air into the intake manifold of a 1:4 scale model using a controlled flow compressor. The main objective of the study was the validation of the original solution of introducing additional air, which leads to superior performance compared to the classical processes. Further objectives were to investigate the influence of the geometric characteristics of the proposed solution, such as the position of the access slot in the intake manifold and the angle of the nozzle. There were also other secondary objectives regarding the geometrical and mechanical characteristics of the secondary air introduction, but having little influence, they are not presented here. This study presents the essential elements of the research performed, the results of which can be generalized to other similar constructions.

III. NUMERICAL DETERMINATIONS

Flow simulation software was used to model airflow through the gas turbine intake manifold model, which features the ability to model laminar and turbulent flows. The Favre-averaged Navier-Stokes equations were used to predict turbulent flows, where time-averaged effects of the flow turbulence on the flow parameters were considered, whereas the other, i.e. large-scale, time-dependent phenomena were directly taken into account. Flow simulation uses transport equations for the turbulent kinetic energy and its dissipation rate [10-11]. The conservation laws for mass [12], angular momentum, and energy in the Cartesian coordinate system rotating with angular velocity about an axis passing through the coordinate system's origin can be written in the conservation form as follows:

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_i}(\rho v_i) = S_M^p \quad (1)$$

$$\begin{aligned} \frac{\partial \rho v_i}{\partial t} + \frac{\partial}{\partial x_i}(\rho v_i v_j) = \\ - \frac{\partial p}{\partial x_i} + \frac{\partial}{\partial x_i}(\tau_{ij} + \tau_{ij}^R) + S_i + S_{li}^p \end{aligned} \quad (2)$$

$$\begin{aligned} \frac{\partial \rho H}{\partial t} + \frac{\partial}{\partial x_i}(\rho v_i H) = \\ \frac{\partial p}{\partial t} + \frac{\partial}{\partial x_i}(v_j(\tau_{ij} + \tau_{ij}^R) + q_i) - \\ \tau_{ij}^R \frac{\partial v_i}{\partial x_j} + \rho \varepsilon + S_i v_i + S_H^p + Q_H \end{aligned} \quad (3)$$

$$H = h + \frac{v^2}{2} + \frac{5}{3}k - \frac{\Omega^2 r^2}{2} - \sum_m h_m^0 y_m \quad (4)$$

where v is the velocity of the fluid, ρ is the density, S_i is the mass-distributed external force, which can be gravitational ($S_i^{grav} = -\rho g_i$, g_i is the component of the gravitational acceleration in the i direction), porous, or inertial due to the rotation of the coordinate system, h is the enthalpy, and S_M^p , S_{li}^p and S_H^p are due to the interaction at the interface with the Euler-Lagrange particles. Q_H is the heat source per unit volume, τ_{ij} is the viscous stress tensor, q_i is the diffusive heat flux, Ω is the angular velocity of the mobile coordinate system in rotational motion, r is the radius from the elementary fluid particle to the

axis of rotation, k is the kinetic energy turbulence, h_m^0 is the enthalpy of the m component of the fluid mixture and y_m is the concentration of the m component of the fluid mixture [10, 13]. The stress tensor for Newtonian fluids is defined as:

$$\tau_{ij} = \mu_t \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \frac{2}{3} \delta_{ij} \frac{\partial u_k}{\partial x_k} \right) \quad (5)$$

In the Boussinesq hypothesis, the Reynolds stress tensor has the form:

$$\tau_{ij}^R = \mu_t \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \frac{2}{3} \delta_{ij} \frac{\partial u_k}{\partial x_k} \right) - \frac{2}{3} \rho k \delta_{ij} \quad (6)$$

where δ_{ij} is the Kronecker symbol or delta function, μ is the dynamic viscosity, μ_t is the turbulent viscosity, k is the turbulent kinetic energy, and ϵ is the turbulent dissipation rate [14-15]. As part of the numerical determinations, simulations were carried out in transient mode, with the operation of only the fan at maximum speed, and several simulations in stationary mode for different geometrical configurations of the secondary air collector gallery, using only air injection. Fan modeling allows the calculation of airflow through the gallery using a geometric domain as the envelope of the fan blades rotating at the fan's angular velocity. This calculation is a way to estimate the airflow flowing through the intake manifold when using the fan and secondary air injection simultaneously. The following variants were considered to estimate the influence of the effect of the introduction of compressed air on the airflow passing through the gallery:

- The variation of the outlet position of the manifold nozzle in the intake manifold compared to the entrance to the modified section (relative to the leading edge of the wall profile)
- The variation of the angle of the nozzle with respect to the longitudinal axis of the intake manifold.

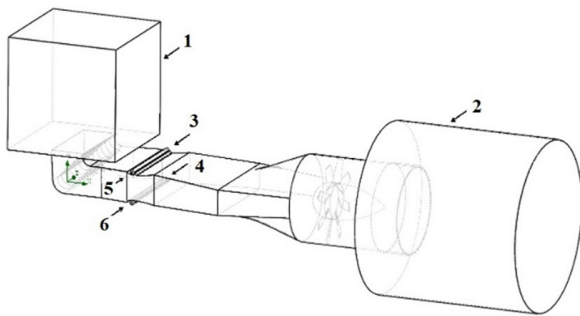


Fig. 2. The flow domain and the boundary conditions. 1,2 – atmospheric pressure surfaces; 3, 4, 5, 6 – compressed air at 0,015 kg/s flow.

In all these simulations, the convergence of the mass flow through the exit surface of the computational domain (the base and the side surface of the cylinder at the exit of the fan) and the convergence of the default parameters were observed. The final solution was chosen according to the geometric configuration that ensured the highest flow rate of air leaving the flow domain, keeping the same static pressure of the secondary air in each case. Figure 2 shows the flow domain

used in the CFD analysis [16-17] and the position of the boundary conditions used.

A. Influence of the Nozzle Position on Air Flow through the Intake Ducting Using Only Secondary Air

The position of the nozzle outlet in the gallery was changed, starting from 30 to 60 mm from the entrance to the modified section, in 5 mm increments, to study the influence of the position between the outlet of the nozzle in the intake ducting compared to the leading edge of the modified section wall on the airflow. The reduction of the distance below 30 mm of the nozzle outlet in the manifold can no longer be achieved for the 12 mm diameter of the manifold (next to this position being the minimum section of the intake manifold) without its geometry intersecting the wall of the gas turbine intake. For these numerical determinations, for the condition of stopping the calculation, the convergence of the calculated mass flow rate on the exit surface of the domain and the total mass flow rate in addition to the usual parameters was followed. Figures 3 and 4 show the velocity field on the plane of symmetry of the gallery, in the area of the modified section, for the exit at 30 and 60 mm from the entrance to the modified section.

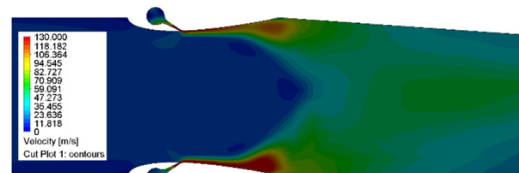


Fig. 3. Exit at 30 mm from the section entrance.

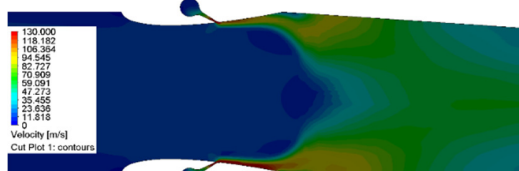


Fig. 4. Exit at 60 mm from the section entrance.

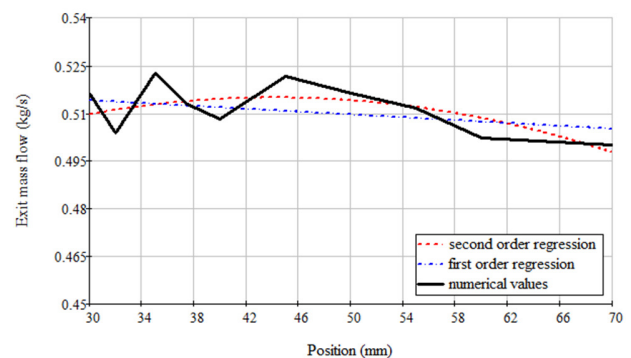


Fig. 5. The variation of the airflow at the outlet of the gallery depending on the distance between the nozzle and the edge of the wall profile.

Figure 5 shows the variation of the mass flow rate in the exit section of the domain depending on the exit position of the nozzle in the gallery. The laws of linear regression and second order are:

$$0.521 - 2.233 \cdot 10^{-4} \cdot x_{\text{position}} \quad (7)$$

$$0.464 + 2.296 \cdot 10^{-3} \cdot x_{\text{position}} - 2.595 \cdot 10^{-5} \cdot x_{\text{position}}^2 \quad (8)$$

B. Influence of Nozzle Angle on Manifold Airflow Using Only Secondary Air

To determine the influence of the angle of the channel with the axis of the gas turbine intake manifold, the geometric configuration with the 12 mm collector manifold, the outlet of the nozzle in the manifold at 35 mm from the leading edge, and the dihedral angle of the nozzle of 8° were used. The angle of inclination of the channel bisector plane relative to the horizontal was varied from 26 to 52° in increments of 2° . When the angle of inclination of the nozzle varied, the condition of its constant length was maintained. Figures 6 and 7 show the velocity field on the plane of symmetry of the gallery, in the area of the modified section, for inclination angles of 26° and 40° . Figure 8 shows the mass airflow rate in the exit section of the domain depending on the angle of the nozzle and the linear regression functions and by the least squares method. The laws of linear regression and second order are:

$$0.627 - 3.62 \cdot 10^{-3} \cdot x_{\text{angle}} \quad (9)$$

$$0.536 + 1.261 \cdot 10^{-3} \cdot x_{\text{angle}} - 6.259 \cdot 10^{-5} \cdot x_{\text{angle}}^2 \quad (10)$$

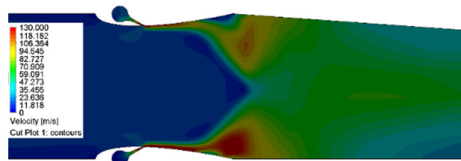


Fig. 6. The 26 degrees nozzle outlet inclination.

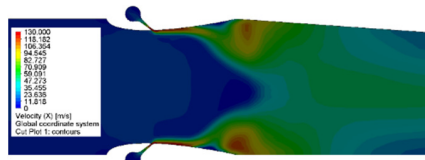


Fig. 7. The 40 degrees nozzle outlet inclination.

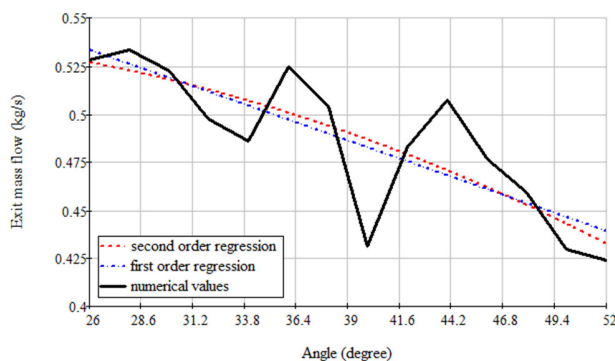


Fig. 8. The variation of mass flow at the outlet of the gallery according to the angle of the nozzle.

C. Airflow Modeling Through the Modified Intake Manifold with Fan and Secondary Air Operation

The geometric configuration used corresponds to the nozzle exit position in the gallery in front of the entrance to the modified section, with the angle of the nozzle axis to the longitudinal direction of the air intake manifold. The calculation was performed in transient mode, for a time interval of two seconds, an interval considered sufficient to equalize the flow through the gallery. The rotation speed of the domain containing the fan blades was 3000 rpm. The law of variation of the airflow rate for the secondary air intakes was implemented in tabular. Figures 9 and 10 show the streamlines as a function of speed and the static pressure distribution on the plane symmetry for the moment of time $t = 2$ s, respectively.

TABLE I. TIME VARIATION OF SECONDARY AIRFLOW RATE

Time (s)	Mass airflow rate (kg/s)
0.0	0.000
2.0	0.000
2.1	0.015
4.0	0.015

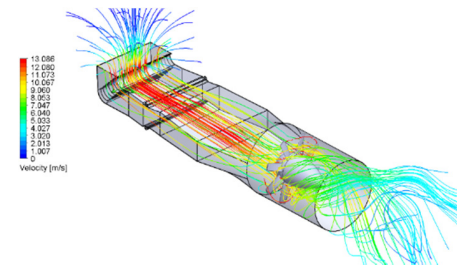


Fig. 9. Velocity distribution on streamlines at $t = 2$ s.

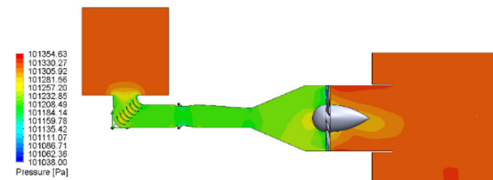


Fig. 10. Pressure distribution on the symmetry plane of the gallery at $t = 2$ s.

Figure 11 shows the speed distribution on the streamlines at the time of 2.4 s. The average value of the flow produced by the fan is 0.401 kg/s and after the injection of the secondary air with a total flow in the four intakes of 0.06 kg/s, an average value of 0.623 kg/s results at the exit, as shown in Figure 12.

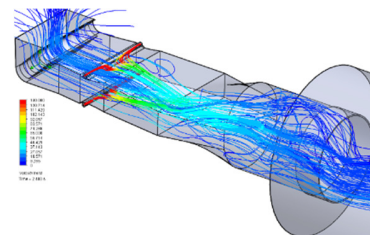


Fig. 11. Streamlines at $t = 2.4$ s.

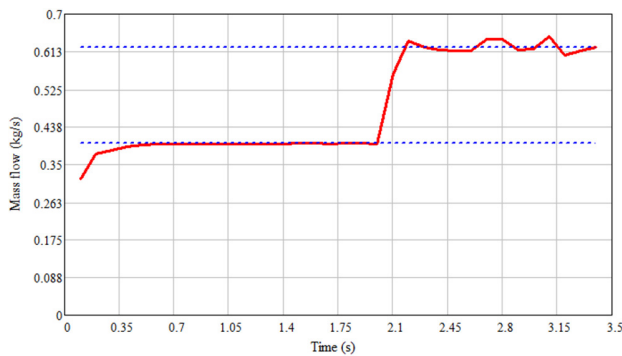


Fig. 12. Time variation of the airflow at the exit of the calculation domain.

IV. EXPERIMENTAL DETERMINATIONS

Experimental determinations were carried out on a 1:4 scale 3D printed model of the intake manifold, as shown in Figure 13. The area in the gallery for which the scale model was made was chosen for constructive reasons to allow the addition of a compressed air installation next to the intake gallery. Secondary air was injected between the TR1 and TR2 intake sections. The scale model of the gallery was made from the entrance of the 90-degree transition volute to the exit of the TR2 intake manifold. The gallery components were made of PLA [18-21] by 3D printing, and assembled by screws [22-24]. The side walls were made of plexiglass plates.

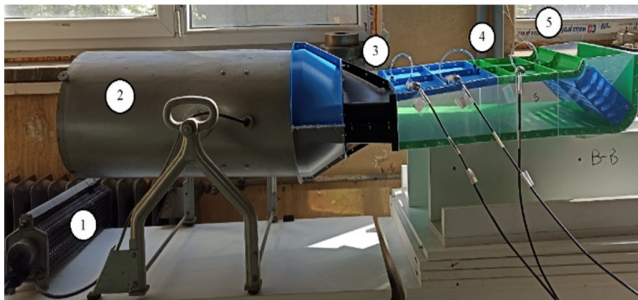


Fig. 13. The 1:4 scale model of the original intake ducting: 1 - rheostat; 2 - axial fan; 3, 4, and 5 - pressure nozzles.

To determine the static pressure, the intake manifold was equipped with static pressure sockets, where the pressure sensors were connected. Pressure socket 1 was at the elbow exit (position 5), pressure socket 2 was at the exit of the section with constant section (position 4), and section 3 was at the exit from section 3 (position 3), respectively, at the entrance to the fan connection. To simulate the flow of the admission air of the gas turbine, a variable speed ducted fan was used, which was coupled to the scale model using a variable cross-section, with the inlet section from the rectangular shape (outlet from TR2) and the outlet section on the circular (entry into the ducted fan). This investigation aimed to determine a geometric configuration that allows air injection into the gas turbine intake manifold using linear nozzles arranged transversely to the gallery walls and to confirm the numerical models used for airflow calculation. The 1:4 scale model of the modified manifold contains a section to which two transverse linear nozzles are added, fed by cylindrical manifolds. The cylindrical

manifolds were supplied with pressurized air from the side of the experimental device having a total of four air intakes. Figure 14 shows the experimental setup, with the secondary air supply system mounted and connected to a compressor. Figure 15 shows the modified intake duct section with the direction of airflow located between TR1 and TR2, and the geometry of the manifold, linear nozzle, and manifold wall in the form of the NACA 0018 profile. The nozzle exit position is downstream with 35 mm against minimal transversal area and the manifold channel and the angle of 30 degrees. The geometry of the collector gallery is cylindrical with a diameter of 12 mm and the angle between the walls of the nozzle is 8°.

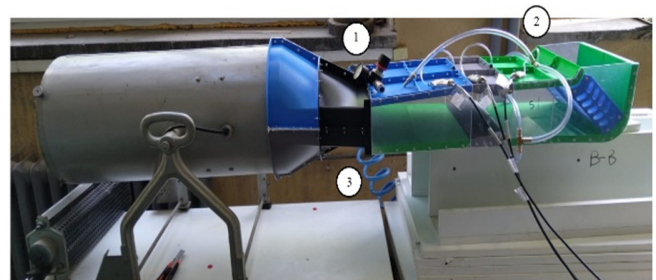


Fig. 14. The 1:4 scale model of the modified intake ducting: 1 - pressure regulator; 2 - compressed air tubing system; 3 - compressed air main supply.

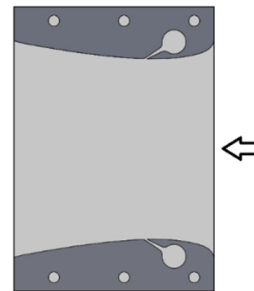


Fig. 15. The modified intake ducting section with flow direction.

To determine static pressure, Kistler piezoresistive differential pressure sensors with a scale of ± 10 kPa were used, which were connected (with the depression socket) to the pressure sockets provided by construction on the scale model. A Pitot tube connected to a pressure sensor described above was used to determine the speed. The useful length of the anemometer rod was 143 mm, and subtracting 3 mm of the thickness of the plexiglass side wall resulted in a measurement length of 140 mm from the side wall. Data acquisition was performed with the QuantumX MX1615B module. The device was used to acquire data from resistive, inductive, or piezoresistive strain gauge bridges in full-bridge, half-bridge, and quarter-bridge configurations. This device is also used for transducers, potentiometers, and resistance thermometers based on strain gauges. The module was equipped with 16 individually configurable sensor inputs. To determine the velocities, the Pitot tube was inserted into the intake manifold, its tip passing through lines 1, 2, and 3. Its position with respect to the reference system was: line 1 at $x = 155$ mm, line 2 at $x = 387$ mm, and line 3 at $x = 572$ mm. For all lines, $y = 59$ mm relative to the coordinate system of Figure 16.

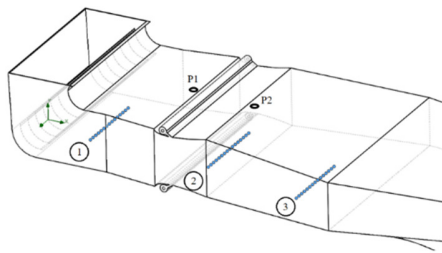


Fig. 16. The measurement lines.

V. RESULTS VERIFICATION AND VALIDATION

Figures 17-22 present the numerical and experimental speed variations for the three lines under the operating conditions of only the fan in the maximum operating mode, with a speed of 3000 rpm (experimental: red, numerical: blue, measured in 10 mm increments). The graphic representations in these Figures show that in the positioning domain of the z coordinate, there is a range where the agreement of the numerical results with the experimental ones is very good. The discrepancy outside this range, although sometimes quantitatively admissible, is caused by the end effects of the range considered.

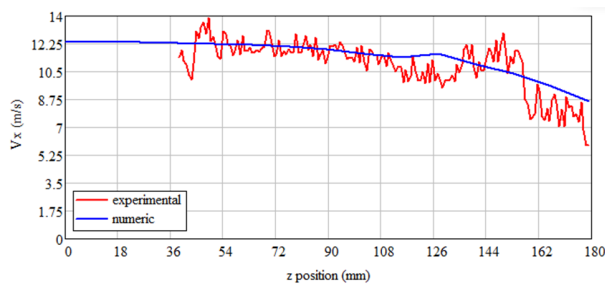


Fig. 17. Speed variation on line 1 on the original gallery model.

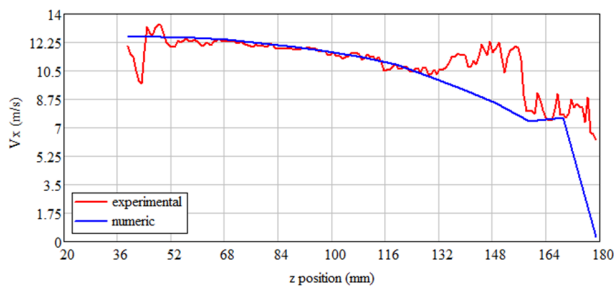


Fig. 18. Speed variation on line 1 on the modified gallery model.

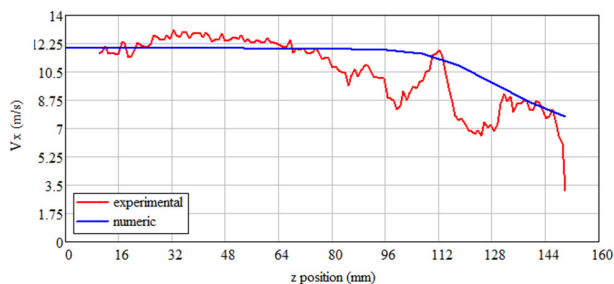


Fig. 19. Speed variation on line 2 on the original gallery model.

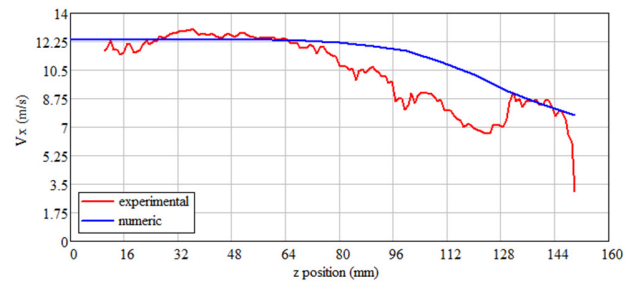


Fig. 20. Speed variation on line 2 on the modified gallery model.

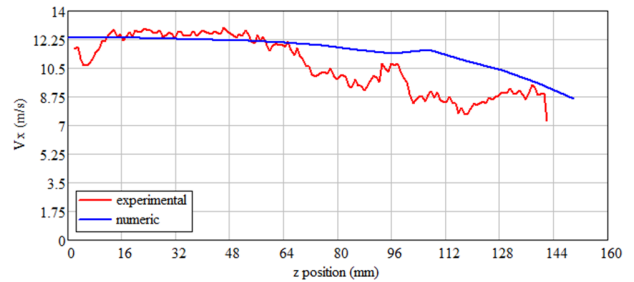


Fig. 21. Speed variation on line 3 on the original gallery model.

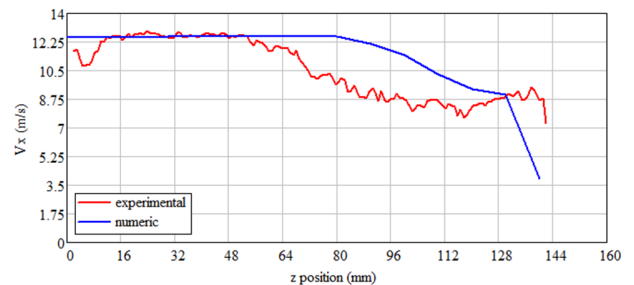


Fig. 22. Speed variation on line 3 on the modified gallery model.

VI. CONCLUSIONS

The concordance of the numerical and experimental results obtained on the model with the modified section validates both the applied solution and the correctness of the numerical modeling. The differences noted in the streamlines are noticeable in the vertical plane, especially after exiting the modified section, in the area located in the vicinity of the plane of symmetry. It was found that for the geometrical configuration considered, using the secondary air injection in the scale model of the gallery section, an additional flow rate of about 7 times higher than the injected flow rate is created at the entrance to the gallery. These results were obtained numerically, under steady flow conditions and for a path shorter than the original gas turbine inlet manifold path, with lower pressure losses, for a 1:4 scale model.

For all geometric configurations considered, in the case of secondary air injection, after exiting the intake manifold nozzle, there is a tendency to concentrate the airflow toward the plane of symmetry of the manifold. Numerical calculations showed a decrease in airflow at the gallery outlet with an increase in the distance from the leading edge of the wall of the modified section to the linear nozzle, according to the first-order regression function. According to this observation, the

maximum effect was obtained for the exit of the nozzle in the turbine intake manifold according to the minimum section of the modified section, at $x = 30$ mm. The quadratic regression function has a maximum at $x = 44.24$ mm. Since these values are indicative, a value in this range can be adopted for the outlet of the nozzle in the gallery. Increasing the angle of inclination of the linear nozzle relative to the horizontal leads to an airflow decrease at the gallery outlet. Both regression laws indicate this, the maximum of the second-order regression being at the tilt angle of approximately 10° , which is practically impossible to achieve. For constructive reasons, an angle of 30° can be adopted. For values lower than this, the manifold is very close to the body of the intake manifold. Therefore, it was assumed that the experimental results confirmed the numerical results for the airflow produced by the fan, both for the scale model of the original gallery and for the scale model of the gallery with the modified section. The differences between the measured velocity values and the experimental ones are mainly due to the positioning error of the anemometer and the alteration of the measurement in the vicinity of the gallery walls. The directions in which the velocities were measured are located in the middle of the gallery, noting the existence of small differences in the velocity profile for the two configurations. This study aimed to constitute a validated method and model, with a high degree of generality for other similar situations.

REFERENCES

- [1] T. Giampaolo, *Gas Turbine Handbook: Principles and Practice, Fifth Edition*, 5th ed. New York, NY, USA: River Publishers, 2020.
- [2] "The Twin TF Marine Gas Turbine Propulsion Package: The Solution for High Performance," *Vericor*. <https://www.vericor.com/success-stories/the-twin-tf-marine-gas-turbine-propulsion-package-the-solution-for-high-performance/>.
- [3] M. Wilcox, R. Baldwin, A. Garcia-Hernandez, and K. Brun, "Guideline for gas turbine inlet air filtration systems," Gas Machinery Research Council - Southwest Research Institute, Apr. 2010.
- [4] G. I. Balan, O. N. Volintiru, I. C. Scurtu, F. Ioniță, M. L. Vasile, and C. Borzea, "Considerations regarding the anti-icing system for the ship propulsion plant with gas turbine," *E3S Web of Conferences*, vol. 286, 2021, Art. no. 04013, <https://doi.org/10.1051/e3sconf/202128604013>.
- [5] F. Niculescu, M. L. Vasile, G. Balan, A. Săvescu, and R. Nicolae, "Virtual Indication of the Torque for a Marine Gas Turbine | Technium: Romanian Journal of Applied Sciences and Technology," *Technium: Romanian Journal of Applied Sciences and Technology*, vol. 3, no. 10, pp. 74–81, Nov. 2021, <https://doi.org/10.47577/technium.v3i10.5140>.
- [6] I. Vlăduță et al., "Automation Control System for Naval Propulsion Retrofitting," in *2021 International Conference on Applied and Theoretical Electricity (ICATE)*, Craiova, Romania, Feb. 2021, pp. 1–6, <https://doi.org/10.1109/ICATE49685.2021.9465065>.
- [7] N. Aye, G. McAndrews, and B. Mendenhall, "Marine Gas Turbine Package for the Korean Navy PKX Program," presented at the ASME Turbo Expo 2008: Power for Land, Sea, and Air, Berlin, Germany, Aug. 2009, pp. 387–394, <https://doi.org/10.1115/GT2008-50098>.
- [8] P. Sun, H. Gao, J. Zhong, and M. Yang, "Aerodynamic Performance Investigation of Modern Marine Gas Turbine Intake System," presented at the ASME Turbo Expo 2012: Turbine Technical Conference and Exposition, Copenhagen, Denmark, Jul. 2013, pp. 367–373, <https://doi.org/10.1115/GT2012-69694>.
- [9] T. Zierer and B. Matyschok, "Design, Development and Verification of Gas Turbine GT24 Air Intake," presented at the ASME 1997 Turbo Asia Conference, Singapore, Dec. 2014, <https://doi.org/10.1115/97-AA-009>.
- [10] E. Avram, V. Nastasescu, and G. Malos, *Metoda Elementelor Finite in Mecanica Fluidelor*. Bucharest, Romania: Editura Academiei Tehnice Militare, 2003.
- [11] S. Stefan and V. Nastasescu, *Simularea fenomenelor mecanice și hidraulice*. Bucharest, Romania: Editura Academiei Tehnice Militare, 2000.
- [12] G. Balan, *Aerogazodinamica*. Chisinau, Moldova: Technica-Info, 2003.
- [13] "Heating, Ventilating, and Air-Conditioning Applications," ASHRAE, Atlanta, GA, USA, 2019.
- [14] V. Nastasescu, "The Using of the Multilayer Plate Concept in the Calculus of Functionally Graded Plates," *Applied Sciences*, vol. 12, no. 21, 2022, <https://doi.org/10.3390/app122110695>.
- [15] V. Nastasescu, A. Stefan, and C. Lupoiu, *Analiza neliniară a structurilor mecanice prin metoda elementelor finite*. Bucharest, Romania: Editura Academiei tehnice Militare, 2002.
- [16] P. K. Kundu and I. M. Cohen, *Fluid Mechanics*, 4th ed. Burlington, MA, USA: Academic Press.
- [17] D. C. Rennels and H. M. Hudson, *Pipe Flow: A Practical and Comprehensive Guide*, 1st ed. Wiley, 2012.
- [18] D. G. Zisopol, I. Nae, and A. I. Portoaca, "Compression Behavior of FFF Printed Parts Obtained by Varying Layer Height and Infill Percentage," *Engineering, Technology & Applied Science Research*, vol. 12, no. 6, pp. 9747–9751, Dec. 2022, <https://doi.org/10.48084/etasr.5488>.
- [19] D. G. Zisopol, A. I. Portoaca, I. Nae, and I. Ramadan, "A Comparative Analysis of the Mechanical Properties of Annealed PLA," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 8978–8981, Aug. 2022, <https://doi.org/10.48084/etasr.5123>.
- [20] D. G. Zisopol, I. Nae, A. I. Portoaca, and I. Ramadan, "A Statistical Approach of the Flexural Strength of PLA and ABS 3D Printed Parts," *Engineering, Technology & Applied Science Research*, vol. 12, no. 2, pp. 8248–8252, Apr. 2022, <https://doi.org/10.48084/etasr.4739>.
- [21] D. G. Zisopol, I. Nae, A. I. Portoaca, and I. Ramadan, "A Theoretical and Experimental Research on the Influence of FDM Parameters on Tensile Strength and Hardness of Parts Made of Polylactic Acid," *Engineering, Technology & Applied Science Research*, vol. 11, no. 4, pp. 7458–7463, Aug. 2021, <https://doi.org/10.48084/etasr.4311>.
- [22] D. G. Zisopol, A. I. Portoaca, and M. Tanase, "Improving the Impact Resistance through Annealing in PLA 3D Printed Parts," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11768–11772, Oct. 2023, <https://doi.org/10.48084/etasr.6281>.
- [23] D. G. Zisopol, A. I. Portoaca, and M. Tanase, "Dimensional Accuracy of 3D Printed Dog-bone Tensile Samples: A Case Study," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11400–11405, Aug. 2023, <https://doi.org/10.48084/etasr.6060>.
- [24] D. G. Zisopol, N. Ion, and A. I. Portoaca, "Comparison of the Charpy Resilience of Two 3D Printed Materials: A Study on the Impact Resistance of Plastic Parts," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10781–10784, Jun. 2023, <https://doi.org/10.48084/etasr.5876>.

Deep Learning Approaches for Age-based Gesture Classification in South Indian Sign Language

Ramesh M. Badiger

Department of Computer Science and Engineering, Tontadarya College of Engineering, India
rameshmbadiger@gmail.com

Rajesh Yakkundimath

Department of Computer Science and Engineering, KLE Institute of Technology, India
rajeshymath@gmail.com

Guruprasad Konnurmath

School of Computer Science and Engineering, KLE Technological University, Hubballi, India
guruprasad.konnurmath@kletech.ac.in

Praveen M. Dhulavvagol

School of Computer Science and Engineering, KLE Technological University, Hubballi, India
praveen.md@kletech.ac.in

Received: 4 January 2024 | Revised: 21 January 2024 | Accepted: 29 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6864>

ABSTRACT

This study focuses on recognizing and categorizing South Indian Sign Language gestures based on different age groups through transfer learning models. Sign language serves as a natural and expressive communication method for individuals with hearing impairments. The intention of this study is to develop deep transfer learning models, namely Inception-V3, VGG-16, and ResNet-50, to accurately identify and classify double-handed gestures in South Indian languages, like Kannada, Tamil, and Telugu. A dataset comprising 30,000 images of double-handed gestures, with 10,000 images for each considered age group (1-7, 8-25, and 25 and above), is utilized to enhance and modify the models for improved classification performance. Amongst the tested models, Inception-V3 achieves best performance with test precision of 95.20% and validation accuracy of 92.45%, demonstrating its effectiveness in accurately categorizing images of double-handed gestures into ten different classes.

Keywords-sign language; age group; gesture identification; transfer learning; Inception-V3; VGG-16; ResNet-50

I. INTRODUCTION

The rapid evolution of AI technologies has opened up significant opportunities to develop the life quality for individuals with disabilities. According to the World Health Organization, approximately 1.2 billion people, or 15% of the worldwide population, live with some type of disability. Recognizing the need for inclusivity, AI solutions to assist the disabled community in their daily activities have been leveraged. Consequently, AI, accessibility, and disabilities have become closely intertwined, with extensive research in AI and machine learning directed towards accommodating the daily lives of disabled individuals. AI has made significant strides in facilitating non-verbal communication for persons with impairments, including the recognition and understanding of

sign language and gestures. Sign language, which relies on hand gestures and body motions, serves as a vital means of communication with impaired people. Advancements in computer vision and pattern recognition have been instrumental in enabling AI systems to interpret and respond to sign language effectively. But, despite the recent progress, the fast and accurate recognition of hand gestures remains a challenging task [1]. Researchers are continuously working to get better performance of AI models in recognizing and categorizing these gestures. Gestures may vary in the speed, orientation, and hand alignment. Age category divisions for hand gesture performance are not rigidly defined, as there is considerable overlap and individual variations in hand gesture

usage. However, we can broadly categorize the latter into the following age groups:

- Early Childhood (1-7 years). Early Childhood refers to the early stages of development in which individuals have limited skills and communication abilities. They use simple gestures like reaching out for objects, waving, and pointing to objects.
- Adolescence (8-25 years). As children advance through middle childhood to adolescence, their gesture vocabulary expands further, and they become more adept at using gestures for social interaction.
- Adulthood (25+ years). Adults typically possess a fully developed gesture repertoire. They use gestures to enhance verbal communication, convey complex ideas, express emotions, and adhere to cultural norms and social cues.

Gesture accuracy gradually changes as individuals grow up. Table I provides a comprehensive list of double-handed gestures considered for the research work, alongside with their respective letters in various South Indian languages based on different age groups.

TABLE I. LABEL ENCODING/NUMERICAL CODE ASSIGNMENTS FOR THE CONSIDERED DOUBLE-HAND GESTURES

S. No.	Word in Kannada/Telugu/Tamil/English	Class label (alphanumeric code) in years		
		1-7	8-25	≥ 25
1	ಮೀನು / చేప / மீன் /Fish	1D	2D	3D
2	ಯಾವುದು / ఏదీ / எப்படி /Which	1E	2E	3E
3	ಅಡಿ / అడుగు / ஆடி /Foot	1F	2F	3F
4	ಚರ್ಮ / చర్మం / தோல் /Skin	1J	2J	3J
5	ನವಿಲು / నేమలీ / மயில் /Peacock	1M	2M	3M
6	ಸಮಯ / సమయం / நேரம் /Time	1N	2N	3N
7	ಉದ್ದ / పొడవు / நீளம் /Length	1P	2P	3P
8	ಚಿಟ್ಟೆ / తేలికబుట్టి / பூறா /Butterfly	1T	2T	3T
9	ಮನೆ / ఇల్లు / வீடு /House	1W	2W	3W
10	ಏನು / ఏమి / என்ன /What	1X	2X	3X

The literature survey on pattern recognition applications in the identification of sign language gestures reveals the following key findings. Authors in [2] employed a time-distributed Convolution Neural Network (CNN) and a Gated Recurrent Unit (GRU) to extract skin features from hand images. The models achieved a performance accuracy of 96.5% in categorizing hand images into 17 age groups ranging from 18 to 75 years old. Authors in [3] utilized MRELBP to extract features from right and left hand images. The images were classified by person, age, and gender, reporting accuracies of 91.4%, 85.9%, and 92.6%, respectively. Authors in [4] presented age detection on an image dataset using combinations of deep learning and image processing techniques, achieving an accuracy of 91%. The literature survey shows that noteworthy research has been explored in the area of sign language and hand gesture identification, with a particular emphasis given on deep learning techniques for recognizing human poses and scenes from images [5-9]. However, there is a noticeable gap in the identification of South Indian Sign Language (SISL) gesture images regarding

different age groups [10]. This research article aims to address this issue and provide a novel method for recognizing SISL gestures across various age groups.

II. PROPOSED METHODOLOGY

The proposed method consists of two primary stages: the first involves organizing the image dataset into various age groups, while the second entails employing deep learning methods to classify the gestures; Figure 1 illustrates a block diagram showcasing the stages involved.

A. Data Preparation

The dataset preparation process involves selecting 10 categories of double-hand gestures based on different age groups. High-resolution images were captured using a Nikon D3300 camera with a resolution of 24.2 megapixels. The images were taken against black or green backgrounds under natural lighting conditions. Each image has an original size of 1080×2400 pixels. For the initial age group (age ranging from 1 to 7), the dataset consists of 5000 images, with 500 images per gesture category. Image extension methods were applied to expand the dataset to 10,000 images, enhancing its diversity. To ensure efficient processing and storage, all images were resized to 300×300 pixels. The dataset was divided into three subsets: training (70%), validation (15%), and testing (15%).

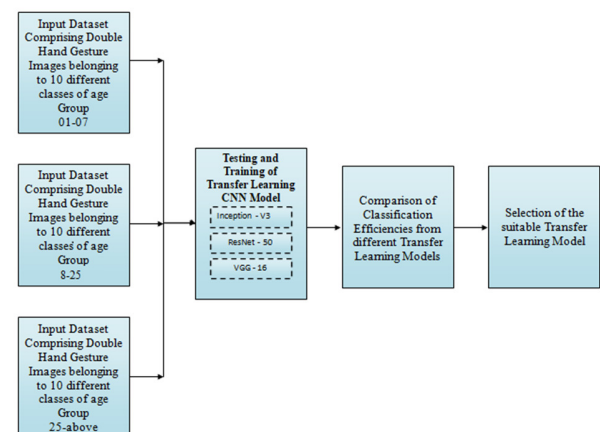


Fig. 1. Block diagram of the proposed methodology.

The training subset contains 7000 images, while the validation and testing subsets contain 1500 images each. This division ensures the acquisition of enough data for the training and reliable estimation of the models. The same methodology was applied to the other age groups. By following this systematic approach, a comprehensive dataset of double-hand gestures is prepared, covering various age groups, facilitating the development, and estimating the accuracy of recognition models. Figure 2 displays a sample of the utilized dataset.

B. CNN Classifiers

In this study, three prominent transfer learning CNN [11] models are utilized for the recognition and classification of double-handed gesture images. These models include Inception-V3 [12], VGG-16 [13], and ResNet-50 [14], acknowledged for their effectiveness in image recognition

tasks. The training process of the CNN models on the prepared dataset follows a specific set of steps, as demonstrated in the following Algorithm.

Algorithm: Training Steps for CNN Models
Inputs: Prepared dataset of double-handed gesture images, Chosen CNN models (Inception-V3, VGG-16, or ResNet-50), and

hyper parameters: learning rate, batch size, number of epochs.

Outputs: Trained CNN model

Start

Step 1. Load the prepared dataset

Step 2. Initialize the chosen CNN model

Step 3. Set the hyper parameters for training

Step 4. Split the dataset into training, validation, and testing subsets

Step 5. Training Phase:

for each epoch in the specific number of epochs

do:

for each batch in the training subset

do:

Load a batch of images and their corresponding labels

Perform forward propagation through the CNN model

Compute the loss using the loss function

Perform backward propagation to calculate the gradients

Update the model's parameters using the optimization algorithm

Step 6. Validation Phase:

for each batch in the validation subset

do:

Load a batch of images and their corresponding labels

Perform forward propagation through the CNN model

Compute the validation loss and accuracy

Step 7. Evaluation:

After training, evaluate the trained model on the testing subset

do:

for each batch in the testing subset

do:

Load a batch of images and their corresponding labels

Perform forward propagation through the trained CNN model

Compute the testing accuracy

Step 8. Fine-tuning and Hyper parameter Tuning:

Adjust hyper parameters and fine-tune the model based on the validation results

Step 9. Repeat steps 5-8 until

satisfactory performance is achieved

Step 10. Save the trained model for future use in gesture recognition and classification tasks

Stop.



Fig. 2. Dataset samples.

III. EXPERIMENTAL RESULTS AND DISCUSSION

The experiments on double-handed gesture classification for different age categories were conducted using the Deep Learning Toolbox given by the MATLAB R2022b platform. The pre-trained CNN models employed in the experiments were imported and prepared for transfer learning by modifying the properties of suitable layers utilizing the Deep Network Designer application. Specifically, the last learnable layer and the output or classification layer were replaced to align with the classes in the newly constructed double-handed gesture image dataset. To control the training process, specific options were set for the CNN models. The initial learning rate, validation frequency, number of epochs, and mini-batch size were initialized to 0.0001, 10, 30, and 35, respectively. These values were chosen based on experimentation and empirical knowledge to achieve optimal training performance. In terms of activation functions, all hidden layers in the CNN models were activated implementing the Rectified Linear Unit (ReLU) function, which has been usually used in deep learning due to its ability to introduce non-linearity and handle vanishing gradients. The output layer, responsible for classification, was activated utilizing the softmax function, which produces a probability distribution over the different classes. To fine-tune the network and optimize its performance, the Stochastic Gradient Descent (SGD) algorithm was employed as the optimization algorithm. The customized CNN models were trained and validated applying the prepared augmented image dataset. The training progress of each CNN model, along with the corresponding validation accuracy and loss, is monitored and visualized in Figures 3 to 5. These figures provide a graphical representation of the training process, allowing for a better knowledge of the model's performance and convergence.

TABLE II. EVALUATION METRICS DERIVED FROM THE CONFUSION MATRICES OF THE CONSIDERED CNN MODELS

CNN model	Average performance metrics across all the single-hand gesture classes			Average validation accuracy (%)	Average test accuracy (%)
	Precision	Recall	F1 Score		
Inception-V3	0.9520	0.9520	0.9500	95.20%	96.10%
ResNet-50	0.9247	0.9253	0.923	92.50%	93.20%
VGG-16	0.9027	0.903	0.901	90.20%	91.70%

It can be seen that the Inception-V3 model demonstrated the best performance among all the considered pre-trained CNN models. It achieved the highest evaluation metrics, as well as the maximum validation and testing accuracies for different age groups. The high accuracy values of the Inception-V3 model indicate its efficacy in precisely classifying double-hand gestures based on age groups. The superior performance of the Inception-V3 model suggests its appropriateness for recognizing and categorizing South Indian Sign Language gestures across different age groups. The performance comparison results of all the considered pre-trained CNN models can be seen in Figure 7.

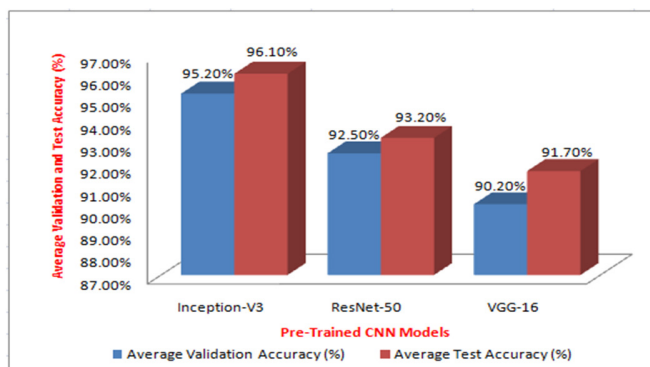


Fig. 7. Performance evaluation results of all the pre-trained CNN models.

IV. CONCLUSION

The current study utilized popular pre-trained CNN models, namely Inception-V3, ResNet-50, and VGG-16, to classify double-hand gesture images across 30 different classes, focusing on 10 signs for different age categories (1–7, 8–25, and 25 and above). These models were customized and fine-tuned to accommodate the specific image classes and improve classification performance. The study results demonstrated impressive performance across all three models. However, Inception-V3 emerged as the top-performing model, achieving an average classification accuracy of 95.20%, indicating its efficiency in accurately classifying double-hand gestures based on age categories. The outcomes of this work have potential applications in constructing automated systems that can identify South Indian sign language gestures from both still images and streaming videos [15–17]. By leveraging these advanced CNN models, communication barriers can be reduced, allowing for easier and more effective communication with the outside world. Future research endeavors could

explore the use of the latest CNN models and incorporate publicly available image datasets to further enhance the image dataset employed in this study.

REFERENCES

- [1] H. G. Doan and N. T. Nguyen, "Fusion Machine Learning Strategies for Multi-modal Sensor-based Hand Gesture Recognition," *Engineering, Technology & Applied Science Research*, vol. 12, no. 3, pp. 8628–8633, Jun. 2022, <https://doi.org/10.48084/etasr.4913>.
- [2] M. A. Abderrahmane, I. Guelzim, and A. A. Abdelouahad, "Human Age Prediction Based on Hand Image using Multiclass Classification," in *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, Sakheer, Bahrain, Jul. 2020, <https://doi.org/10.1109/ICDABI51230.2020.9325591>.
- [3] E. Aydemir and R. Alalawi, "Classification Of Hand Images by Person, Age and Gender with The Median Robust Extended Local Binary Model," *Balkan Journal of Electrical and Computer Engineering*, vol. 11, no. 1, pp. 78–87, Jan. 2023, <https://doi.org/10.17694/bajece.1171905>.
- [4] G. Bakshi, A. Aggarwal, D. Sahu, R. R. Baranwal, G. Dhall, and M. Kapoor, "Age, Gender, and Gesture Classification Using Open-Source Computer Vision," in *Emerging Technologies in Data Mining and Information Security*, Singapore, 2023, pp. 63–73, https://doi.org/10.1007/978-981-19-4052-1_8.
- [5] S. K. Gupta and N. Nain, "Review: Single attribute and multi attribute facial gender and age estimation," *Multimedia Tools and Applications*, vol. 82, no. 1, pp. 1289–1311, Jan. 2023, <https://doi.org/10.1007/s11042-022-12678-6>.
- [6] W.-B. Horng, C.-P. Lee, and C.-W. Chen, "Classification of Age Groups Based on Facial Features," *Tamkang Journal of Science and Engineering*, vol. 4, no. 3, pp. 183–192, Sep. 2001.
- [7] K. R. "Deep Learning for Age Group Classification System," *International Journal of Advances in Signal and Image Sciences*, vol. 4, no. 2, pp. 16–22, Dec. 2018, <https://doi.org/10.29284/ijasis.4.2.2018.16-22>.
- [8] M. Madhvarasan and P. P. Roy, "A Comprehensive Review of Sign Language Recognition: Different Types, Modalities, and Datasets," *arXiv*, Apr. 7, 2022, <https://doi.org/10.48550/arXiv.2204.03328>.
- [9] N. S. Russel and A. Selvaraj, "Gender discrimination, age group classification and carried object recognition from gait energy image using fusion of parallel convolutional neural network," *IET Image Processing*, vol. 15, no. 1, pp. 239–251, 2021, <https://doi.org/10.1049/ipr2.12024>.
- [10] R. M. Badiger and D. Lamani, "Deep Learning Based South Indian Sign Language Recognition by Stacked Autoencoder Model and Ensemble Classifier on Still Images and Videos," *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 21, pp. 6587–6597, 2022.
- [11] N. C. Kundur, B. C. Anil, P. M. Dhulavvagol, R. Ganiger, and B. Ramadoss, "Pneumonia Detection in Chest X-Rays using Transfer Learning and TPUs," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11878–11883, Oct. 2023, <https://doi.org/10.48084/etasr.6335>.
- [12] N. Kumar, A. Hashmi, M. Gupta, and A. Kundu, "Automatic Diagnosis of Covid-19 Related Pneumonia from CXR and CT-Scan Images," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 7993–7997, Feb. 2022, <https://doi.org/10.48084/etasr.4613>.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778, <https://doi.org/10.1109/CVPR.2016.90>.
- [14] S. Khan, I. Ali, F. Ghaffar, and Q. Mazhar-ul-Haq, "Classification of Macromolecules Based on Amino Acid Sequences Using Deep Learning," *Engineering, Technology & Applied Science Research*, vol. 12, no. 6, pp. 9491–9495, Dec. 2022, <https://doi.org/10.48084/etasr.5230>.
- [15] G. Konnurmath and S. Chickerur, "An Investigation into Power Aware Aspects of Rendering 3D Models on Multi-Core Processors," *Procedia*

Computer Science, vol. 218, pp. 887–898, Jan. 2023, <https://doi.org/10.1016/j.procs.2023.01.069>.

- [16] G. Konnurmath and S. Chickerur, "GPU Shader Analysis and Power Optimization Model," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12925–12930, 2024, <https://doi.org/10.48084/etasr.6695>.
- [17] G. Konnurmath and S. Chickerur, "Power-Aware Characteristics of Matrix Operations on Multicores," *Applied Artificial Intelligence*, vol. 35, no. 15, pp. 2102–2123, Dec. 2021, <https://doi.org/10.1080/08839514.2021.1999013>.

Utilizing Extremely Fast Decision Tree (EFDT) Algorithm to Categorize Conflict Flow on a Software-Defined Network (SDN) Controller

Mutaz H. H. Khairi

Future University, Sudan
Taza1040@gmail.com

Bushra Mohammed Ali Abdalla

Faculty of Computer Science and Information Technology, Ibn Sina University, Sudan
bushra0912115@gmail.com (corresponding author)

Mohamed Khalafalla Hassan

University Technology Malaysia, Malaysia | Future University, Sudan
mohamed.khalafalla.hassan@gmail.com

Sharifah H. S. Ariffin

University Technology Malaysia, Malaysia
shafizah@utm.my

Mosab Hamdan

Interdisciplinary Research Center for Intelligent Secure Systems, King Fahd University of Petroleum and Minerals, Saudi Arabia
mosab.mohamed@kfupm.edu.sa

Received: 21 December 2023 | Revised: 11 January 2024 and 27 January 2024 | Accepted: 29 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6793>

ABSTRACT

Software-Defined Networks (SDNs) provide a contemporary approach to networking technology, offering a versatile and dynamically efficient network architecture for enhanced surveillance and performance. However, SDN architectures may encounter flow conflicts. These conflicts arise when modifications are made to specific flow properties, such as priority, match field, and action. Despite the existence of recommended solutions, the process of resolving conflicts in SDN continues to encounter difficulties. This study proposes an Extremely Fast Decision Tree (EFDT) classification technique to detect and categorize conflicts inside the flow table. The novelty of this method is based on the development of an accurate and effective machine-learning technique implemented on the Ryu controller plane and validated using the Mininet simulator. The effectiveness and efficiency of the proposed method were evaluated using various indicators, demonstrating superior performance in recognizing and categorizing conflict flow types in all flow sizes ranging from 10,000 to 100,000.

Keywords-*software-defined network; conflict flow; machine learning; extremely fast decision tree*

I. INTRODUCTION

Software-Defined Networks (SDNs) can improve network performance through dynamic and customizable network design [1]. This architecture enables easy modifications using a centralized control console for network engineers and administrators to meet the evolving business needs. It also enables easy modifications using a centralized control console for network engineers and administrators to meet evolving

business needs [2]. SDNs integrate network technologies to enhance flexibility and agility by separating control functions from forwarding planes [3]. This split could allow independent configuration of the network control plane, giving system specialists complete control over its operations [4-5]. In addition to its fundamental advantages, an SDN is dynamic, controllable, adaptive, and cost-effective, making it a perfect solution for the ever-expanding size and high-bandwidth

applications of the Internet [6]. OpenFlow is a key advancement in SDN. The controller is a vital element in the architecture, enabling the creation of diverse applications through an Application Programming Interface (API). The controller's activities influence the success of an SDN. An OpenFlow switch has multiple flow tables linked to the controller via the OpenFlow protocol to distribute, classify, and assign packets based on flow entries [7-11].

SDN systems can be optimized using Machine Learning (ML) to handle data more effectively when comprehending and extracting patterns from data is difficult [12]. With the increase in available datasets, ML is becoming more common in fields such as medicine and government that require relevant data [13-14]. ML aims to learn from data, and many studies have been conducted to improve such methods [15]. The Transaction Conflict Detection and Resolution (TCDR) framework was created to eliminate flow policy conflicts, which improves controller layer performance while maintaining low cost [16]. In [17], a new technique was proposed to anticipate conflicts while evaluating the effectiveness of an application. The algorithm predicts network failure by analyzing remaining actions, based on the guidelines created by the developer. Conflict prediction identifies unwanted system behavior, and prediction-related capabilities should be incorporated based on the evaluation of the prediction model to ensure practical control applications. In [18], cross-layering between flow components was investigated using OpenFlow table variables, and an accurate approach was recommended by analyzing the entire input table.

This study examines seven types of conflict, including redundancy, shadowing, overlapping, correlation (A and B), generalization, and imbrication using conflict rules derived from relevant studies as a guide [19-21]. Flow conflicts in an OpenFlow switch can be categorized by type, priority, action, protocol, and IP source address. These conflicts can have a significant impact on SDN operations in a variety of scenarios. This study aims to reduce the negative impact of conflicts in the SDN environment. To achieve this, the Extremely Fast Decision Tree (EFDT) algorithm was used, which depends on the decision tree structure. This study introduces several novel principles, including the EFDT algorithm, which identifies flow conflicts in SDN. The algorithm is used with varying amounts of flows to identify and categorize conflicting flows. To the best of our knowledge, this is the first study that applies ML algorithms to classify seven distinct categories of conflict flows.

II. PROBLEM DEFINITION

Multiple conflict types adversely influence efficiency across conventional networks and SDNs [20]. Two primary categories of conflicts are based on their rules and outcomes: Interpretative conflicts (generalization, correlation, imbrication) and intelligible conflicts (redundancy, shadowing, overlapping). This study focuses on discussing the limitations of flow entry in terms of priority, match fields, and action fields. Packet counters and timeout values are not critical in handling flow conflicts. Conflicts in SDN can arise due to changes made to attributes such as action and priority. Conflicts may occur in the controller and flow table depending

on changes made to the flow rule policies or entries. It is essential to identify priorities and take action to create SDN rules and flow entries [10, 22]. Traditional networks and SDNs are widely recognized to differ significantly in features, particularly priority and action [21]. Additionally, the flow table can create conflicts in various situations, including inconsistencies in rules caused by the following factors [20]:

- Conflicting behavior may occur when the network management system addresses similar flows in multiple tables.
- VPN programs that modify header information may unintentionally apply flow rules to a particular flow.
- The injection of different subsystems, which used the northbound as primary sources from a controller, may cause the flow rule to exhibit conflicting behavior for identical flows.

III. OBJECTIVES

This study aims to develop a method for recognizing flow conflicts in SDNs. The proposed algorithm will be capable of identifying and categorizing each flow in the OpenFlow switch, which will help reduce conflicts among the numerous flows. The proposed method aims to achieve the following specific goals:

- Creating and integrating the EFDT algorithm into the SDN Ryu controller.
- Utilizing the conflict classification technique to categorize and identify various conflicting flows.

IV. MATERIALS AND METHODS

A. Environment Setup for SDN Platform

The proposed SDN solution was built using Ryu, a network controller, in combination with the Mininet simulator. Mininet can be operated in VirtualBox, which is one of the virtual network simulators. However, it is crucial to consider and prepare for the specific requirements related to the virtual machine, SDN controller, and OpenFlow switch version. Table I provides a list of specifications for the simulation environment.

TABLE I. SYSTEM AND ENVIRONMENT SPECIFICATION

Software and Hardware	Specifications of Machine
Processor	Core i7
RAM	16 GB
Operating system	Ubuntu 18.04
SDN controller	Ryu
Programming language	Python 2.7
OpenFlow Switch	Version 1.3

B. SDN Topology

The fat-tree topology was utilized and integrated into a mininet connected to the Ryu controller. The Topo.py application established the topology of the switches and hosts. Similar topologies have been used in previous studies on SDN conflict flows, including those undertaken to pursue knowledge [18, 20, 22-23].

C. Dataset

The data set used was the one used in [18]. Iperf is a tool that generates and collects OpenFlow data in a Ryu controller topology, including ten performance servers per host. Each server can listen to a different destination port, and new flows are generated based on source and destination IP addresses and source and destination ports and protocols. The controller is responsible for installing a new flow on the switch, updating policy rules, and capturing all flow entries in a CSV file. The analysis proposed a new method to create flows in an SDN controller. This was necessary because there is a lack of SDN datasets that have important flow entry properties such as priority and action features. The algorithm was developed and executed using OpenFlow Switch version 1.3 and the Ryu controller. The dataset used in the study was created by extracting flow entries from the switch OpenFlow table. These flow entries represent the network flows identified and processed by the switch. The study analyzed this dataset to gain insight into network behavior and identify potential performance issues or anomalies. Figure 1 provides a visual representation of the dataset used in the study.

```
cookie=0x6b, duration=41.248s, table=0, n_packets=3, n_bytes=162, priority=100, tcp, dl_src=00:00:00:00:00:02, dl_dst=00:00:00:00:00:04, nw_src=10.6.1.0/24, nw_dst=10.211.2.65, tp_src=5060, tp_dst=8080 actions=output:"s1-eth2"
cookie=0x6c, duration=41.248s, table=0, n_packets=3, n_bytes=162, priority=100, tcp, dl_src=00:00:00:00:00:02, dl_dst=00:00:00:00:00:04, nw_src=10.6.1.0/24, nw_dst=10.211.2.65, tp_src=5060, tp_dst=8085 actions=output:"s1-eth2"
```

Fig. 1. Flow entries from the OpenFlow table in the switch.

D. Conflict Classification Model

The proposed method involves classifying the conflict categories present in conflict flows across all flow data sizes using the Conflict Classification Model (CCM). The EFDT, combined with the Hoeffding Anytime Tree SEA Generator, is an innovative learning technique that consistently exceeds the traditional decision tree approach. The EFDT has been shown to outperform the Hoeffding Tree version of the Very Fast Decision Tree (VFDT) on multiple standard evaluation workloads, specifically with respect to prequential reliability. This impressive performance highlights the remarkable potential of the EFDT for application in various business and academic settings. The EFDT's ability to produce more reliable outcomes than traditional decision tree methods makes it a compelling option, especially in time-sensitive situations where high accuracy is critical. As such, the EFDT is a valuable addition to the range of data-driven techniques available for decision-making. The CCM comprises four primary activities: compiling conflict flow data, setting up the Ryu controller to use the implemented classification algorithm, learning the EFDT algorithm with various conflicts, and evaluating the algorithm. Figure 2 illustrates the design and implementation of the CCM. Gathering and classifying conflict flows is the first stage in the CCM process. Next, two distinct categories are created from the flows gathered using preprocessing: 70% for training and 30% for testing the EFDT classification algorithm. Then, each of the conflict types is classified using the classification method. The following steps are used to organize

conflict flows and serve as a guide for designing the classification algorithm.

- Step 1: Implement and execute the EFDT classification algorithm.
- Step 2: Start by analyzing the detection flows.
- Step 3: Check the priority features of conflicting flows.
- Step 4: Check the conflicting flows' Internet Protocol (IP) addresses.
- Step 5: Based on the results of Steps 4 and 5, categorize the different types of conflict.

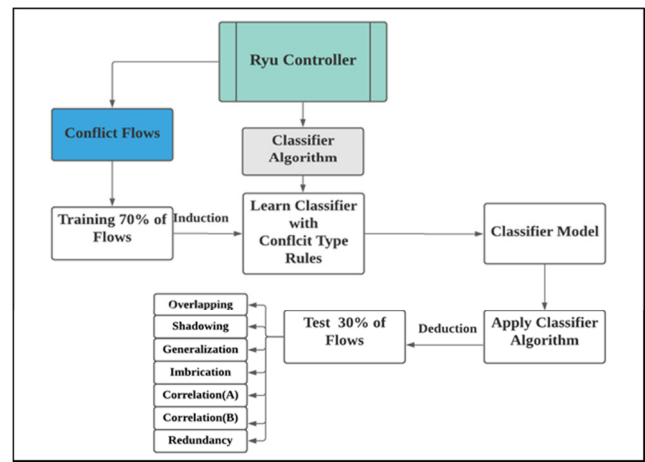


Fig. 2. Conflict classification model.

E. Implementation of the Classification Algorithm

The EFDT algorithm was integrated into the Ryu controller to identify conflicting flows. The algorithm creates a conditional filter that is used to identify flows that may conflict with each other. Once these flows are identified, the Hoeffding tree is used to categorize them into one of the seven conflict types. The criteria used to categorize flows into these conflict types are based on several factors, such as type of traffic, priority of the flows, and current state of the network. The implementation steps of the EFDT algorithm in the Ryu controller can be summarized as follows:

- Create the SEA Generator with decision tree elements.
- Integrate the classifier with the Hoeffding tree.
- Set up the Hoeffding Tree estimator to verify the actions and IP address rules for the produced flows and implement any required adjustments.
- Develop new parameters to supervise the cycle to verify actions and IP address rules.
- Retrieve and include all streams produced in the OpenFlow switch for various stream sizes and make it ready for deployment.
- Utilize 70% of the generated flows to train the EFDT algorithm and 30% to test it.

F. Classification Process

Figure 3 presents the process used to identify seven different types of conflict. The first step is to collect all data flows. Then, the EFDT classification method is applied to the data on conflict flow to classify the different types of conflict at the topology controller level. The conflict flow data are divided into two groups and the algorithm is trained to learn all conflict rules for each type of conflict. This is done to allow the prediction algorithm to be prepared and evaluated.

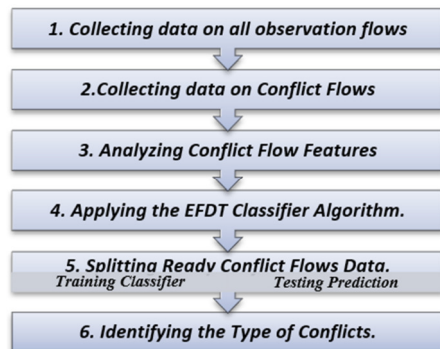


Fig. 3. Classification process.

V. RESULTS AND DISCUSSION

Figure 4 displays the results of the EFDT classifier's production for different flow conflict sizes. The results show that all seven conflict categories were identified correctly, and no conflicts were missed. The classifier accurately identified all conflict flows of various sizes. The two-line sets represent conflicts related to shadowing, overlapping, generalization, imbrication, redundancy, correlation (A), and correlation (B), including shadowing and overlapping conflicts. All conflicts were classified linearly.

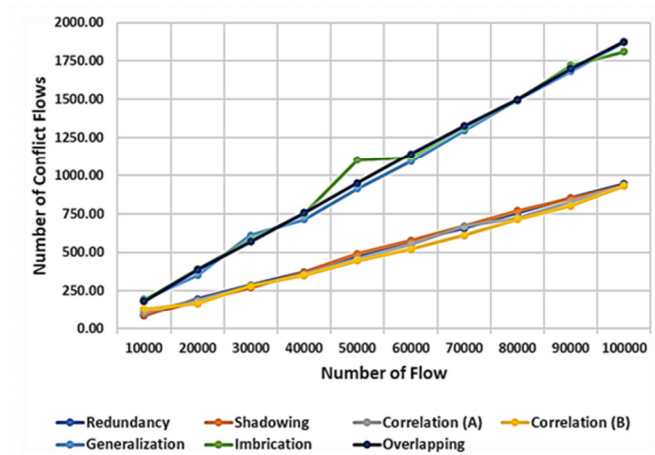


Fig. 4. Types of conflict classified by EFDT.

Figure 4 shows that different types of imbrication conflicts arise in flows containing 50,000 entries. This is a critical observation because the number of conflicts has been increasing during the implementation phase. As a result, the

number of conflicts has also increased to ensure that each conflict flow is handled appropriately. During the OpenFlow transition, the flow entries are transferred to a new set of tables, and the size of the flow tables is significant in this case of 50,000 flows. One of the primary criteria used to identify the connection between two sets of flow tables is the MAC address of the flow entries involved. The MAC address is also used to implement several imbrication conflicts, which are classified into seven conflict types by the EFDT algorithm.

Tables II and III extensively investigate the seven conflict groups identified using the EFDT classification approach. These conflict groups were derived using conflict flow data of different sizes. The initial row of the data table displays the aggregate count of flows managed by the OpenFlow switch. In contrast, the remaining rows show the distribution of the discovered conflicts among the seven distinct conflicts categorized by the EFDT and its classification method. As the volume of flow data increases for each conflict type, the number of conflict types also increases. Moreover, the results indicate that the classification method accurately categorizes all flow data into specific conflict types. This is remarkably accurate when considering the rising number of conflict kinds. Tables II and III present the flows classified by the EFDT algorithm for flow ranges of 10,000-50,000 and 60,000-100,000, respectively. A comprehensive analysis was performed to compare the effectiveness of two algorithms in categorizing various types of conflict: the proposed EFDT and the Brew security [14] algorithms. The performance of these algorithms was evaluated using a dataset ranging from 10,000 to 100,000 flows for learning and classification purposes. The results showed that the EFDT algorithm outperformed the Brew security algorithm in classification. Moreover, the EFDT algorithm could identify and categorize seven distinct forms of conflict, while the prior method could only detect six of them.

TABLE II. NUMBER OF FLOWS CLASSIFIED BY THE EFDT ALGORITHM FOR (10000-50000) FLOWS.

Number of flows	10000	20000	30000	40000	50000
Redundancy	85	198	287	373	477
Shadowing	89	185	272	372	491
Correlation (A)	110	190	282	360	458
Correlation (B)	131	169	283	351	446
Generalization	195	353	611	714	914
Imbrication	190	392	580	751	1011
Overlapping	182	387	569	756	949

TABLE III. NUMBER OF FLOWS CLASSIFIED BY THE EFDT ALGORITHM FOR (60000-100000) FLOWS.

Number of flows	60000	70000	80000	90000	100000
Redundancy	537	656	755	853	943
Shadowing	577	669	770	849	933
Correlation (A)	550	670	722	826	934
Correlation (B)	521	610	714	801	930
Generalization	1093	1297	1493	1683	1873
Imbrication	1117	1319	1496	1721	1811
Overlapping	1141	1325	1498	1700	1877

VI. CONCLUSION

This study introduces a new ML approach, called EFDT, that efficiently classifies conflict flows in the SDN

infrastructure. The algorithm identifies different types of conflict based on flow rule priority, action, protocol, and IP source address. The algorithm was developed using the Hoeffding tree function to enhance its effectiveness and efficiency. A Mininet emulation and the Ryu controller were used to connect a fat tree topology and carry out the experiments. The experiment involved selecting flows from 10,000 to 100,000, increasing 10,000 flows per dataset. According to the results obtained, the EFDT algorithm demonstrated superior classification performance compared to the Brew security algorithm. This means that the EFDT algorithm was better at identifying and categorizing the various conflict types in the dataset. The EFDT algorithm identified and categorized seven distinct forms of conflict, while the Brew security algorithm could only detect six types. This indicates that the EFDT algorithm is more comprehensive and effective in analyzing complex conflict scenarios, which could be helpful in various fields such as social sciences, conflict resolution, and international relations. This study is the first attempt to classify multiple conflict flows in SDN using such methods. Further studies will explore other ML techniques and evaluate them on the same dataset.

REFERENCES

- [1] S. Bera, S. Misra, and A. V. Vasilakos, "Software-Defined Networking for Internet of Things: A Survey," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1994–2008, Sep. 2017, <https://doi.org/10.1109/JIOT.2017.2746186>.
- [2] M. H. H. Khairi, S. H. S. Ariffin, N. M. A. Latiff, A. S. Abdullah, and M. K. Hassan, "A Review of Anomaly Detection Techniques and Distributed Denial of Service (DDoS) on Software Defined Network (SDN)," *Engineering, Technology & Applied Science Research*, vol. 8, no. 2, pp. 2724–2730, Apr. 2018, <https://doi.org/10.48084/etasr.1840>.
- [3] M. Karakus and A. Duresi, "A survey: Control plane scalability issues and approaches in Software-Defined Networking (SDN)," *Computer Networks*, vol. 112, pp. 279–293, Jan. 2017, <https://doi.org/10.1016/j.comnet.2016.11.017>.
- [4] E. T. B. Hong and C. Y. Wey, "An optimized flow management mechanism in OpenFlow network," in *2017 International Conference on Information Networking (ICOIN)*, Da Nang, Vietnam, Jan. 2017, pp. 143–147, <https://doi.org/10.1109/ICOIN.2017.7899493>.
- [5] M. H. H. Khairi, P. I. D. S. H. S. Ariffin, P. M. D. N. M. A. Latiff, D. K. M. Yusof, and M. K. Hassan, "A Review of Flow Conflicts and Solutions in Software Defined Networks (SDN)," *IJUM Engineering Journal*, vol. 22, no. 2, pp. 178–187, Jul. 2021, <https://doi.org/10.31436/iijum.v22i2.1613>.
- [6] P. P. Ray and N. Kumar, "SDN/NFV architectures for edge-cloud oriented IoT: A systematic review," *Computer Communications*, vol. 169, pp. 129–153, Mar. 2021, <https://doi.org/10.1016/j.comcom.2021.01.018>.
- [7] W. Hao, Y. Jiang, and J. Gao, "Detection mechanisms of rule conflicts in SDN based on a path-tree model," in *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, Aug. 2017, pp. 336–339, <https://doi.org/10.1109/ICSESS.2017.8342927>.
- [8] M. S. Tok and M. Demirci, "Security analysis of SDN controller-based DHCP services and attack mitigation with DHCPguard," *Computers & Security*, vol. 109, Oct. 2021, Art. no. 102394, <https://doi.org/10.1016/j.cose.2021.102394>.
- [9] C. N. Tran and V. Danciu, "A General Approach to Conflict Detection in Software-Defined Networks," *SN Computer Science*, vol. 1, no. 1, Jul. 2019, Art. no. 9, <https://doi.org/10.1007/s42979-019-0009-9>.
- [10] M. K. Hassan, S. H. S. Ariffin, S. K. Syed-Yusof, N. E. Ghazali, and K. A. Obeng, "A Short Review on the Dynamic Slice Management in Software-Defined Network Virtualization," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12074–12079, Dec. 2023, <https://doi.org/10.48084/etasr.6394>.
- [11] M. H. H. Khairi *et al.*, "The Impact of conflict flows on TCP And UDP Transfer Rate in Software Defined Network," *Innovative Networking Technologies Series 1*, no. 978, 2022.
- [12] M. K. Hassan, A. Babiker, M. Baker, and M. Hamad, "SLA Management For Virtual Machine Live Migration Using Machine Learning with Modified Kernel and Statistical Approach," *Engineering, Technology & Applied Science Research*, vol. 8, no. 1, pp. 2459–2463, Feb. 2018, <https://doi.org/10.48084/etasr.1692>.
- [13] M. K. Hassan *et al.*, "DLVisor: Dynamic Learning Hypervisor for Software Defined Network," *IEEE Access*, vol. 11, pp. 84144–84167, 2023, <https://doi.org/10.1109/ACCESS.2023.3302266>.
- [14] B. Mahesh, "Machine Learning Algorithms - A Review," *International Journal of Science and Research*, vol. 9, no. 1, pp. 381–386, 2018.
- [15] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A Survey on Bias and Fairness in Machine Learning," *ACM Computing Surveys*, vol. 54, no. 6, Apr. 2021, Art. no. 115, <https://doi.org/10.1145/3457607>.
- [16] J. Cui, S. Zhou, H. Zhong, Y. Xu, and K. Sha, "Transaction-Based Flow Rule Conflict Detection and Resolution in SDN," in *2018 27th International Conference on Computer Communication and Networks (ICCCN)*, Hangzhou, China, Jul. 2018, pp. 1–9, <https://doi.org/10.1109/ICCCN.2018.8487415>.
- [17] V. Danciu and C. N. Tran, "Side-Effects Causing Hidden Conflicts in Software-Defined Networks," *SN Computer Science*, vol. 1, no. 5, Aug. 2020, Art. no. 278, <https://doi.org/10.1007/s42979-020-00282-0>.
- [18] M. H. H. Khairi, S. H. S. Ariffin, N. M. A. Latiff, and K. M. Yusof, "Generation and collection of data for normal and conflicting flows in software defined network flow table," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 1, pp. 307–314, Apr. 2021, <https://doi.org/10.11591/ijeecs.v22.i1.pp307-314>.
- [19] R. Aryan, A. Yazidi, P. E. Engelstad, and Ø. Kure, "A General Formalism for Defining and Detecting OpenFlow Rule Anomalies," in *2017 IEEE 42nd Conference on Local Computer Networks (LCN)*, Singapore, Jul. 2017, pp. 426–434, <https://doi.org/10.1109/LCN.2017.94>.
- [20] S. Pisharody, J. Natarajan, A. Chowdhary, A. Alshalan, and D. Huang, "Brew: A Security Policy Analysis Framework for Distributed SDN-Based Cloud Environments," *IEEE Transactions on Dependable and Secure Computing*, vol. 16, no. 6, pp. 1011–1025, Aug. 2019, <https://doi.org/10.1109/TDSC.2017.2726066>.
- [21] Mutaz Hamed Hussien Khairi, "Flow Conflict Eliminations through Machine Learning for Software Defined Network," Ph.D. dissertation, Universiti Teknologi Malaysia, 2021.
- [22] M. H. H. Khairi *et al.*, "Detection and Classification of Conflict Flows in SDN Using Machine Learning Algorithms," *IEEE Access*, vol. 9, pp. 76024–76037, 2021, <https://doi.org/10.1109/ACCESS.2021.3081629>.
- [23] M. Hamdan *et al.*, "Flow-Aware Elephant Flow Detection for Software-Defined Networks," *IEEE Access*, vol. 8, pp. 72585–72597, 2020, <https://doi.org/10.1109/ACCESS.2020.2987977>.

Hydraulic Jump Characteristics Downstream of a Compound Weir consisting of Two Rectangles with a below Semicircular Gate

Majed O. A. Alsaydalani

Civil Engineering Department, Umm Al Qura University, Saudi Arabia
mosaydalani@uqu.edu.sa (corresponding author)

Received: 16 January 2024 | Revised: 29 January 2024 | Accepted: 1 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6918>

ABSTRACT

Weirs are often used in laboratories, industries, and irrigation channels to measure discharge. The discharge capacity of a structure is vital for its safety and plays an important role in the combined gate-weir flow, which is a complicated phenomenon in hydropower. This study carried out experiments on a combined hydraulic structure, which included a compound sharp-crested weir made up of two rectangles along with an inverted semicircular sharp gate. Installed on a straight channel, this structure served as a control instrument. The study aimed to investigate the downstream hydraulic jump characteristics of this combined structure, specifically, the sequent depth ratio (y_2/y_1), the hydraulic jump height ratio (H_j/y_1), the energy loss ratio through the jump (E_L/E_u), and the jump length ratio (L_j/y_1). The width of the upper rectangle on the weir was set at 20 cm. The width of the lower rectangle (W_2) was set at 5, 7, and 9 cm, while its depths (z) were fixed at 6, 9, and 11 cm. The gate's diameters varied between 8, 12, and 15 cm. These measurements were alternated with varying initial Froude numbers (F_{u1}) ranging between 1.32 and 1.5. The results showed that the dimensions of both the weir and the gate influenced the hydraulic jump characteristics. Empirical formulas were developed to predict y_2/y_1 , H_j/y_1 , E_L/E_u , and L_j/y_1 based on the differing dimensions of the combined structure. The findings and analysis of this study are limited to the range of data that were tested.

Keywords-hydraulic jump; combined weir; semicircular gates; combined structure; open channels; head loss

I. INTRODUCTION

Heading-up structures serve mainly to measure discharge flow and control water levels. Their usage can be either individual or combined, depending on the intended purpose. Although these devices are considered old-fashioned, weirs and sluice gates are commonly used in open channels to measure, divert, and regulate flow. Weirs employ the principle of rapidly varied flow to measure the discharge in such channels. They are typically classified into two main categories: sharp-crested weirs and broad-crested weirs. Sharp-crowned weirs are also referred to as thin-plate weirs or notches and possess a cut on the top, giving them their name. The most common shapes for these weirs are triangular (V-notch), rectangular, and trapezoidal.

Weirs and gates can be used simultaneously and this is widely applicable in industries, laboratories, irrigation processes, and dam instrumentation applications. Utilizing sluice gates or sharp-crowned weirs separately is typically associated with morphological drawbacks, such as scour and sedimentation [1-3]. Additionally, one of the main passive functions of sluice gates is to trap floating materials. This ultimately results in a decrease in the accuracy of the discharge measurement and a decrease in the heading-up performance of the structure. Furthermore, the instability of the entire structure

might be threatened by the local scouring process. Therefore, it is thought that installing a single hydraulic structure that houses both a weir and a sluice gate would be a practical way to optimize the benefits of installing both in straight open channels. Compared to the separate use of a traditional weir and gate, the combined structure conveys more discharge. However, the downward opening reduces the amount of sediment that the flow traps. Consequently, due to the simultaneous flow over the weir and beneath the gate, numerous studies have recently been conducted to accurately estimate discharge, discharge coefficient, and the geometric properties of the combined structures under various hydraulic settings, such as weir and gate angles, upstream water head, and combined structure width and height [4-5]. A governing equation to estimate discharge over a triangular sharp-crowned weir was expressed in [6]. Various empirical equations have been presented to estimate discharge in various weir types [7-9]. In [10], the discharge equation obtained by [11] was used as a basis to estimate the discharge for the inverted V-shaped gate of a combined system consisting of a rectangular weir and a triangular gate, showing that the combined discharge is significantly affected by the triangle gate. Experimental tests on rectangular compound sharp-crowned side weirs demonstrated the negligible relationship between the discharge coefficient and the upstream Froude number [12].

In [13], the numerical results produced by the Flow 3D software were compared with experimental results to examine the flow of a compound system that consisted of a vertical sharp-crowned weir and gate. In [14], the flow characteristics over a combined sharp-crowned rectangular triangular weir were studied. In [15], a combined structure made up of two triangular parts with different notch angles was used to measure the flow rates for a wide range of discharges. In [16], the properties of free flow through a combined triangular weir and rectangular sluice gate were investigated, showing that the vertical distance between the weir and the gate and the discharge coefficient was directly proportional and inversely related to the weir angle. In [17], the flow below an inverted rectangular sharp gate and over a rectangular sharp crested weir was investigated, demonstrating how surface tension and viscosity affected combined discharge. In [18], the flow characteristics downstream of a combination of a rectangular sluice gate and a V-notch weir were studied, showing that an increase in the weir vertex angle increased the conveyed discharge. Additionally, a semi-empirical equation was formulated to calculate the compound system's combined discharge coefficient. In [19], experimental tests were carried out to investigate the flow beneath a rectangular sluice gate and over a trapezoidal sharp-crested weir, showing that raising the separation between the upper and lower edges of the gate and the weir increased the combined structure's discharge coefficient. In [20], the discharge coefficient of a combined structure made up of a semicircular gate and a rectangular sharp-crowned weir was examined. In [21], the combined flow below gates and over sharply crowned weirs was investigated. In [22], a combination of the triangular weir and the rectangular gate structure was used to assess the discharge coefficient. In [23-24], the hydraulic properties of weirs and combined weirs were explored according to the following scenarios: rectangular, V-notch, semicircular, rectangular combined weir with a rectangular gate, V-notch combined weir with a rectangular gate, semicircular combined weir with a rectangular gate, and semicircular combined weir with a semicircular gate. The study investigated the significant effects of the combined structure geometry on downstream flow characteristics. In addition, many studies have presented various empirical equations based on experimental data to estimate the hydraulic jump length, focusing on this parameter as a focal point in the stilling basin design [25-35]. Table I contains a tabulation of these equations.

Very little data are available for the combination of a compound sharp-crested weir consisting of two rectangles and the inverted semicircular gate. Therefore, this study aims to explore the flow pattern and hydraulic jump characteristics downstream of this new shape. Different parameters were studied to formulate a clear understanding of the new combined structure. Data were experimentally examined to formulate empirical equations for predicting the hydraulic jump characteristics downstream of the proposed two-rectangle compound combined with an inverted semicircular gate structure. Three distinct semicircular gate diameters and three distinct lower rectangle heights and widths were used to achieve this purpose.

TABLE I. THE HYDRAULIC JUMP LENGTH, L_j FORMULAS

Reference	L_j , Formula
[25]	$L_j = y_2 \left(4.5 - \frac{V_1}{V_c} \right)$
[26]	$L_j = 5.2 y_2$
[27]	$L_j = 3 y_2$
[28]	$L_j = 6.02 (y_2 - y_1)$
[29]	$L_j = 5 (y_2 - y_1)$
[30]	$L_j = 4.5 - 7 (y_2 - y_1)$
[31]	$L_j = C_j (y_2 - y_1)$
	(a) $C_j = 6.3237 + 0.5974 \frac{y_1}{y_2^{0.1}}$ (b) $C_j = 3.827 + 3.088 \left(\frac{y_1}{y_2} \right)^{0.1}$
[32]	$\frac{L_j}{y_1} = 2.84 - 12.6 F_{r1} + 14.9 \frac{y_2}{y_1}$
[33]	$L_j = C_j (y_2 - y_1)$ $C_j = 10 (F_{r1})^{-0.16}$
[34]	$L_j = C_j (y_2 - y_1)$ $C_j = 8 - 0.05 \frac{y_2}{y_1}$
[35]	$L_j = C_j (y_2 - y_1)$ $C_j = 10.6 (F_{r1}^2)^{-0.185}$

II. THEORETICAL STUDY

Figure 1 shows the combined structure's views and a sketch of the concurrent flow over the weir and below the gate. The hydraulic jump characteristics developed downstream of the combined structure can be ascertained using the list of independent variables as follows:

$$f(E_1, E_2, E_u, y_u, y_1, y_2, H_j, L_j, W_1, W_2, Z, d, W, H, Q, \rho, g, \mu, \sigma) = 0 \quad (1)$$

where E_1 is the initial specific energy, E_2 is the sequent specific energy, E_u is the energy upstream of the combined structure, y_u is the upstream water depth, y_1 is the initial water depth, y_2 is the sequent water depth, H_j is the hydraulic jump height, L_j is the hydraulic jump length, W_1 and W_2 are the upper and lower weir rectangle widths respectively, Z is the height of the lower rectangle, d is the semicircular gate diameter, W and H are the total width and height of the combined structure, Q is the discharge, ρ is the water density, g is the gravitational acceleration, μ is the water dynamic viscosity, and σ is the water surface tension.

Dimensionless groups affecting the hydraulic jump characteristics that form a downstream of the compound weir with a below semicircular gate can be obtained by applying the dimensional analysis π -Buckingham's theory and its characteristics to (1) as follows:

$$f\left(\frac{E_1}{y_1}, \frac{E_2}{y_1}, \frac{E_u}{y_1}, \frac{y_u}{y_1}, \frac{y_2}{y_1}, \frac{H_j}{y_1}, \frac{L_j}{y_1}, \frac{W_1}{y_1}, \frac{W_2}{y_1}, \frac{Z}{y_1}, \frac{d}{y_1}, \frac{W}{y_1}, \frac{H}{y_1}, F_{r1}, R_e, W_e\right) = 0 \quad (2)$$

where F_{r1} , R_e , and W_e are the initial Froude, Reynolds, and Weber numbers, respectively. It was assumed that R_e and W_e have little bearing on the combined structure, except when it reaches a very low head. It should be stated that the non-dimensional groups mentioned above can be combined to create a variety of dimensionless numbers. W_1 , W , and H had constant values throughout this study's experiments. Several

studies investigated the connections between discharge and water depth for gates and weirs [36-37].

$$Q = C_d \left(\frac{2}{3} \sqrt{2g} (W_1 - W_2) h_1^{3/2} + \frac{2}{3} \sqrt{2g} W_2 h_2^{3/2} + \frac{\pi}{8} d^2 \sqrt{2g h_3} \right) \quad (3)$$

where C_d is the discharge coefficient for the combined structure and Q is the total discharge through the combined structure.

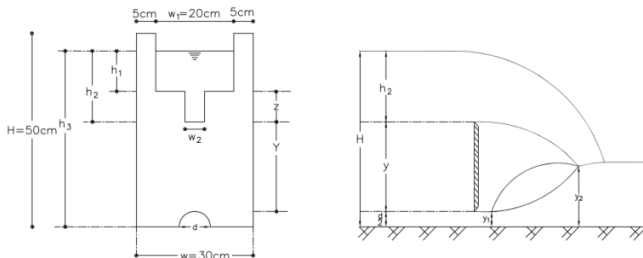


Fig. 1. Sketch of the compound weir with the below semicircular gate.

III. EXPERIMENTAL SETUP

Experimental tests were carried out with a rectangular flume measuring 4.0 m in length, 0.30 m in width, and 0.50 m in height, as presented in Figure 2, using a non-tilting type. The experimental setup uses a closed-water cycle. The effective water circulating system of the flume was powered by a three-horsepower pump. It was easy to observe the experimental tests due to the glassy sheets that form the side walls of the flume. The flume's entry was controlled by placing baffle vertical plates at the channel's entrance to prevent vortex motion and to regulate flow and damp fluctuations. The combined structure's exit water was collected in a hydraulic bench type F13, and the actual discharge was calculated by dividing the volume of water collected in the hydraulic bench by the corresponding time. The actual discharge for each experimental run was the average of three recorded discharge values. The semicircular gate head h_3 and the weir heads h_1 and h_2 were measured using a vernier type gauge with 1 mm accuracy. Calibration was performed before each experiment to prevent instrument errors. It was possible to maintain a steady flow through the channel by carefully adjusting the depth rod to the water surface. Each experiment was conducted with constant discharge. For the experimental study, a rectangular sharp compound weir with a lower semicircular gate, constructed of acrylic glass and having varying lower rectangle widths and depths (W_2 and Z), was used. The below semicircular gate also employed three diameters of 8, 12, and 15 cm, in addition to the case of utilizing the compound weir without the semicircular gate. Three weir lower rectangle widths W_2 of 5, 7, and 9 cm and depths Z of 6, 9, and 11 cm were used, while F_{r1} ranged from 1.3 to 1.5. The aforementioned measurements were applied equally. After the calibration process, the water heads on the weir's upstream side were measured to determine the accuracy of the discharge measurement. A point gauge featuring a 1 mm vernier scale was engaged to measure the water heads. According to [37], the point gauge was fixed four times the maximum head over the weir upstream. As the bottom boundary effect requires the flow through the weir section to be free, the discharge was calculated by measuring the head over the weir in the weir section.



Fig. 2. A model sample during operation for the rectangular flume.

IV. EXPERIMENTAL PROGRAM AND PROCEDURES

A total of fifty-four combined structure models were constructed using acrylic glass sheets, featuring varying semicircular gate diameters d and weir rectangles W_2 , Z . In this study, the semicircular gate with diameters d of 8, 12, and 15 cm was utilized interchangeably with three weir lower rectangle widths W_2 of 5, 7, and 9 cm and depths Z of 6, 9, and 11 cm each. For each setup, six tests were carried out using various water heads. Therefore, 216 tests were performed, as shown in Table II.

TABLE II. THE EXPERIMENTAL PROGRAMS

Semicircular gate diameter, d (cm)	Lower weir width, W_2 (cm)	Lower weir depth, Z (cm)	Initial Froude's number F_{r1}
8 cm	5	6, 9, and 11	Six values (1.3 to 1.5)
	7	6, 9, and 11	Six values (1.3 to 1.5)
	9	6, 9, and 11	Six values (1.3 to 1.5)
12 cm	5	6, 9, and 11	Six values (1.3 to 1.5)
	7	6, 9, and 11	Six values (1.3 to 1.5)
	9	6, 9, and 11	Six values (1.3 to 1.5)
15 cm	5	6, 9, and 11	Six values (1.3 to 1.5)
	7	6, 9, and 11	Six values (1.3 to 1.5)
	9	6, 9, and 11	Six values (1.3 to 1.5)

The procedures mentioned below were followed for every experimental run. At the flume's end, the model was fixed. Water was gradually added to the flume until each head over the weir reached the desired level. The experiment was carried out with a constant flow rate over the weir and below the gate to ensure reaching a steady state. The actual discharge for each experiment was determined by dividing the volume of water collected by the corresponding time using the hydraulic bench. Each experiment was terminated by turning off the pump, and then the following combined structure model was tested.

V. RESULTS AND DISCUSSION

This study focused on demonstrating the effects of the sharp compound weir with two rectangles and the below

semicircular gate as a hydraulic controlling device on the hydraulic jump characteristics downstream of it. These characteristics were the energy loss ratio E_L/E_u , the sequent depth ratio y_2/y_1 , the hydraulic jump height ratio H/y_1 , and the hydraulic jump length ratio L/y_1 . Different lower weir rectangle width and depth values W_2 and Z , respectively, and the below semicircular gate diameter d values were applied throughout the experimental tests. After analyzing the results obtained from these tests, an equation was finally formulated for each hydraulic jump characteristic using regression analysis.

A. Effect of combined Structure Parameters on Energy Loss Ratio

The head loss ratio E_L/E_u through a hydraulic jump is a crucial factor in open-channel flow and water engineering. The hydraulic jump formed downstream of the combined structure causes extra energy to be released as waves and turbulence. The effectiveness of energy dissipation in the hydraulic jump is measured by the head loss ratio, which is defined as the ratio of energy loss to the upstream flow of energy. The result analysis on the effect of W_2/W on E_L/E_u indicates that, as a general outcome, E_L/E_u decreases as F_{r1} increases. For example, for W_2/W of 0.23, z/H of 0.18, and d/H of 0.24, increasing F_{r1} by 13% resulted in a decrease in E_L/E_u by 17.62%. Furthermore, E_L/E_u has an inverse relationship with W_2/W . For instance, for F_{r1} of 1.48, z/H of 0.18, and d/H of 0.24, increasing W_2/W from 0.17 to 0.3 results in decreasing E_L/E_u from 0.741 to 0.694, i.e. a 76.47% increase in W_2/W leads to 6.8% decrease in E_L/E_u , as illustrated in Figure 3.

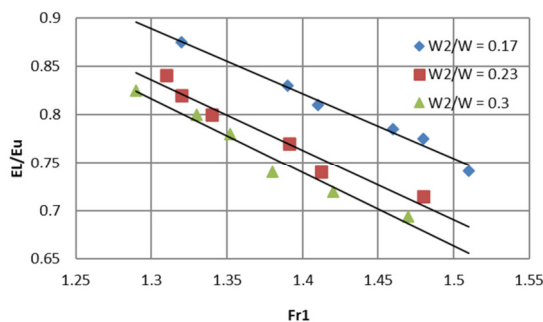


Fig. 3. Relationship between E_L/E_u and F_{r1} at different W_2/W and z/H of 0.18, and d/H of 0.24.

Eighteen experiments were conducted to investigate the effect of the weir lower rectangle height ratio z/H on the energy loss ratio E_L/E_u for given d/H and W_2/W . The results show that E_L/E_u has a direct relationship with z/H . For example, for W_2/W of 0.23, d/H of 0.22, and F_{r1} of 1.48, increasing z/H from 0.12 to 0.22 results in increasing E_L/E_u from 0.804 to 0.842, i.e. a 4.72% increase in E_L/E_u corresponds to 83.33% increase in z/H , as displayed in Figure 4. In addition, eighteen experiments were carried out to show the impact of the below semicircular gate diameter ratio d/H on the energy loss ratio through the formed hydraulic jump E_L/E_u . Figure 5 shows that E_L/E_u increases as d/H decreases for given W_2/W , z/H , and F_{r1} . For example, for F_{r1} of 1.48, W_2/W of 0.23, and z/H of 0.18, if d/H increases from 0.16 to 0.3, E_L/E_u decreases from 0.741 to

0.606. This means that if d/H increases by 87.5%, E_L/E_u decreases by 22.28%. Regression analysis was applied for the energy loss ratio with other independent parameters. So, an equation was formulated for the different discussed cases as follows:

$$\frac{E_L}{E_u} = 2.033 - 0.49 F_{r1} - 0.513 \frac{W_2}{W} - 1.34 \frac{z}{H} - 0.929 \frac{d}{H} \quad (4)$$

Equation 4 was validated using the experimentally measured data versus the calculated values for the energy loss ratio E_L/E_u as portrayed in Figure 6. The findings demonstrate a good agreement between the measured and calculated values.

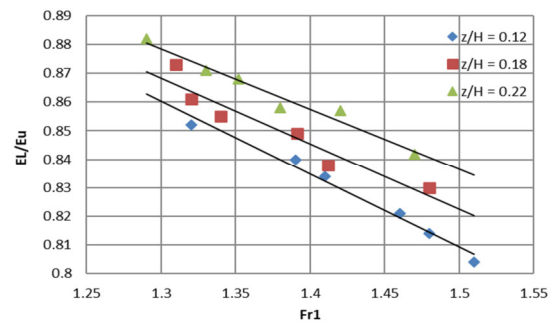


Fig. 4. Relationship between E_L/E_u and F_{r1} at different z/H and W_2/W of 0.23, and d/H of 0.22.

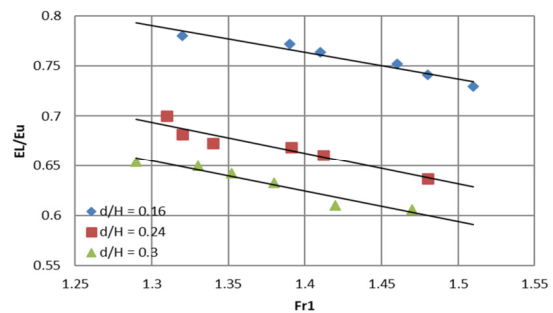


Fig. 5. Relationship between E_L/E_u and F_{r1} at different d/H and W_2/W of 0.23, and z/H of 0.23.

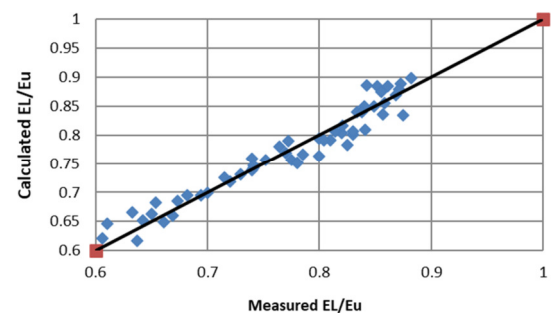


Fig. 6. Comparison between observed and calculated values for E_L/E_u .

B. Effect of the combined Structure Parameters on the Sequent Depth and the Jump Height Ratios

The sequent depth ratio y_2/y_1 and the hydraulic jump height ratio H/y_1 are crucial parameters in the analysis of hydraulic jumps. y_2/y_1 represents the ratio of the downstream hydraulic jump depth y_2 to the upstream hydraulic jump depth y_1 and

H_f/y_1 represents $y_2/y_1 - 1$. Overall, the results of the analysis demonstrating the impact of W_2/W on y_2/y_1 and H_f/y_1 indicate that y_2/y_1 and H_f/y_1 increase as Fr_1 increases. For example, a 13% increase in Fr_1 causes a 100.62% and 151.5% increase in y_2/y_1 and H_f/y_1 respectively, for W_2/W of 0.23, z/H of 0.18, and d/H of 0.24. Furthermore, there are inverse relationships between W_2/W and y_2/y_1 and H_f/y_1 , as shown in Figures 7 and 8. For example, increasing W_2/W from 0.17 to 0.3 causes y_2/y_1 and H_f/y_1 to decrease from 7.99 to 5.32 and from 6.99 to 4.32, respectively, for Fr_1 of 1.48, z/H of 0.18, and d/H of 0.24. This means that a 76.47% increase in W_2/W results in a 50.19% and 61.81% decrease in y_2/y_1 and H_f/y_1 , respectively.

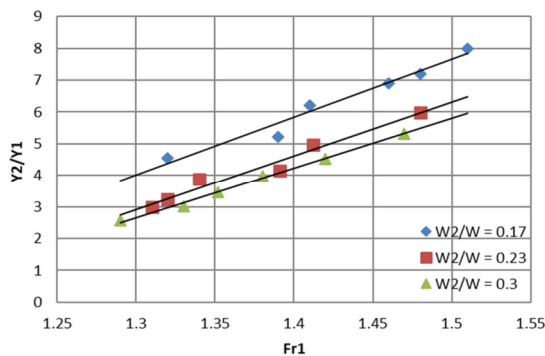


Fig. 7. Relationship between y_2/y_1 and Fr_1 at different W_2/W and z/H of 0.18, and d/H of 0.24.

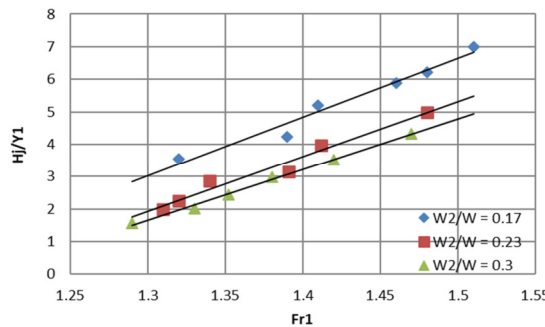


Fig. 8. Relationship between H_f/y_1 and Fr_1 at different W_2/W and z/H of 0.18, and d/H of 0.24.

Eighteen experiments were performed to examine how z/H affected y_2/y_1 and H_f/y_1 for given d/H and W_2/W . The findings indicate a direct relationship between z/H , y_2/y_1 , and H_f/y_1 . For instance, increasing z/H from 0.12 to 0.22 resulted in decreased y_2/y_1 and H_f/y_1 from 8.99 to 6.89 and 7.99 to 5.89, accordingly, i.e. 30.48% and 35.65% increase in y_2/y_1 and H_f/y_1 correspond to an 83.33% decrease in z/H , for W_2/W of 0.23, d/H of 0.22, and Fr_1 of 1.48, as illustrated in Figures 9 and 10.

Additionally, eighteen experiments were carried out to demonstrate how y_2/y_1 and H_f/y_1 are affected by the below semicircular gate diameter ratio d/H . Figures 11 and 12 reveal that for the given W_2/W , z/H , and Fr_1 , y_2/y_1 and H_f/y_1 increase as d/H increases. For instance, if d/H increases from 0.16 to 0.3 for Fr_1 of 1.48, W_2/W of 0.23, and z/H of 0.18, each y_2/y_1 and H_f/y_1 increases from 6.82 to 7.42 and from 5.82 to 6.42, respectively. This indicates that y_2/y_1 and H_f/y_1 increase by

8.8% and 10.31, accordingly, if d/H increases by 87.5%. For each y_2/y_1 and H_f/y_1 , regression analysis was performed along with other independent parameters, producing the following equations:

$$\frac{y_2}{y_1} = -17.57 + 19.28 Fr_1 - 11.26 \frac{W_2}{W} - 20.94 \frac{z}{H} + 10.67 \frac{d}{H} \quad (5)$$

$$\frac{H_f}{y_1} = -18.57 + 19.28 Fr_1 - 11.26 \frac{W_2}{W} - 20.94 \frac{z}{H} + 10.67 \frac{d}{H} \quad (6)$$

As observed in Figures 13 and 14, (5) and (6) were verified by comparing the calculated values for each y_2/y_1 and H_f/y_1 with the experimentally measured data. The calculated and measured values exhibited good agreement.

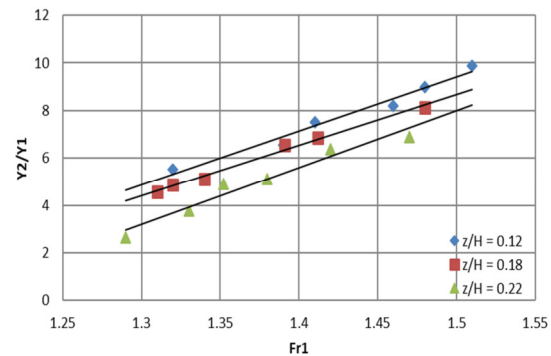


Fig. 9. Relationship between y_2/y_1 and Fr_1 at different z/H and W_2/W of 0.23, and d/H of 0.22.

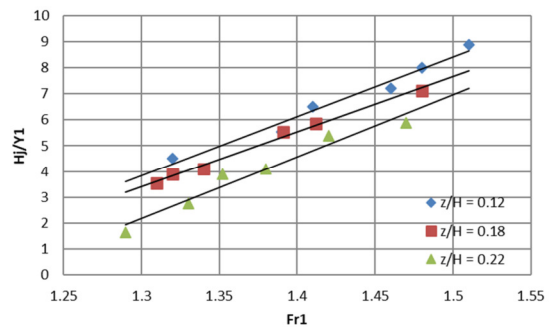


Fig. 10. Relationship between H_f/y_1 and Fr_1 at different z/H and W_2/W of 0.23, and d/H of 0.22.

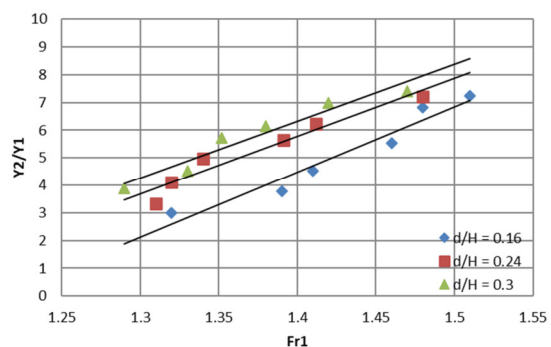


Fig. 11. Relationship between y_2/y_1 and Fr_1 at different d/H and W_2/W of 0.23, and z/H of 0.23.

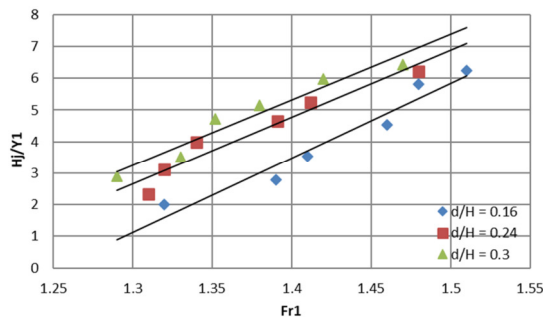


Fig. 12. Relationship between H/y_1 and F_{r1} at different d/H and W_2/W of 0.23, and z/H of 0.23.

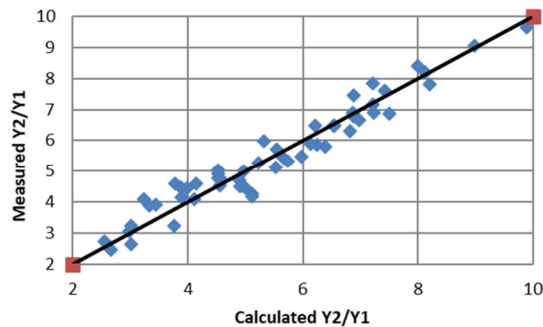


Fig. 13. Comparison between observed and calculated values for y_2/y_1 .

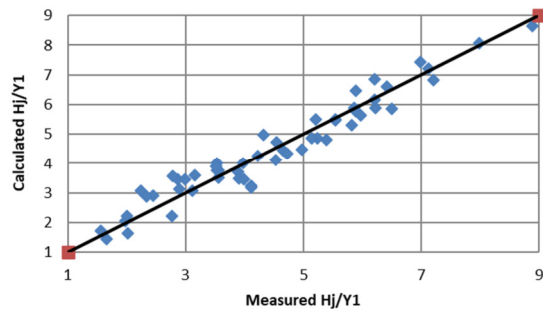


Fig. 14. Comparison between observed and calculated values for H/y_1 .

C. Effect of combined Structure Parameters on the Jump Length Ratio

The findings disclose that L_f/y_1 increases with an increase in F_{r1} . For instance, for W_2/W of 0.23, z/H of 0.18, and d/H of 0.24, a 13% increase in F_{r1} results in an increase of L_f/y_1 by 168.4%. Additionally, Figure 15 shows that there are inverse relationships between W_2/W and L_f/y_1 . For example, increasing W_2/W from 0.17 to 0.3 causes L_f/y_1 to decrease from 25.62 to 18.11, for F_{r1} of 1.48, z/H of 0.18, and d/H of 0.24. This means that a 76.47% increase in W_2/W leads to a 41.46% decrease in L_f/y_1 .

The effects of the weir lower rectangle height ratio z/H were investigated for the hydraulic jump length ratio L_f/y_1 in experimental tests, for given d/H and W_2/W . The results exhibit that L_f/y_1 has an inverse relationship with z/H . For example, when z/H increases from 0.12 to 0.22, L_f/y_1 decreases from 35.12 to 23.45. This means that a decrease of 49.77% in L_f/y_1 corresponds to an 83.33% increase in z/H for W_2/W of 0.23, d/H of 0.22, and F_{r1} of 1.48, as evidenced in Figure 16.

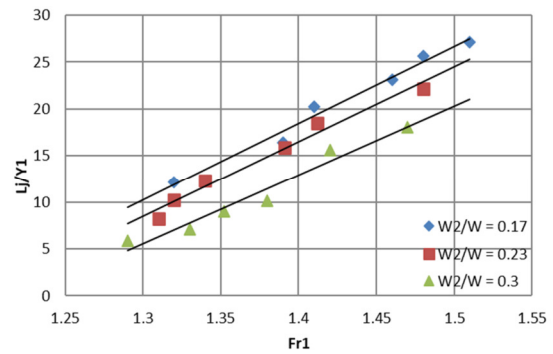


Fig. 15. Relationship between L_f/y_1 and F_{r1} at different W_2/W , z/H of 0.18, and d/H of 0.24.

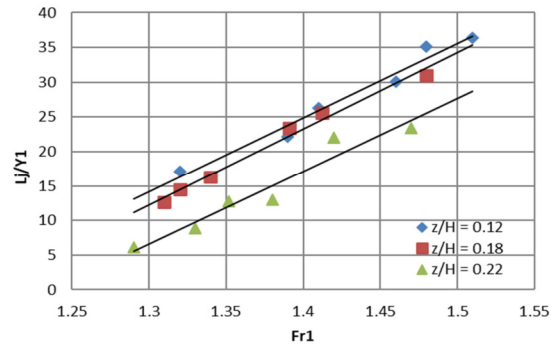


Fig. 16. Relationship between L_f/y_1 and F_{r1} at different z/H , W_2/W of 0.23, and d/H of 0.22.

In addition, experimental tests were conducted to show how the below semicircular gate diameter ratio d/H influences the hydraulic jump length ratio L_f/y_1 . Figure 17 indicates that L_f/y_1 increases as d/H increases for given W_2/W , z/H , and F_{r1} . For example, L_f/y_1 increases from 27.5 to 31.5 if d/H increases from 0.16 to 0.3, for F_{r1} of 1.48, W_2/W of 0.23, and z/H of 0.18. This means that if d/H increases by 87.5%, then L_f/y_1 increases by 14.55%. A regression analysis was carried out for L_f/y_1 , along with other independent parameters, producing the following equation:

$$\frac{L_f}{y_1} = -94.26 + 92.26 F_{r1} - 37.62 \frac{W_2}{W} - 90.77 \frac{z}{H} + 38.88 \frac{d}{H} \quad (7)$$

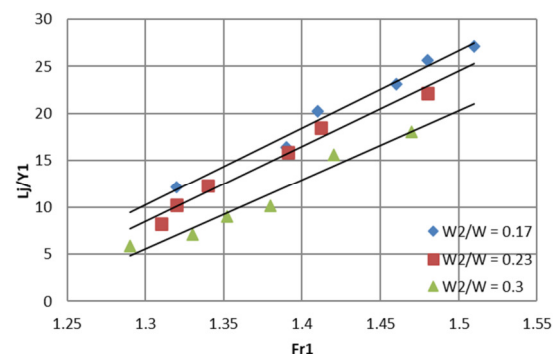


Fig. 17. Relationship between L_f/y_1 and F_{r1} at different W_2/W and z/H of 0.18, and d/H of 0.24.

Figure 18 presents a comparison of (7) with the experimentally measured data for L_j/y_1 . The calculated and measured values exhibited good agreement. To assess the precision of the hydraulic jump length measurements and verify (7), the results were compared with the previously published equations shown in Table 1. Figure 19 displays the plotted results. The results were within the range of the calculated values, and the figure denotes a good correlation and a similar trend to that of earlier studies.

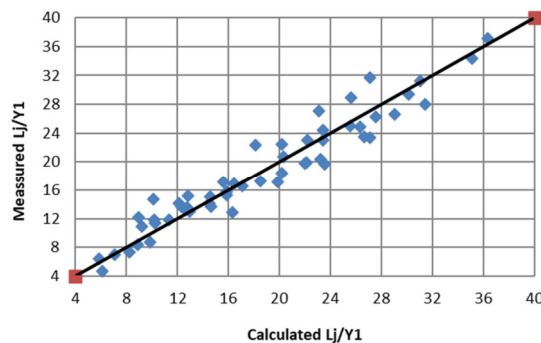


Fig. 18. Comparison between observed and calculated values For L_j/y_1 .

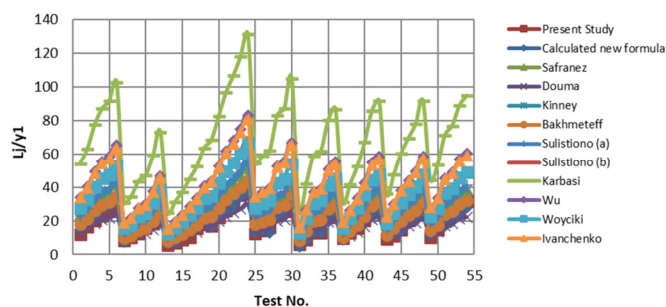


Fig. 19. Comparison between calculated L_j/y_1 of the current study and calculated using other formulas.

VI. CONCLUSION

This study carried out experiments to examine the characteristics of hydraulic jumps that form downstream of a combined hydraulic structure composed of a compound weir and a semicircular gate. The impact of various parameters, width (W_2) and depth (Z) of the weir's lower rectangle, the diameter (d) beneath the gate, and the initial Froude's number (F_{r1}), on the hydraulic jump characteristics (y_2/y_1 , H_j/y_1 , E_L/E_u , and L_j/y_1) were analyzed. The findings are as follows:

- The energy loss ratio through the formed hydraulic jump E_L/E_u decreases as F_{r1} increases.
- Each of the sequent depth ratios y_2/y_1 , hydraulic jump height ratios H_j/y_1 and hydraulic jump length ratios L_j/y_1 increase with increasing F_{r1} .
- The energy loss ratio E_L/E_u has a direct proportional relationship with the lower rectangle height ratio of the weir z/H . On the other hand, E_L/E_u has an inversely proportional relationship with the weir lower rectangle width ratio W_2/W

and the semicircular gate diameter ratio (d/H) for a given F_{r1} .

- Each of y_2/y_1 , H_j/y_1 , and L_j/y_1 increases with d/H and decreases with z/H and W_2/W for a given F_{r1} .
- General equations were formulated to calculate the hydraulic jump characteristics y_2/y_1 , H_j/y_1 , E_L/E_u , and L_j/y_1 for different combined structure parameters.
- The observed hydraulic jump length values were in good agreement with those calculated in previous studies.

This study provides insights into the complex dynamics of hydraulic jumps, particularly in the context of compound weirs, and offers valuable information for the design and optimization of hydraulic structures. The findings of this study have implications for hydraulic engineering, offering a deeper understanding of flow characteristics, and aiding in the development of optimal design techniques for hydraulic structures.

REFERENCES

- [1] F. Rooniyan, "The Effect of Confluence Angle on the Flow Pattern at a Rectangular Open-Channel," *Engineering, Technology & Applied Science Research*, vol. 4, no. 1, pp. 576–580, Feb. 2014, <https://doi.org/10.48084/etasr.395>.
- [2] A. S. Kote and P. B. Nangare, "Hydraulic Model Investigation on Stepped Spillway's Plain and Slotted Roller Bucket," *Engineering, Technology & Applied Science Research*, vol. 9, no. 4, pp. 4419–4422, Aug. 2019, <https://doi.org/10.48084/etasr.2837>.
- [3] S. M. Kori, A. A. Mahessar, M. Channa, A. A. Memon, and A. R. Kori, "Study of Flow Characteristics Over a Rounded Edge Drop Structure in Open Channel," *Engineering, Technology & Applied Science Research*, vol. 9, no. 3, pp. 4136–4139, Jun. 2019, <https://doi.org/10.48084/etasr.2584>.
- [4] F. Granata, F. Di Nunno, R. Gargano, and G. de Marinis, "Equivalent Discharge Coefficient of Side Weirs in Circular Channel—A Lazy Machine Learning Approach," *Water*, vol. 11, no. 11, Nov. 2019, Art. no. 2406, <https://doi.org/10.3390/w11112406>.
- [5] S. Salehi and A. H. Azimi, "Discharge Characteristics of Weir-Orifice and Weir-Gate Structures," *Journal of Irrigation and Drainage Engineering*, vol. 145, no. 11, Nov. 2019, Art. no. 04019025, [https://doi.org/10.1061/\(ASCE\)IR.1943-4774.0001421](https://doi.org/10.1061/(ASCE)IR.1943-4774.0001421).
- [6] "Discharge Measurement Structures," International Institute for Land Reclamation and Improvement, Wageningen, The Netherlands, 1976.
- [7] S. Emami, J. Parsa, H. Emami, and A. Abbaspour, "An ISaDE algorithm combined with support vector regression for estimating discharge coefficient of W-planform weirs," *Water Supply*, vol. 21, no. 7, pp. 3459–3476, Apr. 2021, <https://doi.org/10.2166/ws.2021.112>.
- [8] A. B. Altan-Sakarya, M. A. Kokpinar, and A. Duru, "Numerical modelling of contracted sharp-crested weirs and combined weir and gate systems," *Irrigation and Drainage*, vol. 69, no. 4, pp. 854–864, 2020, <https://doi.org/10.1002/ird.2468>.
- [9] P. and A. J.M. Harrison Ackers, *Weirs and Flumes for Flow Measurement*. Chichester, UK: John Wiley & Sons, 1978.
- [10] A. A. Alhamid and D. Husain, "Discharge equation for simultaneous flow over rectangular weirs and below inverted triangular weirs," *Arab Gulf Journal of Scientific Research*, vol. 14, no. 3, pp. 595–607, 1996.
- [11] N. Rajaratnam and K. Subramanya, "Flow Equation for the Sluice Gate," *Journal of the Irrigation and Drainage Division*, vol. 93, no. 3, pp. 167–186, Sep. 1967, <https://doi.org/10.1061/JRCEA4.0000503>.
- [12] A. Zahiri, H. Md. Azamathulla, and S. Bagheri, "Discharge coefficient for compound sharp crested side weirs in subcritical flow conditions," *Journal of Hydrology*, vol. 480, pp. 162–166, Feb. 2013, <https://doi.org/10.1016/j.jhydrol.2012.12.022>.

- [13] S. A. Sarhan and S. A. Jalil, "Analysis of Simulation Outputs for the Mutual Effect of Flow in Weir and Gate System," *Journal of University of Babylon for Engineering Sciences*, vol. 26, no. 6, pp. 48–59, Apr. 2018, <https://doi.org/10.29196/jubes.v26i6.1050>.
- [14] M. M. Muhammad and S. A. Abdullahi, "Experimental Study of Flow over Sharp Crested Rectangular-Triangular Weir Models," in *Proceedings of Nigeria Engineering Conference*, Zaria, Nigeria, 2014, pp. 34–45.
- [15] M. Piratheepan, N. E. F. Winston, and K. P. P. Pathirana, "Discharge Measurements in Open Channels using Compound Sharp-Crested Weirs," vol. 40, no. 3, Jul. 2007, Art. no. 31, <https://doi.org/10.4038/engineer.v40i3.7144>.
- [16] H. A. A. M. Hayawi, A. A. A.-G. Yahia, and G. A. A. M. Hayawi, "Free Combined Flow Over a Triangular Weir and Under Rectangular Gate," *Damascus University Journal*, vol. 24, no. 1, pp. 9–22, 2008.
- [17] A. A. M. Negm, A. M. Al-Brahim, and A. A. Alhamid, "Combined-free flow over weirs and below gates," *Journal of Hydraulic Research*, vol. 40, no. 3, pp. 359–365, May 2002, <https://doi.org/10.1080/00221680209499950>.
- [18] A. A. Alhamid, A. A. M. Negm, and A. M. Al-Brahim, "Discharge Equation for Proposed Self-cleaning Device," *Journal of King Saud University - Engineering Sciences*, vol. 9, no. 1, pp. 13–23, Jan. 1997, [https://doi.org/10.1016/S1018-3639\(18\)30664-0](https://doi.org/10.1016/S1018-3639(18)30664-0).
- [19] K. C. Okafor, C. C. Udeze, F. N. Ugwoke, O. Ifesinachi, and O. Nnaemeka, "AFIM: A High Level Conceptual ATM Design Using Composite Formal Modelling With Capture Simulation Pattern Matching Technique," *International Journal of Scientific & Engineering Research*, vol. 5, no. 4, pp. 755–761, 2014.
- [20] A. A. G. Alniami, D. G. A. M. Hayawi, and H. A. M. Hayawi, "Coefficient Of Discharge For A Combined Hydraulic Measuring Device," *Al-Rafidain Engineering Journal (AREJ)*, vol. 17, no. 6, pp. 92–100, Dec. 2009, <https://doi.org/10.33899/rengj.2009.43616>.
- [21] J. M. Samani and M. Mazaheri, "Combined Flow over Weir and under Gate," *Journal of Hydraulic Engineering*, vol. 135, no. 3, pp. 224–227, Mar. 2009, [https://doi.org/10.1061/\(ASCE\)0733-9429\(2009\)135:3\(224\)](https://doi.org/10.1061/(ASCE)0733-9429(2009)135:3(224)).
- [22] B. Balouchi and G. Rakhshandehroo, "Using Physical and Soft Computing Models to Evaluate Discharge Coefficient for Combined Weir–Gate Structures Under Free Flow Conditions," *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, vol. 42, no. 4, pp. 427–438, Dec. 2018, <https://doi.org/10.1007/s40996-018-0117-0>.
- [23] A. K. I. Al-Saadi, "Study Coefficient of Discharge for a Combined Free Flow over Weir and under Gate for Multi Cases," *Euphrates Journal of Agriculture Scienc*, vol. 5, no. 4, pp. 26–35, 2013.
- [24] M. A. R. Eltoukhy, F. S. Abdelhaleem, T. H. Nasralla, and S. Shaban, "Effect of Compound Weir and below Circular Gate Geometric Characteristics on its Discharge Coefficient," *International Journal of Scientific & Engineering Research*, vol. 11, no. 10, 2020.
- [25] A. Ludin and D. P. Barnes, "Investigation of the length of the hydraulic jump at Berlin," *Berlin Civil Engineering*, vol. 4, no. 5, 1934.
- [26] K. Safranez, *Untersuchungen über den Wechselsprung*. Im Selbstverlag, 1929.
- [27] C. S. Maxwell, "Study of Stilling-Basin Design," *Transactions of the American Society of Civil Engineers*, vol. 99, no. 1, pp. 490–512, Jan. 1934, <https://doi.org/10.1061/TACEAT.0004556>.
- [28] C. W. Kinney, "Stilling pools for spillways," Ph.D. dissertation, State University of Iowa, 1935.
- [29] B. A. Bakhmeteff and A. E. Matzke, "The Hydraulic Jump in Terms of Dynamic Similarity," *Transactions of the American Society of Civil Engineers*, vol. 101, no. 1, pp. 630–647, Jan. 1936, <https://doi.org/10.1061/TACEAT.0004708>.
- [30] S. M. Woodward and C. J. Posey, *Hydraulics of steady flow in open channels*. New York, NY, USA: John Wiley & Sons, 1941.
- [31] B. Sulistiono and L. Makrup, "Study of Hydraulic Jump Length Coefficient with the Leap Generation by Canal Gate Model," *American Journal of Civil Engineering*, vol. 5, no. 3, pp. 148–154, Apr. 2017, <https://doi.org/10.11648/j.ajce.20170503.14>.
- [32] M. Karbasi, "Estimation of classical hydraulic jump length using teaching–learning based optimization algorithm," *Journal of Materials and Environmental Sciences*, vol. 7, no. 8, pp. 2947–2954, 2016.
- [33] C. K. Wu, "Hydraulic Jump in Rectangular Channels," Ph.D. dissertation, A. & M. College of Texas, 1949.
- [34] K. Wóycicki, "Wassersprung, Deckwalze und Ausfluss unter einer Schütze," ETH Zurich, 1931.
- [35] A. I. Ivanchenko, "Discussion of 'The Hydraulic Jump in Terms of Dynamic Similarity,'" *Transactions of the American Society of Civil Engineers*, vol. 101, pp. 668–669, 1936.
- [36] C. E. Kindsvater and R. W. Carter, "Discharge Characteristics of Rectangular Thin-Plate Weirs," *Journal of the Hydraulics Division*, vol. 83, no. 6, pp. 1453–36, Dec. 1957, <https://doi.org/10.1061/JYCEAJ.0000142>.
- [37] M. O. A. Alsaydalani, "Discharge Coefficient of a Two-Rectangle Compound Weir combined with a Semicircular Gate beneath it under Various Hydraulic and Geometric Conditions," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12587–12594, Feb. 2024, <https://doi.org/10.48084/etasr.6605>

Influence of Supplementary Oxide Layer on Solar Cell Performance

Mihai Oproescu

Faculty of Electronics, Communication and Computers, National University of Science and Technology POLITEHNICA Bucharest, Pitesti University Centre, Romania
mihai.oproescu@upb.ro (corresponding author)

Adriana-Gabriela Schiopu

Faculty of Mechanics and Technology, National University of Science and Technology POLITEHNICA Bucharest, Pitesti University Centre, Romania
gabriela.schiopu@upb.ro (corresponding author)

Valentin Marian Calinescu

Doctoral School Materials Science and Engineering, National University of Science and Technology POLITEHNICA Bucharest, Romania
valentin.calinescu@upit.ro

Vasile-Gabriel Iana

Faculty of Electronics, Communication and Computers, National University of Science and Technology POLITEHNICA Bucharest, Pitesti University Centre, Romania
vasile_gabriel.iana@upb.ro

Nicu Bizon

Faculty of Electronics, Communication and Computers, National University of Science and Technology POLITEHNICA Bucharest, Pitesti University Centre, Romania
nicu.bizon@upit.ro

Mohammed Sallah

Applied Mathematical Physics Research Group, Physics Department, Faculty of Science, Mansoura University, Egypt | Department of Physics, College of Sciences, University of Bisha, Saudi Arabia
msallahd@mans.edu.eg

Received: 8 January 2024 | Revised: 25 January 2024 | Accepted: 29 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6879>

ABSTRACT

The increasing use of solar energy for electricity production has led to a directly proportional growth in the production of solar cells. Photovoltaic (PV) performance of silicon solar cells can be improved by using more efficient technologies, optimizing processes, and changing behavior in order to reduce operational costs and greenhouse gas emissions. In order to propose solutions for commercial solar cell production with better performance, this article presents an experimental assessment on Supplementary Oxide Layers (SOLs) that are deposited on the surface of a solar cell absorber layer. SOLs are typically used to improve the performance of solar cells by passivating surface defects, reducing recombination losses, and improving the electrical contact between the absorber layer and the metal electrodes. The obtained solar cells are tested under natural sunlight conditions, following a variable dynamic electronic charge profile. The experimental results along with the corresponding I-V and P-V curves, are assessed according to the process parameters, the lighting parameters, and the dynamic load scenario. SOLs have been shown to improve the Power Conversion Efficiency (PCE) of solar cells considerably. The proposed method for increasing the energy efficiency of solar cells can be applied to any type of commercial solar cell and it is easy to implement at the industrial or research level by controlling process parameters. The integration of

the whole process, i.e. development of precursor solutions, deposition of thin films, and testing of electrical properties is another contribution of the current study, along with its interdisciplinary character, which involves materials science, electronics, and software programming.

Keywords-solar cells; energy efficiency; spin-coating deposition; metal-oxide film; software acquisition; characterization

I. INTRODUCTION

Starting with the first research in the field of capturing solar energy carried out by Antoine-César Becquerel around 1839, continuing with the development of the first physical model of the Solar Cell (SC) by Charles Fritts in 1883 and up to the present day, the improvement of operating and exploitation conditions, as well as the improvement of their energy efficiency were milestones in scientific research. Even if in 2021, charcoal remained the main fuel for electricity generation (its share being 36%) [1], in the context where worldwide energy production electricity in 2021 increased by 6.2%, a dynamic similar to that observed in 2010 as a result of the financial crisis, wind energy and solar energy reached a share of 10.2% of the total energy generated. It was the first time that wind and solar power have provided more than 10% of global electricity. The economic context, the conditions inflicted by the environmental protection, imposed photovoltaic (PV) energy as a viable source of renewable energy. As it can be seen from the dynamics of solar energy production [1], over 13% of the renewable energy is now produced with the help of SCs. The performance of SCs is influenced by several factors, such as:

- Climate. SCs are exposed to the environment for their entire life-time. Various factors affect the efficiency of the photoelectric conversion of an SC, including light, wind, and temperature [2]. In addition, some of these factors may damage the structure and functions of SC.
- Specific technological parameters, such as manufacturing technology and the types of basic materials used in SCs.
- Parameters of the output electronic converter: the type of electrical load and its dynamics, the type and number of power converters used in the output electronic circuit, the use of control algorithms for obtaining maximum energy - MPPT (Maximum Power Point Tracking).
- Energy efficiency refers to how energy is transformed, stored and utilized, as well as how products and technologies can be designed to reduce energy consumption

[3, 4]. This can be applied to products, systems, and technologies.

A. Polycrystalline Solar Cells

Until now three generations of SCs have been reported. Crystalline silicon cells are the oldest and most developed solar panels. Unlike these, thin film cells are more flexible and 350 times thinner, aimed at industrial application for power energetics capacities. The type of semiconductor materials that are responsible for the photovoltaic (PV) effect represents the essential factor that distinguishes them. In thin film SCs, the semiconductor layer is placed between transparent conducting layers [5, 6]. Even though silicon is sometimes used to produce them, it is not the same as solid silicon wafer types, it is a non-crystalline type of silicon [7]. Typically, thin film SCs are manufactured by successively depositing thin films, with thicknesses between 1 and 10 μm , in a vacuum, on different substrates (such as polymer, glass, metal, and others) using etching technology to produce integrated modules over large areas. Multi-junction cells are produced based on high technology from two or more junctions layered on top of each other. Commercial viability is based on slowly depositing layers in systems, such as concentrated PV or space applications.

Silicon with different structural morphologies is usually employed for SCs. The production technology based on Si-wafers covered around 95% of the total production in 2020. PV technology production is significant because all manufacturing methods are conducted at low temperatures and on thin Si wafers with about 100 μm thickness. The use of silicon in SCs has been reduced significantly to about 3 g/Wp due to increased efficiencies and thinner wafers [5, 7]. The part of mono-crystalline technology is now about 84% of total Si solar cell production. Also, because it is mainly extracted from sand or quartz, it is 100% recyclable. Table I summarizes the characteristics of the best-reported polycrystalline commercial SCs. These commercial SCs have in common the type of applied Anti-Reflection Coating (ARC), namely Si_3N_4 , as well as the front electrode, Ag and back electrode, Al. The differences are efficiency and the output power.

TABLE I. COMMERCIAL SOLAR CELLS CHARACTERISTICS

Type of silicon solar cell	Efficiency (%)	Output max. power (W)	ARC	Back electrode	Front electrode
Polycrystalline	17.8 ~ 18.8	4.33 ~ 4.58	Si_3N_4	Al	Ag
Polycrystalline	19.3 ~ 20.6	4.742 ~ 5.037	Si_3N_4	Al	Ag
Polycrystalline	18.1 ~ 18.9	4.4 ~ 4.57	Si_3N_4	Al	Ag
Polycrystalline	17 ~ 18.8	4.14 ~ 4.53	Si_3N_4	Al	Ag
Polycrystalline	19 ~ 19.6	4.668 ~ 4.816	Si_3N_4	Al	Ag
Polycrystalline	18 ~ 19.1	4.43 ~ 4.69	Si_3N_4	Al	Ag
Polycrystalline	17.8 ~ 18.9	4.33 ~ 4.6	Si_3N_4	Al	Ag, Al
Polycrystalline	18.40	4.478	Si_3N_4	Al	Ag
Polycrystalline	17.8 ~ 18.9	4.33 ~ 4.6	Si_3N_4	Al	Ag, Al
Polycrystalline	16.4 ~ 18.2	3.991 ~ 4.429	Si_3N_4	Al	Ag, Al

B. Metal Oxide Coatings

In recent studies, the use of some metal oxide films (ZnO, TiO₂, SiO₂, Al₂O₃, NiO, MgO) for the fabrication of SCs has been reported due to their superior electric and optical properties [8-15]. The metal oxide films present p or n-type conductivity and an energy gap around 3-7 eV. The deposition methods used are more and more varied: sol-gel, spin coating, chemical vapor deposition, physical vapor deposition, radio frequency-sputtering or printing. One of the best layer candidates is MgO, especially known for its high transmission value of about 91.48% in the visible range [11]. The current trends are to use MgO as an intermediate layer in perovskite, dye sensitized, or polymer SCs [12-14]. For perovskite SCs, a magnesium oxide layer was achieved in [15] via decomposition of magnesium acetate on mesoporous TiO₂. The obtained efficiency increased to 13.9%. Also, a MgO layer was successfully incorporated between SnO₂ and F-doped tin oxide to increase the PCE from 16.4% to 18.23% in [16]. Nanostructured MgO films, using spray pyrolysis on glass substrate show a good transmittance of 90% [17]. The optical band gap between 3.64 and 3.7 eV make these films usable for protective layers in SCs, a anti-reflective coating MgO of 86 nm thickness reduces reflectivity to 30% compared to a normal solar cell [18]. In [19], MgO nanostructured films were deposited on a glass substrate using Nd-YAg laser ablation of a methanol solution. The uniform obtained films of a p-MgO/n-Si solar cell structure had a good optical transparency and a low refractive index with increasing wavelength [19]. Enhancement in the photocurrent and the voltages was found in solid-state dye sensitized cells fabricated with MgO coatings on nanocrystalline SnO₂, as anode and CuI as cathode [20].

II. MATERIALS AND METHODS

A. Choice of Oxide

Magnesium oxide was chosen as the precursor material due to its nontoxicity, physical stability, high transmission value in visible range, and the wide direct band gap of 7.8 eV. It is very stable at high temperatures in oxidizing atmospheres up to 2300 °C and reducing atmospheres at 1700 °C, respectively, while it crystallizes in the cubic rock-salt type (cell parameter 4.21 Å) [21]. The hygroscopic nature of MgO and its transition to Mg(OH)₂ are often mentioned in experimental research with the statement that MgO is generally stable.

B. Choice of Solvents

For the realization of MgO coatings, 3 solvents were chosen: ethanol (EtOH), ethanolamine (EA), and ethylene glycol (EG). MgO is not soluble in any of the chosen solvents and can be safely dispersed without dissolution. EtOH has lower polarity and evaporates the fastest. EtOH evaporates completely after 25-30 min. EA has a low vapor pressure at room temperature, and the rate of evaporation increases with increasing temperature. EG is a symmetrical molecule and it can be believed that it has a permanent dipole moment. However, its C-OH bonds rotate, which favors dipoles. It is also a dynamic molecule, which undergoes constant rotations and vibrations, and can form or receive hydrogen bonds thanks to its two OH groups. EG evaporates completely at 197.2 °C.

To obtain the precursor solutions, 2g MgO (Roth reagent, analytical grade) was dispersed in 100 ml of each solvent: EtOH, EA, and EG, on ultrasonic bath, for 90 s, generating different solutions. The pH of the precursor solutions is in the basic range, having the following values: 11 corresponding to the dispersion of MgO in EtOH, 12.75 corresponding to the dispersion of MgO in EA, respectively, 10.33 for the dispersion of MgO in EG. No binder was used because it can lead to changes in the properties of the coatings.

C. Spin Coating Procedure

The spin coating method was chosen for developing coatings because it can be implemented in a wide variety of productions and technology sectors. Its primary advantage over the other methods is its ability to fast and easily produce very uniform and adherent coatings. Commercial polycrystalline silicon SC (poly-Si) pieces for experimental procedure with dimensions of 5.2×1.9 cm were cleaned in the ultrasonic bath in ethanol and deionized water for 180 s for each cleaning agent. The cleaning process aimed to remove impurities (e.g. grease, dust). At the first stage of the deposition, 100 µl of each precursor solution was added to a small rotation on the top of the polycrystalline silicon SC (dynamic spin coating). The rotation speed of the SCV-10 spin-coater was initially set to 25 rps, i.e. 1500 rpm. In the second stage, 4 drops of 50 µl, at intervals of 30 s, at 2000 rpm, were added. In the third stage, the covered commercial SC was heat treated between 850 °C and 1200 °C for 10 min on the hotplate MS-H280-Pro. The characteristic parameters of deposition are presented in Table II and the deposition process is schematically shown in Figure 1.

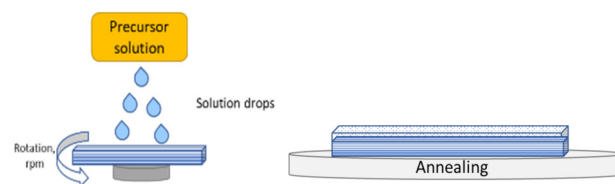


Fig. 1. Deposition process.

TABLE II. CHARACTERISTICS OF SPIN-COATING DEPOSITION OF MGO COATINGS

Solar cell code	Synthesis parameters	Total volume of precursor solution (µl)	Time process (min)
poly-Si/MgO/EtOH/0.3 poly-Si/MgO/EA/0.3 poly-Si/MgO/EG/0.3	1 drop of 100 µl at 1500 rpm, time 30 s 4 drops of 50 µl at 30 s range, 2000 rpm	300	3
poly-Si/MgO/EA/1.1	1 drop of 100 µl at 1500 rpm, time 30 s 10 drops of 100 µl at 30 s range, 2000 rpm	1100	6
poly-Si/MgO/EG/1	1 drop of 100 µl at 1500 rpm, time 30 s 9 drops of 100 µl at 30 s range, 2000 rpm	1000	6

D. Microscopic Characterization

The way of crystallization and formation of the SOL was observed by the optical microscopic method using the digital microscope Hayer Hy-2070. The results of the microscopic analysis are presented in Figure 2.

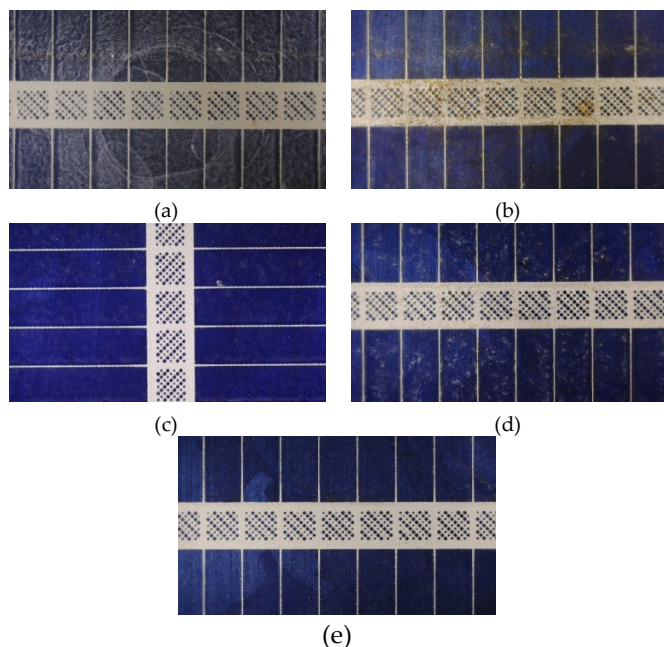


Fig. 2. Microscopic analysis of layers: (a) poly-Si/MgO/EtOH/0.3, (b) poly-Si/MgO/EA/0.3, (c) poly-Si/MgO/EG/0.3, (d) poly-Si/MgO/EA/1.1, (e) poly-Si/MgO/EG/1.

Microscopy analysis indicates the formation of some crystallization nuclei agglomerated towards the center of the cell corresponding to small amounts of precursor solutions. At the same time, it was noticed that when adding larger volumes (1 ml), the deposits WERE uniformly crystallized.

E. Optical characterization

The three dispersions OF MgO in EtOH, EA, and EG were subjected to UV-Visible spectroscopy by Ocean Optics HR2000+ in the range of 200-800 nm in order to get maximum absorbance, as can be seen in Figure 3.

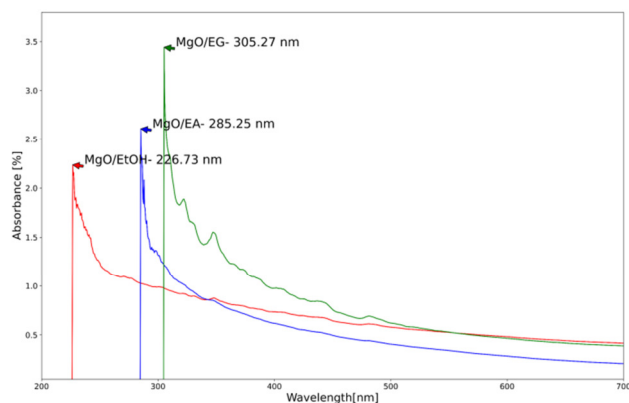


Fig. 3. UV-VIS spectra of MgO in EtOH, EA, EG.

MgO particles in EtOH exhibit a characteristic absorption peak in the ultraviolet (UV) region of the spectrum, typically around 226 nm. The UV-visible absorption spectrum of MgO particles in EA is similar to that in EtOH, with a characteristic absorption peak in the UV region of the spectrum, typically around 285 nm. This peak arises from the excitation of electrons from MgO's valence band to its conduction band. EA is a more polar molecule than EtOH, and this led to stronger interactions between the MgO particles and the EA molecules, which affect the absorption properties of the particles. As can be seen from the spectrum, the absorption peak around 285 nm is still quite pronounced, indicating strong UV absorption. The steep rise in absorption at shorter wavelengths suggests that MgO particles in EA absorb UV light efficiently. The UV-Vis absorption spectrum of MgO particles in EG is similar to that in EtOH and EA, with a characteristic absorption peak in the UV region of the spectrum, around 305 nm. EG is a more viscous solvent than EtOH and EA, and this may make it more difficult for the MgO particles to diffuse freely in the solvent. However, the absorption coefficient is slightly higher than that in EtOH and EA. The absorption properties of MgO nanoparticles in EG are of interest for SOL elaboration on poly-crystalline SCs.

III. SOLAR CELL TESTING SYSTEM

In accordance with the main objective of the experimental research, increasing the energy efficiency by obtaining SC covered with advanced SOL, a hardware & software system was designed and developed for the control of operating parameters and real-time monitoring of SC [22, 23]. Thus, the system of data acquisition and control of parameters was developed, in accordance with the software application for the fusion of data from the variation of ambient parameters (temperature, lighting) and functional electrical parameters of the SC (no-load voltage, load voltage, current short circuit).

A. Solar Cell Testing System

The proposed system for the control and acquisition of parameters that influence the operation of SCs is displayed in Figure 4. The testing system consists of:

- Solar cell block: In this block we find the SC type UFY002914 [22].
- Controlled electronic load block: It is implemented through a current source that fixes the current through the solar cell in the 0-30 mA range.
- Acquisition and control bloc.
- Stabilized, controlled power supply: This source provides the continuous voltages used by the other building blocks.

Each SC is connected to a controlled electronic load. All six electronic loads are synchronously controlled, using the same control parameters. For the experiments, 100 values were imposed for the load current through each solar cell, between 0 mA (open circuit) and 50 mA. The acquisition started 1 s after each value of the current was set. For each set value of the current, 50 values were considered for the solar cell voltage and 50 values for the current were drawn by the SC. The experiment was conducted as see in Figure 5.

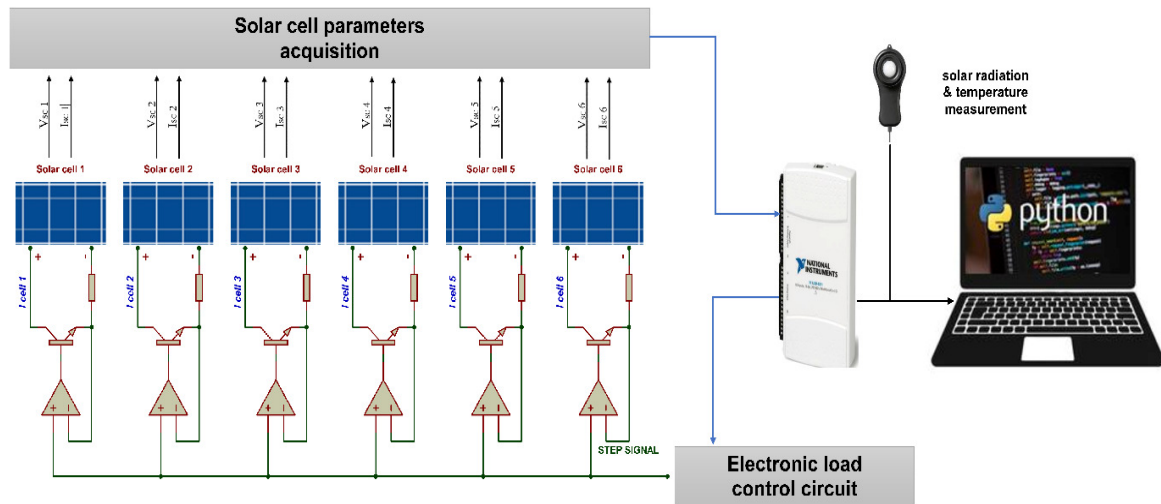


Fig. 4. Control system and acquisition of parameters from SCs using real atmospheric conditions.

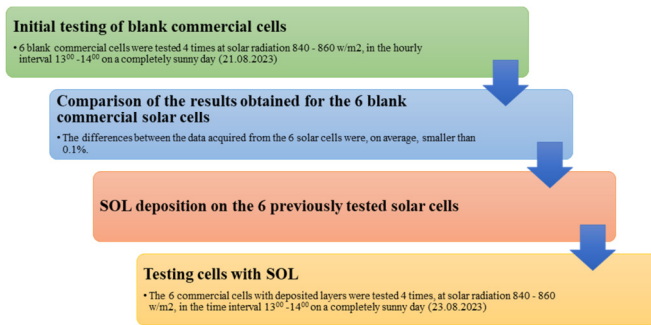


Fig. 5. The experimental process.

The entire management regarding the establishment of the parameters of the lighting level, the level of electronic load and the acquisition of SC parameters (V , I) in accordance with the set parameters, was carried out with the help of the acquisition and control board NI USB 6211 [22]. Its purchasing interface has, according to the manufacturer's data, the following input/output ports: 16 analog input channels (16 bits, 250 kS/s), 2 analog output channels (250 kS/s), and 8 digital input/output channels. In the case of the experiments carried out, the following of these were used: 2-analog inputs for the acquisition of the voltage of the solar cell and 1 analog output for controlling the electric output load. At the same time, the connection with the Python software application was made through the NI USB 6211. Python software was implemented to generate comparative graphs between the parameters of 2, 3, or more SCs, and the loaded vectors already having the optimal structure.

IV. RESULTS AND DISCUSSION

A. Comparative Analysis of I-V, P-V Characteristics

Each SC was subjected to tests in order to establish the I-V and P-V capacities, one without additional layers deposited and one after the deposition of oxide layers. The conditions for the electric charge at the output have always been the same, with the current through the load ranging from 0 to the maximum value I_{sc} . In the first stage, testing of the 6 standard p-type

commercial SCs (SC_{com}), without MgO coatings, was carried out (see Figure 6). No significant differences were found (variation of less than 0.2%) between the purchased parameters. Subsequently, the 6 p-type SCs covered with MgO films, from different solutions and different process parameters, were tested at the same conditions, to obtain information about the energy efficiency evolution of the coatings. In order to have the most accurate picture of the acquired values, each test was repeated 5 times, under approximately similar conditions, obtaining 5 value vectors for each performed test. Table III presents the correlation between the tests performed and the characteristics of the tested cells.

TABLE III. CORRELATION BETWEEN THE TESTS PERFORMED AND THE CHARACTERISTICS OF THE TESTED CELLS

Test cod	Solar cell sample
TEST 1	Average value for 6 SC _{com} – SC _{com}
TEST 2	Poly-Si/MgO/EG/0.3
TEST 3	Poly-Si/MgO/EA/1.1
TEST 4	Poly-Si/MgO/EA/0.3
TEST 5	Poly-Si/MgO/EG/0.5
TEST 6	Poly-Si/MgO/EtOH/0.3
TEST 7	Poly-Si/MgO/EG/1.0

The next stage was to conduct comparative analysis using the results of two or three tests as input values. For accuracy, this analysis was carried out taking into account the synthesis parameters of MgO coatings. The results of the testing led to the realization of a database, with 12000 values for each cell tested, including the values of the output voltage V , the value of the current by load I , and the value of the output power P . Analyzing the results (Figure 7) for SC_{com} versus the poly-Si/MgO/EtOH/0.3 cell, it can be revealed that the current intensity value through the SC and the output power measured, are higher for the SC_{com}. These characterizations conclude that the EtOH solvent does not advantage the deposition of a MgO coating considering as goal the increasing of SC energy efficiency. The comparative analysis of the efficiency of MgO coatings obtained with EA and EG, at different solution volumes, is shown in Figure 7.

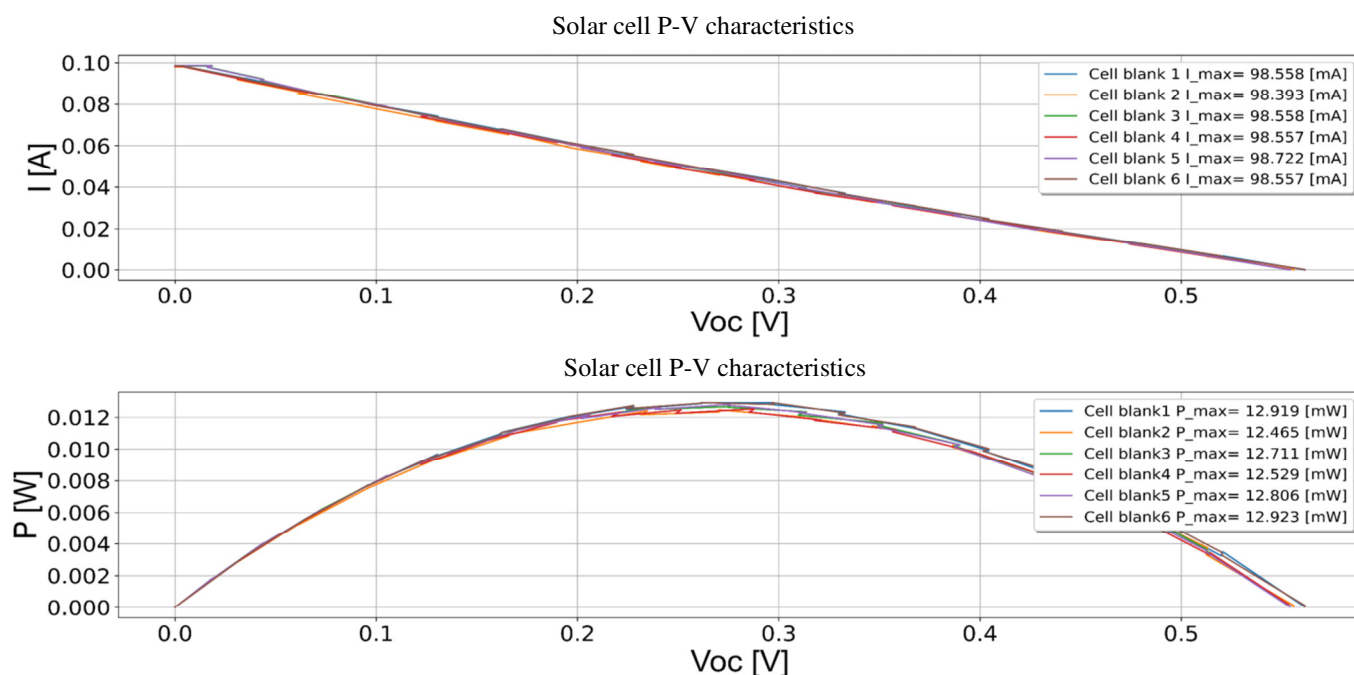


Fig. 6. Comparative analysis of I-V and P-V for six commercial cells before the deposition of additional layers.

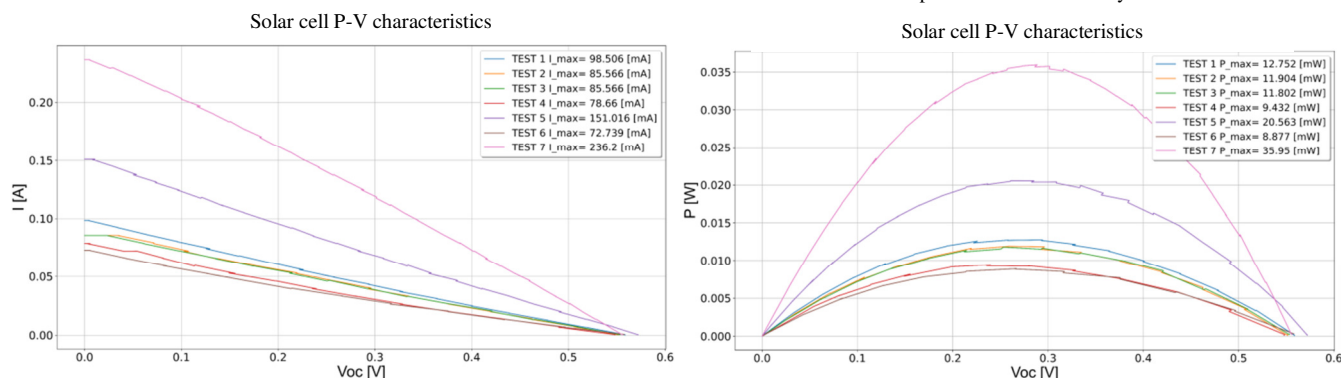


Fig. 7. Comparative analysis I-V, P-V between six commercial cells (after the deposition of additional layers) and the average value before the deposition of additional layers, TEST 1 - average value for 6 commercial SCs (blue), TEST 2 - Poly-Si/MgO/EG/0.3 (green), TEST 3 - Poly-Si/MgO/EA/1.1 (orange), TEST 4 - Poly-Si/MgO/EA/0.3 (red), TEST 5 - Poly-Si/MgO/EG/0.5 (magenta), TEST 6 - Poly-Si/MgO/EtOH/0.3 (brown), TEST 7 - Poly-Si/MgO/EG/1.0 (pink).

Comparing the results of testing the SC_{com} versus the poly-Si/MgO/EA/0.3 cell, it can be said that the current intensity value through the cell and the measured output power, are also higher in the case of SC_{com}. The average value (the average of the values recorded in the 5 acquisitions of the same test) of the maximum power determined is about 37% (3.53 W/m^2) higher in the case of SC_{com} versus the poly-Si/MgO/EA/0.3 cell. Also, the value of the open voltage (without load) for the poly-Si/MgO/EA/0.3 cell is smaller than that of the SC_{com} cell. These characterizations conclude that EA does not significantly favor increasing the energy efficiency of the MgO coatings on SCs, for the synthesis conditions, therefore the experimental protocol has been modified. To verify its influence on the electrical properties, volume solution was increased.

The intensity of the current through the SC and the output power measured for SC_{com} versus poly-Si/MgO/EA/1.1 cell, are smaller in the case of poly-Si/MgO/EA/1.1 cell with an

average decrease value of 14.22 mA. The average power value of poly-Si/MgO/EA/1.1 cell is smaller than the power value of the SC_{com} by 9.49%. Even if the average value of the electrical power offered by poly-Si/MgO/EA/1.1 is still lower than in the case of SC_{com}, a substantial increase in the power of poly-Si/MgO/EA/1.1 can be observed, compared to the SC covered with the same solvent, but with a smaller volume (poly-Si/MgO/EA/0.3), with $2,399 \text{ W/m}^2$ (25%). The results confirm that the EA solvent favors the formation of MgO coatings that can offer increasing energy efficiency at increased volumes (greater than 1 ml).

The same comparative analyzes were performed for the EG solvent, for different solution volumes. The maximum value of the current intensity for SC_{com} cell versus poly-Si/MgO/EG/0.3 cell was about the same, but higher for the SC_{com}. In the case of the poly-Si/MgO/EG/0.3 cell, an approximately constant value of the maximum current was acquired for all measurements. The average power value of

poly-Si/MgO/EG/0.3 cell is less than the power of the SC_{com}, but the differences are not significant, with percentage differences around 4% (4.98 W/m²).

Following these characterizations, it is observed that the EG solvent favors the increase of the current intensity value, compared to the other two solvents. This result agrees with the fact that the absorption coefficient of MgO in EG is slightly higher than that in EtOH and EA and can lead to the formation of more absorbent layers in the UV region. Respecting the previous reasoning of research, the experimental protocol was modified in terms of the volume of the dropped solution.

The output power measured for the SC_{com} was lower than that of the poly-Si/MgO/EG/0.5, the average difference being 49.5 mA. The power of the poly-Si/MgO/EG/0.5 cell, is greater than the power of the SC_{com} by an average of 59.59%. After obtaining these results, the increase of the volume of solvent deposited on the SC_{com} (1 ml) was continued. The output power measured for poly-Si/MgO/EG/1 cell was higher than that of SC_{com} with an average difference of 135.6 mA (182%). Compared to the same solvent, but for a smaller quantity (0.5 ml - poly-Si/MgO/EG/0.5), poly-Si/MgO/EG/1 offers an average increase of 36.36 W/m².

When using the same increased amount (1 ml) of precursor solution, but changing the type of solvent (EA vs. EG) the following results were achieved. The average value of the power obtained in the case of poly-Si/MgO/EG/1 is higher than in the case of the test poly-Si/MgO/EA/1.1 by 24.7 W/m². It is obvious that the use of the EG solvent leads to obtaining an MgO coating characterized by increased electric power. In average value, the resulting power for poly-Si/MgO/EG/1 cell is 206.8% higher than for poly-Si/MgO/EA/1.1 cell.

B. Comparative Energy Efficiency Analysis for the Tested Cells

Energy efficiency refers to how much electricity can be generated by an SC in a given time frame. The energy efficiency of SCs may vary depending on the type and quality of the SCs used, but also on other external factors. The efficiency of commercial SCs generally varies between 10 and 23%, depending on their type and quality. The efficiency of polycrystalline SCs is measured by the ratio between the electrical power generated by the cell and the sunlight energy absorbed by it. In general, polycrystalline SCs have a lower efficiency than monocrystalline SC, around 15-16%. Also, the energy efficiency of SCs can be improved by using special materials or by applying advanced technologies. High-performance SCs have been obtained under laboratory conditions with efficiencies of more than 26% [9]. In this paper, for the calculation of the energy efficiency of the solar cell we used the following relationship:

$$\eta_{solar\ cell} = \frac{P_{max-solar\ cell}}{Surface \times incident\ radiation\ flux} \times 100 \quad (1)$$

where $Surface = 0,000988\ m^2$, $P_{max-solar\ cell}$ [W] are the values acquired from the tested SCs, and incident radiation flux was measured with Voltcraft PL-110SM digital pyranometer, considering 3 scenarios: under Standard Test Conditions (STC - Irradiance 1000 W/m², Cell Temperature 25 °C, AM1.5), under Nominal Module Operating Temperature (NMOT - Irradiance

800 W/m², Ambient Temperature 20 °C, AM1.5, Wind Speed 1 m/s), and under REAL operating conditions (Irradiance 840 W/m², Ambient Temperature 28 °C, AM1.5). The comparative analysis of the energy efficiency of SCs coated with MgO layers obtained from different precursor solutions is portrayed in Figures 8 and 9.

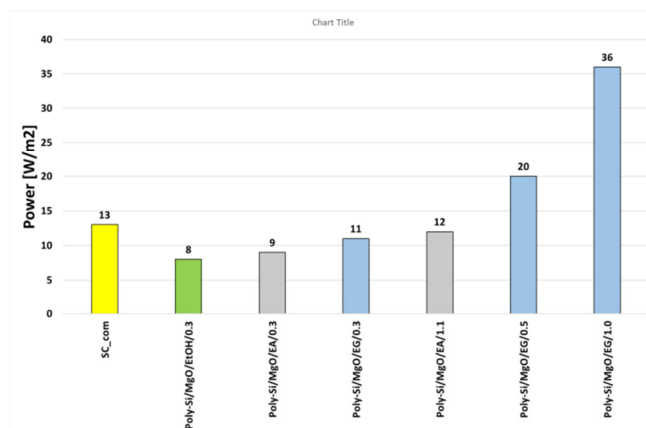


Fig. 8. Power energy of MgO coatings, obtained from different solvents [W/m²].

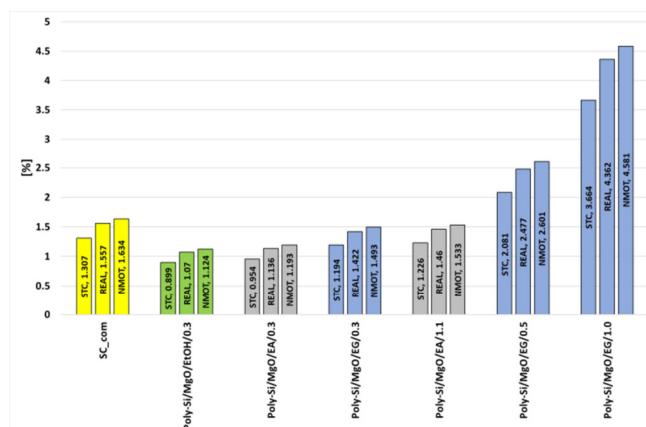


Fig. 9. Energy efficiency of MgO coatings, obtained from different solvents and for different operating conditions (yellow – blank SC, green – EtOH, gray – EA, blue – EG) in [%].

As in the case of the power flown by the tested SCs and in the case of energy efficiency (there is a direct dependence between these parameters), the best results were obtained for volumes of 1 ml solvent (for EA and EG), (see Figures 8, 9). Using the EtOH, it is observed that the efficiency of MgO deposited SCs was significantly lowered compared to the uncoated SCs. Making a comparison between polycrystalline SCs coated with additional coatings with the uncoated cells (Figure 10), where the uncovered cell represents the zero axis, we note that only poly-Si/MgO/EG/0.5 and poly-Si/MgO/EG/1 cells offer improved energy efficiency. If we utilize the same conditions as in the case of poly-Si/MgO/EG/1 cell to cover a commercial solar panel Sharp ND-RB270 [24, 25] of 164.9 W/m² and 16.5% efficiency, the new output power, on the STC favorable conditions (no shading, temperatures in the nominal operating range) increases up to 262.4 W/m² and 26.2%.

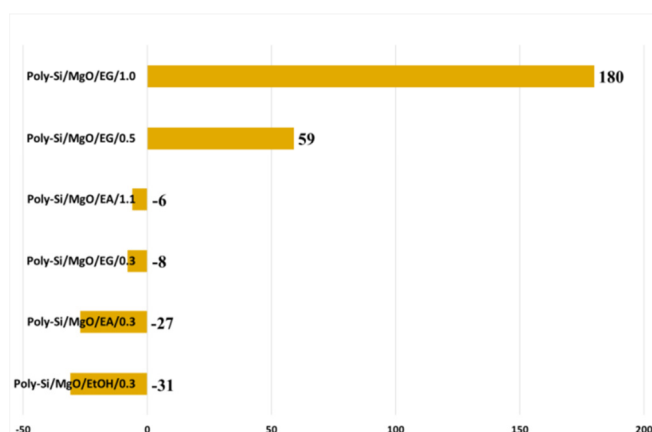


Fig. 10. Increase / decrease of energy efficiency percentage compared to the average efficiency value of commercial blank cells.

V. CONCLUSIONS

The novelty of this paper is supported by the following aspects: (a) The presented method for increasing the energy efficiency of Solar Cells (SCs) is applicable for any type of commercial silicon SCs. (b) For certain process parameters (type of solvent used and volume of solution applied), an electric power value almost twice that of the same cell without additional layer was obtained. (c) The implementation of the deposition process at an industrial level has low costs.

The main contributions of the paper are represented by the development of additional oxide layers deposited on commercial SCs, starting from MgO, dispersed in different solvents, which lead to the increase of electrical power, compared to the same SCs without additional layers and the realization of a complex system for testing SCs.

- MgO coatings have been successfully obtained on commercial polycrystalline SCs by spin-coating deposition starting from the dispersion of MgO into different solvents. Testing of electrical properties of the coated SCs in real conditions indicates that additional MgO coatings can increase energy efficiency. The carried out comparative analysis of the tests shows the importance of the deposition process of additional layers to improve the electrical performance of SCs. The experimental characterization was performed for the over 12000 recorded values of current intensity and output power, specific to each tested cell.
- Under the conditions of using ethanolamine solvent, a slight improvement of power was observed only when the volume of the deposited solution reached 1 ml. For smaller solution volumes, the obtained coating did not provide a substantial conversion of light into electricity.
- The current intensity through the solar cell and the output power measured are higher in the case of poly-Si/MgO/EG/1, corresponding to MgO coating by spun of 1 ml ethylene glycol precursor solution. The power of these is greater than the power of the commercial cell by 180%.
- Regarding the spin-coating method of synthesizing MgO coatings, it is recommended to use volumes larger than 1 ml

to achieve a uniform film at a high rotational speed (2000 rpm), with 30 s time interval between droplets, and total process time of 6 min.

Future research will focus on testing the influence of adding additional layers on commercial polycrystalline cells, testing the energy efficiency of the layers added by the spraying method combined with spin coating, the use of other types of metal oxides (e.g. ZnO), and defining the life span of a cell coated with MgO films. The use of magnesium oxide nanoparticles (through solar energy through SPVD - Solar Physical Vapor Deposition) in different volumes of solvent and at different parameters of the deposition process is also a concern for further studies.

REFERENCES

- [1] "Renewable electricity generation," *Our World in Data*, <https://ourworldindata.org/grapher/modern-renewable-energy-consumption>.
- [2] M. Jomaa, M. Abbes, F. Tadeo, and A. Mami, "Greenhouse Modeling, Validation and Climate Control based on Fuzzy Logic," *Engineering, Technology & Applied Science Research*, vol. 9, no. 4, pp. 4405–4410, Aug. 2019, <https://doi.org/10.48084/etasr.2871>.
- [3] A. Alanazi, "Optimization of Concentrated Solar Power Systems with Thermal Storage for Enhanced Efficiency and Cost-Effectiveness in Thermal Power Plants," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12115–12129, Dec. 2023, <https://doi.org/10.48084/etasr.6381>.
- [4] F. Mavromatakis, Y. Franghiadakis, and F. Vignola, "Modeling Photovoltaic Power," *Engineering, Technology & Applied Science Research*, vol. 6, no. 5, pp. 1115–1118, Oct. 2016, <https://doi.org/10.48084/etasr.612>.
- [5] S. Ho, "A Review of Metal Oxide Thin Films in Solar Cell Applications," *International Journal of Thin Film Science and Technology*, vol. 11, no. 1, pp. 37–45, Jan. 2022, <https://doi.org/10.18576/ijtfst/110105>.
- [6] M. Jlassi, I. Sta, M. Hajji, and H. Ezzaouia, "NiO thin films synthesized by sol-gel: Potentiality for the realization of antireflection layer for silicon based solar cell applications," *Surfaces and Interfaces*, vol. 6, pp. 218–222, Mar. 2017, <https://doi.org/10.1016/j.surfin.2016.10.006>.
- [7] Y. Lee, C. Park, N. Balaji, Y.-J. Lee, and V. A. Dao, "High-efficiency Silicon Solar Cells: A Review," *Israel Journal of Chemistry*, vol. 55, no. 10, pp. 1050–1063, 2015, <https://doi.org/10.1002/ijch.201400210>.
- [8] I. Zeghib and A. Chaker, "Efficiency of a Solar Hydronic Space Heating System under the Algerian Climate," *Engineering, Technology & Applied Science Research*, vol. 6, no. 6, pp. 1274–1279, Dec. 2016, <https://doi.org/10.48084/etasr.875>.
- [9] S. Kim, J. Park, P. D. Phong, C. Shin, S. M. Iftiqar, and J. Yi, "Improving the efficiency of rear emitter silicon solar cell using an optimized n-type silicon oxide front surface field layer," *Scientific Reports*, vol. 8, no. 1, Jul. 2018, Art. no. 10657, <https://doi.org/10.1038/s41598-018-28823-x>.
- [10] A. M. Mouafki, F. Bouaicha, A. Hedibi, and A. Gueddim, "Porous Silicon Antireflective Coatings for Silicon Solar Cells," *Engineering, Technology & Applied Science Research*, vol. 12, no. 2, pp. 8354–8358, Apr. 2022, <https://doi.org/10.48084/etasr.4803>.
- [11] B. R. Ali, "Synthesis And Characterization Of Nanoparticles Mgo Films On Silicon Substrates For Solar Cells Applications," *Journal of Multidisciplinary Engineering Science Studies*, vol. 2, no. 7, pp. 610–612, 2016.
- [12] A. Kulkarni, A. K. Jena, H.-W. Chen, Y. Sanehira, M. Ikegami, and T. Miyasaka, "Revealing and reducing the possible recombination loss within TiO₂ compact layer by incorporating MgO layer in perovskite solar cells," *Solar Energy*, vol. 136, pp. 379–384, Oct. 2016, <https://doi.org/10.1016/j.solener.2016.07.019>.

- [13] S. Huang, B. Kang, L. Duan, and D. Zhang, "Highly efficient inverted polymer solar cells by using solution processed MgO/ZnO composite interfacial layers," *Journal of Colloid and Interface Science*, vol. 583, pp. 178–187, Feb. 2021, <https://doi.org/10.1016/j.jcis.2020.09.047>.
- [14] J. Dagar, S. Castro-Hermosa, G. Lucarelli, F. Cacialli, and T. M. Brown, "Highly efficient perovskite solar cells for light harvesting under indoor illumination via solution processed SnO₂/MgO composite electron transport layers," *Nano Energy*, vol. 49, pp. 290–299, Jul. 2018, <https://doi.org/10.1016/j.nanoen.2018.04.027>.
- [15] X. Guo, H. Dong, W. Li, N. Li, and L. Wang, "Multifunctional MgO Layer in Perovskite Solar Cells," *ChemPhysChem*, vol. 16, no. 8, pp. 1727–1732, 2015, <https://doi.org/10.1002/cphc.201500163>.
- [16] J. Ma *et al.*, "MgO Nanoparticle Modified Anode for Highly Efficient SnO₂-Based Planar Perovskite Solar Cells," *Advanced Science*, vol. 4, no. 9, 2017, Art. no. 1700031, <https://doi.org/10.1002/advs.201700031>.
- [17] O. Diachenko *et al.*, "Surface Morphology, Structural and Optical Properties of MgO Films Obtained by Spray Pyrolysis Technique," *Acta Physica Polonica A*, vol. 3, no. 130, pp. 805–810, 2016, <https://doi.org/10.12693/APhysPolA.130.805>.
- [18] B. Dhamodharan and S. Periyasamy, "Analysis of Solar Cell with MGO Anti-Reflective Coating," *International Journal for Scientific Research & Development*, vol. 4, no. 4, pp. 415–417, Jan. 2016.
- [19] J. You *et al.*, "Improved air stability of perovskite solar cells via solution-processed metal oxide transport layers," *Nature Nanotechnology*, vol. 11, no. 1, pp. 75–81, Jan. 2016, <https://doi.org/10.1038/nnano.2015.230>.
- [20] S. Perera *et al.*, "The Effect of MgO on the Enhancement of the Efficiency in Solid-State Dye Sensitized Photocells Fabricated with SnO₂ and CuI," *Bulletin of the Chemical Society of Japan*, vol. 76, no. 3, pp. 659–662, Mar. 2003, <https://doi.org/10.1246/bcsj.76.659>.
- [21] T. Huma, N. Hakimi, M. Younis, T. Huma, Z. Ge, and J. Feng, "MgO Heterostructures: From Synthesis to Applications," *Nanomaterials*, vol. 12, no. 15, Aug. 2022, Art. no. 2668, <https://doi.org/10.3390/nano12152668>.
- [22] V. Calinescu, M. Oproescu, G.-V. Iana, O. C. Novac, and M. C. Novac, "Efficiency of Nanostructured Layers Deposited on Solar Cells - hardware system proposal," in *14th International Conference on Electronics, Computers and Artificial Intelligence*, Ploiesti, Romania, Jul. 2022, pp. 1–6, <https://doi.org/10.1109/ECAI54874.2022.9847309>.
- [23] A. Elgharbi, D. Mezghani, A. Mami, and A. Gharbi, "Intelligent Control of a Photovoltaic Pumping System," *Engineering, Technology and Applied Science Research*, vol. Vol.9, pp. 4689–4694, Oct. 2019, <https://doi.org/10.48084/etasr.2982>.
- [24] "Sharp ND-RB270," https://www.europe-solarstore.com/sharp-nd-rb270.html?gad_source=1&gclid=CjwKCAiAyp-sBhBSEiwAWWzTnqRIPK-ur7_DQeIKmS18s-d_D7ZgrJ5yw7we5dk_b9TCQSPnrfOXnBoCN0EQAvD_BwE.
- [25] M. E. Bendib and A. Mekias, "Solar Panel and Wireless Power Transmission System as a Smart Grid for Electric Vehicles," *Engineering, Technology & Applied Science Research*, vol. 10, no. 3, pp. 5683–5688, Jun. 2020, <https://doi.org/10.48084/etasr.3473>.

Evaluation of the Effect of Access Point Density on the Safety of Primary Roads. A Case Study

Ashar Ahmed

Department of Urban and Infrastructure Engineering, NED University of Engineering and Technology, Karachi, Pakistan
aahmed@cloud.neduet.edu.pk (corresponding author)

Md. Kamrul Islam

Department of Civil and Environmental Engineering, College of Engineering, King Faisal University, Al-Ahsa, Saudi Arabia
maislam@kfu.edu.sa

Ahmad Farhan Mohd Sadullah

Universiti Putra Malaysia, Malaysia
cefrhn@upm.edu.my

Uneb Gazder

Department of Civil Engineering, College of Engineering, University of Bahrain, Bahrain
ugazder@uob.edu.bh

Received: 13 January 2024 | Revised: 25 January 2024 | Accepted: 28 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6905>

ABSTRACT

This paper investigates the nuanced exploration of access point density's influence on accident frequency, specifically focusing on primary roads in Malaysia. The analysis is multifaceted, considering geographic variations, land use patterns, and the density of access points per km. This investigation scrutinizes the direct relationship between the number of access points per km and the corresponding accident frequency. A critical threshold value for access point density is identified, revealing its consequential impact on average accident frequency. The observed direct proportionality between access point density and accident frequency is a pivotal discovery. Moreover, the role of land use parameters emerges as a key determinant in understanding how accident frequency varies with access point density, particularly on specific road types. This establishes eight access points per km as a potential threshold value for ensuring optimal access point density within a road network. In summary, this study provides insights into the intricate dynamics of access point density and its consequential impact on road safety. The identified threshold value and the recognition of the role of land use contribute valuable perspectives for informed decision-making in road network planning and management.

Keywords-crash frequency; road geometry; land use; unsignalized intersection; access point per kilometer

I. INTRODUCTION

Access commonly refers to the provision of a path through which a vehicle can enter or exit a road or street. Formally an Access Point (AP) in a transportation network includes driveways, street connections [1, 2], and unsignalized intersections. As per their functionality, there is no significant difference between an AP and an unsignalized intersection but the former can be referred to as a facility that provides access to minor-streets, side-streets, and private driveways from a

building to a major road [3], while the latter is primarily referred to as a facility that occurs when two roads meet each other, with one road classified as the minor and the other as the major. Unsignalized intersections can be three-legged or T-junctions and four-legged or crossroads, whereas APs are mostly T-junctions. A driveway can be further divided into several types, such as major, minor, commercial, residential, and industrial/institutional [3]. These driveway types are parts of every road network and development schemes, therefore, evaluation of their contribution to the safety of the road

network is necessary. It is a known fact in the field of transportation engineering that access and mobility are inversely proportional to each other. An increase in access will decrease the mobility and vice versa, as shown in Figure 1. The increase in access, on the other hand, increases the exposure to risk, because the addition of any AP adds to the number of merging and diverging movements through traffic [4]. Thus, the AP density, measured in terms of the number of APs per km, affects the safety of highways and streets. It has been reported [5] that an increase of one AP per km will increase the number of accidents by 4% for a given volume on a particular road. There is a positive association of AP density with the accident rate per million vehicle mile travel [6].

The roads which fall under the federal government are termed as primary roads [7] and can be classified as arterial roads. This paper presents the results of the field survey conducted on 28 roads in Penang, Malaysia. Geometric parameters like the width of major and minor roads, the length of the entire major road, the number of intersection legs, lane marking, and shoulder width, control parameters like stop signs/line or no control at all, and social parameters like land use pertinent to the APs on the roads under consideration, were monitored. The effect of the number of APs per km on the number of accidents was analyzed. A threshold value for the AP density was determined and its impact on the average accident frequency is presented.

II. LITERATURE REVIEW

The literature on road safety management and APs reveals a comprehensive understanding of the factors influencing accident frequency, particularly in the context of AP density. Several studies contribute valuable insights into the geometric and functional design of APs, shedding light on their impact on safety. Authors in [8] conducted a comprehensive study on the management of uncontrolled APs, emphasizing the importance of considering design criteria and risk factors associated with these road network elements. The research focused on APs without traffic control features, such as traffic lights or stop signs, highlighting safety issues related to their location, design, and visibility. By analyzing the geometric and functional design of uncontrolled accesses, the study aimed to provide insights into defining more adequate design criteria for improving safety. Authors in [9] delved into the association between driveway land use and safety performance on rural highways, particularly in the context of county roads in Michigan. While prior research had explored the safety impacts of driveway density, this study extended the focus to examine the effects of different classifications of driveway land utilization. The findings indicated that commercial driveways had a more pronounced effect on crash occurrence compared to other land use types, contributing valuable insights into the nuanced relationship between land use, driveway density, and road safety. Authors in [10] evaluated the safety impacts of access management techniques in Utah, addressing the growing traffic volumes and congestion on arterial streets. The study employed a geographic information system (GIS)-enabled data almanac to assess crash data and identify high crash locations. Although the study did not uniformly observe reduced crash rates, it highlighted safety improvements, particularly in eliminating the severity of collisions. The findings emphasized the need to consider a range of safety outcomes beyond simple crash rates when evaluating access management techniques. Authors in [11] focused on the effect of midblock APs on traffic accidents on state highways in New Jersey. Through statistical analysis, the study identified access density as a contributing factor to accidents, emphasizing that while it was a significant factor, it was not the sole determinant. The research provided valuable insights into the distribution of accidents along state highways, distinguishing between section accidents and those occurring at signalized intersections. Authors in [12] expanded the focus beyond access management characteristics, examining the relationships between various physical characteristics of arterial roadways and safety. This research

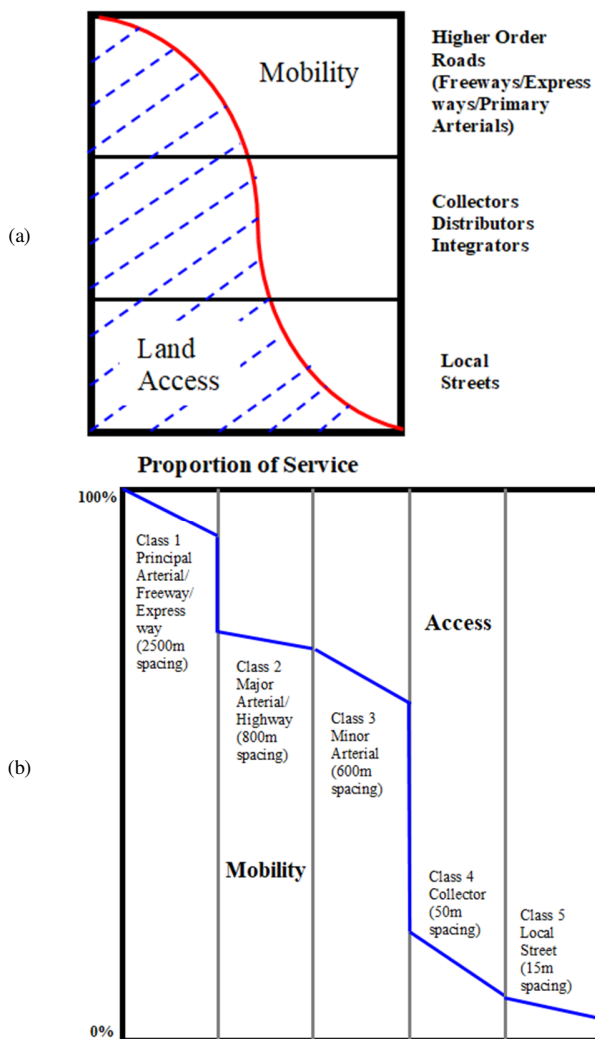


Fig. 1. (a) Effect of mobility and accessibility on highway classification, (b) mobility versus accessibility in highway classification.

The importance given to the APs or unsignalized intersection density per km in the literature and their contribution in decreasing road safety provided the motivation for this study. This paper aims to explore the effect of AP density on the accident frequency at primary roads in Malaysia.

brought attention to the positive relationship between specific access management-related characteristics and increased crash rates and severity. Additionally, it highlighted the role of land use in influencing the safety of arterial roadways, particularly the association between adjacent commercial land use and higher crash rates. The impact of access density on the number of crashes at commercial roadways, considering various access management techniques was explored in [13]. The study, conducted in Oregon, involved a comprehensive analysis of crash incidence per mile in relation to access density. The results indicated a correlation between crashes and access density, suggesting that factors, such as average daily traffic and number of lanes may also influence the number of segment crashes per mile. Authors in [14] presented a quantitative framework for evaluating highway access safety programs, emphasizing the importance of performance metrics and a decision-aiding framework. The framework applied multicriteria analysis and cost-benefit analysis with parameter uncertainties, offering a systematic approach to assess the relative needs at existing APs. This research demonstrated the applicability of the framework across large corridor networks and diverse time horizons. Authors in [15] investigated the effect of access-point density on traffic crashes on a rural two-lane highway, specifically the Ahar-Tabriz Highway. The study, conducted through a case study approach, provided valuable insights into the relationship between AP density, spacing, geometric features, and crash rates. The results indicated a direct correlation between AP density and accident rates, emphasizing the role of AP specifications in highway safety.

III. STUDY AREA AND DATA COLLECTION

The state of Penang, Malaysia, chosen as the study area for this research, represents a compelling and dynamic case study for the evaluation of the effect of AP density on the safety of primary roads. Nestled on the northwest coast of Peninsular Malaysia, Penang is a state comprising the island of Penang and a portion of the mainland known as Seberang Perai. Penang presents a dynamic and diverse landscape that encompasses both urban and rural types of environment. The region offers a unique blend of economic vitality and cultural diversity. The arterial roads of Penang Island weave through densely populated cities, catering to a spectrum of economic activities, while Seberang Perai introduces elements of rural landscapes into the road network. With a mix of commercial hubs, industrial zones, and agricultural areas, Penang's road network reflects its economic significance and multicultural population. The interconnected highways serve not only as conduits for urban commuters, but also as vital links for rural communities. As Penang undergoes rapid urbanization and economic growth, the study seeks to delve into the safety implications of AP density, examining the ways the varied road network influences safety across urban and rural contexts. This dynamic study area is poised to provide valuable insights into the interplay between AP density and road safety, contributing to local policy considerations and also offering lessons of broader significance in the context of global urban development and transportation challenges.

The methodology formulated for the field data collection was simple and required minimal equipment and manpower. Regarding the survey, only a safety vest, a measuring wheel and a set of noting pad and pens were needed. A single person can easily conduct the survey of a particular roadway during normal working hours. The procedure comprised the following steps:

- Identifying the survey location using Google Maps.
- Taking screenshots of the survey location from the beginning to the end of the road.
- Arranging the screenshots according to their sequence.
- Start the survey from the beginning of the road by walking. Spot any unsignalized intersection that comes first on the way. Mark its position on the map and give it a serial number for identification.
- Measuring the widths of the minor and major roads as shown in Figure 2.

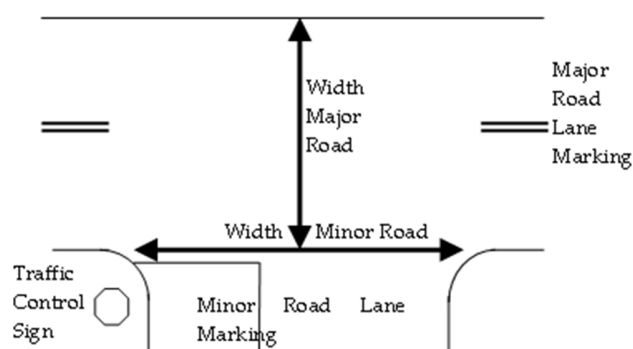


Fig. 2. Widths of the minor and major roads.

The survey was carried out between May and July 2013. A total of 28 roads were surveyed in the state of Penang, Malaysia. Table I displays a list of the parameters used in this study on the roads under consideration. As mentioned above, the state of Penang consists of two different geographical parts. One is located on the west side of Peninsular Malaysia and is a part of the mainland, while the other is an island. The parts have different topography, land use, sociology, and economy. Therefore, the roads surveyed were divided based on their location and a comparative analysis of all the parameters collected was performed with respect to geographical location. The names and lengths of the roadways are presented in Table II.

TABLE I. PARAMETER LIST

Parameter type	Parameter
Geometric	Width of major roads
	Width of minor roads
	Length of the entire major road
	Number of legs of intersection
	Shoulder width
Control	Stop sign/line
	Lane marking
Social	Land use: city, town or rural area

A. Descriptive Analysis

The data collected for this study consisted of two parts. The first part was related to the number of APs or unsignalized intersections on a particular road and the geometry of each AP. These data were acquired through field surveys as explained above. The second part comprised the accident data which were acquired from the Malaysian Institute of Road Safety Research (MIROS). The data consisted of the total number of accidents that occurred along each road for a period of six years, from 2006 to 2011. Descriptive analysis of the data was performed and the effect of various parameters, such as geographical boundaries, land use, number of APs per km on the frequency of road accidents were analyzed. The descriptive statistics of the accident data acquired from MIROS and geometric parameters obtained through field survey, are presented in Table III.

TABLE II. NAME AND LENGTH OF SURVEYED ROADS WITH RESPECT TO GEOGRAPHICAL LOCATION

S. No.	Name of Road	Road length (km)
Island		
1	Lebuhraya Thean Teik	3.1
2	Jalan Paya Terubong	8.4
3	Jalan Persiaran Gurney	1.9
4	Jalan Cantonment	1.1
5	Jalan Air Itam	0.9
6	Jalan Padang Tembak	1
7	Jalan Burma	3.6
8	Jalan Jelotong	2.3
9	Jalan Bukit Gambir	5.6
10	Jalan Sri Tan Hi Ewe Lim	1.2
11	Jalan Tanjung Bungah	2.6
12	Jalan Batu Gantung	0.8
13	Jalan Penaga Z0372	0.55
14	Jalan Sultan Azlan Shah	11.5
Mainland		
1	Jalan Raja Uda	4.1
2	Jalan Bagan Ajam	0.75
3	Jalan Pongsu Seribu	2.7
4	Jalan Besar Nibong Tebal	0.65
5	Lorong Mak Mandin 5	1.2
6	Jalan Bagan Lalang	2
7	Jalan Pengkalan Macang	0.07
8	KM22 Jalan Besar Simpang Ampat	0.16
9	Jalan Bertam 1, Kepala Batas	0.75
10	Jalan Dua Kepala Batas	1.1
11	Jalan Merbau Kudung	0.15
12	Jalan Telaga Air	0.8
13	Jalan Padang Bengali	4.7
14	Jalan Penaga P001	4.3

TABLE III. DESCRIPTIVE STATISTICS OF THE ACCIDENT AND FIELD SURVEY DATA

	Mean	Min	Max	SD
Number of accidents	8.71	2	31	7.57
APs per km	9.58	3	43	7.77
Average road width (m)	10.54	6.6	14.7	2.05
Average AP width (m)	13.61	10.1	21.4	2.69
Average shoulder width (m)	0.98	0	3.75	0.94

IV. RESULTS AND DISCUSSION

Access commonly refers to the provision of a path through which a vehicle can enter or exit a road or street. Formally an AP in a transportation network includes driveways, street connections, minor unsignalized T-intersections [22]. This section provides a discussion about the effect of the geographical distribution on the geometry of APs, the effect of land use followed by the effect of AP density on accidents. A threshold for the number of APs per km was determined and the way it influences the average accident frequency was examined.

A. Effect of Geography on Geometric Parameters

The geometry of an AP plays an important role in the severity and frequency of accidents that occur on a particular road. In this study, each road encompassed several APs, thus, the geometric characteristics of all the APs on each road were averaged. Table IV provides a quick reference to their values. Since the socio-economic, demographical, and geographical conditions of the mainland and island were different, the geometric characteristics of their access points also varied.

TABLE IV. GEOMETRIC PARAMETERS WITH RESPECT TO GEOGRAPHICAL LOCATION

Parameter	Mainland	Island
Average length of road (km)	1.67	3.18
Average major road width (m)	10.92	10.16
Average minor road width (m)	12.74	14.48
Average number of 3-leg APs	11.57	19.36
Average number of 4-leg APs	1.36	1.64
Average number of APs with double line lane marking	4.86	6.86
Average number of APs with single line lane marking	5.36	4.29
Average left shoulder width (m)	1.06	1.05
Average right shoulder width (m)	1.12	0.78
Average number of accidents	8.50	8.93

The average lengths of the roads on the island were greater than those on the mainland as they encompassed a larger area and connected several points stretched along the road. Higher widths of major roads were observed on mainland, which depicts that land acquisition is more difficult on the island due to its higher land value. For the same reason, the shoulder widths of the APs along the major roads were more on the mainland. Contrary to major roads, the widths of minor roads were greater on the island, which indicates that the minor roads of the APs on the island are required to cater for more traffic as compared to the mainland. Both the average numbers of 3-leg unsignalized intersections or APs as well as 4-leg unsignalized intersections were higher on the island. Consequently, it can be inferred that the density of APs in general is higher on the island. The effect of the density of unsignalized intersection/AP per km is discussed in detail in the subsequent section. The primary purpose of lane markings provided along the middle of the road is to separate the traffic that moves in opposite directions, but they also serve as a surrogate measure for traffic volume. Double line lane markings are usually provided on roads that have a higher volume while single line lane markings are provided on roads that have a comparatively lower volume. This could be the probable reason why the average numbers of

APs with double line lane markings on the major road were higher on the island. Similarly, the average numbers of APs with single line lane markings on the major road were higher on the mainland. The effects of all the geometric parameters discussed above explain the reason behind the higher frequency of accidents observed on the APs located in the island.

B. Effect of Land Use

Various types of land use generate different magnitudes of trip [16], which can affect crash frequency. Thus, the risk of accident due to APs that lie on major roads situated in different locations can be attributed to their land use. Variation in land use in turn causes variation in the density of APs per km. This is evident in [17], where it was reported that the Department of Transportation in the state of Minnesota in USA uses area type, similar to land use type, as one of its selection criteria for spacing between APs [18]. Therefore, locations which are composed of a greater percentage of high trip land use may experience a higher crash frequency. Figure 3 shows the distribution of land use in the two geographical boundaries of Penang state. The results of the survey indicated that the part that was located in the island was more urbanized and had a higher percentage of cities and towns while areas that can be classified as small towns were only 7% with no presence of rural land use at all. Dissimilarly, the part which was in the mainland was less urbanized and had a higher percentage of small towns and rural land use. Therefore, the higher frequency of accidents observed on the APs in the island, as mentioned in Table II, can be attributed to the high number of merging and weaving maneuvers that take place due to a greater number of trips generated because of the greater percentage of urbanized land use. This result is in accordance with the findings in [19].

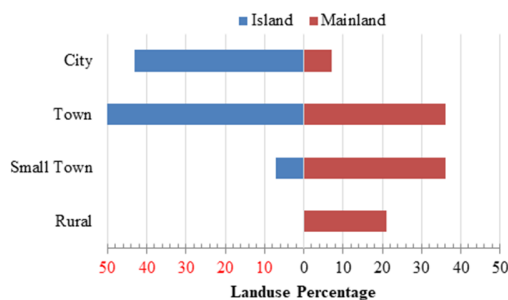


Fig. 3. Distribution of land use with respect to the geographical location.

C. Effect of Unsignalized Intersection/Access Point Density

There was a remarkable difference between the density of APs in the island and the mainland. The average number of unsignalized intersections/APs observed on each road in the island were 1.7 times higher than in the mainland. Figure 4 depicts a comparison between the AP density and the accident frequency. These values were obtained by dividing the total number of APs observed on all the roads surveyed in each part of the state with the total number of the roads in each part. Thus, they do not represent the density in terms of Aps per km. It can be seen that the increase in the number of unsignalized intersections or APs raises with the number of accidents.

D. Access Point Density Threshold

The density of APs per km was calculated by dividing the total number of unsignalized intersections observed on each road with the total length of the road. The general density of APs per km for the entire state was found to be 9.6, which is close to the findings of [20], in which relevant data were provided and the calculated value was 8.55. The accidents were further grouped with respect to the number of APs per km and the respected cumulative frequency percentage curve was plotted. From this curve, the value of the number of Aps per km that accounted for 50% of the cumulative accidents was found, as illustrated in Figure 5. This is called the 50th percentile value and is a common statistical parameter related to grouped data. It bisects the data into two categories with respect to the independent variable and helps further examination by evaluating the relationship of the dependent variable with each category. The graph indicated that an AP density of 7.1 to 8 per km is critical in relation to the accident frequency. The 50th percentile value of APs per kilometer lied between 7.1 to 8, so, the threshold value of 8 APs per km was chosen and a further analysis with respect to accident frequency was performed.

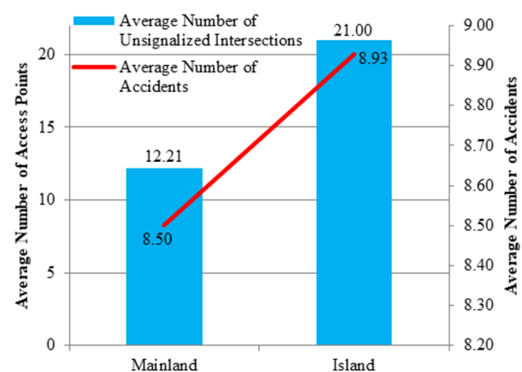


Fig. 4. Average number of APs and accidents with respect to the geographical location.

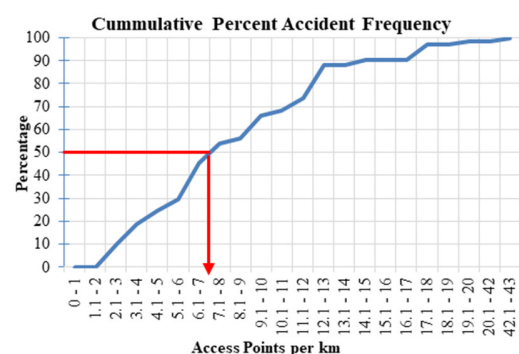


Fig. 5. AP density threshold.

E. Access Point Density Threshold and Average Accident Frequency

The threshold value of 8 APs per km was found from the analysis of the cumulative percentage of the number of accidents. Therefore, the data regarding the number of

accidents on each road were grouped into two categories. The first group involved the roads that had less than 8 APs/km and the second group covered the roads that had at least 8 APs/km, as portrayed in Figure 6. Their respective average accident frequency was calculated by dividing the total number of accidents observed on all roads in each group with the total number of roads. The overall analysis of the concentration of unsignalized intersections or Aps/km of all the 28 surveyed roads, indicated that accident frequency was directly proportional to the concentration of APs. This is a very important finding, which is also evident from the results reported in other studies and the conclusions drawn from them [17, 21, 22]. Authors in [23] disclosed a decrease in the risk of accidents with the increase in AP density, but accepted that this result is not conformal with other studies [24] in which a positive association was found. It has been documented [24] that higher AP densities and the presence of bus stops tend to increase the risk of severe crashes. Incorporating the effect of speed, authors in [17] announced that the spacing between APs should be 45.72 m for roads having speed limit of 56 km/hr, as applied by the DOTs of Georgia, West Virginia, and New Jersey. This equals a density of 22 AP/km. In the same study, it was recorded that the spacing between APs should be 109.72 m for roads having speed limit of 72 km/hr, as applied by the DOTs of Minnesota and Texas. This equals a density of 9 APs/km.

It should be noted that AP density is one parameter that affects the frequency of accidents along various classes of roads. The literature indicates that apart from this parameter, other factors that affect accident occurrence are over-speeding, negligent driving, alcohol intake, and violation of traffic laws [25, 26, 27]. As a result, these factors are required to be explored in combination with the AP density.

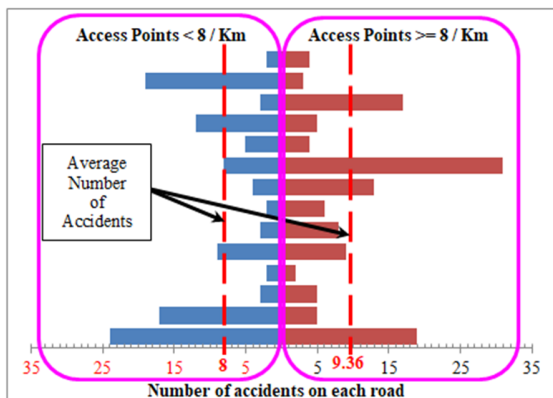


Fig. 6. Average number of accidents with respect to AP density per km.

V. CONCLUSION

In conclusion, this study underscores the direct proportional relationship between access point density and accident frequency. A value of eight access points per kilometer has been identified as the threshold after which an increase in accident frequency can be expected on primary roads. Land use, particularly in areas generating high traffic, further amplifies this impact. It is recommended that road design and

management strategies consider limiting access points, especially in regions with commercial, educational, industrial, or high-density residential land use. At the micro level, the findings reveal that, on an individual road basis, the number of accidents does not exhibit a clear relationship with access point density. However, at the macro level, a positive association emerges when comparing roads with fewer than 8 access points per kilometer to those with 8 or more. This macro-level insight aligns with the findings in the literature, supporting the notion that access point density values exceeding 8 can compromise road safety. Moving forward, it is crucial for future investigations to incorporate additional microscopic traffic parameters, such as spot speed and spacing between vehicles. Acknowledging this study limitations, further research is anticipated to refine the understanding of the intricate relationship between access point density and accident frequency, providing valuable insights for road safety, design, and urban planning.

ACKNOWLEDGMENT

This work was financially supported by the Deanship of Scientific Research at the King Faisal University, Saudi Arabia (Grant: 5536). The authors are also obliged to the Malaysian Institute of Road Safety Research (MIROS) and the Royal Malaysian Police for providing the accident data for this study and acknowledge their support.

REFERENCES

- [1] M. Williamson and H. Zhou, "A Study of Safety Impacts of Different Types of Driveways and their Density," *Procedia - Social and Behavioral Sciences*, vol. 138, pp. 576–583, Jul. 2014, <https://doi.org/10.1016/j.sbspro.2014.07.241>.
- [2] *Access Management Manual*. Design Division (DES), Texas Department of Transportation, 2011.
- [3] *Highway Safety Manual*. Washington DC, USA: American Association of State and Highway Transportation Officials, 2010.
- [4] X. Xu, A. Kouhpanejade, and Ž. Šarić, "Analysis on Influencing Factors Identification of Crash Rates Using Tobit Model with Endogenous Variable," *Promet - Traffic & Transportation*, vol. 25, no. 3, pp. 217–224, Jun. 2013, <https://doi.org/10.7307/ptt.v25i3.291>.
- [5] R. Elvik, "A synthesis of studies of access point density as a risk factor for road accidents," *Accident Analysis & Prevention*, vol. 107, pp. 1–10, Oct. 2017, <https://doi.org/10.1016/j.aap.2017.07.006>.
- [6] "Introduction: Intersection Safety Briefing Sheets," 2004, Available: https://www.academia.edu/5591388/Introduction_Intersection_Safety_Briefing_Sheets.
- [7] M. M. Abdul Manan and A. Várhelyi, "Motorcycle fatalities in Malaysia," *IATSS Research*, vol. 36, no. 1, pp. 30–39, Jul. 2012, <https://doi.org/10.1016/j.iatssr.2012.02.005>.
- [8] G. Perri and R. Vaiana, "Road Safety Management of Uncontrolled Access Points: Design Criteria and Insights into Risk Factors," *Applied Sciences*, vol. 12, no. 24, Jan. 2022, Art. no. 12661, <https://doi.org/10.3390/app122412661>.
- [9] M. Chakraborty and T. J. Gates, "Association between Driveway Land Use and Safety Performance on Rural Highways," *Transportation Research Record*, vol. 2675, no. 1, pp. 114–124, Jan. 2021, <https://doi.org/10.1177/0361198120965232>.
- [10] G. G. Schultz, J. S. Lewis, and T. Boschert, "Safety Impacts of Access Management Techniques in Utah," *Transportation Research Record*, vol. 1994, no. 1, pp. 35–42, Jan. 2007, <https://doi.org/10.3141/1994-05>.
- [11] K. C. Mouskos, W. Sun, S. I. Chien, A. Eisdorfer, and T. Qu, "Effect of Midblock Access Points on Traffic Accidents on State Highways in New Jersey," *Transportation Research Record*, vol. 1665, no. 1, pp. 75–83, Jan. 1999, <https://doi.org/10.3141/1665-11>.

- [12] G. G. Schultz, K. T. Braley, and T. Boschert, "Relationship between Access Management and Other Physical Roadway Characteristics and Safety," *Journal of Transportation Engineering*, vol. 136, no. 2, pp. 141–148, Feb. 2010, [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000085](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000085).
- [13] X. Yi, "Impact of access density on number of crashes at commercial roadways," M.S. thesis, Oregon State University, Corvallis, OR, USA, 2009.
- [14] J. Xu, J. H. Lambert, and C. J. Tucker, "Highway Access Safety Program Evaluation with Uncertain Parameters," *Journal of Transportation Engineering*, vol. 140, no. 2, Feb. 2014, Art. no. 04013010, [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000631](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000631).
- [15] A. Khavandi and S. O. Rezaeeyan, "Investigating the Effect of Access-point density on Traffic Crash on Rural Two-Lane Highways (Case study: Ahar- Tabriz Highway)," *Quarterly Journal of Transportation Engineering*, vol. 9, no. 4, pp. 619–630, Jun. 2018.
- [16] D. M. Priyantha Wedagama, R. N. Bird, and A. V. Metcalfe, "The influence of urban land-use on non-motorised transport casualties," *Accident Analysis & Prevention*, vol. 38, no. 6, pp. 1049–1057, Nov. 2006, <https://doi.org/10.1016/j.aap.2006.01.006>.
- [17] C. C. Minh, N. Huynh, M. Chowdhury, J. H. Ogle, W. A. Sarasua, and W. J. Davis, "Impact of Minimum Driveway Spacing Policies on Safety Performance: An Integrated Traffic Micro-Simulation and Automated Conflict Analysis," *International Journal of Transportation Science and Technology*, vol. 3, no. 3, pp. 249–264, Sep. 2014, <https://doi.org/10.1260/2046-0430.3.3.249>.
- [18] "Guidelines for Public Street and Driveway Connections," in *Mn/DOT Access Management Manual*, Minnesota Department of Transportation, 2008.
- [19] A. Ahmed, A. Sadullah, and A. Yahya, "Evaluating the contribution of physical parameters on the safety of unsignalized intersections," *Journal of Engineering Science and Technology*, vol. 10, no. 5, pp. 654–666, May 2015.
- [20] M. M. Abdul Manan, T. Jonsson, and A. Várhelyi, "Development of a safety performance function for motorcycle accident fatalities on Malaysian primary roads," *Safety Science*, vol. 60, pp. 13–20, Dec. 2013, <https://doi.org/10.1016/j.ssci.2013.06.005>.
- [21] S. Cafiso, A. Di Graziano, G. Di Silvestro, G. La Cava, and B. Persaud, "Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency and context variables," *Accident Analysis & Prevention*, vol. 42, no. 4, pp. 1072–1079, Jul. 2010, <https://doi.org/10.1016/j.aap.2009.12.015>.
- [22] B. Huang, Y. Zhang, L. Lu, and J. J. Lu, "A new access density definition and its correlation with crash rates by microscopic traffic simulation method," *Accident Analysis & Prevention*, vol. 64, pp. 111–122, Mar. 2014, <https://doi.org/10.1016/j.aap.2013.11.014>.
- [23] P. Prajapati and G. Tiwari, "Evaluating Safety of Urban Arterial Roads of Medium Sized Indian City," in *Proceedings of the Eastern Asia Society for Transportation Studies*, Taipei, Taiwan, Oct. 2013, vol. 9.
- [24] M. Ma, X. Yan, M. Abdel-Aty, H. Huang, and X. Wang, "Safety Analysis of Urban Arterials under Mixed-Traffic Patterns in Beijing," *Transportation Research Record*, vol. 2193, no. 1, pp. 105–115, Jan. 2010, <https://doi.org/10.3141/2193-13>.
- [25] R. Abdulla, B. Qader, and K. Sdiq, "Traffic Accident Traits and Driver Characteristics Implication on Road Accidents using Descriptive Analysis: A Cross Sectional Study in Sulaymaniyah, Iraq," *Engineering, Technology & Applied Science Research*, vol. 13, no. 2, pp. 10372–10376, Apr. 2023, <https://doi.org/10.48084/etasr.5669>.
- [26] M. Angin and S. I. A. Ali, "Analysis of Factors Affecting Road Traffic Accidents in North Cyprus," *Engineering, Technology & Applied Science Research*, vol. 11, no. 6, pp. 7938–7943, Dec. 2021, <https://doi.org/10.48084/etasr.4547>.
- [27] M. Touahmia, "Identification of Risk Factors Influencing Road Traffic Accidents," *Engineering, Technology & Applied Science Research*, vol. 8, no. 1, pp. 2417–2421, Feb. 2018, <https://doi.org/10.48084/etasr.1615>.

Real-Time Inspection of Fire Safety Equipment using Computer Vision and Deep Learning

Asmaa Alayed

College of Computing, Umm Al-Qura University, Saudi Arabia
asayed@uqu.edu.sa (corresponding author)

Rehab Alidrisi

College of Computing, Umm Al-Qura University, Saudi Arabia
s439016154@st.uqu.edu.sa

Ekram Feras

College of Computing, Umm Al-Qura University, Saudi Arabia
s441015984@st.uqu.edu.sa

Shahad Aboukazzana

College of Computing, Umm Al-Qura University, Saudi Arabia
s441003855@st.uqu.edu.sa

Alaa Alomayri

College of Computing, Umm Al-Qura University, Saudi Arabia
s441008331@st.uqu.edu.sa

Received: 15 December 2023 | Revised: 6 January 2024 and 27 January 2024 | Accepted: 5 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6753>

ABSTRACT

The number of accidental fires in buildings has been significantly increased in recent years in Saudi Arabia. Fire Safety Equipment (FSE) plays a crucial role in reducing fire risks. However, this equipment is prone to defects and requires periodic checks and maintenance. Fire safety inspectors are responsible for visual inspection of safety equipment and reporting defects. As the traditional approach of manually checking each piece of equipment can be time-consuming and inaccurate, this study aims to improve the inspection processes of safety equipment. Using computer vision and deep learning techniques, a detection model was trained to visually inspect fire extinguishers and identify defects. Fire extinguisher images were collected, annotated, and augmented to create a dataset of 7,633 images with 16,092 labeled instances. Then, experiments were carried out using YOLOv5, YOLOv7, YOLOv8, and RT-DETR. Pre-trained models were used for transfer learning. A comparative analysis was performed to evaluate these models in terms of accuracy, speed, and model size. The results of YOLOv5n, YOLOv7, YOLOv8n, YOLOv8m, and RT-DETR indicated satisfactory accuracy, ranging between 83.1% and 87.2%. YOLOv8n was chosen as the most suitable due to its fastest inference time of 2.7 ms, its highest mAP0.5 of 87.2%, and its compact model size, making it ideal for real-time mobile applications.

Keywords-Fire Safety Equipment (FSE); fire safety inspection; visual inspection; deep learning; computer vision; object detection; You-Only-Look-Once (YOLO)

I. INTRODUCTION

The risk of accidental fire in buildings is one of the most dangerous ones due to the potentially of incurring huge amounts of damage and costs. According to the Saudi Civil Defense statistics, economic losses have recently increased by 61% [1]. In Saudi Arabia, fires are common, with an average of 42 fires per day [1]. The most dangerous fires are those that

occur in crowded areas, as they can cause a high number of injuries and deaths, such as the Mena accident in 1997 during the Hajj season in Makkah, Saudi Arabia, which caused more than 340 deaths and 1,500 casualties. Thus, not only is the development and implementation of fire safety protocols in buildings required by the civil defense laws and regulations, but it is also crucial for the safety of all those individuals being in a building during a fire emergency [2].

In Saudi Arabia, most current fire protection measures are prescriptive and based on the requirements of the Saudi Building Code for Fire Protection [3]. Buildings are equipped with fire safety devices and procedures to minimize the risk of fire and its impact [4]. However, the mere existence of this equipment does not necessarily mean that the latter will be effective, as it is prone to failures and impairments for many reasons, such as human errors, mechanical failure, neglect, and the environment [5]. It is estimated that a third of Fire Safety Equipment (FSE) would not work properly in an emergency [6]. The Saudi government has made significant efforts to reduce the risk of fire, especially at events and during religious seasons, to protect the lives of people and minimize damage to buildings. One of these efforts is the implementation of fire safety inspections conducted by safety and preventive supervision teams in the Civil Defense. The inspection and maintenance of the FSE are designed to ensure that it is always in excellent operating condition [7]. FSE inspection is divided into several modes. Visual inspection is part of the FSE inspection process and aims to evaluate its condition (damaged or not) as well as the suitability of its configuration [7]. It is now mandatory to regularly inspect and document the presence, location, and working order of all FSEs over their lifespan according to regulations and modern safety standards [5]. The regulations related to the visual inspection of fire extinguishers are as follows:

- The fire extinguisher should be visible and free of obstructions.
- The pressure gauge should be pointed to the green zone.
- There should be no rust or damage to the extinguisher.
- Check the condition of the hose connected to the fire extinguisher.
- Ensure the presence of the safety pin.
- Ensure the existence of labels that show the expiration date of the extinguisher.

Currently, the visual approach to detect defects in safety equipment is the standard method to evaluate its quality [8]. However, due to inaccuracies in identifying defects or damage from photographic evidence, visual inspection solutions are labor-intensive, time-consuming, and error-prone [9]. Thus, conventional human visual inspection is a difficult-to-measure process with variable and subjective results [8]. Several Saudi government applications offer solutions for the inspection process, such as Madani [10], and Salamti [11], which allow the public to file safety violation reports so that preventive supervision teams can take action to avert the occurrence of accidents. However, there are limitations to these applications, including the possibility of reporting incorrect or inaccurate violations due to the public's inexperience in the inspection process. Furthermore, it is time-consuming for preventive supervision teams to personally validate each violation.

Deep learning is a popular technology used in different domains, including text processing, medical diagnosis, weather forecasting, and climate change analysis. With deep learning, computer vision has produced encouraging results, like in weed

detection [12], agriculture, satellite image analysis [13], and fire detection [14]. YOLO (You Only Look Once) is a deep learning algorithm that shows high speed and accuracy. Therefore, YOLO is widely used in UAV image detection [15-16] and real-time applications [17]. Several studies have explored the use of deep learning for FSE inspection. In [18], a custom Convolutional Neural Network (CNN) approach was proposed for object recognition in fire safety systems. In [19], the Single Shot Detector (SSD) was proposed to locate fire equipment in 3D building layouts, achieving 76% accuracy for extinguisher detection. In [20-21], YOLOv5 and YOLOv7 were used for FSE object detection. In [22], an advanced solution was proposed that could do more than detect the presence of FSE, such as fire extinguishers, by identifying its condition as defective or non-defective. In this study, MobileNet V2 SSDLite, FPN Resnet-50 SSD, and Inception Resnet v2 Faster R-CNN achieved an accuracy rate of 86%, but some inspection criteria, like gauge checking, were missing. Additionally, the lack of a dataset can lead to lower accuracy in fire extinguisher detection.

This study aims to solve the pressing problems of FSE inspection and defect detection automation and intelligence levels using deep-learning computer vision. This study trained defect inspection models to inspect various specific fire extinguisher conditions according to [3]. As shown in Figure 1, a mobile application will use the model to identify defective equipment, document it, and send it to the relevant authority to quickly determine the appropriate action procedure. This study intended to:

- Create and annotate a comprehensive data set for both defective and non-defective fire extinguishers.
- Evaluate the effectiveness of various versions of the YOLO algorithm alongside transformers, and compare precision, speed, and model size.

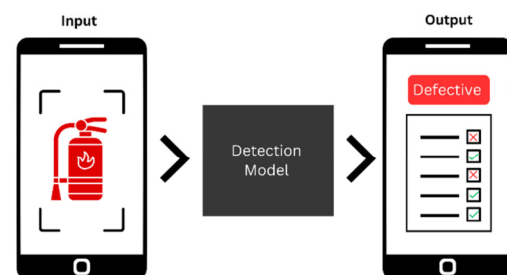


Fig. 1. Overview of a mobile application that uses the proposed model to inspect an extinguisher and run through the checklist automatically.

II. DATA COLLECTION AND RESEARCH METHOD

A. Dataset Collection

High-quality image datasets are essential for training deep learning models [23]. In the context of the fire extinguisher inspection, the dataset must encompass both non-defective and defective fire extinguishers, covering all potential reasons for defects. After conducting a thorough search, no existing dataset was found to meet these specific requirements. In addition, the

available extinguisher datasets were relatively small and included extinguisher models that are not commonly used in Saudi Arabia. As a result, a dataset of fire extinguishers was built, including different types of defective and non-defective fire extinguishers. This study focused on specific types of extinguishers: powder, water, and foam. These types are similar in shape and size and are most commonly used in buildings. Images were gathered from various facilities, including the university, schools, shops, restaurants, residential buildings, and hospitals. To collect these data, a technique of capturing videos of fire extinguishers and then extracting images from these videos was adopted. This approach allowed us to generate a high volume of data while varying the angles and distances. The video speed was slow enough to avoid blurry images, but fast enough to prevent image duplication. This adjustment significantly affected the quality of the dataset. The best way to capture videos was reached after several rounds of attempts. On average, the collected video time was 6.35 seconds and the images were extracted at 5-15 fps. Finding an adequate number of fire extinguishers was not difficult since they are available in different facilities and public places. However, it was a challenge to locate fire extinguishers that were truly rusted or had defective gauges. This was due to the rigorous safety laws and regulations in Saudi Arabia that require the replacement of defective fire extinguishers with fully functional ones. Therefore, several fire safety centers were contacted to help in finding a wide variety of defective extinguishers to add such images to the dataset. As a result, 3,592 images were collected.

The next step was the annotation process that involved manual labeling of the classes and localization of the objects in the collected images. The class labels were selected based on five key criteria informed by the regulations that govern the inspection and appearance of fire extinguishers. First, the presence of a hose was only noted when the hose existed. Non-existent hoses cannot be detected by the model. The same is true for fire safety pins and expiration date labels. For the defective body, only the rusty parts were annotated, allowing the model to recognize the rust characteristics instead of attempting to identify the overall appearance of the rusted extinguishers. Additionally, the gauge was annotated as either good or bad based on its indicator position. If the indicator fell within the green area, the gauge was labeled as good, whereas a position in the red area indicated a bad gauge. Finally, the dataset consisted of six classes: "hose_exist", "pin_exist", "rust", "gauge_good", "gauge_bad", and "expire_date". Figure 2 displays a sample of the collected images. During the annotation process, it was found that individual images often contained multiple objects of interest. An instance segmentation approach was used to accurately delineate each separate object by surrounding them with polygons [24]. Roboflow [25], was employed to segment objects. Figure 3 presents the outcome of this segmentation process. The latter proved invaluable for object detection tasks and also expanded the utility of the dataset beyond mere detection. The images and videos collected were taken on various mobile devices, resulting in nonuniform sizes. Standardizing the input size and maintaining consistent resolution were crucial for the dataset. To achieve this, during the preprocessing step, all images were

resized to a uniform resolution of 640×640 pixels. It is important to note that this size was chosen carefully to strike a balance: neither too small to accommodate small objects, nor too large to slow down the training process. The rectangular images were filled with black padding, ensuring that the extinguishers were not distorted or sheared.



Fig. 2. Dataset examples of defective and non-defective fire extinguishers: (a) Non-defective extinguisher meeting all conditions; (b) Defective extinguisher lacking both safety pin and hose, and the gauge indicator being in the red area; (c) Defective extinguisher showing rust and absence of the hose.

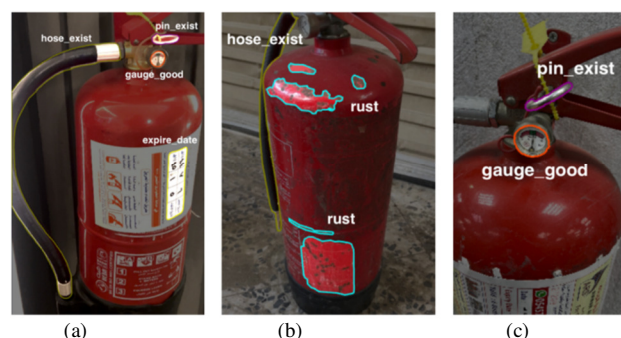


Fig. 3. Examples of annotated images: (a) All parts of the extinguisher; (b) Rust detection; (c) Closed image to small objects.

Data augmentation techniques were applied to diversify, generalize, and increase the dataset. These techniques included rotations of up to 15°. The rotation aimed to expose the model to objects from various angles. Other techniques were brightness adjustments of 25% and saturation modifications of 25%. These can simulate different lighting conditions, thereby enhancing the model's ability to classify objects accurately. Eventually, the dataset consisted of 7,663 fire extinguisher images with 16,092 annotations, with an average of 2.1 objects per image. Figure 4 illustrates the number of instances per class. The dataset was sorted and randomly split into training sets, validation sets, and test sets, according to an 8:1:1 proportion. Empirical analysis has proven that 80% of training yields the highest results [26]. Inspecting an extinguisher using a single photo can be challenging since certain parts may not always be fully visible, such as when the safety pin is positioned at the back or when there is a tag covering the gauge, as shown in Figure 5. Employing a real-time solution can be advantageous as it allows users to receive instant feedback when the details are not clearly visible, enhancing the effectiveness of the inspection process.

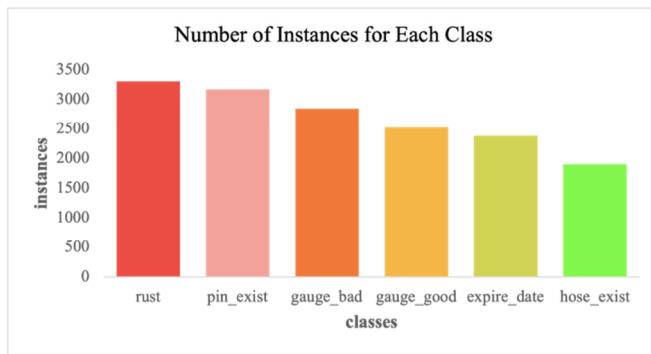


Fig. 4. Number of instances for each class.

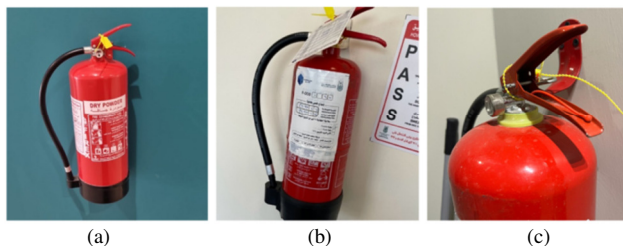


Fig. 5. Some parts of extinguishers can be hidden: (a) The gauge is not clear because it is too far away; (b) The gauge is covered by the tag; (c) The angle of the shot does not clearly show the safety pin.

B. YOLO Algorithms

Computer vision projects using deep learning approaches have produced remarkable results [27]. One such powerful example is the YOLO algorithm [28-29], which has become a cutting-edge algorithm in object detection. YOLO works by layering a grid onto an image and passing it through a neural network once only. This makes YOLO the fastest available object detection algorithm and an excellent choice for real-time applications [30]. The YOLO algorithm has been continuously improved and modified over the years, resulting in several versions. This study selected the most recent ones, which are YOLOv5, YOLOv7, and YOLOv8. These versions provide lightweight models, offering higher accuracy and faster processing times [31].

1) YOLOv5

YOLOv5n is the lightest, smallest, and fastest model in terms of detection speed [31-32]. This study considered model efficiency, accuracy, and size. Based on the YOLOv5n architecture, an improved design was developed to detect defective fire extinguishers. The YOLOv5n architecture includes three components: the backbone network, the neck network, and the detection network. In the backbone, features are extracted from input images, which are essential for object detection, while in the neck, three different scales of feature maps are created, which are used by the detection head [33].

2) YOLOv7

YOLOv7 is a real-time object detector currently driving a profound transformation in computer vision. YOLOv7 proposed the Efficient Layer Aggregation Networks (ELAN) architecture, which can improve the self-learning capacity of the network without destroying the original gradient path [34].

3) YOLOv8

One of the most significant changes in YOLOv8 is the use of free anchor boxes. Anchor boxes are typically employed in object detection models to help them predict the location and size of objects in an image. This feature reduces detection time by speeding up Non-Maximum Suppression (NMS), a post-processing step that sifts through candidate detections after inference [35]. Another key improvement in YOLOv8 is that it utilizes new loss functions for bounding box loss and classification loss. These functions improve its performance, particularly when dealing with smaller objects [32].

C. RT-DETR Model

Existing real-time detectors typically use CNN-based architectures to achieve a reasonable trade-off between detection speed and accuracy. Transformer-based detectors (DE-TRs) have recently demonstrated exceptional performance. However, their high computational cost results in slow real-time object detection. Real-Time Detection Transformer (RT-DETR) was developed [36], which is an end-to-end object detector that offers real-time performance and high accuracy. It engages a hybrid encoder that processes multiscale features by decoupling intra-scale interaction and cross-scale fusion. This one-of-a-kind vision transformer-based design reduces computational cost. RT-DETR is highly adaptable, allowing for flexible inference speed adjustments without retraining. This model outperforms current real-time detectors in accuracy and speed, requires no postprocessing, ensures stable inference speed, and fully utilizes an end-to-end detection pipeline advantage [36].

D. Training Methodology

This study aimed to perform inspection tasks with the highest performance while maintaining a minimal size suitable for use in a mobile application. YOLO offers different sizes (nano, small, medium, large, and xlarge) that provide various trade-offs between speed and accuracy. Several experiments were carried out to evaluate the previously mentioned YOLO versions for nano size, YOLOv5n, YOLOv7-tiny, and YOLOv8n. Additionally, YOLOv8m was investigated for improved accuracy. The evaluation was extended to include RT-DETR. This allows us to comprehensively assess and compare the efficacy of various models in the search for superior inspection capabilities.

As the dataset was not big enough, transfer learning was implemented to fine-tune pre-trained model weights on new data. Therefore, all the previously mentioned models were pre-trained on the Common Objects in Context (COCO) dataset, which is one of the largest datasets made by Microsoft [37]. These models were trained using a set of hyperparameters, including Adam optimizer, a learning rate of 0.001, and batch size of 64 and 50 epochs with a training dataset of 6130 images, a testing set of 766 images, and a validation set of 766 images. Throughout the training process, all models were configured consistently, except for RT-DETR's batch size, which due to its large size could only handle a batch size of 16. Table I portrays the hyperparameters.

TABLE I. THE HYPERPARAMETERS SET DURING MODEL TRAINING

Hyperparameter	Value for YOLO	Value for RT-DETR
Input image size	640	640
Epochs	50	50
Batch size	64	16
Optimizer	ADAM	ADAM
Initial learning rate	0.001	0.001
Final learning rate	0.00001	0.00001
Momentum	0.937	0.937
Weight decay	0.0005	0.0005

E. Training Environment

Training a model requires higher computational resources such as GPUs. Google Colab, which is a cloud-based platform to execute Python code, was used to train the models. The resources that Google Colab offers for free include a GPU T4 graphics card with 12GB of VRAM. The paid version was utilized for some slow experiments. Colab Pro offers more options for powerful GPUs, including GPU V100 and GPU A100, which are faster than GPU T4. Using the upgraded version considerably speeded up the training process.

F. Evaluation Metrics

Several metrics, such as precision (P), recall (R), and mAP , are applied to assess the performance of a model and its ability on the detection task. These metrics provide a comprehensive knowledge of how accurate and reliable the model is at detecting defective fire extinguishers. These metrics are calculated using a confusion matrix that consists of four parts:

1. True Positives (TP): The number of instances that belong to the positive class and are correctly classified as positive by the model.
2. False Positives (FP): The number of instances that belong to the negative class but are incorrectly classified as positive by the model.
3. True Negatives (TN): The number of instances that belong to the negative class and are correctly classified as negative by the model.
4. False Negatives (FN): The number of instances that belong to the positive class but are incorrectly classified as negative by the model.

From these values:

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

Another important term for object detection is the Intersection over Union (IoU) [24]. This metric measures the accuracy of each bounding box by taking the ratio of overlapping areas between the actual (B_{gt}) and predicted (B_{pr}) bounding boxes to the area of their union. Equation 3 shows its formula and Figure 6 illustrates it.

$$IoU = \frac{B_{pr} \cap B_{gt}}{B_{pr} \cup B_{gt}} \quad (3)$$

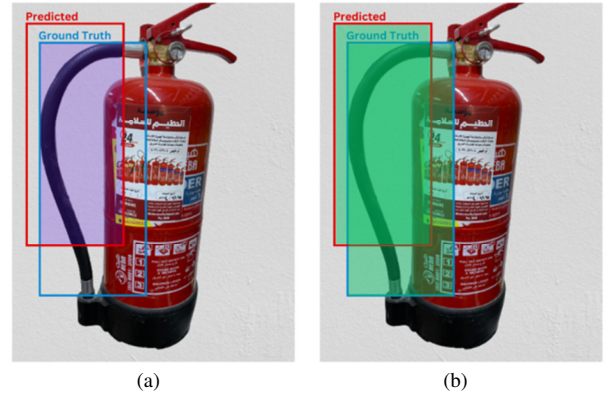


Fig. 6. IoU is the ratio of the intersection area over the union area: (a) Intersection area; (b) Union area.

The Mean Average Precision at 0.5 ($mAP_{0.5}$) is a comprehensive metric that consolidates these various aspects. First, it involves the Average Precision (AP) which is the area under the precision-recall curve [24]. The individual-class AP values are then averaged to derive mAP . $mAP_{0.5}$ indicates that this calculation uses an IoU threshold of 0.5. These properties make $mAP_{0.5}$ a suitable metric for most detection applications.

III. RESULTS AND DISCUSSION

After conducting a series of experiments, a comparative analysis was performed to evaluate the models and identify their optimal performance based on accuracy, speed, and model size.

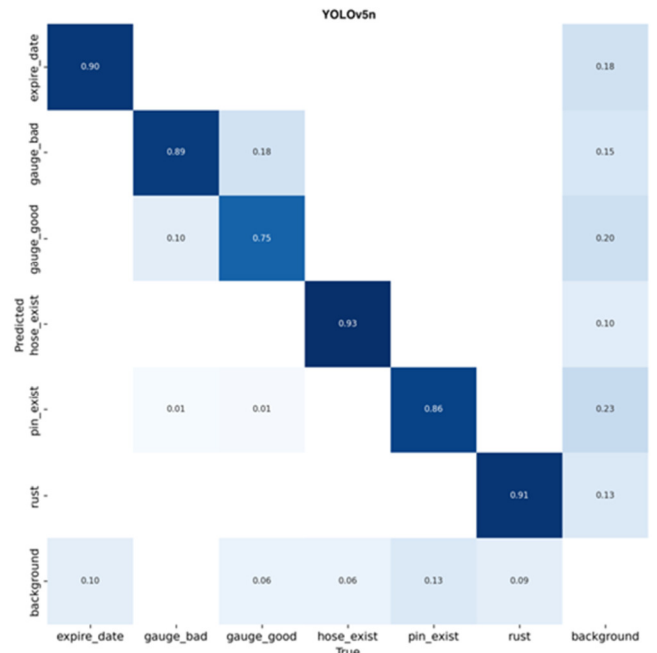


Fig. 7. Confusion matrix for YOLOv5n.

A. Confusion Matrix

The results presented in Figures 7-11 show the confusion matrices for each model at a confidence level of 0.25. In the confusion matrix, the diagonal line shows instances correctly classified by the model, offering a visual representation of its accuracy. Gauge classes are difficult to detect, leading to the lowest accuracy levels. In addition, there was some degree of misclassification between gauge good and bad.

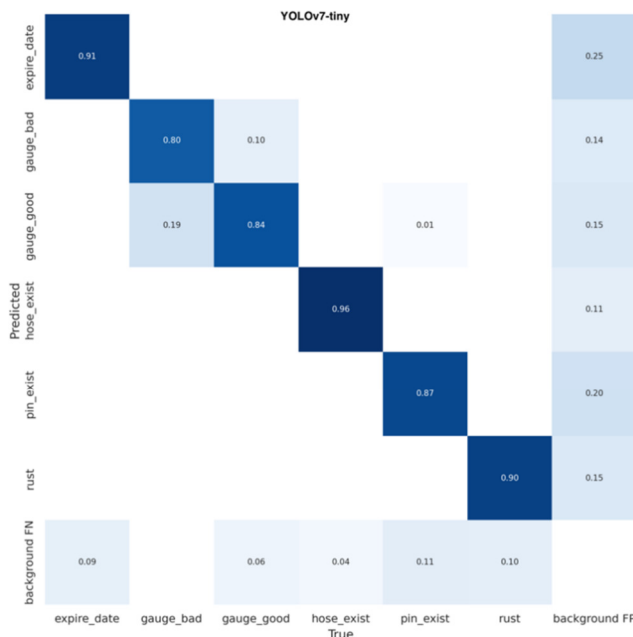


Fig. 8. Confusion matrix for YOLOv7tiny.

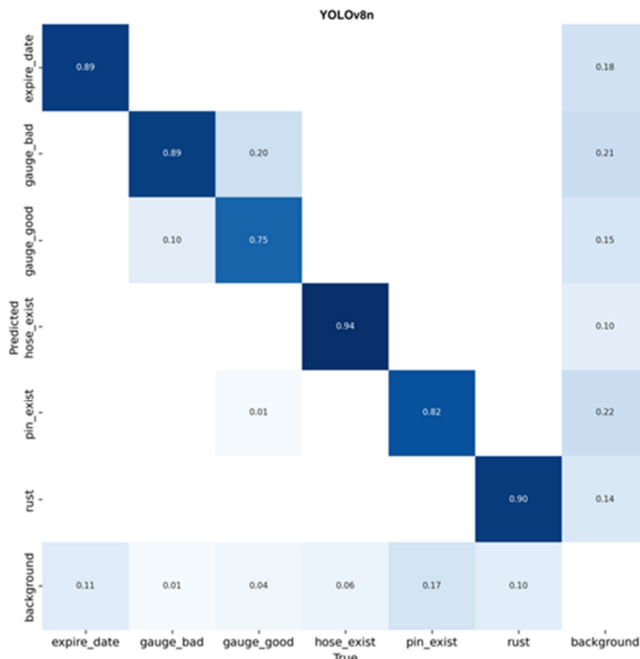


Fig. 9. Confusion matrix for YOLOv8n.

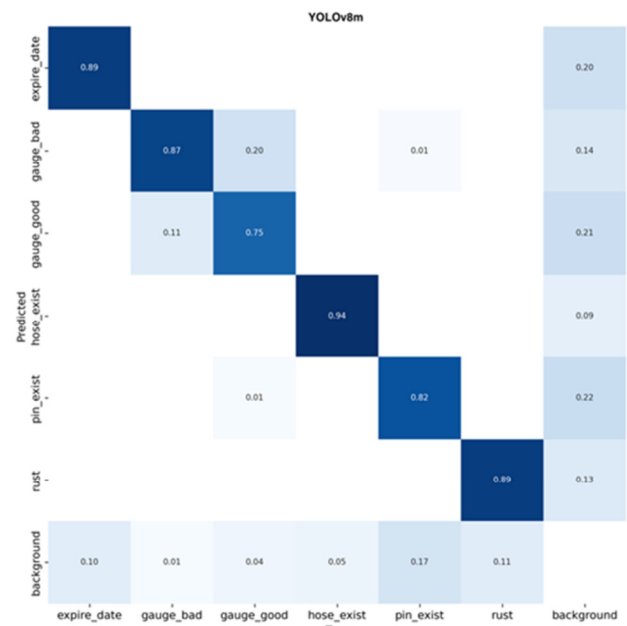


Fig. 10. Confusion matrix for YOLOv8m.

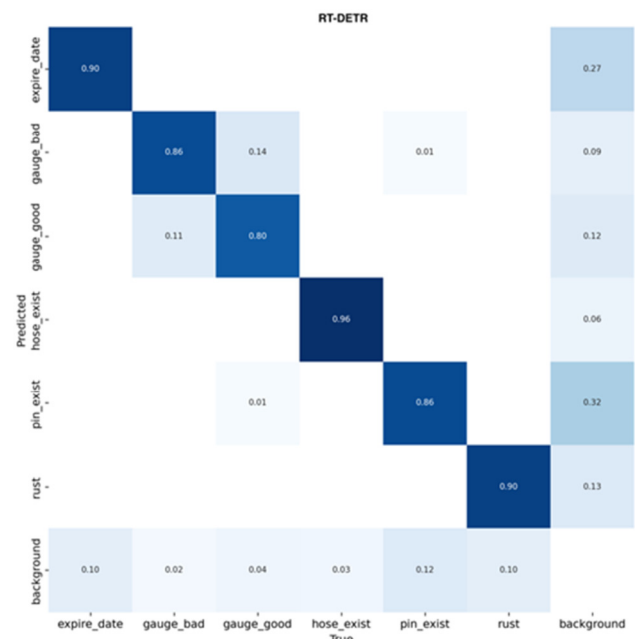


Fig. 11. Confusion matrix for RT-DETR.

B. Mean Average Precision and Model Size

Table II presents the metrics for each model, along with their respective sizes. Compared to other models, the YOLOv8n model excelled in terms of mAP , achieving 87.2%. On the contrary, RT-DETR achieved the lowest mAP of 83.1%. When considering model size, YOLOv5n stood out by attaining the most modest model size among other models, while RT-DETR had the largest model size, which is unsuitable for a mobile application. Figure 12 illustrates the $mAP_{0.5}$ of each model over 50 epochs.

TABLE II. MODEL EVALUATION METRICS COMPARISON.

Model	Model Size (MB)	P (%)	R (%)	mAP0.5 (%)
YOLOv5n	3.9	82.1	82.9	86.1
YOLOv7- tiny	12.3	84.5	82.1	85.8
YOLOv8n	6.3	82.5	83.7	87.2
YOLOv8m	52.0	83.8	83.5	86.6
RT-DETR	66.2	82.7	83.8	83.1

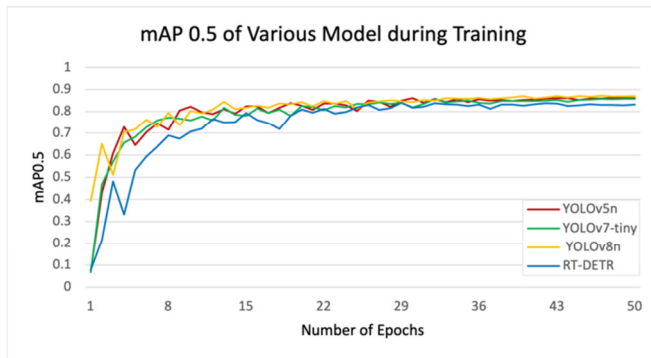


Fig. 12. mAP0.5 of each model over 50 epochs.

C. Speed

As the proposed solution operates in real-time, temporal efficiency is of paramount importance. Table III provides a breakdown of time across two phases, inference (time taken for passing the image through the neural network), and post-processing required for the NMS algorithm. YOLOv8n exhibited the fastest overall time at 2.7ms, whereas YOLOv5n had the slowest performance. Notably, RT-DETR stands out for its commendable post-processing time, as its key feature. However, the inference time hinders the achievement of an expedited total detection time.

TABLE III. COMPARISON OF DETECTION TIME OF EACH MODEL (USING T4 GPU)

Model	Inference (ms)	Postprocessing (ms)	Total Time (ms)
YOLOv5n	43.1	1.5	44.6
YOLOv7- tiny	4.9	1.5	6.4
YOLOv8n	2	0.7	2.7
YOLOv8m	23.2	2	25.2
RT-DETR	16.5	0.3	16.8

D. Discussion

The results demonstrate that YOLOv8n fulfills the requirements in terms of overall performance, given the high importance of attaining real-time solutions. Although YOLOv8n did not achieve the smallest model size, it successfully met the requirement of accurately detecting defective equipment at high speed. Table IV displays the precision, recall, and mAP for each class. The "hose_exist" class reveals the highest results because it is easy to see from any angle or distance. However, "gauge_bad" exhibits lower precision due to more mistakes. This happens when the gauge is far away or in dark light, making it hard to detect if it is in the green or red area. So, the model sometimes predicts both good and bad situations, leading to less accurate results. Figure 13 depicts a sample of testing results.

TABLE IV. YOLOV8N RESULTS FOR EACH CLASS

Class	P (%)	R (%)	mAP0.5 (%)
all	82.5	83.7	87.2
expire_date	82.6	85.0	81.4
gauge_bad	67.9	89.0	88.9
gauge_good	81.9	69.6	84.3
hose_exist	93.6	93.1	97.5
pin_exist	78.9	78.3	80.7
rust	89.9	86.9	90.3

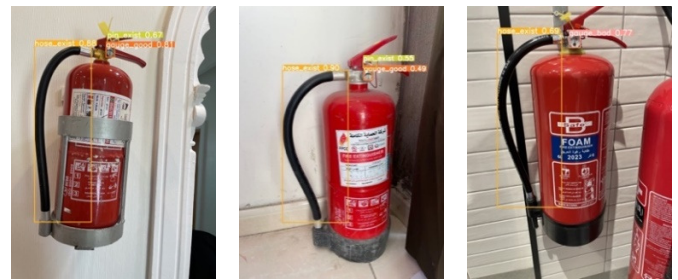


Fig. 13. Sample of testing results of YOLOv8n.

In contrast, the poor efficiency and large model size of RT-DETR make it an unsuitable choice for the detection of fire equipment in complex environments. Additionally, the slower performance speeds of both YOLOv5n and RT-DETR further diminish their suitability for such applications. This study demonstrated better results than [22], which is attributed to the use of the latest version of the YOLO algorithm (YOLOv8n). Furthermore, the models in this study benefited significantly from the rich dataset, which contributed to their improved performance.

The application of YOLOv8n was put to a practical test by expert inspectors, who offered insightful comments. Their first-hand experience confirmed the efficiency, effectiveness, and usefulness of the proposed model, adding substantial value to inspections with technologies previously unavailable. The experts also highlighted the future potential of the model, emphasizing how well it may integrate with other technologies to increase effectiveness. Moreover, they pointed out its potential to improve accuracy, speed up the inspection process, and reduce errors.

Potential errors originating from the model can be addressed by providing administrators with access to fire extinguisher images to analyze them and enhance performance. This contributes to the system's reliability and competence and increases the safety levels of individuals and properties. Consequently, this would increase the trust of society in the safety inspection mechanisms used by the safety inspectors.

IV. CONCLUSION AND FUTURE WORK

This study addressed a critical issue in Saudi Arabia, the challenges faced by safety equipment inspectors. Taking advantage of computer vision and deep learning techniques, this research presented a system dedicated to inspecting fire extinguishers and identifying defects in real-time. To achieve this, a rich dataset was built, comprising 7,663 images with 16,092 instances of fire extinguisher defects, each labeled with classes like gauge bad, expiration date, gauge good, hose exist,

pin exist, and rust. Subsequently, experiments were carried out, evaluating RT-DETR and different versions of the YOLO algorithm, including YOLOv5n, YOLOv7-tiny, and YOLOv8n. Among these models, the YOLOv8n algorithm emerged as the best performer, achieving a $mAP_{0.5}$ score of 87.2%. Therefore, this model was selected as the most suitable option for the proposed system due to its efficient inference time, making it ideal for real-time mobile applications. Although YOLOv5 outperformed YOLOv8n in terms of its compact model size, it is less favorable due to its long inference time. Additionally, the performance of YOLOv7-tiny did not provide a significant advantage over YOLOv8n. In contrast, RT-DETR exhibited significantly lower detection accuracy, larger model size, and longer inference times, further indicating its inability to identify fire equipment defects, especially in complex environments.

Despite its success, the limitations of the proposed system are acknowledged. The model was trained exclusively to identify specific types of extinguishers, namely, powder, foam, and water extinguishers, all of a particular size. Regarding the detection of expiration data, it is important to note that the model's current functionality is limited to identifying the presence of an expiration date on the extinguisher. However, it cannot determine whether the extinguisher has expired due to limitations arising from Python's library support for extracting expiration dates from Arabic language text. Future improvements would address this constraint. Regarding research directions, expanding training datasets to include other sizes and types of extinguishers, such as carbon dioxide gas extinguishers, water extinguishers, and others, would widen the scope of the model. The proposed approach is scalable, with the ability to evolve and adapt to detect new classes of defective safety equipment, including but not limited to electrical extensions and smoke detectors. These improvements would further enhance the model's capacity to identify safety equipment violations across a wide spectrum. Moreover, this system can be seamlessly integrated with a mobile application, thereby expanding its usability and effectiveness. Ultimately, this technological progress can make a substantial impact, propelling the improvement of safety inspection practices, not only in Saudi Arabia but also on a global scale.

REFERENCES

- [1] "Annual Statistical Report," Civil Defense Directorate, Saudi Arabia, 2020.
- [2] "Why is Fire Safety Important? | Alsco." <https://alsco.com/resources/why-is-fire-safety-important/>.
- [3] "Saudi Fire Protection Code Fire Protection Requirements," Saudi Building Code National Committee, SBC 801, 2018.
- [4] V. Kodur, P. Kumar, and M. M. Rafi, "Fire hazard in buildings: review, assessment and strategies for improving fire safety," *PSU Research Review*, vol. 4, no. 1, pp. 1–23, Jan. 2019, <https://doi.org/10.1108/PRR-12-2018-0033>.
- [5] D. E. Della-Giustina, *Fire Safety Management Handbook*, 3rd ed. Boca Ration, FL, USA: CRC Press, 2014.
- [6] D. Dieken, "Inspection, testing and maintenance of fire protection systems at industrial plants," *Process Safety Progress*, vol. 18, no. 3, pp. 151–155, 1999, <https://doi.org/10.1002/prs.680180306>.
- [7] Y. J. Chen, Y. S. Lai, and Y. H. Lin, "BIM-based augmented reality inspection and maintenance of fire safety equipment," *Automation in Construction*, vol. 110, Feb. 2020, Art. no. 103041, <https://doi.org/10.1016/j.autcon.2019.103041>.
- [8] Y. C. How, A. F. A. Nasir, K. F. Muhammad, A. P. P. A. Majeed, M. A. M. Razman, and M. A. Zakaria, "Glove Defect Detection Via YOLO V5," *MEKATRONIKA*, vol. 3, no. 2, pp. 25–30, 2021, <https://doi.org/10.15282/mekatronika.v3i2.7342>.
- [9] S. H. Hsu, H. T. Hung, Y. Q. Lin, and C. M. Chang, "Defect inspection of indoor components in buildings using deep learning object detection and augmented reality," *Earthquake Engineering and Engineering Vibration*, vol. 22, no. 1, pp. 41–54, Jan. 2023, <https://doi.org/10.1007/s11803-023-2152-5>.
- [10] "Madani Application," *AppStore*. <https://apps.apple.com/sa/app/%D9%85%D8%AF%D9%86%D9%8A/id1596908770>.
- [11] "Salamti Application - Saudi Civil Defense," 2017. <https://my.998.gov.sa/app/salamati>.
- [12] M. Anul Haq, "CNN Based Automated Weed Detection System Using UAV Imagery," *Computer Systems Science and Engineering*, vol. 42, no. 2, pp. 837–849, 2022, <https://doi.org/10.32604/csse.2022.023016>.
- [13] M. Anul Haq, "Planetscope Nanosatellites Image Classification Using Machine Learning," *Computer Systems Science and Engineering*, vol. 42, no. 3, pp. 1031–1046, 2022, <https://doi.org/10.32604/csse.2022.023221>.
- [14] F. M. Talaat and H. ZainEldin, "An improved fire detection approach based on YOLO-v8 for smart cities," *Neural Computing and Applications*, vol. 35, no. 28, pp. 20939–20954, Oct. 2023, <https://doi.org/10.1007/s00521-023-08809-1>.
- [15] A. Jawaharlalnehru *et al.*, "Target Object Detection from Unmanned Aerial Vehicle (UAV) Images Based on Improved YOLO Algorithm," *Electronics*, vol. 11, no. 15, Jan. 2022, Art. no. 2343, <https://doi.org/10.3390/electronics11152343>.
- [16] M. A. Haq, G. Rahaman, P. Baral, and A. Ghosh, "Deep Learning Based Supervised Image Classification Using UAV Images for Forest Areas Classification," *Journal of the Indian Society of Remote Sensing*, vol. 49, no. 3, pp. 601–606, Mar. 2021, <https://doi.org/10.1007/s12524-020-01231-3>.
- [17] B. Xiao, M. Nguyen, and W. Q. Yan, "Fruit ripeness identification using YOLOv8 model," *Multimedia Tools and Applications*, Aug. 2023, <https://doi.org/10.1007/s11042-023-16570-9>.
- [18] A. Corneli, B. Naticchia, M. Vaccarini, F. Bosché, and A. Carbonari, "Training of YOLO Neural Network for the Detection of Fire Emergency Assets," *ISARC Proceedings*, pp. 836–843, Oct. 2020.
- [19] R. Kostoeva, R. Upadhyay, Y. Sagar, and A. Zakhor, "Indoor 3D Interactive Asset Detection Using a Smartphone," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2-W13, pp. 811–817, Jun. 2019, <https://doi.org/10.5194/isprs-archives-XLII-2-W13-811-2019>.
- [20] H. Bayer and A. Aziz, "Object Detection of Fire Safety Equipment in Images and Videos Using YOLOv5 Neural Network," in *Forum Bauinformatik*, Munich, Germany, 2022, pp. 62–69.
- [21] J. H. Heinbach and A. Aziz, "Visual partial inspection of fire safety equipment using machine learning," presented at the 34th Forum Bauinformatik, Bochum, Germany, Sep. 2023, <https://doi.org/10.13154/294-10096>.
- [22] B. G. Ferreira, B. G. Lima, and T. F. Vieira, "Visual Inspection of Collective Protection Equipment Conditions with Mobile Deep Learning Models," presented at the 1st International Conference on Deep Learning Theory and Applications, Feb. 2024, pp. 76–83, <https://doi.org/10.5220/0009834600760083>.
- [23] H. Bichri, A. Chergui, and M. Hain, "Image Classification with Transfer Learning Using a Custom Dataset: Comparative Study," *Procedia Computer Science*, vol. 220, pp. 48–54, Jan. 2023, <https://doi.org/10.1016/j.procs.2023.03.009>.
- [24] R. Szeliski, *Computer Vision: Algorithms and Applications*. Cham, Switzerland: Springer International Publishing, 2022.
- [25] "Roboflow: Give your software the power to see objects in images and video." <https://roboflow.com/>.

- [26] A. Gholamy, V. Kreinovich, and O. Kosheleva, "Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation," The University of Texas at El Paso, UTEP-CS-18-09, Feb. 2018. [Online]. Available: https://scholarworks.utep.edu/cs_techrep/1209.
- [27] L. Petricca, T. Moss, G. Figueroa, and S. Broen, "Corrosion Detection Using A.I: A Comparison of Standard Computer Vision Techniques and Deep Learning Model," in *CS & IT Conference Proceedings*, May 2016, vol. 6.
- [28] R. Rajamohanam and B. C. Latha, "An Optimized YOLO v5 Model for Tomato Leaf Disease Classification with Field Dataset," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12033–12038, Dec. 2023, <https://doi.org/10.48084/etasr.6377>.
- [29] M. Salemdaeb and S. Erturk, "Multi-national and Multi-language License Plate Detection using Convolutional Neural Networks," *Engineering, Technology & Applied Science Research*, vol. 10, no. 4, pp. 5979–5985, Aug. 2020, <https://doi.org/10.48084/etasr.3573>.
- [30] J. Du, "Understanding of Object Detection Based on CNN Family and YOLO," *Journal of Physics: Conference Series*, vol. 1004, no. 1, Dec. 2018, Art. no. 012029, <https://doi.org/10.1088/1742-6596/1004/1/012029>.
- [31] J. Terven and D. Cordova-Esparza, "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680–1716, Nov. 2023, <https://doi.org/10.3390/make5040083>.
- [32] T. Saidani, "Deep Learning Approach: YOLOv5-based Custom Object Detection," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12158–12163, Dec. 2023, <https://doi.org/10.48084/etasr.6397>.
- [33] H. Gong *et al.*, "Swin-Transformer-Enabled YOLOv5 with Attention Mechanism for Small Object Detection on Satellite Images," *Remote Sensing*, vol. 14, no. 12, Jan. 2022, Art. no. 2861, <https://doi.org/10.3390/rs14122861>.
- [34] C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 7464–7475.
- [35] J. Solawetz, F. JAN 11, and 2023 12 Min Read, "What is YOLOv8? The Ultimate Guide.," *Roboflow Blog*, Jan. 11, 2023. <https://blog.roboflow.com/whats-new-in-yolov8/>.
- [36] W. Lv *et al.*, "DETRs Beat YOLOs on Real-time Object Detection." *arXiv*, Jul. 06, 2023, <https://doi.org/10.48550/arXiv.2304.08069>.
- [37] T. Y. Lin *et al.*, "Microsoft COCO: Common Objects in Context," in *Computer Vision – ECCV 2014*, Zurich, Switzerland, 2014, pp. 740–755, https://doi.org/10.1007/978-3-319-10602-1_48.

Improving Electric Vehicle Autonomy in the Smart City Concept

Ahmed Saad Eddine Souissi

Department of Industrial Engineering, College of Engineering, Northern Border University, Saudi Arabia
ahmed.souissi@nbu.edu.sa

Habib Kraiem

Department of Electrical Engineering, College of Engineering, Northern Border University, Saudi Arabia
| Processes, Energy, Environment and Electrical Systems (Code: LR18ES34), National Engineering
School of Gabes, University of Gabes, Tunisia
alhabeeb.kareem@nbu.edu.sa

Aymen Flah

Processes, Energy, Environment, and Electrical Systems (code: LR18ES34), National Engineering
School of Gabes, University of Gabes, Tunisia | MEU Research Unit, Middle East University, Amman,
Jordan | College of Engineering, University of Business and Technology (UBT), Jeddah 21448, Saudi
Arabia | Private Higher School of Applied Sciences and Technologies of Gabes, University of Gabes,
Tunisia | Applied Science Research Center, Applied Science Private University, Amman, 11931, Jordan
flahaymening@yahoo.fr (corresponding author)

Al Walid Amjad Almadani

Department of Industrial Engineering, College of Engineering, Northern Border University, Saudi Arabia
amjad.madani@yahoo.fr

Received: 21 January 2024 | Revised: 30 January 2024 | Accepted: 6 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6941>

ABSTRACT

Organizing automobiles in a city is challenging due to the sensitive data that need to be disclosed. Information that can be utilized to identify a car and provide some useful characteristics about it is among the large amount of data that can be collected from an automobile. This operation will be easier if the vehicles are placed on a specific platform based on the smart city concept. Even if sensors and cameras are installed around the roads and the city, having the vehicle information will be more useful. The current study tries to demonstrate how it is feasible to improve vehicle autonomy by initially enhancing the vehicle's energetic performance, based on the smart city idea. Intelligent control topology serves as the foundation for the exposed energy management protocol. The suggested concept is created and the associated results are displayed using the Matlab Simulink platform.

Keywords-*electric vehicle; power management; communication; optimization; neural networks; drive cycle mode; smart city*

I. INTRODUCTION

Electric Vehicles (EVs) are rising to prominence as a vital mode of mobility. Electrical transportation equipment includes a wide range of types and categories, such as the double-wheel electric scooter, the four-wheel electric vehicle for personal transportation, and electric buses and rails [1, 2]. A specification that characterizes all these electrified transportation tools is the energy storage element, which basically is an electric storage battery. The latter is the most important factor in the establishment of vehicle autonomy and

of the possible distance that can be made when using any EV [3, 4]. On the other hand, this type of battery entails certain risks during transportation, such as the risk of fire or that of an explosion in case high temperatures are reached. All these weaknesses deteriorate the EV usage as a transportation tool. Therefore, researchers strive to avoid the aforementioned problems by providing more solutions to them and employing technologies to make any EV's use secure and improve its battery autonomy. In fact, these two objectives were treated in more than one manner and methods. Considering the vehicle autonomy issue, some works presented several solutions for

enhancing the former. Some strategies concentrated on minimizing the aerodynamic effect on the vehicle body to decrease the air reaction at high speeds. Some researchers tried to minimize vehicle weight, so that it could be possible to save a greater amount of power for the traction phase [5]. Engineers have been also looking for ways to construct traction motors with increased power, which consume less energy with higher rentability. For this purpose, specialized AC machines were proved to be effective transportation tools [6, 7]. Different forms of power management have appeared in the literature. Some researchers contributed to the electrical traction way. Also, certain modified vector control strategies, or direct torque control topologies, engaged to get a better response from the traction machine and to decrease power loss in the transition phases, have been presented [8, 9]. Other solutions to manage the recharge and discharge phases in the Hybrid Electric Vehicle (HEV) models were provided in [10]. Those techniques, based on intelligent control, have demonstrated satisfactory energy consumption and greater autonomy [11-13]. Numerous industries have been highly benefited from smart cities in that they have managed to ease traffic congestion and shorten public transit wait times in the transportation sector. Centralized video networks improve security by making it easier to identify and monitor suspects. With more accurate estimates of monthly energy requirements and more effective power management, the energy sector is benefited as well. Multiple studies have highlighted that the accident ratio has been decreased as a result of the use of smart city technologies. Specifically, a research carried out in Barcelona, Spain, discovered that the implementation of smart city infrastructure, such as cameras and sensors, resulted in a 20% decrease in traffic accidents. In addition, a study in Singapore revealed that the employment of autonomous vehicles and smart traffic management systems led to a 20% decrease in traffic accidents. Furthermore, a study in Los Angeles disclosed that the implementation of a smart traffic management system led to a 16% reduction in traffic accidents. These statistics demonstrate the potential for smart city infrastructure to significantly eliminate the accident ratio and improve safety [14, 15].

An intelligent communication protocol between vehicles was presented in [16, 17], aiming to produce a database for several vehicles, which includes each vehicle's energy experience. Following that, a learning phase will be generated to create an adaptive model that can provide each trajectory with a better drive approach [18, 19]. All the related parameters and variables were defined and associated to formulate a mathematical problem.

II. THE CONCEPT OF CONNECTED ELECTRIC VEHICLE

Two versions of EVs have been developed: HEVs and Total Electric Vehicles (TEVs). An electrical motor that can function as either the primary active motor or a backup active motor should be positioned. The HEV version allows for this [9]. There are two types of cars: one with an electrical power system and the other without. The hybrid variant is the only one with an ICE and a fuel tank. A HEV can be classified as series or as parallel depending on where the hybrid electrical motor is situated [10].

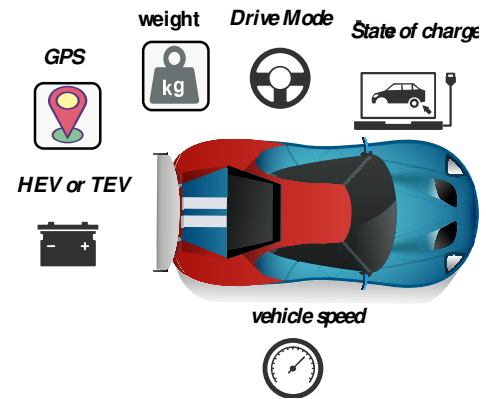


Fig. 1. Concept of the connected vehicle with information to be exchanged.

The concept of a vehicle can be modeled in two different ways when the vehicle is on the road, namely with and without speeds [20]. The EV typically travels a distance s during each stop-and-go phase by accelerating, maintaining a constant speed of v , and decelerating. The related energy formula during the constant velocity period can be expressed as:

$$E_2 = \int_0^t (Au^2 + (B + Cu))u dt \quad (1)$$

where A , B , and C are defined as follows:

$$\begin{aligned} A &= \frac{1}{2} \rho A_f C_D \\ B &= C_r mg \\ C &= C_u mg \end{aligned} \quad (2)$$

In the acceleration form, the revolted energy consumption model can be expressed as in (2), where a is the acceleration ratio and u is the speed of the car.

$$\begin{aligned} E_1 &= \int_0^{u/a} (ma + A(at)^2 + (B + Cat))at dt \\ &= \frac{1}{2} mu^2 + \frac{Au^4}{4a} + \frac{Bu^2}{2a} + \frac{Cu^3}{3a} \end{aligned} \quad (3)$$

So, the global energy consumption model regroups the two previous energy equations. The corresponding model for multiple stops and for a giving drive cycle can be expressed as in (4), where n is the vehicle stop number, m is the vehicle mass, g is a constant force factor, C_r and C_u are the resistive torque and the acceleration torque, respectively [11]:

$$E(u, n) = (n + 1)(E_1 + E_3) + E_2(u, n) \quad (4)$$

III. THE CONNECTED VEHICLES IN THE SMART CITY CONCEPT

A. Smart City Concept (Database) and Connected Vehicles

The concept of smart city and connected vehicles has several fields to be discovered and ameliorated. The relationship between these two concepts can be discussed for improving road safety or enhancing vehicle and driver security.

It can be also used for delivery service and time optimization or similar to road trajectory optimization and energy inside vehicle management.

This study aims to demonstrate that the integration of EVs within smart cities can enhance vehicle autonomy. In reality, every car has a different energy consumption pattern depending on internal and external circumstances. A vast information database that gathers more than vehicle energy management experience can be found in vehicles that drive on the same road. An optimization system will then generate the best energy experience, which will be shared with other cars on the road in order to reduce the energy cost.

1) Principle of Information sharing between Vehicle and Database

The principle of exchanging information between a vehicle passing through a database center, which regroups the energetic identity of the vehicle through sensors installed around and on vehicles, is described in Figure 2. More than one vehicle will cross the same trajectory on the map. Those vehicles were classified by the time of entrance to the trajectory, and the oldest one was the first vehicle that entered the trajectory. The last vehicle which enters the road will be the newest car, and it is the vehicle that needs to know the best way to manage power on this road.

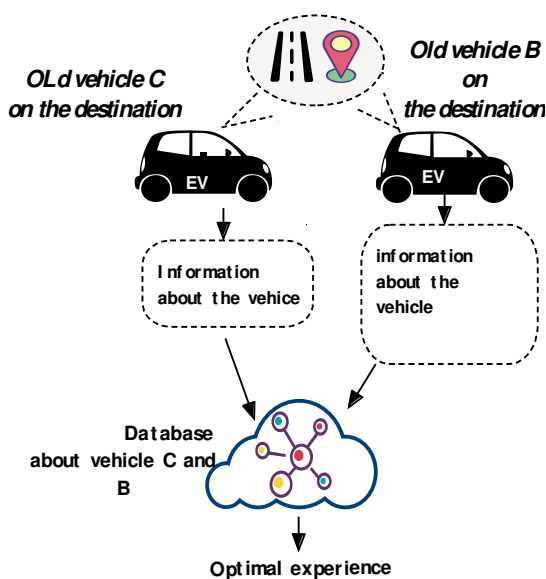


Fig. 2. Principle of information sharing between vehicles on the same trajectory.

2) Vehicle and Smart City Relationship

In this concept, the vehicle is supposed to be inside the city limits, where the smart city concept has all the required apparatus for communication protocols and basic sensors and equipment. Mixing the Internet of Things (IoT) and vehicle-to-vehicle becomes a solution for managing the power of the vehicle and increasing the security factor inside the city. Much information about the vehicle and its surroundings will be available. Knowing the vehicle's GPS position will help determine its slope and any potential needs. It will be easier to

see the conductor mode of the car by using the face detection system. Knowing the status of the batteries and the fuel will also help decide whether or not to start the hybrid mode.

Knowing the relevant torque on the machine is just as crucial as the vehicle weight. Each piece of information will, therefore, be essential for reducing power use. Knowing the road situation and the road traffic will also be valuable for obtaining information regarding future decisions. Those data will be regrouped and classified from each connected vehicle for building an Identification Energetic Label (IEL) for each vehicle. Creating a flawless archive is difficult since similar car models with the same fitted sensors and the same architecture are not readily available in the same zone. In this initial study, it will be assumed that all linked automobiles have exactly the same attributes in order to simplify the application. More than just a car connected to the cloud database is needed in order to start developing a database. Each car will display its information, as seen in Figure 3.

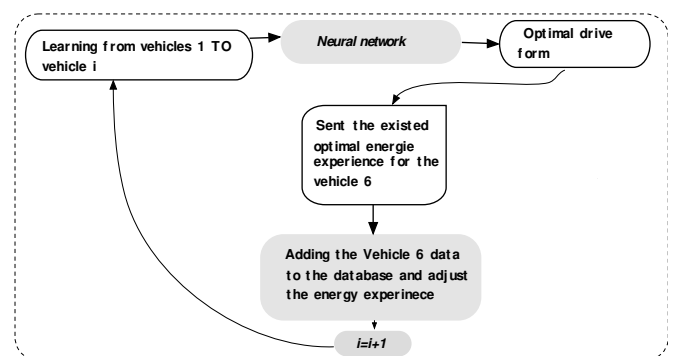


Fig. 3. Description of the algorithm function.

3) Communication Protocols

The system will initially inquire about the car model, the vehicle position, and the target trajectory if the original vehicle is brand-new for this Database. The vehicle will be placed in the appropriate category. The employed algorithm will compare the vehicle information with other existing information inside the Database. If there is a similarity between two scenarios for the vehicle state in the Database, the algorithm will select the best scenario from those that are already possible. In the absence of new information, the system will update the Database to reflect this vehicle category and begin a new learning phase. After that, a brand-new energy-optimal condition will exist. Until a different optimal condition is provided by the system, this vehicle will retain its energy experience. For the hybrid vehicle category, the automobile will be questioned about its energy usage, including the way it consumes fuel and electricity and the quantity of energy it stores. The speed of the car will also be questioned. All that data must be reachable from a specific km distance. If the vehicle is not categorized as a pure electric vehicle, the system will essentially inquire about the amount of electricity consumed, the speed, and the acceleration of the vehicle over the last specific km distance. The car requests to join the

Database if one of those critical variables of information is missing.

B. The Learning Algorithm Discerption

Since the Database contains information from several sources, including multiple connected vehicles, the system will aggregate most of the vehicle-identified data into numerous vector inputs. Equation (5) presents an information vector that contains the desired data for the pure electric vehicle number i . The vector involves information on the vehicle's weight, speed, and State Of Charge (SOC). Equation (6) exposes the desired output signal that contains the necessary information to be used by the algorithm. All these vectors will be utilized in order to reduce the global equation system exposed in (7).

$$X_{TEV}(i) = [W(i) \ V(i) \ SOC(i) \ D_{A \rightarrow B}] \quad (5)$$

$$Y_{TEV}(i) = [ACC^*(i) \ V_{low}(i) \ V_{max}(i)] \quad (6)$$

The selected neural network algorithm is based on the given architecture in Table I.

TABLE I. NEURAL NETWORK ARCHITECTURE FOR THE TRAINING PHASE

Input layer	Hidden layers	Output layer	Learning function
5	2	1	Sigmoid

C. The Objective Function

Finding the optimal driving cycle that reduces the EV's whole trip energy consumption is the ideal issue that can be expressed as follows:

$$\min E(v, n) = f(a) \quad (7)$$

Therefore, the energy consumption factor can be decreased by adjusting the acceleration ratio. It is crucial to note that, for the chosen trajectory, the vehicle speed cannot be higher than the vehicle limit speed set by the other cars.

IV. PROTOTYPE DESCRIPTION AND RESULTS

Six linked cars will be employed in an attempt to describe the working algorithm. Five cars will build the Database and the sixth will act as a visitor who will request an optimal energy experience. Assumedly, the trajectory for each

automobile is the same. The learning system will make use of the data from the five known vehicles, as displayed in Figure 3. The program will first examine those cars in an effort to generate the best possible energy experience. The learning system will look for the best energy experience and determine the best driving technique for a new automobile when one is present and requests it in order to drive it in the desired trajectory. Following that, the algorithm will resume and attempt to enhance its database by adding the automobile data. The learning algorithm will make use of the saved data to provide the best possible energy experience and then the corresponding best drive method.

Figure 3 depicts four pieces of car information in relation to the acceleration form and its corresponding speed behavior. It is clear that some differences in the drive mode exist at 100, 200, 500 s for the three cars. Also, at 1500 s, it can be observed that each car has a different drive method. The car's weight is 1200 kg and their trajectories are fixed at 2500 s, while their weight is assumed to be the same.

With this data, the neural network algorithm will attempt to understand how this vehicle type behaves for a specific acceleration. The learning phase was developed over 1500 iterations, which corresponds to the smallest learning error and an execution duration that is feasible. As the database information is large enough, in relation to the vehicle drive evolution in time function, the rest of the result analysis will be kept on a zoom section, for instance, 200 s to 340 s. As it is required to follow the energy consumption factor, Figure 5 shows the related energy consumption form for the given drive section illustrated in Figure 4. It can be noticed that some differences exist in the energy curves for three vehicle cases.

The energetic experience in the database is presented in Figure 5. The zoomed zones disclose the EV feedback in relation to the modified driven cycle form that was utilized to operate the car. Approximately, all four energy experiences are the same, with just minor variations. For instance, it appears that the driver of the first electrical car operated the vehicle differently from the other examples, particularly in the 200–300 s range. Every car in the other vehicles was driven using a unique driving technique. For the learning phase, each of these experiences and their SOC forms were employed.

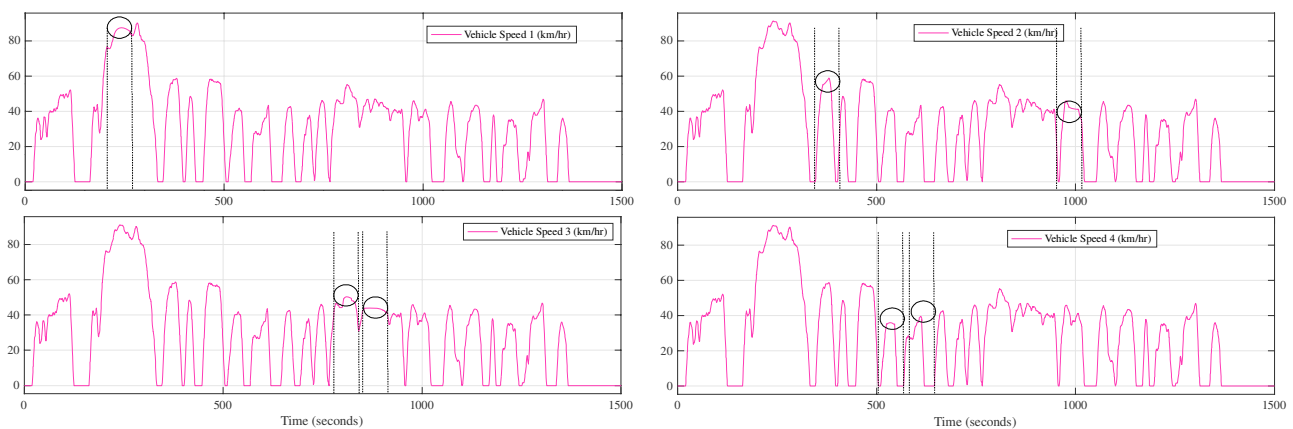


Fig. 4. Acceleration forms for four cars in the Database for the same trajectory with similar speed limits and some modifications on the drive method.

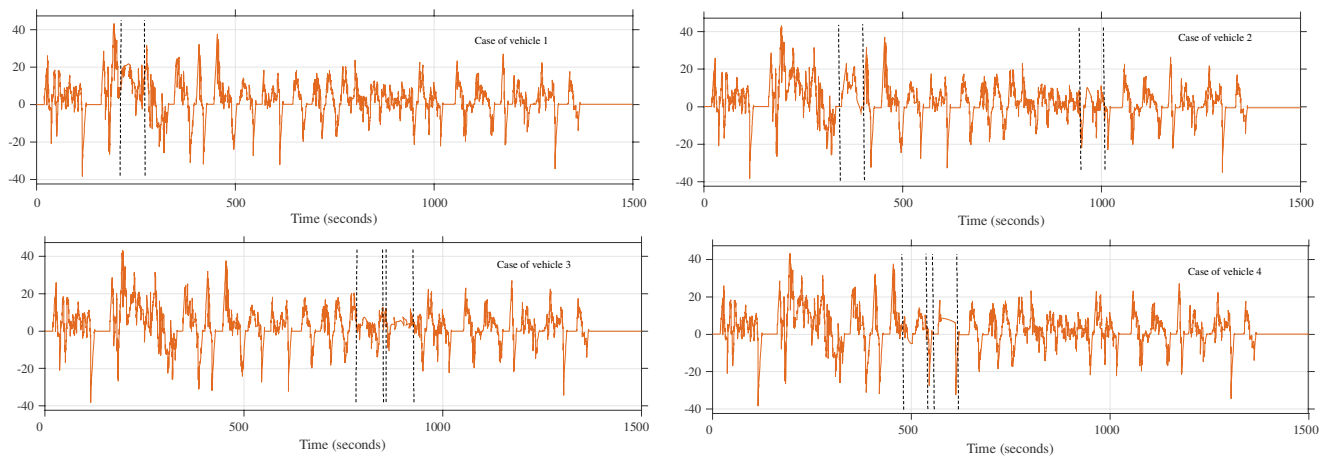


Fig. 5. Energetic experience of the four vehicles.

After running the corresponding neural block, the overall algorithm will have finished. It is easy now for any vehicle that asks for an optimal energy experience to have what is needed. The results related to the guest vehicle can be seen in Figure 6. The adaptability and the efficiency of the obtained neural network bloc will be tested on a new drive cycle form, as is in Figure 6. This Figure demonstrates the optimal trajectory drive cycle obtained by the neural network as well as two different

cars using the same trajectory with a different acceleration form. The related electrical motor speed form can also be spotted, while few modifications are observed. When a vehicle employs its own drive cycle form or the best drive cycle model, Figure 6 displays the associated SOC form for each scenario. In conclusion, it is evident that there is a difference in the SOC and if the best course is taken, the SOC will end up being greater than in the other scenarios.

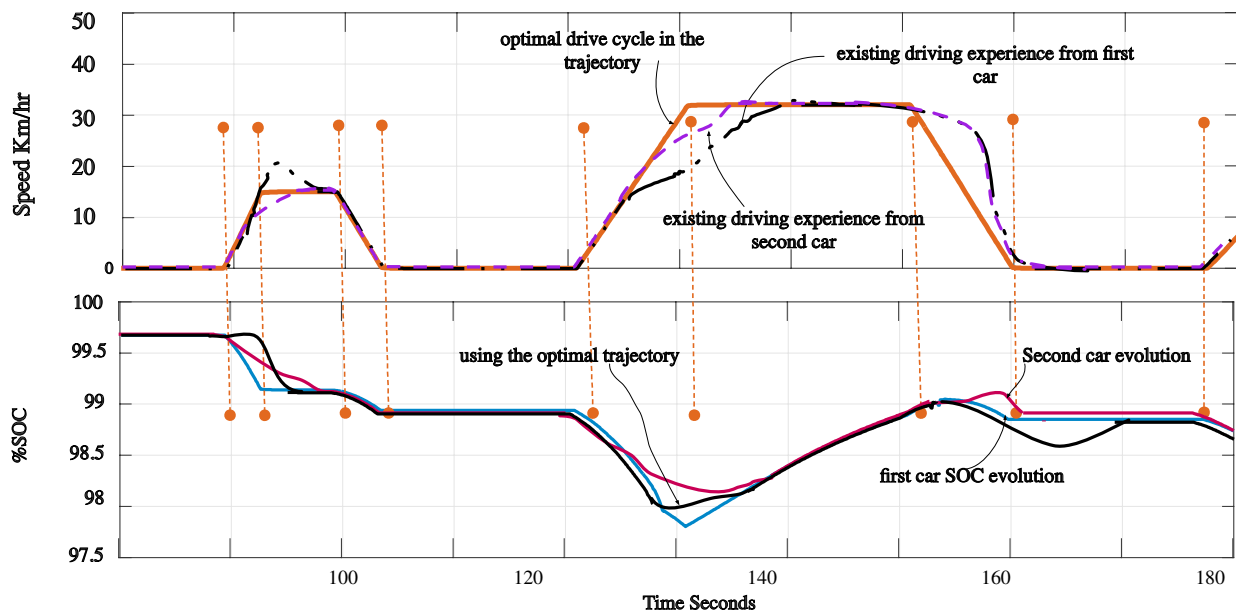


Fig. 6. Optimal drive cycle from two existing drive cycle forms.

Figure 6 exposes the related battery SOC for three different cars. One of them has used the optimal drive cycle given by the neural network and the others have implemented their own drive method. It can be seen that the SOC is different between the three cases, especially in the up-slop, at 10 s, 50 s, 90 s and from 150 to 180. This is due to the intervention of the optimization block, which has concluded that in these instances, the acceleration slope is not necessary to be

maximum and can be decreased. The optimization algorithm successfully applied to various slopes, maintained the battery's SOC at its initial value by the end of the simulation period. This consistency is partly due to the car model's energy regeneration feature on down-slope roads, which explains the increases in battery SOC during certain periods.

V. CONCLUSION

In this paper, a new energy management topology is proposed, which can improve vehicle autonomy and enhance the use of electric vehicles. The proposed concept has been based on the smart city concept, where more information acquisition tools exist and information exchange is easy. The idea is to create a database for every car, containing data from various sensors, both inside and outside the vehicle. This database will be organized and utilized during the training process to train a neural network that can determine the best drive cycle strategy for each unique set of road conditions. A multiple drive cycle form was employed to construct the concept, which was then evaluated in a specific drive cycle example.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the approval and the support of this research study by the grant no. ENGA-2022-11-1449 from the Deanship of Scientific Research at Northern Border University, Arar, K.S.A.

REFERENCES

- [1] J. Javorski Eckert, L. Corrêa de Alkmin e Silva, F. Mazzariol Santiciolli, E. dos Santos Costa, F. C. Corrêa, and F. Giuseppe Dedini, "Energy storage and control optimization for an electric vehicle," *International Journal of Energy Research*, vol. 42, no. 11, pp. 3506–3523, 2018, <https://doi.org/10.1002/er.4089>.
- [2] N. Mutoh and T. Inoue, "A Control Method to Charge Series-Connected Ultraelectric Double-Layer Capacitors Suitable for Photovoltaic Generation Systems Combining MPPT Control Method," *IEEE Transactions on Industrial Electronics*, vol. 54, no. 1, pp. 374–383, Feb. 2007, <https://doi.org/10.1109/TIE.2006.885149>.
- [3] F. Musavi, M. Edington, and W. Eberle, "Wireless power transfer: A survey of EV battery charging technologies," in *2012 IEEE Energy Conversion Congress and Exposition (ECCE)*, Raleigh, NC, USA, Sep. 2012, pp. 1804–1810, <https://doi.org/10.1109/ECCE.2012.6342593>.
- [4] Y. Cao *et al.*, "An Optimized EV Charging Model Considering TOU Price and SOC Curve," *IEEE Transactions on Smart Grid*, vol. 3, no. 1, pp. 388–393, Mar. 2012, <https://doi.org/10.1109/TSG.2011.2159630>.
- [5] I. W. Cox, "Electric vehicle traction control system and method," US20150005988A1, Jan. 01, 2015.
- [6] F. Un-Noor, S. Padmanaban, L. Mihet-Popa, M. N. Mollah, and E. Hossain, "A Comprehensive Study of Key Electric Vehicle (EV) Components, Technologies, Challenges, Impacts, and Future Direction of Development," *Energies*, vol. 10, no. 8, Aug. 2017, Art. no. 1217, <https://doi.org/10.3390/en10081217>.
- [7] S. A. A. Tarusan, A. Jidin, M. L. M. Jamil, K. A. Karim, and T. Sutikno, "A review of direct torque control development in various multilevel inverter applications," *International Journal of Power Electronics and Drive Systems (IJPEDS)*, vol. 11, no. 3, pp. 1675–1688, Sep. 2020, <https://doi.org/10.11591/ijpeds.v11.i3.pp1675-1688>.
- [8] T. Ahmed, H. Kada, and A. Ahmed, "New DTC strategy of multi-machines single-inverter systems for electric vehicle traction applications," *International Journal of Power Electronics and Drive Systems*, vol. 11, no. 2, pp. 641–650, Jun. 2020, <https://doi.org/10.11591/ijpeds.v11.i2.pp641-650>.
- [9] D. Casadei, F. Profumo, G. Serra, and A. Tani, "FOC and DTC: two viable schemes for induction motors torque control," *IEEE Transactions on Power Electronics*, vol. 17, no. 5, pp. 779–787, Sep. 2002, <https://doi.org/10.1109/TPEL.2002.802183>.
- [10] L. Xi, X. Zhang, C. Sun, Z. Wang, X. Hou, and J. Zhang, "Intelligent Energy Management Control for Extended Range Electric Vehicles Based on Dynamic Programming and Neural Network," *Energies*, vol. 10, no. 11, Nov. 2017, Art. no. 1871, <https://doi.org/10.3390/en10111871>.
- [11] M. Catak and B. Celikkaya, "Limits of Number of Motes at Smart City Scenarios for LoRaWAN," *Engineering, Technology & Applied Science Research*, vol. 9, no. 3, pp. 4116–4119, Jun. 2019, <https://doi.org/10.48084/etasr.2693>.
- [12] A. Khadhraoui, T. Selmi, and A. Cherif, "Energy Management of a Hybrid Electric Vehicle," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 8916–8921, Aug. 2022, <https://doi.org/10.48084/etasr.5058>.
- [13] N. T. Diep and N. K. Trung, "Transmitting Side Power Control for Dynamic Wireless Charging System of Electric Vehicles," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 9042–9047, Aug. 2022, <https://doi.org/10.48084/etasr.4988>.
- [14] Z. Chen, C. B. Sivaparthipan, and B. Muthu, "IoT based smart and intelligent smart city energy optimization," *Sustainable Energy Technologies and Assessments*, vol. 49, Feb. 2022, Art. no. 101724, <https://doi.org/10.1016/j.seta.2021.101724>.
- [15] L. Zhang *et al.*, "Research on the orderly charging and discharging mechanism of electric vehicles considering travel characteristics and carbon quota," *IEEE Transactions on Transportation Electrification*, pp. 1–1, 2023, <https://doi.org/10.1109/TTE.2023.3296964>.
- [16] K. Fahem, D. E. Chariag, and L. Sbata, "On-board bidirectional battery chargers topologies for plug-in hybrid electric vehicles," in *2017 International Conference on Green Energy Conversion Systems (GECS)*, Hammamet, Tunisia, Mar. 2017, pp. 1–6, <https://doi.org/10.1109/GECS.2017.8066189>.
- [17] Y. Bian, Y. Zheng, W. Ren, S. E. Li, J. Wang, and K. Li, "Reducing time headway for platooning of connected vehicles via V2V communication," *Transportation Research Part C: Emerging Technologies*, vol. 102, pp. 87–105, May 2019, <https://doi.org/10.1016/j.trc.2019.03.002>.
- [18] A. Flah and C. Mahmoudi, "Design and analysis of a novel power management approach, applied on a connected vehicle as V2V, V2B/I, and V2N," *International Journal of Energy Research*, vol. 43, no. 13, pp. 6869–6889, 2019, <https://doi.org/10.1002/er.4701>.
- [19] H. Kraiem *et al.*, "Decreasing the Battery Recharge Time if Using a Fuzzy Based Power Management Loop for an Isolated Micro-Grid Farm," *Sustainability*, vol. 14, no. 5, Jan. 2022, Art. no. 2870, <https://doi.org/10.3390/su14052870>.
- [20] B. R. Siddharth, D. J. Pradeep, Y. V. P. Kumar, C. P. Reddy, and A. Flah, "Dynamic performance analysis of front-wheel drive hybrid electric vehicle architectures under different real-time operating conditions," *International Journal of Powertrains*, vol. 11, no. 1, pp. 62–89, Jan. 2022, <https://doi.org/10.1504/IJPT.2022.121974>.

Optimizing Unit Scheduling with Fuzzy Logic: A Strategic Approach for Efficient Power Network Operations

Sahbi Marrouchi

LaTICE Laboratory, National Higher Engineering School of Tunis, University of Tunis, Tunisia | Higher Institute of Technological Studies of Kef, Tunisia
sahbimarrouchi@yahoo.fr (corresponding author)

Moez ben Hessine

Higher Institute of Applied Sciences and Technologies of Gafsa, University of Gafsa, Tunisia
moezbenhessine@yahoo.com

Souad Chebbi

LaTICE Laboratory, National Higher Engineering School of Tunis, University of Tunis, Tunisia
chebbi.souad@gmail.com

Received: 12 January 2024 | Revised: 27 January 2024 | Accepted: 5 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6894>

ABSTRACT

This study delves into addressing the challenge of resolving the Unit Commitment (UC) problem, which focuses on enhancing the efficiency of production units and devising their operational schedules to accommodate fluctuations in consumption spanning from a day to a month. Given the intricate, combinatorial, and nonlinear constraints associated with each production unit, this study advocates an optimization approach rooted in fuzzy logic. A Lagrangian function was established to simplify the UCP and to transform the different inequality into a linear unconstrained problem. The choice of fuzzy inputs was established using the partial derivatives of a Lagrangian function as a function of the powers injected into each node of the electrical network. This combination of the Lagrangian function and the input of the fuzzy regulator made it possible to control the different constraints in the total production cost function and to improve the operating efficiency of the different production units. This method was effectively applied to a 14-bus IEEE power network encompassing 5 generating units, to address the UC problem by optimizing generator load capacity (LCG) and minimizing Incremental Losses (IL). The numerical processing of the fuzzy linguistic variables was implemented using Mamdani-type fuzzy rules. This strategy stands out for its robust exploratory capability, facilitating the identification of optimal solutions to reduce production costs while ensuring optimal planning of production units.

Keywords-unit commitment; optimization; fuzzy logic; production unit; energy management

I. INTRODUCTION

The production of electricity must be compatible with consumption since the electricity is not stored. Therefore, a power company must plan the generators to start and organize the moments to connect them to the network as well as their duration of operation. The Unit Commitment (UC) is the best solution in the field of modern power grid planning, enabling both the optimization of the day-to-day operational planning of the grids and the reduction of the total production cost through the improvement of units and respect of the schedules. The main objective is to program the production units to meet the consumption demand with the minimum cost. Generation scheduling includes the determination of commissioning and

generation level for each unit during a given planning period. In addition, each unit has its own production limits and a minimum restart and shutdown time. It is therefore a complex, combinatorial, and non-linear optimization issue [1-6].

The UC problem is directly related to unit scheduling and economic dispatch, while knowing that the system is subjected to several constraints. Equilibrium offers demand and limitation of the minimum durations from top to bottom of the thermal production unit and ensures that the powers generated are within the allowable margins. However, this matter includes a multitude of difficulties, namely the large size of the network studied, the presence of coupling constraints, the presence of operational constraints, and the time constraint that must be small with regard to the size of the problem [5-6]. In

this context, production unit operation planning should be established to select which of the production units to be available to supply the forecast load of the system over a future period. Many numerical optimization techniques have been proposed to address the UC problem, like dynamic programming [6-9], the Lagrangian relaxation method [10-13], mixed variable programming [14-15], and the branch-and-bound method [16]. The dynamic programming method is simple but has a rather long computation time to converge to the optimal solution. The branch-and-bound method adopts a linear function to represent fuel consumption and start-up costs as a function of time. The disadvantage of this method is that its required execution time increases rapidly for large-scale UC problems [17-18]. The method of programming in mixed variables uses linear programming to reach an optimal solution. This method has been applied to small UC problems and has required major assumptions that limit the margin of the solutions. On the other hand, the time factor has an advantage for the Lagrangian relaxation method, but the latter suffers from the quality of the optimal solutions obtained.

Several numerical techniques have been applied to the UC problem, such as fuzzy logic [19-22], artificial neural networks [23], simulated annealing [24-27], Tabu search [26, 28], and the genetic algorithm [29-30]. These methods can take into account more complex constraints and are claimed to have a better solution quality. In this context, several studies presented a genetic approach to determine the order of priority of production units [31-32]. These studies have examined the feasibility of using genetic algorithms to optimize production costs and presented effective simulation results. The use of genetic algorithms to solve the UC problem dynamically evaluates the priority of units, taking into account the parameters of the system, the operating constraints, and the load profile requested for a well-defined period. In [4], a hybrid optimization method was proposed to solve the UC problem. This method combined the Particle Swarm Optimization (PSO) method, the Sequential Quadratic Programming technique (SQP), and the Tabu-Search (TS) method. The combinatorial part of the UC problem was solved using the TS method. In [33], a method was utilized for an employee's recovery method to eliminate UC. This technique served to increase the likelihood of generating feasible solutions and significantly reduce the time elapsed for finding unrealizable solutions. In [34], a fuzzy logic approach was implemented to produce a logical and feasible solution for each period, considering the many uncertainties involved in the planning and operation of the electrical grid. The load request and the reserve margin were treated as fuzzy variables.

This study adopted a new strategy to solve the UC problem based on the fuzzy approach, which allows the optimization of the production cost (CP) while guaranteeing adequate planning of the production units using a good selection of the fuzzy inputs and fuzzy rules. The proposed approach relies on a Lagrangian function chosen as the objective function to determine the Load Capacity of the Generator (LCG) and the Incremental Losses (IL), which were chosen as fuzzy input variables. These parameters are essential to minimize the total CP, which was chosen as the fuzzy output variable.

II. PROBLEM FORMULATION

Many studies have been based on an analytical statement of the UC problem [5, 6, 35-36]. This study presents a mathematical model of the UC problem with limited security, which has been adapted in [36-37]. This model is mixed linear and constrained.

$$\text{Min} [F_T(P_{ih}, U_{ih}) = \sum_{i=1}^{N_g} \sum_{h=1}^H [a_i P_{ih}^2 + b_i P_{ih} + c_i] U_{ih} + \sum_{i=1}^{N_g} \sum_{h=1}^H [ST_i (1 - U_{i(h-1)})] U_{ih}] \quad (1)$$

The objective of the UC problem is to establish the best production unit plan that will be available to minimize the total operating cost of the generating units and to supply the forecasted load over a period H [37-38]. ST_i is the starting cost of the i^{th} unit, defined by:

$$ST_i = \begin{cases} HSC_i & \text{if } MDT_i \leq \tau_i^{OFF} \leq MDT_i + SC_i \\ CSC_i & \text{if } \tau_i^{OFF} > MDT_i + SC_i \end{cases} \quad (2)$$

The minimization of the objective function is provided with the following constraints.

A. System Constraints

- Power balance constraint:

$$\sum_{i=1}^{N_g} P_{ih} U_{ih} = P_{dh} \quad (3)$$

- Spinning reserve constraint:

$$P_{dh} + P_{rh} - \sum_{i=1}^{N_g} U_{ih} P_{ih} \leq 0 \quad (4)$$

B. Unit Constraints

- Generation limits:

$$P_i^{\min} U_i \leq P_{ih} \leq P_i^{\max} U_i \quad (5)$$

- Minimum uptime constraint:

$$U_{ih} = 1 \quad \text{for } \sum_{t=h-upt_i}^{h-1} U_{it} \leq MUT_i \quad (6)$$

- Minimum downtime constraint:

$$U_{ih} = 0 \quad \text{for } \sum_{t=h-downt_i}^{h-1} U_{it} \leq MDT_i \quad (7)$$

Therefore, to simplify the UC problem and to transform the different inequality into a linear unconstrained problem, the following Lagrangian function was considered:

$$L(P_{ih}, U_{ih}, \lambda_i) = \sum_{i=1}^{N_g} \sum_{h=1}^H [\varphi_i(P_{ih}) + ST_i (1 - U_{i(h-1)})] U_{ih} + \lambda_i \cdot (P_d - \sum_{i=1}^{N_g} P_{ih} U_{ih}) \quad (8)$$

where λ_i is the Lagrangian coefficient.

III. METHODOLOGY OF RESOLUTION

Not only does fuzzy logic provide a meaningful and powerful representation for the measurement of uncertainties, but also a meaningful representation of the fuzzy notion expressed in normal language. Fuzzy logic is a mathematical theory that encompasses the idea of vagueness when defining a concept or a meaning. For example, there is uncertainty in expressions, such as "low" or "high", since these expressions are imprecise and relative. Thus, such variables are called

"fuzzy" as opposed to "net". Fuzzy is simply a way to describe the uncertainty. Generally, solving such a problem via fuzzy logic is based on the following three steps [37, 39-41]:

- Fuzzification transforms numerical values of net input into fuzzy variables.
- Fuzzy inference consists of a set of fuzzy logic rules.
- Defuzzification allows the transformation of the fuzzy variables into net real output.

Such ideas can be easily applied to solve the UC problem. Fuzzy logic allows a qualitative description of the behavior of a certain system, its characteristics, and its response without the need for an exact mathematical formulation. This study applied a new optimization strategy based on the fuzzy approach, which allows for taking into account many uncertainties involved in the planning and operation of electrical networks.

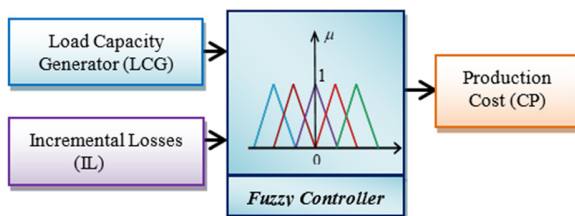


Fig. 1. Block diagram of the fuzzy controller.

Since the amount of expected power is inaccurate, it should be presented as a fuzzy quantity. In previous studies, the associated fuzzy variables for solving the UC problem were the load capacity of the generators, the cost of starting the units, and CP. However, this study considered the LCG and IL. The output variable is the CP. The following steps were taken to properly select the fuzzy variables that can lead to a minimal production cost. Starting from the condition that the partial derivatives of the Lagrange function (8) for each of the controllable variables are zero, the following equations were obtained:

$$\frac{\partial L}{\partial P_{ih}} = \frac{\partial[\varphi_i(P_{ih})]}{\partial P_{ih}} - \lambda \left(\frac{\partial P_{Lh}}{\partial P_{ih}} - U_{ih} \right) = 0 \quad (9)$$

The partial derivative of the Lagrange function for the controllable variable λ is as follows:

$$\frac{\partial L}{\partial \lambda} = P_{dh} - P_{Lh} - \sum_{i=1}^{N_g} P_i U_{ih} = 0 \quad (10)$$

These conditions arise from the fact that to have a local or global minimum for such a linear function, these optimums correspond to the points where the partial derivative of the considered function is equal to zero. Equations (9) and (10) represent the optimality conditions necessary to solve (1) and (8) without resorting to inequality constraints (5), (6) and (7). Equation (9) can be written as follows:

$$\lambda = \frac{\frac{\partial[\varphi_i(P_{ih})]}{\partial P_{ih}}}{\frac{\partial P_{Lh}}{\partial P_{ih}} - U_{ih}}; \quad i = \{1, \dots, N_g\}; \quad h = \{1, \dots, H\} \quad (11)$$

The term $\frac{\partial[\varphi_i(P_{ih})]}{\partial P_{ih}}$ represents the incremental cost of each unit i , and $\frac{\partial P_{Lh}}{\partial P_{ih}}$ represents the incremental losses. To establish an effective strategy leading to the minimization of CP, it was necessary to take a better account of the losses already mentioned. The current expression of losses is translated as follows:

$$P_L = [P_1 \ P_2 \ \dots \ P_{Nn}] \begin{bmatrix} B_{11} & B_{12} & \dots & B_{1Nn} \\ B_{21} & B_{22} & \dots & B_{2Nn} \\ \dots & \dots & \dots & \dots \\ B_{Nn1} & B_{Nn2} & \dots & B_{NnNn} \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ \dots \\ P_{Nn} \end{bmatrix} + [B_{10} \ B_{20} \ \dots \ B_{Nn0}] \begin{bmatrix} P_1 \\ P_2 \\ \dots \\ P_{Nn} \end{bmatrix} + B_{00} \quad (12)$$

leading to:

$$P_L = \sum_{i=1}^{Nn} \sum_{j=1}^{Nn} P_i B_{ij} P_j + \sum_{i=1}^{Nn} B_{i0} P_i + B_{00} \quad (13)$$

where B_{ij} is the element (i, j) of the square matrix, B_{i0} is the element connected to the generator i , and B_{00} is the constant designating system losses if the power demand is equal to zero. These coefficients depend on the amplitude and the phase of the voltage and the reactive power injected into the electrical network at the bus i . The evaluation of (11) and (13) leads to the following equation of incremental losses:

$$IL_i = \frac{\partial P_{Lh}}{\partial P_{gi}} = 2 \sum_{j=1}^{Nn} B_{ij} P_j + B_{i0} \quad (14)$$

Therefore the various variables (LCG and IL) can be established, forming the basis of the proposed strategy which can be used to reduce the total cost of production. These terms intervene as fuzzy variables associated with the UC problem-solving strategy. It should be noted that the strategy is based on the integration of a fuzzy controller to optimize CP while guaranteeing adequate planning of production units. In the current formulation, the fuzzy input variables associated with the UC problem are:

- LCG, which is based on the load to be served.
- IL, since losses can cause changes in the total CP and vary over the overall architecture of the power grid.
- The CP of the system is treated as a fuzzy variable because it is directly proportional to the hourly load demand.

The fuzzy sets defining these variables are selected and standardized between 0 and 1 [19-20]. The processing of the variables by a fuzzy controller requires three steps, namely a so-called fuzzification step, a second step named fuzzy inferences, and a last stage designated by defuzzification, consisting of transforming a fuzzy variable into a non-fuzzy variable. The process of processing variables using fuzzy logic is established as shown in Figure 2.

A. Membership Functions

The numerical processing of the linguistic variables of a fuzzy corrector requires the use of membership functions. These variables were characterized by fuzzy sets: Low (L), below average (BAV), average (AV), above average (AAV),

and High (H). Several membership functions can be associated with these fuzzy sets. Based on the fuzzy sets cited, the membership functions are chosen for each fuzzy input and output variable as displayed in Figure 3. The fuzzy sets defining the input variables were adopted as follows: *LCG* and *IL* are indicated by the following fuzzy sets:

$$LCG = \{L, BAV, AV, AAV, H\}$$

$$IL = \{L, BAV, AV, AAV, H\}$$

The CP, taken as an objective function, is:

$$CP = \{L, BAV, AV, AAV, H\}$$

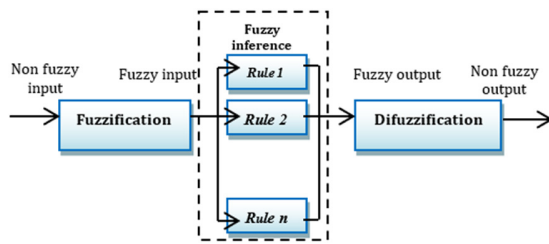


Fig. 2. Resolution process through fuzzy logic.

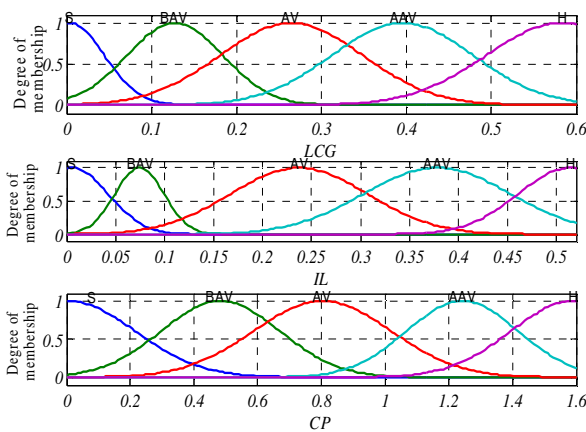


Fig. 3. Membership degrees of *LCG*, *IL*, and *CP*.

B. Fuzzification

Triangular and trapezoidal functions were used for the fuzzification of the input variables for two reasons: they do not require much calculation time in their evaluations, and they are easy to implement. Figure 4 shows the fuzzification of the input variable *IL* to assign a degree of belonging μ for this variable.

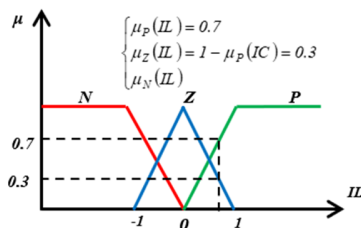


Fig. 4. Fuzzification of the input variable *IL*.

C. Fuzzy Inferences

Fuzzy deductions represent the set of conclusions to be drawn or the actions to be performed for each rule, following the determination of the membership functions and the input and output variables of a system already modeled. The decisions are taken at the level of the control vector of the fuzzy regulator, once the information on the inputs of the regulator is assembled. Based on the aforementioned fuzzy sets, the membership functions are chosen for each fuzzy input and output variable. For simplicity, a triangular shape is employed to illustrate the membership functions considered here. Once these sets are established, the input variables are then linked to the output variable by if-then rules, as illustrated in Figure 5.

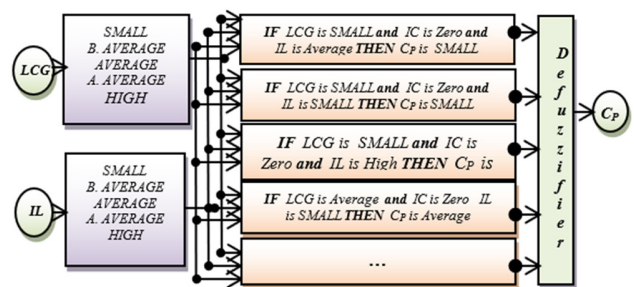


Fig. 5. Configuration of the strategy based on fuzzy logic.

D. Defuzzification

The inference step produced a control law expressed in terms of a membership function and, therefore, a fuzzy control law. Since the system to be controlled can only accept a defined command, this membership function must be translated into a non-fuzzy command. This process is known as defuzzification. Several methods of defuzzification exist in the literature: the best known is defuzzification by center of gravity. This method calculates the non-fuzzy control law U_C^* as an abscissa of the center of gravity of the membership function of the control law $\mu(U_C)$. This study utilized Mamdani-type fuzzy rules to formulate conditional statements that include fuzzy logic. The proposed fuzzy rules are designed to optimize *CP* through a correct evaluation of the *LCG* and *IL* variables. In this respect, according to the fuzzy sets associated with each input variable, 25 fuzzy rules were established ($5 \times 5 = 25$), as shown in Table I. Each rule represents the correspondence between the input and output spaces.

TABLE I. FUZZY RULES LINKING FUZZY INPUT/OUTPUT VARIABLES

Rule	LCG	IL	CP	Rules	LCG	IL	CP	Rules	LCG	IL	CP
1	S	S	S	10	BAV	H	AV	19	AAV	AAV	AAV
2	S	BAV	S	11	AV	S	BAV	20	AAV	H	BAV
3	S	AV	S	12	AV	BAV	BAV	21	H	S	AV
4	S	AAV	S	13	AV	AV	BAV	22	H	BAV	AAV
5	S	H	S	14	AV	AAV	AV	23	H	AV	AAV
6	BAV	S	S	15	AV	H	AV	24	H	AAV	H
7	BAV	BAV	BAV	16	AAV	S	AV	25	H	H	H
8	BAV	AV	BAV	17	AAV	BAV	AV				
9	BAV	AAV	BAV	18	AAV	AV	AAV				

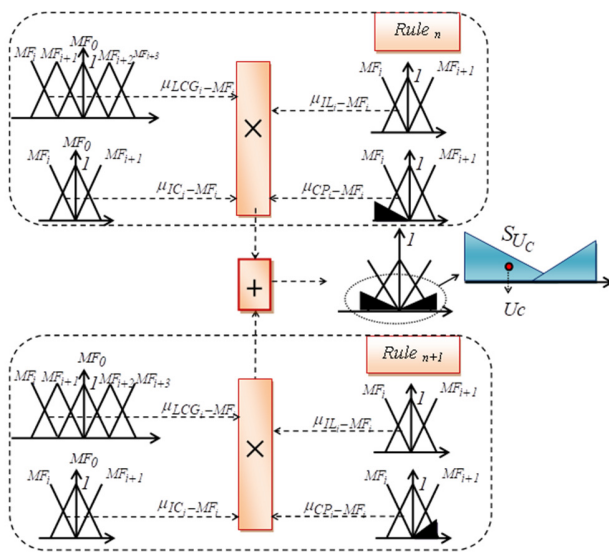


Fig. 6. Fuzzy deductions.

Once the fuzzy rules are defined, the results must be defuzzified to achieve exact values in the desired margins. The defuzzification method applied consists of determining the abscissa of the center of gravity of the surface swept by the fuzzy deductions given by the following equation:

$$C_p = \frac{\int_{-1}^1 C_p \cdot \mu(C_p) \cdot dC_p}{\int_{-1}^1 \mu(C_p) \cdot dC_p} \quad (15)$$

where $\mu(U_C)$ is the degree of membership of the production cost vector. The rule surface shows the output value of CP for every combination of the inputs LCG and IL . The strategy to optimize the proposed UC problem was tested on the IEEE 14 bus test system.

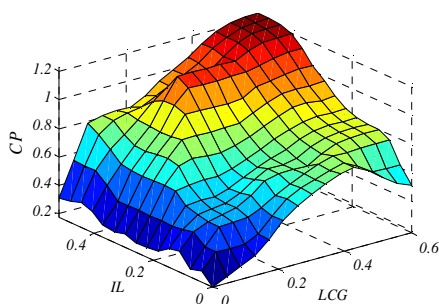


Fig. 7. Fuzzy controller rule surface.

IV. RESULTS AND DISCUSSIONS

The proposed mathematical modeling of the UC problem was simulated in Matlab. These simulations were applied on an IEEE electrical grid 14-bus test system with 5 generators [35-37, 42] over a 24-hour horizon. The strategy was started at time $t = 40$ s and the on-off states and the power quantity generated by each unit were taken every 3h. Tables II and III display the characteristics of the various production units and the expected load demand over a 24-hour horizon. The main successive periods were considered to establish the temporal evolution of the power demand.

TABLE II. CHARACTERISTICS OF PRODUCTION UNITS

U	Pmax (MW)	Pmin (MW)	c	b	a	MUT	MDT	Hot start-up cost (\$)	Cold start-up cost (\$)
1	582	110	379.2	30.36	0.0756	8	8	4500	9000
2	55	15	606.6	27.3	0.2274	3	3	170	340
3	53	10	454.8	22.74	0.2274	3	3	170	340
4	23	8	151.8	22.5	0.1518	1	1	30	60
5	23	8	303.6	22.74	0.1518	1	1	30	60

TABLE III. AMOUNT OF LOAD REQUIRED

Hour	3	6	9	12	15	18	21	24
Demand (MW)	259	200	300	450	527	610	480	320

The simulations aimed to verify the dynamics of the system and to predict whether the proposed strategy provides adequate planning of the production units to minimize CP . In addition, the proposed strategy aimed to reduce the time required to resolve the UC problem. Figure 8 depicts the temporal evolution of the powers generated by each production unit and the optimal power quantities estimated by the proposed method. It should be noted that the generated powers follow the optimal power quantities provided by the proposed optimization algorithm. This demonstrates the high performance of the control algorithms adopted for the supervision of the studied system and proves the efficiency of the regulation loops for the different production units. Furthermore, the proposed strategy aims to obtain sufficient and rapid planning in terms of convergence. The proposed strategy managed to optimize optimal solutions to reduce CP .

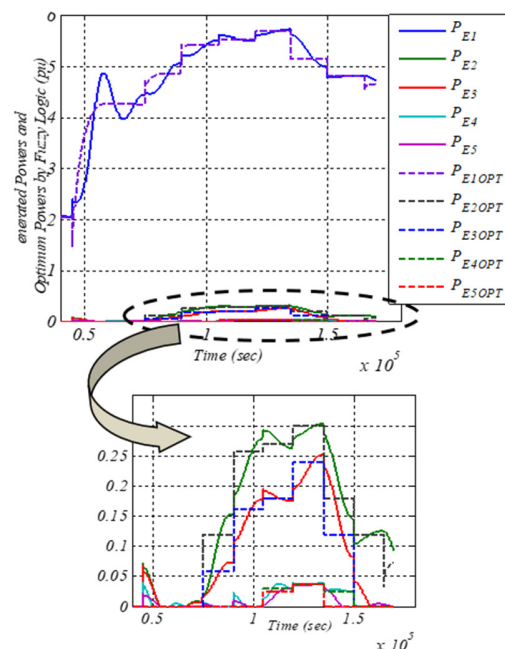


Fig. 8. Scheduling and amount of produced power based on the proposed fuzzy logic strategy.

Based on the results illustrated in Table IV, the proposed method solved the UC problem while addressing the planning of on-off states of the production units, which complied with

the constraints of each unit (minimum start-up MUT_i and shut-down times MDT_i). Moreover, the power produced by the most powerful machine (615 MVA) remains unchanged throughout the 24 hours, while the other production units vary to produce the amount of power demanded by the network.

TABLE IV. RESULTS OF THE RESOLUTION OF THE UC PROBLEM BY FUZZY LOGIC STRATEGY

H	Power demand (MW)	Powers generated by each production (MW)					Total Production Cost (\$)	Optimal scheduling
		615 MVA	60 MVA	60_Bis MVA	25 MVA	25_Bis MVA		
1	259	271.4	0	0	0	0	6820	10000
2	259	287.5	0	0	0	0	7148	10000
3	259	345.6	0	0	0	0	8361	10000
4	200	396.1	0	0	0	0	9457	10000
5	200	441.3	0	0	0	0	10470	10000
6	200	464.9	0	0	0	0	11012	10000
7	300	473.7	12.5	0	0	0	11561	11000
8	300	481.4	14.9	0	0	0	11808	11000
9	300	496.3	17.3	0	0	0	12226	11000
10	450	512.2	18.7	11.57	0	0	12936	11100
11	450	520.3	21.3	16.32	0	0	13342	11100
12	450	521.7	28.0	17.65	0	0	13608	11100
13	527	528.9	29.7	18.94	3.12	3	14011	11111
14	527	536	28.4	18.11	3.75	3.09	14137	11111
15	527	546.2	28.5	18.44	4.46	4.23	14413	11111
16	610	551.3	28.7	20.42	7.982	6.86	14786	11111
17	610	569.2	30.0	23.12	4.87	4.17	15208	11111
18	610	551.7	30.3	26.52	3.92	3.72	14870	11111
19	480	549.5	27.6	18.53	3.75	0	14384	11110
20	480	536.2	22.1	13.32	3.12	0	13732	11110
21	480	502.1	18.3	8.11	2.5	0	12648	11110
22	320	478.5	13.5	0	0	0	11702	11000
23	320	480.4	12.6	0	0	0	11721	11000
24	320	482.1	11.8	0	0	0	11739	11000
Total cost (\$)							2.9210 e+05	
Total power demand (MW)							12192	
Time (s)							7.34	

The production costs established for solving the UC problem when compared with the algorithms in [43-44], demonstrate the high performance of the proposed strategy. Comparing the CP obtained through the proposed approach with that obtained via the genetic algorithm and gradient-PSO (Table V), indicated that the proposed approach was reliable and allowed a gain of 1% of the total CP . However, the strategy based on the use of the hybrid method gradient-genetic algorithm was the most efficient and revealed high performance at the CP level as well as the ability to converge towards the global optimum.

TABLE V. COMPARISON OF PERFORMANCE BETWEEN OPTIMIZATION METHODS

	Genetic Algorithm [24]	Fuzzy Logic	Gradient-PSO [35]
Production cost (\$)	2.9457 e+05	2.9210 e+05	2.9125e+05
Execution time (sec)	10.21	7.34	8.653

By contrast, both genetic and gradient-genetic algorithm methods did not exhibit efficacy-time resolution, since each method requires enough time to reach the optimal solution. This is mainly due to the choice of the initial population. The proposed fuzzy logic strategy was more efficient than the other

two algorithms in terms of execution time and convergence efficiency. This shows the relativity between execution time and the CP value of every production unit.

The proposed strategy had the shortest calculation time compared to other algorithms, such as the Artificial Bee Colony (ABC) and PSO [45]. The calculation time of the UC problem by ABC was about 40.74 s and by PSO 49.03 s [31], while the proposed strategy had a calculation time of 7.34 s. Table IV discloses that the proposed hybrid optimization method was able to organize the running and shutdown states of various production units. This is done by using an estimate of the amount of charge desired by the electrical network while taking into account the permissible constraints. However, fairly optimal planning allows us to benefit from a minimal CP . The superiority of the resolution through fuzzy logic theory is obvious. This strategy works better than other stochastic methods in terms of the planning of the on/off states of the various units and thus optimizes the total CP . Indeed, starting from the combination probability equation for such planning:

$$P_{Combination} = (2^n - 1)^m \quad (16)$$

where n is the number of units and m is the discretized time. For this case study, $P_{Combination}$ is worth about 6.2^{35} combinations. This number suggests the ability of the proposed method to choose a perfect plan that guarantees the balance between supply and demand and a fairly minimal CP . Considering the technical constraints specific to each production unit (power generation limit, minimum up-time constraint, minimum operating time before shutdown), the proposed strategy enables the planning of the on-off states of the various units while optimizing their produced power within the allowable margins. Solving the UC problem proves to be reliable for a problem involving identical production units, which is not the case for dynamic programming theory [46]. Figure 8 displays that the operational state planning based on fuzzy theory is promising, can be remarkable in the temporal evolution of the power produced by the most powerful machine (615 MVA), and suggests its resolution efficiency, especially when confronted with systems that are not difficult to model. This is only guaranteed with great consideration of the limitations taken for the amount of charge produced by each generator per hour and the permissible margins of the voltage levels for each node of the electrical network.

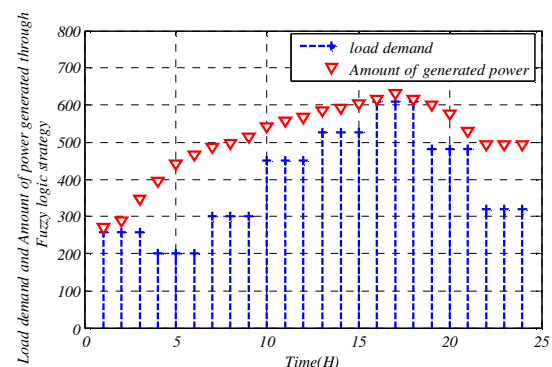


Fig. 9. Load demand and amount of power generated through the proposed strategy.

The improvement in CP of the proposed method depends on the number of fuzzy rules in the resolution. Optimization of CP by a genetic algorithm requires an adequate selection of parameters that vary from one system to another. The method using gradient-PSO [38] for the UC resolution suffers from convergence. In PSO, the swarm can prematurely converge, and the main cause of this problem is for the best global solution search. Furthermore, particles converge to a single point located between the best local and global solutions. This is not guaranteed for a local optimum. Another reason for this problem is the rapid rate of information transmitted between particles, leading to the creation of similar particles, and this results in a loss of diversity, which increases the possibility of falling into local optima. Therefore, it is difficult to reduce the calculation time and the CP for both of the methods cited. The proposed strategy proves to be the most promising since it has resulted in a better combination of the operating states of the production units, leading at an optimal cost and having a very competitive convergence time.

When integrating fuzzy logic into the UC problem, its resolution can be affected by several factors:

- Computational complexity: Fuzzy logic models can introduce additional computational complexity to the UC problem. The resolution of the problem can be affected by the efficiency of the algorithms used to solve the optimization model evoked by the objective function (8).
- Data quality: The quality of input data, including the amount of the consumed power at the h^{th} hour P_{dh} and unit parameters of each production unit, plays a crucial role in solving the UC problem with fuzzy logic. Inaccurate or unreliable data can lead to suboptimal solutions.
- Controller design: The choice of the fuzzy sets and the membership (triangular or trapezoidal) functions for input and output variables can influence the performance of the fuzzy logic model, improve the resolution of the UC problem, and reduce the computing time.

An integration with other metaheuristic algorithms or traditional optimization methods can affect the resolution of the problem.

V. CONCLUSION

This paper outlined the resolution of the UC problem by employing an optimization strategy based on fuzzy logic. The simulation results underscored the reliability of the proposed approach, with an execution time of 7.34 s and notable convergence efficiency. Emphasizing the crucial role of judiciously chosen controller input variables, the strategy facilitates effective planning of unit on-off states while optimizing power generation. The proposed approach was based on a Lagrangian function, chosen as an objective function to determine LCG and IL , which were chosen as fuzzy input variables according to partial derivatives for the different injected powers P_{ih} at each bus. These parameters are essential to minimize the total CP . The findings underscore the effectiveness of the strategy, qualifying it to address UC challenges in complex systems with a variable number of units. Future studies could extend this work to encompass a larger

number of production units, focusing on data quality and intricate power system modeling. Additionally, ongoing efforts focus on refining CP reduction and minimizing computing time by integrating metaheuristic methods with fuzzy logic, aiming to further enhance the sought optimal solutions.

REFERENCES

- [1] D. N. Simopoulos, S. D. Kavatzia, and C. D. Vournas, "Unit commitment by an enhanced simulated annealing algorithm," *IEEE Transactions on Power Systems*, vol. 21, no. 1, pp. 68–76, Jan. 2006, <https://doi.org/10.1109/TPWRS.2005.860922>.
- [2] M. Premkumar *et al.*, "An efficient and reliable scheduling algorithm for unit commitment scheme in microgrid systems using enhanced mixed integer particle swarm optimizer considering uncertainties," *Energy Reports*, vol. 9, pp. 1029–1053, Dec. 2023, <https://doi.org/10.1016/j.egyrs.2022.12.024>.
- [3] A. Abu-Jasser and M. M. Hijjo, "Fuzzy-Logic-Based Approach to Solve the Unit-Commitment Problem," *International Journal of Engineering Research and Development*, vol. 2, no. 1, pp. 22–29, Jul. 2012.
- [4] T. A. A. Victoire and A. E. Jeyakumar, "Unit commitment by a tabu-search-based hybrid-optimisation technique," *IEEE Proceedings - Generation, Transmission and Distribution*, vol. 152, no. 4, pp. 563–574, Jul. 2005.
- [5] C. C. A. Rajan and M. R. Mohan, "An evolutionary programming-based tabu search method for solving the unit commitment problem," *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 577–585, Feb. 2004, <https://doi.org/10.1109/TPWRS.2003.821472>.
- [6] S. Kharbush and W. Wang, "All-Terminal Network Reliability Optimization in Fading Environment via Cross Entropy Method," in *2010 IEEE International Conference on Communications*, May 2010, pp. 1–5, <https://doi.org/10.1109/ICC.2010.5501918>.
- [7] I. Marouani, A. Boudjemline, T. Guesmi, and H. H. Abdallah, "A Modified Artificial Bee Colony for the Non-Smooth Dynamic Economic/Environmental Dispatch," *Engineering, Technology & Applied Science Research*, vol. 8, no. 5, pp. 3321–3328, Oct. 2018, <https://doi.org/10.48084/etasr.2098>.
- [8] S. Marrouchi and S. Chebbi, "New strategy based on fuzzy logic approach to solve the unit-commitment problem," presented at the International Conference on Control, Engineering Information Technology (CEIT' 14), Sousse, Tunisia, Mar. 2014.
- [9] C. Abbey, F. Katiraei, C. Brothers, L. Dignard-Bailey, and G. Joos, "Integration of distributed generation and wind energy in Canada," in *2006 IEEE Power Engineering Society General Meeting*, Jun. 2006, <https://doi.org/10.1109/PES.2006.1709430>.
- [10] X. Guan, P. B. Luh, H. Yan, and J. A. Amalfi, "An optimization-based method for unit commitment," *International Journal of Electrical Power & Energy Systems*, vol. 14, no. 1, pp. 9–17, Feb. 1992, [https://doi.org/10.1016/0142-0615\(92\)90003-R](https://doi.org/10.1016/0142-0615(92)90003-R).
- [11] Z. Ouyang and S. M. Shahidehpour, "An intelligent dynamic programming for unit commitment application," *IEEE Transactions on Power Systems*, vol. 6, no. 3, pp. 1203–1209, Aug. 1991, <https://doi.org/10.1109/59.119267>.
- [12] A. Ahmad, "Unit Commitment Using Hybrid Approaches," Ph.D. dissertation, University of Engineering and Technology Taxila, Pakistan, 2010.
- [13] S. Marrouchi, N. Amor, M. B. Hessine, and S. Chebbi, "Theoretical Investigation of Combined Use of PSO, Tabu Search and Lagrangian Relaxation methods to solve the Unit Commitment Problem," *Advances in Science, Technology and Engineering Systems Journal*, vol. 3, no. 1, pp. 357–365, Feb. 2018, <https://doi.org/10.25046/aj030144>.
- [14] J. M. Arroyo and A. J. Conejo, "Optimal response of a thermal unit to an electricity spot market," *IEEE Transactions on Power Systems*, vol. 15, no. 3, pp. 1098–1104, Aug. 2000, <https://doi.org/10.1109/59.871739>.
- [15] B. Li *et al.*, "Fast Solution Method for the Large-Scale Unit Commitment Problem with Long-Term Storage," *Chinese Journal of Electrical Engineering*, vol. 9, no. 3, pp. 39–49, Sep. 2023, <https://doi.org/10.23919/CJEE.2023.000033>.

- [16] A. Merlin and P. Sandrin, "A New Method for Unit Commitment at Electricite De France," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-102, no. 5, pp. 1218–1225, May 1983, <https://doi.org/10.1109/TPAS.1983.318063>.
- [17] C. L. Chen and S. C. Wang, "Branch-and-bound scheduling for thermal generating units," *IEEE Transactions on Energy Conversion*, vol. 8, no. 2, pp. 184–189, Jun. 1993, <https://doi.org/10.1109/60.222703>.
- [18] T. Senjyu, T. Miyagi, A. Y. Saber, N. Urasaki, and T. Funabashi, "Emerging solution of large-scale unit commitment problem by Stochastic Priority List," *Electric Power Systems Research*, vol. 76, no. 5, pp. 283–292, Mar. 2006, <https://doi.org/10.1016/j.epsr.2005.07.002>.
- [19] S. S. Kumar and V. Palanisamy, "A hybrid fuzzy dynamic programming approach to unit commitment," *Journal-Institution of Engineers India*, vol. 88, Mar. 2008.
- [20] A. Abu-Jasser, "Solving the Unit Commitment Problem Using Fuzzy Logic," *International Journal of Computer and Electrical Engineering*, vol. 3, no. 6, pp. 824–829, Dec. 2011, <https://doi.org/10.7763/IJCEE.2011.V3.427>.
- [21] K. Venkatesan and C. C. Asir Rajan, "A Simulated Annealing Method for Solving Multi-Area Unit Commitment Problem," in *2011 International Conference on Process Automation, Control and Computing*, Jul. 2011, pp. 1–7, <https://doi.org/10.1109/PACC.2011.6047365>.
- [22] S. Yin Wa Wong, "An enhanced simulated annealing approach to unit commitment," *International Journal of Electrical Power & Energy Systems*, vol. 20, no. 5, pp. 359–368, Jun. 1998, [https://doi.org/10.1016/S0142-0615\(97\)00062-8](https://doi.org/10.1016/S0142-0615(97)00062-8).
- [23] A. Torchani, A. Boudjemline, H. Gasmi, Y. Bouazzi, and T. Guesmi, "Dynamic Economic/Environmental Dispatch Problem Considering Prohibited Operating Zones," *Engineering, Technology & Applied Science Research*, vol. 9, no. 5, pp. 4586–4590, Oct. 2019, <https://doi.org/10.48084/etasr.2904>.
- [24] C. C. A. Rajan, M. R. Mohan, and K. Manivannan, "Refined simulated annealing method for solving unit commitment problem," in *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290)*, Honolulu, HI, USA, May 2002, vol. 1, pp. 333–338, vol.1, <https://doi.org/10.1109/IJCNN.2002.1005493>.
- [25] M. Sudhakaran and P. Raj, "Integrating genetic algorithms and tabu search for unit commitment problem," *International Journal of Engineering, Science and Technology*, vol. 2, no. 1, pp. 57–69, 2010, <https://doi.org/10.4314/ijest.v2i1.59085>.
- [26] G. VenkataSubba Reddy, V. Ganesh, and C. SrinivasaRao, "Cost Reduction in Clustering Based Unit Commitment Employing Hybrid Genetic-Simulated Annealing Technique," *Journal of Electrical Engineering & Technology*, vol. 14, no. 1, pp. 27–35, Jan. 2019, <https://doi.org/10.1007/s42835-018-00038-0>.
- [27] C. C. Asir Rajan, "Hydro-thermal unit commitment problem using simulated annealing embedded evolutionary programming approach," *International Journal of Electrical Power & Energy Systems*, vol. 33, no. 4, pp. 939–946, May 2011, <https://doi.org/10.1016/j.ijepes.2011.01.002>.
- [28] L. Montero, A. Bello, and J. Reneses, "A Review on the Unit Commitment Problem: Approaches, Techniques, and Resolution Methods," *Energies*, vol. 15, no. 4, Jan. 2022, Art. no. 1296, <https://doi.org/10.3390/en15041296>.
- [29] C. P. Cheng, C. W. Liu, and C. C. Liu, "Unit commitment by annealing-genetic algorithm," *International Journal of Electrical Power & Energy Systems*, vol. 24, no. 2, pp. 149–158, Feb. 2002, [https://doi.org/10.1016/S0142-0615\(01\)00024-2](https://doi.org/10.1016/S0142-0615(01)00024-2).
- [30] L. Abdelhakem-Koridak and M. Rahli, "Optimisation d'un dispatching environnement/économique de la production d'énergie électrique par un algorithme génétique," presented at the Quatrième Conférence Internationale sur le Génie Electrique CIGE'10, Nov. 2010.
- [31] T. T. Maifeld and G. B. Sheble, "Genetic-based unit commitment algorithm," *IEEE Transactions on Power Systems*, vol. 11, no. 3, pp. 1359–1370, Aug. 1996, <https://doi.org/10.1109/59.536120>.
- [32] A. Rudolf and R. Bayrleithner, "A genetic algorithm for solving the unit commitment problem of a hydro-thermal power system," *IEEE Transactions on Power Systems*, vol. 14, no. 4, pp. 1460–1468, Nov. 1999, <https://doi.org/10.1109/59.801929>.
- [33] F. T. Lin, C. Y. Kao, and C. C. Hsu, "Applying the genetic approach to simulated annealing in solving some NP-hard problems," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 6, pp. 1752–1767, Nov. 1993, <https://doi.org/10.1109/21.257766>.
- [34] N. P. Padhy, "Unit commitment using hybrid models: a comparative study for dynamic programming, expert system, fuzzy system and genetic algorithms," *International Journal of Electrical Power & Energy Systems*, vol. 23, no. 8, pp. 827–836, Nov. 2001, [https://doi.org/10.1016/S0142-0615\(00\)00090-9](https://doi.org/10.1016/S0142-0615(00)00090-9).
- [35] S. Marrouchi, M. Ben Hessine, and S. Chebbi, "New strategy based on combined use of Particle Swarm Optimization and Gradient methods to solve the unit commitment problem," in *2015 IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC)*, Rome, Italy, Jun. 2015, pp. 1362–1367, <https://doi.org/10.1109/EEEIC.2015.7165368>.
- [36] K. Alqunun, "Optimal Unit Commitment Problem Considering Stochastic Wind Energy Penetration," *Engineering, Technology & Applied Science Research*, vol. 10, no. 5, pp. 6316–6322, Oct. 2020, <https://doi.org/10.48084/etasr.3795>.
- [37] S. Marrouchi and S. Ben Saber, "A Comparative Study of Fuzzy Logic, Genetic Algorithm, and Gradient-Genetic Algorithm Optimization Methods for Solving the Unit Commitment Problem," *Mathematical Problems in Engineering*, vol. 2014, Jul. 2014, Art. no. e708275, <https://doi.org/10.1155/2014/708275>.
- [38] S. Marrouchi and S. Chebbi, "Unit Commitment Optimization Using Gradient-Genetic Algorithm and Fuzzy Logic Approaches," in *Complex System Modelling and Control Through Intelligent Soft Computations*, Q. Zhu and A. T. Azar, Eds. Cham, Switzerland: Springer International Publishing, 2015, pp. 687–710.
- [39] A. H. Mantawy and Y. L. Abdel-Magid, "A New Fuzzy Unit Commitment Model and Solution," presented at the 14th PSCC, Seville, Spain, 2002.
- [40] A. C. G. C. Rajan, "Neural Based Tabu Search method for solving unit commitment problem with cooling-banking constraints," *Serbian Journal of Electrical Engineering*, vol. 6, no. 1, pp. 57–74, 2009.
- [41] H. Q. Truong and C. Jeenanunta, "Fuzzy mixed integer linear programming model for national level monthly unit commitment under price-based uncertainty: A case study in Thailand," *Electric Power Systems Research*, vol. 209, Aug. 2022, Art. no. 107963, <https://doi.org/10.1016/j.epsr.2022.107963>.
- [42] P. S. Kundur, "Power System Stability," in *Power System Stability and Control*, 3rd ed., CRC Press, 2012.
- [43] P. Wei, N. H. Li, H. Wu, Y. Zhang, X. Wang, and B. Zhu, "Daily generation scheduling based on genetic algorithm," *Automation of Electric Power Systems*, vol. 23, no. 10, pp. 23–27, 1999.
- [44] C. Cai and Y. Cai, "Optimization of unit commitment by genetic algorithm," *Power System Technology*, pp. 44–47, 1997.
- [45] K. Chandrasekaran, S. Hemamalini, S. P. Simon, and N. P. Padhy, "Thermal unit commitment using binary/real coded artificial bee colony algorithm," *Electric Power Systems Research*, vol. 84, no. 1, pp. 109–119, Mar. 2012, <https://doi.org/10.1016/j.epsr.2011.09.022>.
- [46] S. Dekrajangpetch, G. B. Sheble, and A. Conejo, "Auction implementation problems using Lagrangian relaxation," *IEEE Transactions on Power Systems*, vol. 14, no. 1, pp. 82–88, Feb. 1999, <https://doi.org/10.1109/59.744488>.

Influence of Crumb Rubber and Recycled Steel Fibers on Hybrid Layered Reinforced Concrete Columns under Compression Load

Hasan A. Alasmari

Civil Engineering Department, Faculty of Engineering, Taif University, Saudi Arabia
h.alasmari@tu.edu.sa (corresponding author)

Received: 7 January 2024 | Revised: 2 February 2024 | Accepted: 4 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6876>

ABSTRACT

Waste tires pose an environmental issue that causes health problems when discarded by either land burial or burning. The current study investigated the properties and characteristics of different-length hybrid layered columns of Recycled Steel Fibers (RSF) at a fixed ratio (0.6%) of the volume fraction with utilized fixed content (15%) of Crumb Rubber (CR). Nine square column specimens were prepared and tested under axial compression load to reveal the effect of RSF and CR content on hybrid layered reinforcement columns. The results revealed that as the RSF content increased concrete's properties were enhanced. However, the inclusion of CR at the top layer resulted in performance reduction. Additionally, the layered structure with CR and RSF has higher characteristic properties, including higher load capacity and displacement. Moreover, adding 0.6% RSF to both layers and CR at the top led to a 120% increase in the toughness of the concrete loading capacity. This was followed by a reasonable improvement in displacement and ductility.

Keywords-RSF; CR; layered structure; fixed ratio; axial compression

I. INTRODUCTION

Massive numbers of old and damaged tires are typically burned, buried, or disposed of in landfills causing numerous complicated environmental problems [1]. Therefore, many environmental agents have been striving for environmentally sustainable solutions to important issues such as the disposal of rubber from scrap tires. The latter are becoming a major cause of health problems (mosquito-borne diseases, air pollution, etc.). Insufficient dumping areas and hesitancy in using scrap tires have led to huge excesses of scrap tires in many countries. Authors in [1] note that the recycling of waste tires has received growing attention, with engineers trying to find new recycling approaches. Conversely, there is an increasing demand for concrete, which consumes Natural Aggregates (NA) [2]. Consequently, researchers were intrigued by the prospect of reusing old tire components like Crumb Rubber (CR) and Recycled Steel Fibers (RSF) in a variety of disciplines to reduce their environmental impact and natural resource consumption [3]. It has been shown that utilizing steel fiber in concrete leads to an increase in concrete strength, ductility, and toughness. Authors in [4] investigated the mechanical properties caused by the incorporation of waste metallic fibers of different lengths in concrete and reported that this leads to the best load-carrying capacity with improved concrete properties. Authors in [5] experimented with different volume fractions of RSF to assess its impact on compressive, tensile, and flexural strength. The results demonstrated that the highest compressive strength was acquired for 0.4% volume

fraction, whereas the lowest strength was acquired for 0.6%. In contrast, incorporating CR in concrete led to a negative impact, as it reduced the concrete compressive strength [6]. The similarities in the mechanical properties of concrete with the same impact were found in [7]. According to [8], there is a major problem when using CR in the mixtures: it tends to segregate as the rubber absorbs water, causing increased air spaces, which leads to a fragile matrix by increasing stress concentration. The density of CR is lower compared to NA, with a weak bond between the cement paste and CR [9]. In contrast, there are positive effects of incorporating CR in concrete, notably enhancing earthquake resistance by dissipating the impact through energy absorption [8]. Another positive effect is that rubber delays crack initiation at the tension zone [10]. Several studies have employed different techniques and trials to reduce the CR effect on concrete properties [11-13], including partial replacement of fine or coarse aggregates. Most studies suggest an ideal replacement of CR between 10% and 15% of fine aggregates to achieve a homogenous interaction with the rest of the concrete composite and thus obtain the best results [11-14]. To minimize the CR effects in concrete structure, a layered structure can minimize the impact [15-16]. Some layered structures can be named as hybrid concrete with different materials inserted in each layer controlling the reduction of load capacity due to the concentration of waste material. Authors in [17] conducted comparative studies of the rubberized and hybrid rubberized concrete with double-layer reinforced concrete beams. At the top layer rubberized reinforcement concrete was placed with

sand replacement of 10%, 12.5%, and 15% by CR, while at the bottom normal reinforcement concrete was used. The beam characteristics were evaluated and it was discovered that the performance of the hybrid beams improved in ultimate load, stiffness, and failure pattern. Also, it was confirmed that these applications can be used in sustainable structures. In this regard, authors in [16] investigated the isolated hybrid structure columns with two layers of up to 30% depth in their sections. The normal concrete was on top, and the scrap rubber was at the bottom. This concrete was tested under resonant excitation. The results showed that these kinds of structures can reduce the structure's response when exposed to load. Authors in [18] considered 5%, 10%, and 20% CR as a partial replacement for sand in three kinds of $50 \times 100 \times 400$ mm³ prism beams, namely plain concrete, rubberized concrete, and hybrid rubberized concrete with a double layer (rubberized concrete top and plain concrete bottom) under three-point bending static load. The results disclosed that the peak value of bending load (N) of hybrid rubberized concrete was higher than that of rubberized concrete beams and lower than that of plain concrete. An extensive literature review was conducted; some tests were on RSF and rubberized concrete on different scale specimens. On the contrary, there were limited studies that have investigated the characteristics of layered columns with recycled waste materials (RSF and CR).

In this study, nine columns with RSF and CR in two layers were examined. One layer had rubberized concrete with 15% CR replacing fine aggregates, while the other layer had RSF at 0.6% volume fraction, with different lengths, ranging from 2 cm to 6 cm. The layered concrete columns were subjected to axial compression load until failure. To explore the benefit of using a hybrid approach with CR and mitigate their negative effects, load capacity, displacement, ductility, and failure modes were determined and compared. This investigation will help reduce the cost of structures, ensure the sustainability of buildings through the repurposing of materials, and raise confidence in these applications.

II. MATERIALS AND METHODS

This research investigated the effects of using waste tires having 15% CR as partial replacement of fine aggregates with a fixed ratio of RSF (0.6%) in a hybrid layered structure for columns, cubes, and cylinders. All the specimens, i.e. the control concrete, fibrous rubberized concrete, layered rubberized concrete, fibrous concrete, and layered fibrous rubberized and fibrous concrete were subjected to axial compression tests.

A. Materials

RSF 0.6% (Figure 1(a)) and CR (Figure 1(b)) were obtained from a local company in Saudi Arabia (Kenzy Rubber Factory). The mix diameter of the CR was 4.75 mm, the computed specific gravity and water absorption for CR were 1.26 and 1.03, respectively, whereas the fineness modulus was 4.5. The RSF was extracted from Recycled Steel Bead Wires (RSBW) and consisted of straight fibers with lengths ranging from 30 mm to 60 mm and diameters ranging from 0.2 mm to 0.3 mm, with specific gravity of 2.08 and strength of 2140 MPa, according to the supplier data sheet. Natural sand with a

fineness modulus of 3.0 and crushed sand with a Maximum Nominal Size (MNS) of 12.5" were utilized as fine and coarse aggregates, respectively. For improving the workability of the concrete mixture, superplasticizer (SP), with a specific gravity of 1.06, conplast SP430 was implemented. With an initial settling period of more than 45 min, all prepared concrete mixes were cast using Ordinary Portland Concrete (OPC) Type 1. The main longitudinal steel bars had a diameter of 10 mm, while the stirrups were 8 mm. The properties of the bars are listed in Table I.



Fig. 1. (a) RSF, (b) CR.

TABLE I. PROPERTIES OF STEEL BARS

Diameter	Yield strength (MPa)	Tensile strength (MPa)	Tensile modulus (GPa)
8	420	580	207
10	425	610	205

B. Column Preparation and Details

The nine fabrication square column specimens utilized in this investigation were intended to have at least a 1% longitudinal reinforcement ratio. The columns have a cross-sectional size of 200 mm and a height of 500 mm. As shown in Figure 2, they are reinforced with longitudinal steel bars (4Ø10 mm) and link bars (6Ø8 mm), placed 75 mm apart.

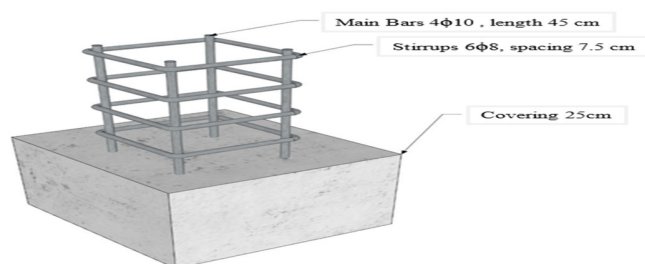


Fig. 2. Reinforcement column 3D sketch details.

C. Mix Design

The columns in Table II are classified into three categories based on RSF and CR with 0.6% and 15% content, respectively. The controlling mixtures (RC) were adopted from previous works with a targeted compressive strength of 35 MPa as displayed in Table II [17]. The RC mix properties are 3.15 relative density, 3 fineness modulus, 1% absorption sand, typical OPC (Type 1), and naturally crushed coarse aggregates with a maximum size of 14 mm were utilized. Three mixtures were prepared and tested, and the RC mixtures were prepared

without any inclusion of CR or RSF. The H1 and H2 were cast into two equal layers as depicted in Table II. The top layer of H1 was rubberized concrete with 15% CR as a fine replacement, whereas the bottom layer was normal concrete. H2 consisted of rubberized fibrous concrete at the top layer,

while the bottom layer contained fibrous concrete at a fixed ratio. Nine square columns, along with 100 mm - edge cubes and cylinders, were cured in water for 28 days and then kept at room temperature until tested.

TABLE II. MIXTURE PROPORTIONS FOR 1 m³ OF CONTROLLED AND HYBRID LAYER

Mix ID	CR (%)		W/C (-)	Fine Aggregates (kg)		Coarse Aggregates (kg)	RSF (%)		Water (kg)	Cement (kg)	SP430 (kg)	
RC	0		0.48	713		1060	-		172.8	360	3.6	
Mix ID	Top layer	Bottom layer	W/C (-)	Top layer	Bottom layer	Coarse Aggregates (kg)	Top layer	Bottom layer	Water (kg)	Cement (kg)	Top layer	Bottom layer
H1	15	-	0.48	668	713	1060	-	0.6	172.8	360	2.1	2.7
H2	15	-	0.48	668	713	1060	0.6	0.6	172.8	360	4.8	2.7

III. TEST SETUP AND INSTRUMENTATION

A hydraulic Universal Testing Machine (UTM) with a capacity of 200 tons was used to apply compression stress to the columns, as portrayed in Figure 3. Two steel plates were mounted on the upper and lower surfaces of the columns to distribute the axial force evenly throughout the column surfaces. Before testing, the column's top surface was polished with a Gipsom layer (Figure 4). The nine columns were tested under concentric axial compression load until they attained actual axial capacities with failure states at a displacement rate of 0.5 mm/min and until the appearance of hairline cracks. The average of the three-column specimens was considered. The characteristics for all of the columns, such as the axial load and vertical displacement data, were recorded using a computerized system linked to the machine. Additionally, the vertical load-displacement curve was plotted, allowing the three important mechanical properties of the column, i.e. ultimate load, displacement, and ductility, to be determined. Throughout the experiment, the pattern of the cracks and the failure modes were determined.



Fig. 3. Testing of concrete columns by UTM load compression.



Fig. 4. Pushing gypsum with a flat glass square.

IV. TEST RESULTS

A. Concrete Properties

Table III summarizes the properties of concrete for each mix, including density, compressive, static modulus, and split tensile strength. In the hybrid structure layer (H1 and H2), there was a decrease in the density by 3.55% and 3.35%, respectively, when the fibrous rubberized concrete was placed on the top and the fibrous or normal concrete at the bottom, compared with RC of 2417.67 kg/m³. This occurred because the relative density of sand was double (2.65) the one of CR (1.22) [17]. The compressive strength was obtained by testing three specimens from each mixture after 28 days and calculating the average strength. Regarding H1, the hybrid layered cubical structure containing rubberized concrete on the top and RSF concrete at the bottom, a slight decrease was observed in compression strength (roughly around 2.99%), whereas the changes in the arrangement (swap up) in H2 led to a slight increase of 1.12% compared to RC. This could occur because when at the top, CR with RSF, could absorb more load, causing delayed cracking, whereas when at the bottom, the delayed cracking could be controlled by the addition of RSF in the same ratio. In terms of splitting tensile strength, the average of three cylindrical concrete specimens was obtained. The splitting tensile strength results followed the reductions of compressive strength. For H1, it decreased by 7.08% when the rubber content was placed at the top and fibrous content was placed at the bottom. Conversely, when rubber content was placed at top and fibrous at the bottom (H2), splitting tensile strength increased by 16.14%.

B. Column Characteristics Results

Table IV illustrates the column axial characteristics derived by testing the axial load of the columns of all mixtures. Each type of mixture abbreviation represents the average of three specimens. The characteristics considered were first crack load (P_{cr}), cracking displacement (Δ_{cr}), ultimate load (P_u), yield load (P_{yield}), yield displacement (Δ_{yield}), ultimate displacement (Δ_u), total energy (E_{total}), secant stiffness (K), and ductility index (μ). Some characteristics were obtained by naked-eye monitoring during the axial test as well as through the area enclosed by the displacement curve, and others were acquired using calculations from (1) and (2).

$$\mu = \frac{\Delta_u}{\Delta_{yield}} \quad (1)$$

$$K = \frac{P_u}{\Delta u} \quad (2)$$

TABLE III. CONCRETE PROPERTIES

Mix. ID	Sp.No.	dcu (kg/m ³)	μdcu (%)	fcu (MPa)	μcu (%)	ftu (MPa)	μtu (%)
RC	1	2405	0	48	0	2.01	0
	2	2437		50		3.35	
	3	2411		47		3.25	
	Average	2417		48		2.87	
H1	1	2317	96	45	96	2.2	92
	2	2332		47		3.4	
	3	2346		49		2.4	
	Average	2331		47		2.67	
H2	1	2327	96	47	101	3.4	116
	2	2352		53		4.1	
	3	2329		46		2.5	
	Average	2336		49		3.33	

dcu = cube density, fcu = compressive strength, ftu = tensile strength, μdcu = average ratio of density, μcu = average ratio of compressive strength, μtu = average ratio of tensile strength

The toughness (total energy) for each mixture was obtained by measuring and adding all the area enclosed by the displacement curve of the reinforced concrete during the axial test until failure. There was an increase in the deformability area for H2, which shows an augmentation in toughness, rising to more than 120% of RC's. The stiffness values declined with the addition of rubberized concrete. On average, H1 and H2 columns had 58.67% and 6.93% reductions, respectively,

compared to RC, as illustrated in Table IV. The reduction in stiffness increases deformability and enhances the ability of the beams to absorb more energy by making the beams more ductile when exposed to high load [20, 23]. The ductility index (μ) in general increased with the incorporation of CR. Therefore, the ductility of H1 and H2 columns significantly increased to 1.16 and 1.13, compared to control RC samples (1.09), representing a raise of 6.42%, and 3.66%, accordingly. The highest values of the ultimate load (P_u) were observed in H2, which included RSF in both layers.

B. Cracking Behavior and Failure Modes

Table V depicts the characteristics of the cracking patterns (number, angles, width, mix, and minimum spacing) for all the tested columns. In H1, the addition of 15% CR at the top layer led to an average increase of 240% in the number of cracks compared to RC, because the modulus of the rubber aggregate's elasticity is lower than that of the fine aggregates, causing an increase in cracking [15-20]. Conversely, as RSF was included in both layers, the cracking amount in H2 was 15.15% lower than that of H1 columns. In the same line, the widths appeared to be smaller when RSF was used, whether in one or two layers, reduced by 12.5% and 37.5% with wider spacing of about 700% and 877% between cracks, for H1 and H2, respectively, when compared to RC. The presence of fibers helped the cracks bridge, delayed the crack growth, and minimized the effect of rubber inclusion in concrete [19].

TABLE IV. COLUMN CHARACTERISTICS RESULTS

Mix ID	Sp.No.	Pcr (KN)	Δcr (mm)	Pyield (KN)	Δyield (mm)	Pu (KN)	Δu (mm)	Ettotal (KN.mm)	K (KN/mm)	μ
RC	1.00	111.00	3.40	1110.00	3.50	1182.00	3.84	1846	307.81	1.10
	2.00	390.00	2.45	560.00	3.70	600.40	4.40	1229	136.45	1.19
	3.00	550.00	2.55	811.00	3.80	890.50	3.79	1640	234.96	1.00
	Average	350.33	2.80	827.00	3.67	890.97	4.01	1571	226.41	1.09
H1	1.00	198.00	1.95	400.00	4.20	453.00	5.00	1185	90.60	1.19
	2.00	245.00	2.37	360.00	3.80	415.00	4.30	983	96.51	1.13
	3.00	170.00	2.10	630.00	4.80	658.00	5.50	1185	119.64	1.15
	Average	204.33	2.14	463.33	4.27	508.67	4.93	1118	93.56	1.16
H2	1.00	615.00	2.75	810.00	3.56	853.10	3.45	1504	247.28	0.97
	2.00	465.00	4.21	1400.00	5.36	1457.90	7.60	7149	191.83	1.42
	3.00	550.00	3.50	810.00	4.50	876.40	4.54	1734	193.04	1.01
	Average	543.33	3.49	1006.67	4.47	1062.47	5.20	3462	210.71	1.13

TABLE V. COLUMN CRACKING PATTERNS

Mix	Specimen no.	No. of spiral	Max crack width (mm)	Angle of crack at failure (deg)	Minimum spacing (mm)	Maximum spacing (mm)	Failure modes
RC	1	2	6	80	1	7	CC
	2	9	5	60	1	35	
	3	6	5	65	1	12	
	Average	5.67	5.33	68.33	1	18	
H1	1	19	4	55	1	125	CC, SB
	2	18	4	55	1	138	
	3	21	6	65	1	175	
	Average	19.33	4.66	58.33	1	146	
H2	1	16	4	65	1	350	CC
	2	17	3	55	1	55	
	3	16	3	45	1	124	
	Average	16.33	3.33	55	1	176.33	

CC= concrete crushing, SB= steel buckling

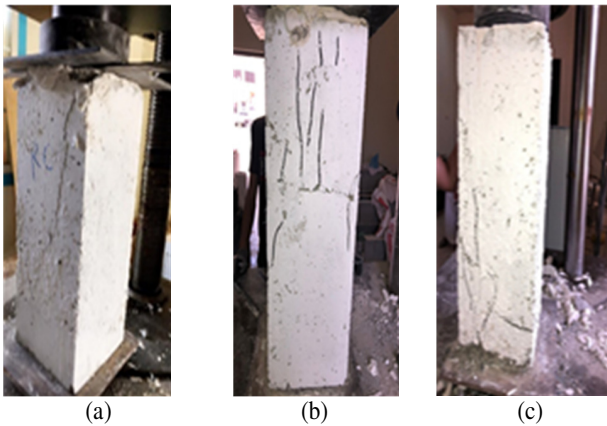


Fig. 5. Failure modes: (a) RC, (b) H1, (c) H2.

In terms of failure mode, the specimens cast with normal concrete (RC) were crushed (CC) (Figure 5(a)). In general, the failure modes of columns with crumb rubber showed Concrete Crushing (CC), followed by Steel Buckling (SB) (Figure 5(b)). Conversely, as RSF was included at 0.6%, for both layers, the failure mode was CC only and enhanced due to the increased load capacity of the specimens.

C. Load-displacement Behaviors

Figure 6 presents the curves of the load-displacement response of the reinforced tested columns with and without 15% CR and 0.6% RSF. It was noticed that the slope started increasing by the linear segment before the first cracks occurred when the displacement was recorded and then the nonlinear slope started gradually decreasing. The linear segment was initiated after the cracking and led to a gradual decrease in the slope, coupled with an increase in the number of cracks until failure. In [15, 18], it was reported that the reason for this is that the steel bar began to yield until it exceeded the yielding point by an additional load, after which the slope of the curves crucially decreased to reach the ultimate load through the crushing concrete. In hybrid layered columns, H1 with rubberized concrete on the top and RSF at the bottom exhibited a higher displacement (22.94% average) compared to control (Figure 6). The concentration of CR in one layer close to the load cell limited the stiffness reduction and raised axial strength. Changing the structure of the hybrid layered columns in H2 by adding fiber concrete at the bottom layer and rubberized concrete with RSF on the top increased the displacement and raised the ultimate load. The potential reason for this could be that as RSF and CR were exposed to the load directly on the top, they started the absorbing the load, which might have led to an increased displacement. Meanwhile, at the bottom, RSF helped control the cracks [21, 22].

V. CONCLUSION

In this paper, the mechanical properties and axial behavior of the nine specimens, including control and layered structure columns, were investigated for axial compression loads, and the following conclusions were drawn:

- In comparison with RC, the mechanical properties of H1 slightly decreased by around 2.99% and 7.08% in

compression tensile strength. However, when steel fibers were utilized in both layers with CR inclusion at the top in H2, the compressive strength and the tensile strength values were higher by around 1.12% and 16.14%, respectively.

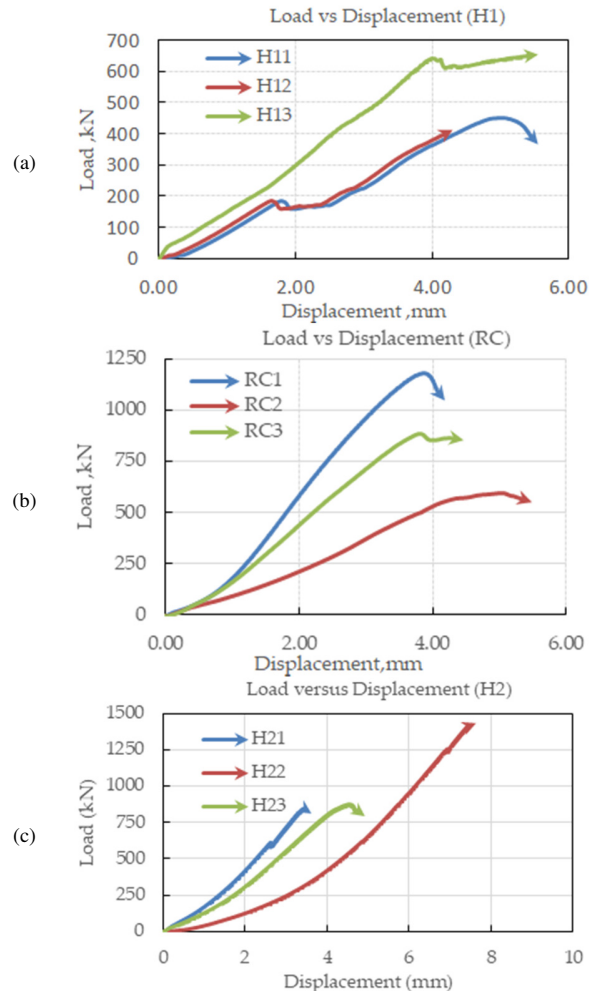


Fig. 6. Load-displacement curves of (a) RC, (b) H1, (c) H2 columns.

- Generally, numerous characteristics of the hybrid layered columns, including toughness, ultimate load and deflection, as well as total energy, seemed to be better than those of the non-layered concrete, especially in H2. The toughness in H2, rising to more than 120% of the control mixtures due to the rubber's ability to absorb the load led to increased values of ultimate load around by 19.32%.
- The maximum crack width of the spiral crack pattern improved with the inclusion of CR. Furthermore, the maximum spacing between spirals increased more than 800 times, which indicates that RSF in both layers of the columns (H2) was able to minimize the effect of CR.

ACKNOWLEDGMENT

The researchers would like to acknowledge the Deanship of Scientific Research, Taif University for funding this work.

REFERENCES

- [1] B. S. Thomas and R. C. Gupta, "A comprehensive review on the applications of waste tire rubber in cement concrete," *Renewable and Sustainable Energy Reviews*, vol. 54, pp. 1323–1333, Feb. 2016, <https://doi.org/10.1016/j.rser.2015.10.092>.
- [2] F. M. Z. Hossain, Md. Shahjalal, K. Islam, M. Tiznobaik, and M. S. Alam, "Mechanical properties of recycled aggregate concrete containing crumb rubber and polypropylene fiber," *Construction and Building Materials*, vol. 225, pp. 983–996, Nov. 2019, <https://doi.org/10.1016/j.conbuildmat.2019.07.245>.
- [3] J. Xu, Z. Yao, G. Yang, and Q. Han, "Research on crumb rubber concrete: From a multi-scale review," *Construction and Building Materials*, vol. 232, Jan. 2020, Art. no. 117282, <https://doi.org/10.1016/j.conbuildmat.2019.117282>.
- [4] M. S. Meddah and M. Bencheikh, "Properties of concrete reinforced with different kinds of industrial waste fibre materials," *Construction and Building Materials*, vol. 23, no. 10, pp. 3196–3205, Oct. 2009, <https://doi.org/10.1016/j.conbuildmat.2009.06.017>.
- [5] S. A. Rossli and I. S. Ibrahim, *Mechanical Properties of Recycled Steel Tire Fibres in Concrete*. University of Technology, 2012.
- [6] J. Xue and M. Shinozuka, "Rubberized concrete: A green structural material with enhanced energy-dissipation capability," *Construction and Building Materials*, vol. 42, pp. 196–204, May 2013, <https://doi.org/10.1016/j.conbuildmat.2013.01.005>.
- [7] K. Bisht and P. V. Ramana, "Waste to resource conversion of crumb rubber for production of sulphuric acid resistant concrete," *Construction and Building Materials*, vol. 194, pp. 276–286, Jan. 2019, <https://doi.org/10.1016/j.conbuildmat.2018.11.040>.
- [8] K. B. Najim and M. R. Hall, "A review of the fresh/hardened properties and applications for plain- (PRC) and self-compacting rubberised concrete (SCRC)," *Construction and Building Materials*, vol. 24, no. 11, pp. 2043–2051, Nov. 2010, <https://doi.org/10.1016/j.conbuildmat.2010.04.056>.
- [9] R. Alwi Assaggaf, S. Uthman Al-Dulaijan, M. Maslehuddin, O. S. Baghabra Al-Amoudi, S. Ahmad, and M. Ibrahim, "Effect of different treatments of crumb rubber on the durability characteristics of rubberized concrete," *Construction and Building Materials*, vol. 318, Feb. 2022, Art. no. 126030, <https://doi.org/10.1016/j.conbuildmat.2021.126030>.
- [10] A. Y. Kamal, "Ductility of simply supported rubberized concrete beams," *Challenge Journal of Concrete Research Letters*, vol. 12, no. 2, Jun. 2021, Art. no. 49, <https://doi.org/10.20528/cjcr.2021.02.002>.
- [11] R. H. Ghedan and D. K. Hamza, "Effect of Rubber Treatment on Compressive Strength and Thermal Conductivity of Modified Rubberized Concrete," *Journal of Engineering and Sustainable Development*, vol. 15, no. 4, pp. 21–29, Jan. 2011.
- [12] E. Güneyisi, "Fresh properties of self-compacting rubberized concrete incorporated with fly ash," *Materials and Structures*, vol. 43, no. 8, pp. 1037–1048, Oct. 2010, <https://doi.org/10.1617/s11527-009-9564-1>.
- [13] M. Elchalakani, "High strength rubberized concrete containing silica fume for the construction of sustainable road side barriers," *Structures*, vol. 1, pp. 20–38, Feb. 2015, <https://doi.org/10.1016/j.istruc.2014.06.001>.
- [14] A. M. Rashad, "A comprehensive overview about recycling rubber as fine aggregate replacement in traditional cementitious materials," *International Journal of Sustainable Built Environment*, vol. 5, no. 1, pp. 46–82, Jun. 2016, <https://doi.org/10.1016/j.ijsbe.2015.11.003>.
- [15] H. A. Alasmari, B. H. A. Bakar, and H. M. Akil, "Influence of rubberized-fibrous concrete on flexural behavior of hybrid reinforced beam," *AIP Conference Proceedings*, vol. 2213, no. 1, Mar. 2020, Art. no. 020284, <https://doi.org/10.1063/5.0000461>.
- [16] L. Li, S. Ruan, and L. Zeng, "Mechanical properties and constitutive equations of concrete containing a low volume of tire rubber particles," *Construction and Building Materials*, vol. 70, pp. 291–308, Nov. 2014, <https://doi.org/10.1016/j.conbuildmat.2014.07.105>.
- [17] H. A. Alasmari, B. H. A. Bakar, and A. T. Noaman, "A Comparative Study on the Flexural Behaviour of Rubberized and Hybrid Rubberized Reinforced Concrete Beams," *Civil Engineering Journal*, vol. 5, no. 5, pp. 1052–1067, May 2019, <https://doi.org/10.28991/cej-2019-03091311>.
- [18] M. M. Al-Tayeb, B. H. Abu Bakar, H. M. Akil, and H. Ismail, "Effect of Partial Replacements of Sand and Cement by Waste Rubber on the Fracture Characteristics of Concrete," *Polymer-Plastics Technology and Engineering*, vol. 51, no. 6, pp. 583–589, Apr. 2012, <https://doi.org/10.1080/03602559.2012.659307>.
- [19] A. T. Noaman, B. H. A. Bakar, and H. M. Akil, "Investigation on the mechanical properties of rubberized steel fiber concrete," *Engineering Structures and Technologies*, vol. 9, no. 2, pp. 79–92, Jun. 2017, <https://doi.org/10.3846/2029882X.2017.1309301>.
- [20] A. T. Noaman, B. H. A. Bakar, H. Md. Akil, and A. H. Alani, "Fracture characteristics of plain and steel fibre reinforced rubberized concrete," *Construction and Building Materials*, vol. 152, pp. 414–423, Oct. 2017, <https://doi.org/10.1016/j.conbuildmat.2017.06.127>.
- [21] B. H. A. Bakar, A. T. Noaman, and H. M. Akil, "Cumulative Effect of Crumb Rubber and Steel Fiber on the Flexural Toughness of Concrete," *Engineering, Technology & Applied Science Research*, vol. 7, no. 1, pp. 1345–1352, Feb. 2017, <https://doi.org/10.48084/etasr.854>.
- [22] N. A. Memon, M. A. Memon, N. A. Lakho, F. A. Memon, M. A. Keerio, and A. N. Memon, "A Review on Self Compacting Concrete with Cementitious Materials and Fibers," *Engineering, Technology & Applied Science Research*, vol. 8, no. 3, pp. 2969–2974, Jun. 2018, <https://doi.org/10.48084/etasr.2006>.
- [23] M. Badawi, A. G. Ahmed, T. A. Eldamaty, and M. M. Helal, "Properties of Recycled Concrete utilizing Waste Rubber," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11451–11458, Aug. 2023, <https://doi.org/10.48084/etasr.5918>.

Effect of Ground Granulated Blast Slag and Temperature Curing on the Strength of Fly Ash-based Geopolymer Concrete

Anil Kumar

Motihari College of Engineering, Motihari, India
anil.20688@gmail.com (corresponding author)

Rajkishor

Bhagalpur College of Engineering, Bhagalpur, India
rk.bce22@gmail.com

Niraj Kumar

Motihari College of Engineering, Motihari, India
nirajdsi10@gmail.com

Anil Kumar Chhotu

Motihari College of Engineering, Motihari, India
akcjucivil@gmail.com

Bhushan Kumar

Government Engineering College, Vaishali, India
bhushanmit08@gmail.com

Received: 6 January 2024 | Revised: 21 January 2024 | Accepted: 23 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6874>

ABSTRACT

Concrete is used most extensively after water to meet construction requirements. Since the population is increasing day by day, the demand for concrete will always increase, hence, the demand for cement will also increase. The production of cement requires a lot of energy and emits greenhouse gases into the environment. Therefore, an alternative material for cement concrete is required. Geopolymer concrete (GPC) is an alternative to cement made of aluminosilicate materials such as fly ash, Ground Granulated Blast Slag (GGBS), silica fume, metakaolin, etc. If these materials are activated with an alkaline activator, then a bond that is responsible for the strength develops. GPC made with fly ash needs temperature curing to develop its strength, which limits its use on a large scale. In this study, a mix ratio of GPC equivalent to conventional M20 concrete was obtained at ambient curing conditions. The effect of temperature curing was also studied. GPC was prepared in three different mixes. In each mix, the binder content was changed by varying the fly ash and GGBS content. Two sets of cube, beam, and cylindrical samples were prepared from each mixture. One set was cured at ambient temperatures and the other at increased temperatures. The temperature-cured specimens provided higher strength than the ambient-cured. If a strength equivalent to conventional M20 concrete is required for ambient curing, then the mix should be 70% fly ash and 30% GGBS, and the ratio of binder, fine aggregate, and coarse aggregate should be 1:1.5:3.

Keywords-ambient curing; compressive strength; GGBS; geopolymer concrete; fly ash; temperature curing

I. INTRODUCTION

The demand for a substitute for cement is increasing due to the environmental pollution associated with its production. Cement production is responsible for 5-9% of global CO₂ emissions [1-5]. Cement production requires raw materials,

such as limestone, which is a non-renewable resource. With the increasing development of infrastructure, concrete has become the second most consumed material after water [6]. At the same time, the world faces environmental issues due to the disposal of industrial waste such as fly ash, Ground Granulated Blast Slag (GGBS), silica fume, rice husk ash, etc. Disposed

industrial waste causes land degradation. Therefore, it is important to find an application for this industrial waste in the construction sector, as such an application will promote sustainable development. One of the alternate materials for cement concrete is geopolymer concrete (GPC), which uses industrial waste such as fly ash, GGBS, and silica fume as binding materials (aluminosilicate materials). GPC has good engineering properties [7-11]. GPC is prepared with fine aggregates, coarse aggregates, aluminosilicate materials, and alkaline activators. The best combination of alkaline activators is the combination of sodium hydroxide and sodium silicate solution [12]. The concentration of alkaline activators also affects the properties of GPC [13], along with curing time, curing temperature, and admixture content [14-17]. In the case of conventional concrete, the C-S-H gel is responsible for the formation of a bond. However, in the case of GPCs, polymerization is responsible for the formation of the bond. GPC has good performance even at elevated temperatures [18-20]. Since temperature-curing improves the strength of GPC [21-27] and GPC has less strength at ambient curing, this may be a drawback to its use on a large scale. Therefore, if GPC is to be used on a large scale, it is necessary to prepare it to have good strength even at ambient curing. If GGBS is used along with fly ash as a binding material, then the strength of GPC may be good even at ambient curing.

In this study, a mixture of GPC was obtained that had a strength similar to conventional M20 grade concrete even in ambient curing, using fly ash and GGBS as binding materials. Three different mixes were prepared using varying percentages of fly ash and GGBS. Cubical (150×150×150 mm), beam (100×100×500 mm), and cylindrical (150×300 mm) samples were prepared using the different mixtures. Half of the samples from each mix were provided ambient-curing and the other half were provided temperature-curing for comparison. After 28 days, compressive strength, flexural strength, and split tensile strength were evaluated for the samples prepared from each mix. The results showed that the mix with 70% fly ash and 30% GGBS had a strength similar to the conventional M20 grade concrete. As expected, the temperature-cured samples had higher strength compared to ambient-cured samples.

II. MATERIALS

A. Fly Ash

The combustion of pulverized coal in power plants produces a thin, powdery waste known as fly ash. Fly ash is made up of mineral granules that are expelled from the boiler by the hot gases. Electrostatic precipitators and cloth filter baghouses are used to capture and recycle these particles. Using fly ash instead of Portland cement, can reduce the environmental impact of building materials such as concrete and greenhouse gas emissions. Fly ash can be used to make bricks, blocks, and even road bases. For this study, fly ash was obtained from the Kanti thermal power plant, in Muzaffarpur, Bihar, India.

B. Ground Granulated Blast Slag (GGBS)

GGBS is a waste product of the blast furnace process, which involves melting iron ore, coke, and limestone into molten iron. Glassy GGBSs are created when the molten slag is

rapidly quenched with water. Most GGBS is made up of calcium-based silicates and aluminosilicates. Lime, alumina, and some other elements are present at trace levels. Physically distinct from Portland cement, it shares many of its chemical characteristics. The use of GGBS leads to improved durability and strength properties.

C. Aggregates and Plasticizer

Fine aggregates from naturally available sand satisfying the criteria of zone II were used. Coarse aggregates having a nominal size of 20 mm (60%) and 10 mm (40%) were used. Since GPC has low workability compared to conventional concrete [24], a plasticizer was used to improve workability.

D. Alkaline Activator

A mixture of sodium hydroxide (NaOH) solution and sodium silicate (Na_2SiO_3) solution was used as an alkaline activator. For the preparation of the alkaline activator, NaOH was mixed uniformly with water so that 480 g (12 M molarity) NaOH produced 1 l solution, as shown in Figure 1. The weight of sodium NaOH was taken and 2.5 times Na_2SiO_3 solution was added.



Fig. 1. Preparation of NaOH solution.

III. METHODS

A. Mixing

The ratio of binder, fine aggregates, and coarse aggregates was 1:1.5:3. Fly ash and GGBS were used as binders with varying percentages so that the total binder was the same in all mixes (Figure 2). The alkaline activator to binder ratio was 0.55. Table I shows the details of the mix. A sample ID was assigned for each mix. FA90G10 indicates that fly ash and GGBS were 90% and 10% of the total binder content, respectively. The plasticizer was 1% of the binder content in each mix. Figure 3 shows one of the prepared mix.

B. Casting and Curing

The GPC mix was transferred to cube, beam, and cylinder molds, as shown in Figure 4. From each mixture, 6 cube, 6 beam, and 6 cylinder samples were prepared. Manual and table vibrations were given to the samples to ensure proper compaction. The samples were left to cure for 24 hr. After 24 hr the samples were de-molded and half of each type was temperature-cured for 4 hr at 150 °C (Figure 5). Half of the other samples were left for ambient curing (Figure 6).



Fig. 2. Preparation of dry mix.



Fig. 3. GPC mix.



Fig. 4. Casted GPC specimens.



Fig. 5. Temperature curing of GPC specimens.



Fig. 6. Ambient curing of GPC specimens.

TABLE I. MIX DETAILS OF GPC SPECIMENS

Sample ID	NaOH (kg/m ³)	Na ₂ SiO ₃ (kg/m ³)	Binder (kg/m ³)		Fine aggregates (kg/m ³)	Coarse aggregates (kg/m ³)	
			FA	GGBS		20 mm	10 mm
FA90G10	85	212	485	54	808	970	646
FA80G20	85	212	431	108	808	970	646
FA70G30	85	212	377	162	808	970	646

C. Testing of GPC Specimens

After 28 days, the samples were subjected to tests to evaluate their compression, flexural, and split tensile strength. The compressive strength test was performed following IS 516:1959. Ambient-cured and temperature-cured cubes were placed in a digital testing machine so that normal stress would develop on the cube. The load was applied at 140 kg/cm²/min. As the cube sample failed, the failure stress of the samples was taken. The IS 516:2002 guidelines were followed to evaluate the flexural strength. The beams were placed and two-point loading was applied such that the rate of total load application was 180 kg/min. As the samples failed, the failure stresses were taken from the digital flexural testing machine. The IS 5816:1999 was followed to evaluate the split tensile strength of the samples. The cylindrical samples were placed horizontally so that the load would pass through their center. The rate of load application was kept at 2.4 N/mm²/min. As the sample failed, failure load was taken and split tensile strength was evaluated as:

$$\sigma = \frac{2P}{\pi DL} \quad (1)$$

where σ is the split tensile strength (Mpa), P is the failure load (N), D is the diameter of the cylinder (mm), and L is the length of cylinder (mm). Figure 7 shows the test setups.

IV. RESULTS AND DISCUSSION

Table II shows the test results of the GPC specimens tested after 28 days. Figures 8-10 show a similar trend in compressive strength, flexural strength, and split tensile strength mixes. As the percentage of GGBS increased, the strength of cube, beam, and cylinder specimens increased. Considering the curing

conditions, the specimens subjected to temperature curing showed higher strength than those subjected to ambient curing. For ambient cured specimens, the specimen having 70% fly ash and 30% GGBS exhibited a compressive strength similar to the conventional M20 concrete prepared with Ordinary Portland Concrete (OPC).

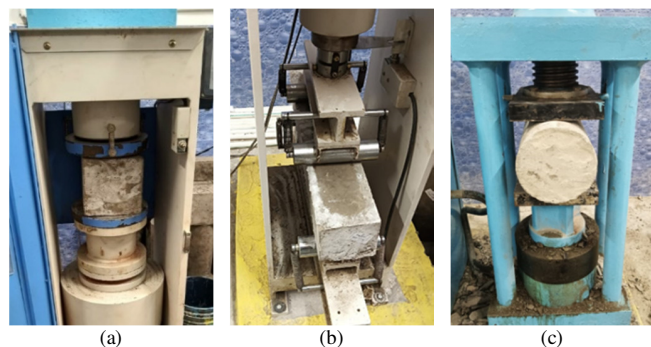


Fig. 7. Testing the GPC specimens: (a) Compressive strength test, (b) flexural strength test, (c) split tensile strength test.

TABLE II. STRENGTH OF GPC SPECIMENS AFTER 28 DAYS

Sample ID	Compressive strength (Mpa)		Flexural strength (Mpa)		Split tensile strength (Mpa)	
	TC	AC	TC	AC	TC	AC
FA90G10	22.04	17.78	5.85	5.34	2.36	2.10
FA80G20	33.18	20.37	6.36	6.03	3.30	2.26
FA70G30	35.77	24.74	6.97	6.11	3.44	2.69

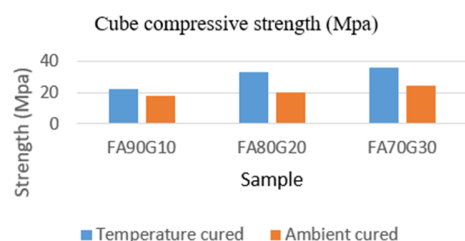


Fig. 8. Comparison of compressive strength of temperature-cured and ambient-cured GPC.

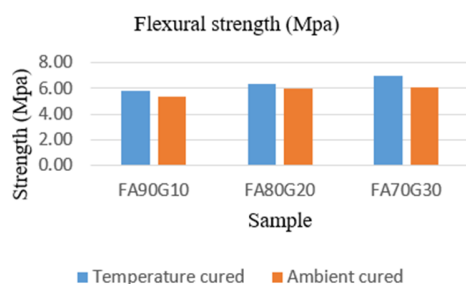


Fig. 9. Comparison of flexural strength of temperature-cured and ambient-cured GPC.

Compressive strength equivalent to conventional M20 concrete can be obtained at ambient temperature if using 70% fly ash and 30% GGBS and keeping the ratio of binder, fine aggregate, and coarse aggregate at 1:1.5:3, respectively. Split tensile strength was found similar to conventional concrete. However, the flexural tensile strength was found to be better

(50% higher) than conventional concrete. If temperature curing is not required for GPC, then it can be used on a large scale in the construction industry which will minimize this industrial waste disposal problem and, as a result, reduce environmental pollution. The high strength of GPC prepared under ambient conditions will promote its use and reduce the use of conventional concrete, which will ultimately minimize the use of cement and the emission of greenhouse gases from the cement industry.

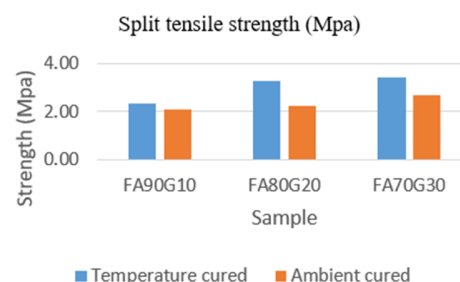


Fig. 10. Comparison of split tensile strength of temperature-cured and ambient-cured GPC.

V. CONCLUSION

Based on the experimental work presented, the following conclusions can be drawn:

- Ambient-cured GPC can give a good result if using 70% fly ash and 30% GGBS as a binder. A compressive strength comparable to the M20 grade of conventional concrete can be obtained by keeping the same ratio of nominal mix (1:1.5:3) as conventional concrete.
- If a higher strength of GPC is required at low GGBS content, then temperature-curing should be provided.
- The strength of GPC increases with increasing GGBS content as the binding material.
- The flexural strength of GPC is better (around 50% higher) than conventional concrete with similar compressive strength.
- Its split tensile strength is similar to that of conventional concrete.
- If GPC provides better results even at ambient curing, it will solve the fly ash disposal problem and reduce environmental pollution and greenhouse gas emissions caused by the cement industry.

CONFLICTS OF INTEREST

The authors declare that there is no conflict of interest regarding this study.

FUNDING STATEMENT AND ACKNOWLEDGEMENT

The authors declare that there was no funding for the publication of this paper. The authors acknowledge the Department of Civil Engineering, MCE, Motihari, for providing the lab facilities for the experimental work.

REFERENCES

- [1] N. Mahasen, S. Smith, and K. Humphreys, "The Cement Industry and Global Climate Change: Current and Potential Future Cement Industry CO₂ Emissions," in *Greenhouse Gas Control Technologies - 6th International Conference*, Kyoto, Japan, Jan. 2003, pp. 995–1000, <https://doi.org/10.1016/B978-008044276-1/50157-4>.
- [2] V. Khozin, O. Khokhryakov, and R. Nizamov, "A «carbon footprint» of low water demand cements and cement-based concrete," *IOP Conference Series: Materials Science and Engineering*, vol. 890, no. 1, Apr. 2020, Art. no. 012105, <https://doi.org/10.1088/1757-899X/890/1/012105>.
- [3] L. Black, "Low clinker cement as a sustainable construction material," in *Sustainability of Construction Materials (Second Edition)*, J. M. Khatib, Ed. Woodhead Publishing, 2016, pp. 415–457.
- [4] C. Lorea, "CO₂ emission from cement industry, what's the best estimate?," Dec. 2023. [Online]. Available: http://www.emccement.com/pdf/GCA_CO2_emission_from_cement_industry.pdf.
- [5] A. Talaie, D. Pier, A. V. Iyer, M. Ahiduzzaman, and A. Kumar, "Assessment of long-term energy efficiency improvement and greenhouse gas emissions mitigation options for the cement industry," *Energy*, vol. 170, pp. 1051–1066, Mar. 2019, <https://doi.org/10.1016/j.energy.2018.12.088>.
- [6] N. Makul, "Advanced smart concrete - A review of current progress, benefits and challenges," *Journal of Cleaner Production*, vol. 274, Nov. 2020, Art. no. 122899, <https://doi.org/10.1016/j.jclepro.2020.122899>.
- [7] A. C. Ganesh and D. M. Muthukannan, "A review of recent developments in geopolymers concrete," *International Journal of Engineering & Technology*, vol. 7, no. 4.5, Sep. 2018, Art. no. 696, <https://doi.org/10.14419/ijet.v7i4.5.25061>.
- [8] "Effect of Polypropylene fibers over GGBS based Geopolymer Concrete Under Ambient Curing," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 2S2, pp. 89–92, Dec. 2019, <https://doi.org/10.35940/ijitee.B1022.1292S219>.
- [9] C. Ganesh and M. Muthukannan, "Investigation on the glass fiber reinforced geopolymer concrete made of M-sand," *Journal of Materials and Engineering Structures*, vol. 6, no. 4, pp. 501–512, Dec. 2019.
- [10] A. C. Ganesh, M. Muthukannan, S. Aakash, Prasad, and B. Subramanian, "Energy efficient production of geopolymer bricks using industrial waste," *IOP Conference Series: Materials Science and Engineering*, vol. 872, no. 1, Mar. 2020, Art. no. 012154, <https://doi.org/10.1088/1757-899X/872/1/012154>.
- [11] A. C. Ganesh, K. Sowmiya, and M. Muthukannan, "Investigation on the effect of steel fibers in geopolymer concrete," *IOP Conference Series: Materials Science and Engineering*, vol. 872, no. 1, Mar. 2020, Art. no. 012156, <https://doi.org/10.1088/1757-899X/872/1/012156>.
- [12] C. L. Hwang and T. P. Huynh, "Effect of alkali-activator and rice husk ash content on strength development of fly ash and residual rice husk ash-based geopolymers," *Construction and Building Materials*, vol. 101, pp. 1–9, Dec. 2015, <https://doi.org/10.1016/j.conbuildmat.2015.10.025>.
- [13] M. M. A. Bakri, H. Mohammed, H. Kamarudin, I. K. Niza, and Y. Zarina, "Review on fly ash-based geopolymer concrete without Portland Cement," *Journal of Engineering and Technology Research*, vol. 3, no. 1, Jan. 2011.
- [14] P. R. Vora and U. V. Dave, "Parametric Studies on Compressive Strength of Geopolymer Concrete," *Procedia Engineering*, vol. 51, pp. 210–219, Jan. 2013, <https://doi.org/10.1016/j.proeng.2013.01.030>.
- [15] S. Saloma, H. Hanafiah, D. O. Elysandi, and D. G. Meykan, "Effect of Na₂SiO₃/NaOH on mechanical properties and microstructure of geopolymer mortar using fly ash and rice husk ash as precursor," *AIP Conference Proceedings*, vol. 1903, no. 1, Nov. 2017, Art. no. 050013, <https://doi.org/10.1063/1.5011552>.
- [16] M. N. S. Hadi, H. Zhang, and S. Parkinson, "Optimum mix design of geopolymer pastes and concretes cured in ambient condition based on compressive strength, setting time and workability," *Journal of Building Engineering*, vol. 23, pp. 301–313, May 2019, <https://doi.org/10.1016/j.jobe.2019.02.006>.
- [17] A. L. Han and J. J. Ekaputri, "The influence of molarity variations to the mechanical behavior of geopolymer concrete," *MATEC Web of Conferences*, vol. 195, 2018, Art. no. 01010, <https://doi.org/10.1051/mateconf/201819501010>.
- [18] A. Hassan, M. Arif, and M. Shariq, "Mechanical Behaviour and Microstructural Investigation of Geopolymer Concrete After Exposure to Elevated Temperatures," *Arabian Journal for Science and Engineering*, vol. 45, no. 5, pp. 3843–3861, May 2020, <https://doi.org/10.1007/s13369-019-04269-9>.
- [19] S. Li *et al.*, "Properties of concrete with waste glass after exposure to elevated temperatures," *Journal of Building Engineering*, vol. 57, Oct. 2022, Art. no. 104822, <https://doi.org/10.1016/j.jobe.2022.104822>.
- [20] S. Mane, "Investigation of geopolymer mortar and concrete under high temperature," *International Journal of Emerging Technology and Advanced Engineering*, vol. 2, no. 12, pp. 384–390, Dec. 2012.
- [21] M. B. Satpute, M. R. Wakchaure, and S. V. Patankar, "Effect of Duration and Temperature of Curing on Compressive Strength of Geopolymer Concrete," *International Journal of Engineering and Innovative Technology*, vol. 1, no. 5, pp. 152–155, May 2012.
- [22] M. R. Nagral, T. Ostwal, and M. V. Chitawadagi, "Effect Of Curing Temperature And Curing Hours On The Properties Of Geo-Polymer Concrete," *International Journal of Computational Engineering Research*, vol. 4, no. 9, Sep. 2014.
- [23] K. Vijai, R. Kumutha, and B. G. Vishnuram, "Effect of types of curing on strength of geopolymer concrete," *International Journal of the Physical Sciences*, vol. 5, no. 9, pp. 1419–1423, Aug. 2010.
- [24] R. M. Waqas, F. Butt, X. Zhu, T. Jiang, and R. F. Tufail, "A Comprehensive Study on the Factors Affecting the Workability and Mechanical Properties of Ambient Cured Fly Ash and Slag Based Geopolymer Concrete," *Applied Sciences*, vol. 11, no. 18, Jan. 2021, Art. no. 8722, <https://doi.org/10.3390/app11188722>.
- [25] N. Kumar and R. Jha, "GIS-based Flood Risk Mapping: The Case Study of Kosi River Basin, Bihar, India," *Engineering, Technology & Applied Science Research*, vol. 13, no. 1, pp. 9830–9836, Feb. 2023, <https://doi.org/10.48084/etasr.5377>.
- [26] R. Sangi, B. S. Sreenivas, and K. Shanker, "Mix Design of Fly Ash and GGBS based Geopolymer Concrete activated with Water Glass," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11884–11889, Oct. 2023, <https://doi.org/10.48084/etasr.6216>.
- [27] Z. A. Tunio, F. U. R. Abro, T. Ali, A. S. Buller, and M. A. Abbasi, "Influence of Coarse Aggregate Gradation on the Mechanical Properties of Concrete, Part I: No-Fines Concrete," *Engineering, Technology & Applied Science Research*, vol. 9, no. 5, pp. 4612–4615, Oct. 2019, <https://doi.org/10.48084/etasr.3046>.

Experimental Study of the Flame Retardancy of PMMA-Graphene Composite Materials

Jawdat Al-Jarrah

Fire and Safety Engineering Department, Prince Al-Hussein Bin Abdullah II Academy for Civil Protection, Al-Balqa Applied University, Jordan
jawdat.jarrah@bau.edu.jo

Diana Rbeht

Fire and Safety Engineering Department, Prince Al-Hussein Bin Abdullah II Academy for Civil Protection, Al-Balqa Applied University, Jordan
dianasr2004@yahoo.com

Mohammed S. El-Ali Al-Waqfi

Fire and Safety Engineering Department, Prince Al-Hussein Bin Abdullah II Academy for Civil Protection, Al-Balqa Applied University, Jordan
mohammad.waqfi@bau.edu.jo

Yarub Al-Jahmany

Fire and Safety Engineering Department, Prince Al-Hussein Bin Abdullah II Academy for Civil Protection, Al-Balqa Applied University, Jordan
jahmany@gmail.com

Received: 9 January 2024 | Revised: 31 January 2024 | Accepted: 6 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6883>

ABSTRACT

In this paper, Polymethyl methacrylate (PMMA)-graphene nano-composites were prepared and tested with the use of a cone calorimeter. Graphene was added to PMMA in limited weight percentages to improve the flame retardancy of PMMA. Two samples of PMMA-graphene, namely 1 and 3 wt%, were investigated. The combustion properties of the tested samples of PMMA-graphene composites, mass loss rate, heat release rate, and time to ignition were measured and calculated. It was found that the peak heat release rate of PMMA-graphene composites reduced by 17% when 3 wt% graphene was added to pure PMMA. Adding graphene to PMMA improves the thermal stability of PMMA by reducing the time of ignition. Also, the presence of graphene enhanced the formation of a continuous carbonized layer at the surface of the burned PMMA.

Keywords-nano-composites; graphene; PMMA; flame retardancy; polymers; heat release rate

I. INTRODUCTION

Polymers are promising materials that can be widely used in a variety of applications and low prices [1, 2]. Polymethyl methacrylate (PMMA) is an example of polymer which has been used in various applications [3]. The most desired properties of the PMMA are its low density, low thermal conductivity, low water absorption, high chemical corrosion resistance, and high flexibility, while its main disadvantages are its low melting temperature and decomposition to volatile combustible products when exposed to heat [4]. In some fields, where flame resistance is required, PMMA is used with additives to enhance its flame retardant properties [5]. The mechanical and thermal properties of PMMA can be enhanced

by adding reinforcements [5]. There are many additives used with PMMA to enhance its flame retardancy [6]. However, the choice of flame retardants as additives depends on the specific applications and degree of retardants required [7, 8]. Graphene is often added as a nano-filler of PMMA matrices to enhance flame retardancy [9], since the addition of graphene even at very low loadings up to 0.7 wt.% enhances the flame retardancy of materials [10]. PMMA-graphene as a nano-composite has a widely range of applications [11, 12]. Graphene has a high surface area to volume ratio, and is considered as the thinnest material to date [13, 14]. The presence of graphene as nano-sheets in polymer composites enhances flame retardancy by producing intumescent char during their combustion [15]. When polymer burns, graphene

forms a dense carbonized layer, which halts the burning of polymer matrices [16]. Also, incorporation of graphene reduces mass loss rate significantly by altering the diffusion path of pyrolysis products [10].

Flame retardants are used to reduce the fire impact and fire growth and spreading [8]. This study aims to reduce the heat release rate from the combustion of polymers like PMMA. Graphene is selected to be used as a nano filler to retard flame spreading during the combustion of PMMA.

II. PREPERATION OF SAMPLES

In this study, PMMA is used as a base material while graphene is used as nano-filler. Pure PMMA supplied by the SABIC factory was used as a row material. Figure 1 shows the pure form of used PMMA. Graphene is selected as nano-filler to PMMA matrices due to its thermal stability, Graphene will not be a combustion fuel, and can be found in the burned char.



Fig. 1. The used PMMA.

Brabender plastograph EC plus instrument was used to mix the graphene with the base material (PMMA) to get a uniformly homogenous mixture. A total of 30 g of PMMA with two different weight percentages (1 wt% and 3 wt%) of graphene were used as a reinforcement. Graphene and pure PMMA were added to the mixer and processed at 180 °C and mixing speed of 50 min⁻¹ to prepare the nano-composites. The high mixing temperature and the suitable torque leads to good dispersion and homogeneous distribution of graphene inside the PMMA matrix. Also, pure samples of 30 g weight were prepared with the Brabender plastograph EC plus mixer under the same parameters that used to prepare the PMMA-graphene nano-composites. After the samples got mixed by the Brabender plastograph EC plus mixer, they were put in a stainless steel mold and were pressed by a hot press machine. A stainless-steel mold with dimensions of 100×100×3 mm³ has been cut and polished to meet the requirements of the sample holder used in the cone calorimeter. The prepared sample was introduced to the sample holder and was tested inside the cone calorimeter which is calibrated according to ISO 5660. The flat square 100×100 mm² sample with 3 mm thickness was exposed to heat flux of 50 kW/m². The heat release rate was measured with the Clayton Huggett method which is in accordance with ISO 5660. Oxygen consumption was used to determine the heat release rate. For each sample, Peak Heat Release Rate (PHRR), Effective Heat Release Rate (EHRR), and Total Heat Release Rate (THRR), time to ignition t_{ig} , and time to reach the PHRR, t_{i-peak} , were measured and recorded. Heat release and mass of the sample were measured and recorded every 5 s.

III. RESULTS AND DISCUSSION

Many methods are used to enhance the flame retardancy of materials, one of these is by char formation [17]. Some materials like graphene and its derivatives when exposed to heat or flame form a char. This leads to the formation of a layer of carbonized material that can act as a barrier between the flame and the underlying material, preventing or slowing further combustion [18]. The addition of small load of graphite up to 5 wt% can significantly enhance the flame retardancy and mechanical properties [3].

Three samples of pure PMMA and PMMA-graphene composites were prepared and tested by the cone calorimeter. Figure 2 shows the prepared samples of pure PMMA and its composites.

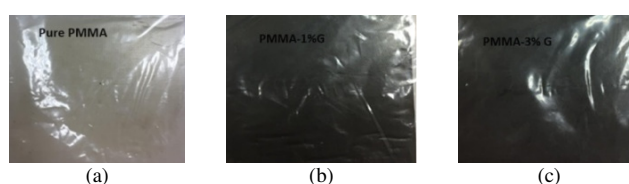


Fig. 2. The prepared samples: (a) Pure PMMA, (b) PMMA-1 % graphene, (c) PMMA-3% graphene.

Figure 3 shows the results of heat release rate from the pure samples. The results of pure PMMA and PMMA-graphene composites indicate the consistency of the tests. Also, it can be concluded that the samples resulted from Brabender plastograph EC plus instrument are homogeneous and uniformly mixed, where the heat release rate is approximately the same for the three samples taken from the mixed batch. The average measured values are shown in Table I. For PMMA-graphene composites, two weight percentages of graphene were considered (1 wt% and 3 wt%).

TABLE I. AVERAGE RESULTS OF THE CONE CALORIMETER TEST

Sample	PHRR (kW/m ²)	EHRR (kW/m ²)	THRR (MJ/m ²)	t_{ig} (s)	t_{i-peak} (s)
Pure PMMA	964	95.94	104.26	41	138.3
PMMA-graphene (1 wt%)	858.16	46.77	102.9	23	140
PMMA-graphene (3 wt%)	801.3	47.22	100.5	20.6	131.6

Figure 4 shows the average values of heat release rate for pure PMMA and PMMA-graphene composites from the cone calorimeter test. Pure PMMA has larger PHRR (964 kW/m²) than PMMA-graphene composites. However, the addition of 1 wt% of graphene to PMMA reduced the PHRR by 11%. With further increment of the addition of graphene to 3 wt%, the PHRR was further reduced to 17% of pure PMMA's. Authors in [19] found that addition of 1 wt% of Graphene Nano Plates (GNPs) leads to 12% reduction in peak release rate compared to pure PMMA. Also, authors in [5] found that the PHRR reduces from 950 to about 780 kW/m², when 0.5 ml/mg of carbon nano tubes were added to PMMA [19]. The mechanism of reducing the PHRR of PMMA in the presence of graphene is attributed to the formation of a carbon layer when the

composites are ignited. This low conductivity of carbon layers covered the surface of the matrix, reducing the heat release of the material [20]. Authors in [21] found the same results when adding grapheme oxide to polypropylene where the PHRR reduced from 1140 to 936 kW/m² when 1 wt% of grapheme oxide was added. However, the time to reach the PHRR has the same importance as the PHRR itself. It was found that addition of 3 wt% of grapheme to PMMA reduces this time from 140 s for pure PMMA to 131 s for PMMA-3 wt% grapheme nanocomposite, as shown in Table I.

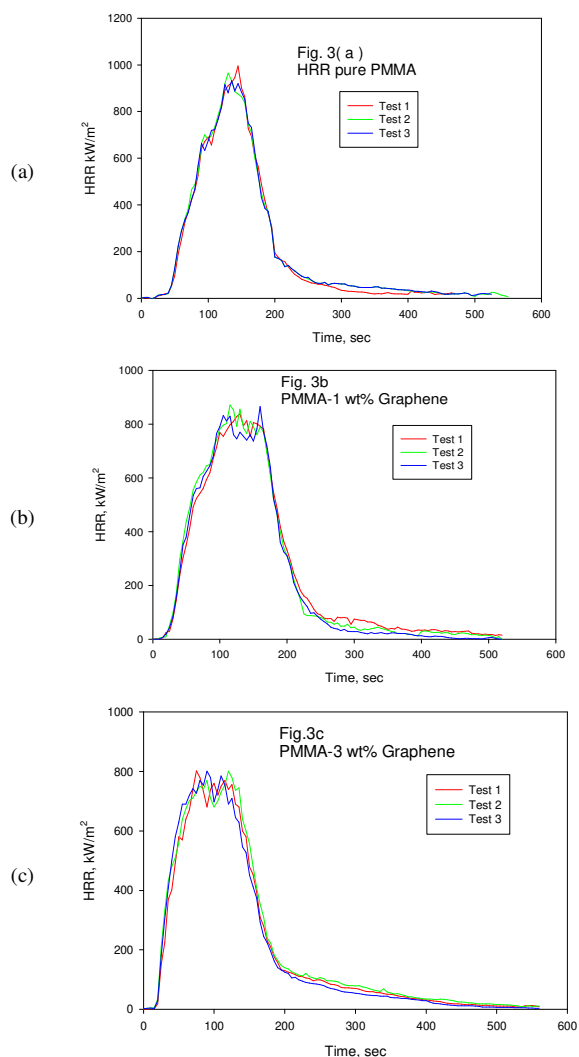


Fig. 3. Heat release rate from: (a) pure PMMA, (b) PMMA-1 wt% graphene, and (c) PMMA- 3 wt% graphene.

Figure 5 shows the images of char residues of PMMA and PMMA – 3 wt% grapheme composite. The char of pure PMMA contains white and black areas. However, PMMA-3 wt% grapheme has exterior black char as shown. This black char slows the combustion of interior layers and prevents the pyrolysis gases to release. Figure 6 shows the microstructure of pure PMMA char and PMMA-3 wt% grapheme char. The presence of graphene enhances the formation of a continuous

carbonized layer at the surface of PMMA-3 wt% grapheme. This compact continuous carbonized layer protects the sublayers of the nanocomposite in two ways: by hindering the pyrolysis gases to transfer to the upper layer and acts as an insulator reducing the amount of heat transferred from the surface toward the sublayers. It has been found that the presence of graphene in the epoxy matrix produces a continuous and compacted char layer which exhibits lower thermal conductivity from pure epoxy [22].

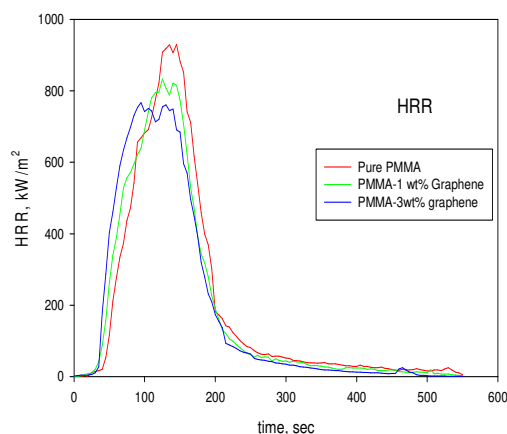


Fig. 4. Average heat release rate for pure PMMA and PMMA-graphene composites.

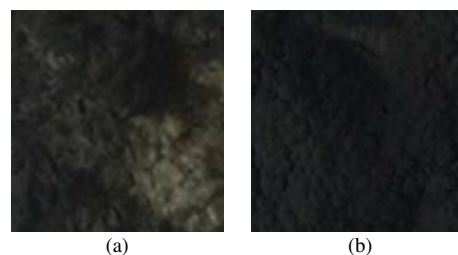


Fig. 5. Char images from (a) PMMA and (b) PMMA-3 wt% graphene.

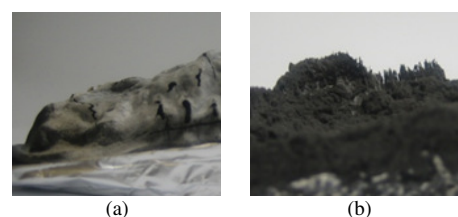


Fig. 6. Char microstructure of: (a) PMMA and (b) PMMA-3 wt% graphene.

Table I reveals that the induction of graphene in low percentages (1 wt% and 3 wt%) has less effect in reducing the THRR values of PMMA-graphene composites. This is expected, because the heat generated from the combustion of 1 g of composites will not change, so all the samples approximately have the same values of THRR. PHRR has more influence on the fire spreading. So, for the same THRR, a lower PHRR means lower fire hazard and flame spreading [20].

Graphene and its products in general improve the thermal stability of the polymers and reduce the time to ignition t_{ig} as shown in Table I [21]. Time to ignition is recorded when the HRR exceeds 10 kW/m^2 for the first time. Adding 1 wt% graphene reduces t_{ig} from 41 s for pure PMMA to 23 s. Authors in [23] reported that the addition of 1 wt% graphene oxide to polypropylene decreases the t_{ig} to 39 s from 49 s.

Figure 7 shows the mass loss rate for pure PMMA and PMMA-graphene composites materials recorded in the cone calorimeter tests. The heat release rate from the samples depends on the amount of gas volatiles released from the sample. So, the mass loss rate curves have identical trends as the heat released rate curves. PMMA has the largest Peak Mass Loss Rate (PMLR) of $33 \text{ g/m}^2 \text{ s}$. The PMLR of PMMA-1 wt% graphene composite is reduced by 22% and even more PMLR reduction was obtained (30%) when graphene increased to 3 wt%.

The thermal conductivity of PMMA is reduced by the addition of graphene [24]. The presence of graphene in PMMA suppresses the transfer of the heat generated by the cone calorimeter from the surface layer of the tested sample to the inner layers, reducing the pyrolysis of PMMA. As a result, the volatile gases are reduced and the mass loss rate is affected. In addition, the mass loss rate depends on the rate of volatile gases released from the surface of the sample. The presence of nano fillers in polymers reduces gas permeability [25]. Authors in [26] found that the gas permeability of PMMA matrix reduced by 50% when 1 wt% oxide graphene was inserted and with further addition to 10 wt%, the nanocomposite became impermeable. However, the gas permeability of nanocomposites depends on the nanofiller aspect ratio, volume fraction, and the distribution uniformity inside the polymer. Graphene as a single layer of graphite has a high aspect ratio of $2630 \text{ m}^2/\text{g}$ [27].

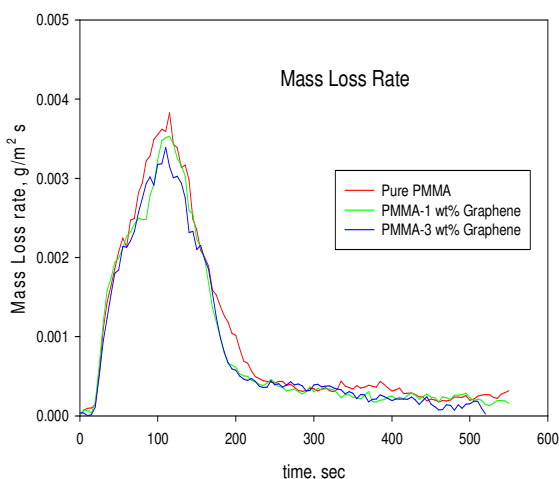


Fig. 7. Mass loss rate of PMMA and its composites.

The effect of the addition of graphene to PMMA to produce composite materials can be observed from the rate production of CO and CO_2 generated from the combustion of volatile gases. Figure 8 shows the production rate of CO and CO_2 from

pure PMMA and PMMA-graphene composites. It seems that the production of CO and CO_2 shows the same trends of heat release rate.

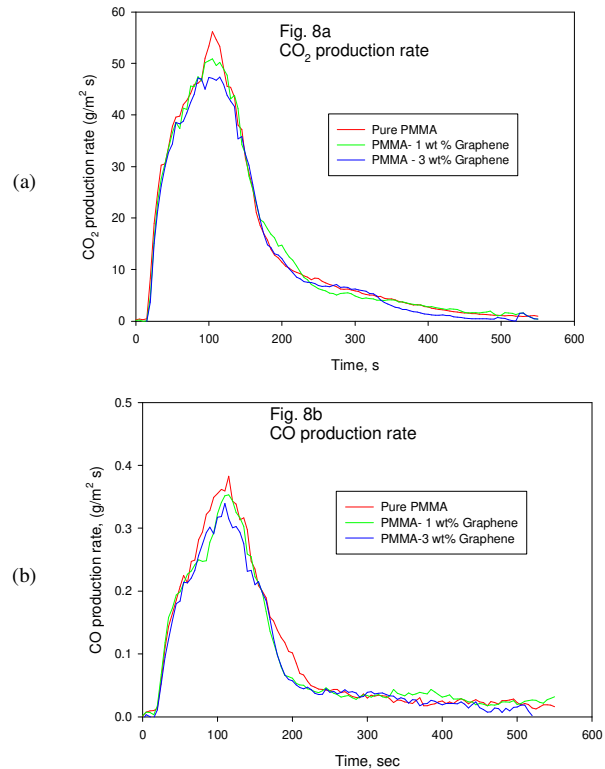


Fig. 8. Production rate of (a) CO_2 and (b) CO during combustion.

IV. CONCLUSIONS

In this study, two levels of graphene particles were mixed with pure PMMA and pressed in a prepared mold, and the resulting composite materials were studied with regard to their flame retardancy. The main conclusions of the current study are:

- The addition of graphene particles enhances the flame retardancy of PMMA
- The peak heat release rate of composite materials (PMMA-graphene) is less than pure PMMA's, while there is no change between the total heat release rate.
- The time to ignition decreases from 40 s for pure PMMA to 20 s for 3 wt% added graphene to PMMA.
- Mass loss rate decreases when graphene is added to PMMA.
- The production of CO_2 and CO from the combustion of pure PMMA and PMMA-graphene nanocomposites have the same trends.

REFERENCES

- [1] Y. Zhang, X. Yu, and Z. Cheng, "Research on the Application of Synthetic Polymer Materials in Contemporary Public Art," *Polymers*,

- vol. 14, no. 6, Jan. 2022, Art. no. 1208, <https://doi.org/10.3390/polym14061208>.
- [2] M. Danikas and S. Morsalin, "A Short Review on Polymer Nanocomposites for Enamelled Wires: Possibilities and Perspectives," *Engineering, Technology & Applied Science Research*, vol. 9, no. 3, pp. 4079–4084, Jun. 2019, <https://doi.org/10.48084/etasr.2678>.
- [3] T. Rajkumar, N. Muthupandiyar, and C. T. Vijayakumar, "Synthesis and investigation of thermal properties of PMMA-maleimide-functionalized reduced graphene oxide nanocomposites," *Journal of Thermoplastic Composite Materials*, vol. 33, no. 1, pp. 85–96, Jan. 2020, <https://doi.org/10.1177/0892705718804595>.
- [4] B. Sang, Z. Li, X. Li, L. Yu, and Z. Zhang, "Graphene-based flame retardants: a review," *Journal of Materials Science*, vol. 51, no. 18, pp. 8271–8295, Sep. 2016, <https://doi.org/10.1007/s10853-016-0124-0>.
- [5] L. Xu *et al.*, "Preparation and Study on the Flame-Retardant Properties of CNTs/PMMA Microspheres," *ACS Omega*, vol. 7, no. 1, pp. 1347–1356, Jan. 2022, <https://doi.org/10.1021/acsomega.1c05606>.
- [6] Q. Liu *et al.*, "Recent advances in the flame retardancy role of graphene and its derivatives in epoxy resin materials," *Composites Part A: Applied Science and Manufacturing*, vol. 149, Oct. 2021, Art. no. 106539, <https://doi.org/10.1016/j.compositesa.2021.106539>.
- [7] J. Shen, J. Liang, X. Lin, H. Lin, J. Yu, and S. Wang, "The Flame-Retardant Mechanisms and Preparation of Polymer Composites and Their Potential Application in Construction Engineering," *Polymers*, vol. 14, no. 1, Jan. 2022, Art. no. 82, <https://doi.org/10.3390/polym14010082>.
- [8] L. Xu, J. Jiang, L. Ni, Z. Chen, and C. Li, "Preparation and study of the flame retardant properties of C 60 /PMMA microspheres," *RSC Advances*, vol. 12, no. 35, pp. 22623–22630, 2022, <https://doi.org/10.1039/D2RA03642H>.
- [9] Z. Wang, P. Wei, Y. Qian, and J. Liu, "The synthesis of a novel graphene-based inorganic–organic hybrid flame retardant and its application in epoxy resin," *Composites Part B: Engineering*, vol. 60, pp. 341–349, Apr. 2014, <https://doi.org/10.1016/j.compositesb.2013.12.033>.
- [10] A. Kausar, "Poly(methyl methacrylate) nanocomposite reinforced with graphene, graphene oxide, and graphite: a review," *Polymer-Plastics Technology and Materials*, vol. 58, no. 8, pp. 821–842, May 2019, <https://doi.org/10.1080/25740881.2018.1563112>.
- [11] A. S. Alghamdi, "Synthesis and Mechanical Characterization of High Density Polyethylene/Graphene Nanocomposites," *Engineering, Technology & Applied Science Research*, vol. 8, no. 2, pp. 2814–2817, Apr. 2018, <https://doi.org/10.48084/etasr.1961>.
- [12] M. J. Allen, V. C. Tung, and R. B. Kaner, "Honeycomb Carbon: A Review of Graphene," *Chemical Reviews*, vol. 110, no. 1, pp. 132–145, Jan. 2010, <https://doi.org/10.1021/cr900070d>.
- [13] G. Huang, S. Wang, P. Song, C. Wu, S. Chen, and X. Wang, "Combination effect of carbon nanotubes with graphene on intumescent flame-retardant polypropylene nanocomposites," *Composites Part A: Applied Science and Manufacturing*, vol. 59, pp. 18–25, Apr. 2014, <https://doi.org/10.1016/j.compositesa.2013.12.010>.
- [14] R. N. Singh and C. S. Sharma, "Preparation of Bimetallic Pd-Co Nanoparticles on Graphene Support for Use as Methanol Tolerant Oxygen Reduction Electrocatalysts," *Engineering, Technology & Applied Science Research*, vol. 2, no. 6, pp. 295–301, Dec. 2012, <https://doi.org/10.48084/etasr.215>.
- [15] B. Tawiah *et al.*, "Highly efficient flame retardant and smoke suppression mechanism of boron modified graphene Oxide/Poly(Lactic acid) nanocomposites," *Carbon*, vol. 150, pp. 8–20, Sep. 2019, <https://doi.org/10.1016/j.carbon.2019.05.002>.
- [16] X. Wang, E. N. Kalali, J.-T. Wan, and D.-Y. Wang, "Carbon-family materials for flame retardant polymeric materials," *Progress in Polymer Science*, vol. 69, pp. 22–46, Jun. 2017, <https://doi.org/10.1016/j.progpolymsci.2017.02.001>.
- [17] W. Liu, D.-Q. Chen, Y.-Z. Wang, D.-Y. Wang, and M.-H. Qu, "Char-forming mechanism of a novel polymeric flame retardant with char agent," *Polymer Degradation and Stability*, vol. 92, no. 6, pp. 1046–1052, Jun. 2007, <https://doi.org/10.1016/j.polymdegradstab.2007.02.009>.
- [18] P. Yang, H. Wu, F. Yang, J. Yang, R. Wang, and Z. Zhu, "A Novel Self-Assembled Graphene-Based Flame Retardant: Synthesis and Flame Retardant Performance in PLA," *Polymers*, vol. 13, no. 23, Jan. 2021, Art. no. 4216, <https://doi.org/10.3390/polym13234216>.
- [19] Z. Wang, S. H. Xu, L. X. Wu, and D. X. Zhuo, "Flammability and Thermal Degradation of PMMA/Graphene Composites," *Advanced Materials Research*, vol. 910, pp. 31–34, 2014, <https://doi.org/10.4028/www.scientific.net/AMR.910.31>.
- [20] M. M. Hirschler, "Flame retardants and heat release: review of traditional studies on products and on groups of polymers," *Fire and Materials*, vol. 39, no. 3, pp. 207–231, 2015, <https://doi.org/10.1002/fam.2243>.
- [21] M. Sabet, H. Soleimani, E. Mohammadian, and S. Hosseini, "The Effect of Graphene Oxide on Flame Retardancy of Polypropylene and Polystyrene," *Materials Performance and Characterization*, vol. 9, no. 1, pp. 284–292, Jul. 2020, <https://doi.org/10.1520/MPC20190256>.
- [22] X. Wang, W. Xing, X. Feng, B. Yu, L. Song, and Y. Hu, "Functionalization of graphene with grafted polyphosphamide for flame retardant epoxy composites: synthesis, flammability and mechanism," *Polymer Chemistry*, vol. 5, no. 4, pp. 1145–1154, Jan. 2014, <https://doi.org/10.1039/C3PY00963G>.
- [23] M. Sabet and H. Soleiman, "Graphene Impact on Thermal Characteristics of LDPE," *Polymer Science, Series A*, vol. 61, no. 6, pp. 922–930, Nov. 2019, <https://doi.org/10.1134/S0965545X20010095>.
- [24] D. Mohanakshmi, S. Duggal, V. V. Nandini, and D. Charles, "Thermal conductivity of graphene incorporated heat activated polymethyl methacrylate: A pilot study," *The Journal of Prosthetic and Implant Dentistry*, vol. 3, no. 1, pp. 51–56, 2019.
- [25] M. A. Priolo, D. Gamboa, K. M. Holder, and J. C. Grunlan, "Super Gas Barrier of Transparent Polymer–Clay Multilayer Ultrathin Films," *Nano Letters*, vol. 10, no. 12, pp. 4970–4974, Dec. 2010, <https://doi.org/10.1021/nl103047k>.
- [26] S. Morimune, T. Nishino, and T. Goto, "Ecological Approach to Graphene Oxide Reinforced Poly (methyl methacrylate) Nanocomposites," *ACS Applied Materials & Interfaces*, vol. 4, no. 7, pp. 3596–3601, Jul. 2012, <https://doi.org/10.1021/am3006687>.
- [27] Y. Arao, Y. Mizuno, K. Araki, and M. Kubouchi, "Mass production of high-aspect-ratio few-layer-graphene by high-speed laminar flow," *Carbon*, vol. 102, pp. 330–338, Jun. 2016, <https://doi.org/10.1016/j.carbon.2016.02.046>.

Interleaved Bidirectional DC-DC Converter for Renewable Energy Application based on a Multiple Storage System

Yehya I. Mesalam

Department of Industrial Engineering, College of Engineering, Northern Border University, Saudi Arabia
| Industrial Engineering Department, Faculty of Engineering, Zagazig University, Egypt
yehya.mesalam@nbu.edu.sa | ymesalam@yahoo.com

Shaaban Awdallah

Department of Electrical Engineering, College of Engineering, Northern Border University, Saudi Arabia
| Department of Engineering Basic Science, Faculty of Engineering, Menoufia University, Egypt
shaban.awdallah@nbu.edu.sa

Hajer Gaied

National Engineering School of Gabes, University of Gabes, Tunisia
hajer.giaed@yahoo.fr

Aymen Flah

Processes, Energy, Environment, and Electrical Systems (code: LR18ES34), National Engineering School of Gabes, University of Gabes, Tunisia | MEU Research Unit, Middle East University, Amman, Jordan | College of Engineering, University of Business and Technology (UBT), Jeddah 21448, Saudi Arabia | Private Higher School of Applied Sciences and Technologies of Gabes, University of Gabes, Tunisia | Applied Science Research Center, Applied Science Private University, Amman, 11931, Jordan
flahaymening@yahoo.fr (corresponding author)

Received: 21 January 2024 | Revised: 30 January 2024 | Accepted: 7 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6944>

ABSTRACT

Due to its fewer components, the DC-DC three-phase converter has a simpler design and could be less expensive. However, it can present challenges in terms of precise voltage regulation and current balancing, due to the limited number of switching phases. On the other hand, the three-phase converter offers more precise voltage regulation and improved current balance owing to its higher number of phases. Although this results in increased design complexity and potentially higher cost, it allows for a more uniform distribution of current load among MOSFETs. The particular needs of the application, acceptable trade-offs between complexity, cost, and performance, as well as the requirement for precise voltage regulation and ideal current balancing, can determine which option is the best. This work investigates a three-phase interlaced DC converter with a parallel MOSFET. A two-way DC-DC converter was used to assess PWM when charging and discharging a battery. The results demonstrate a great DC voltage gain without a very high cycle load.

Keywords-*interleaved bidirectional DC-DC converter; parallel-connected MOSFET; buck-boost DC-DC converter; battery storage; SOC*

I. INTRODUCTION

DC-DC converters must be used in renewable energy applications to successfully incorporate renewable energy sources into power systems. Some significant state-of-the-art parts for DC-DC converters in this industry are listed in [1-2]. Voltage conversion is necessary to integrate photovoltaic solar

panels into the electrical grid, even though they generate constant voltage. DC-DC converters, such as buck or boost converters, regulate the voltage of solar panels according to the requirements of the system. To optimize conversion efficiency, DC-DC converters also employ advanced techniques such as Maximum Power Point Tracking (MPPT) [3]. DC-DC converters, like batteries, are also utilized in energy storage

systems to manage charge, discharge, and voltage conversion. These converters help optimize battery charging/discharge efficiency, control voltage, and manage compatibility with the rest of the system [4-12]. Therefore, DC-DC converters play an essential role in renewable energy applications by allowing efficient conversion and control of DC voltage generated by renewable energy sources, integrating energy storage systems, and optimizing energy system performance.

Research and development in this area focus on improving energy efficiency, miniaturizing converters, and integrating renewable energy sources into power systems. The interleaved DC-DC converter, often referred to as the cross-switched DC-DC converter, is one kind of uninsulated converter implemented to change a DC voltage from one level to another. Its primary function is to alter the voltage of a DC power source to satisfy the demands of an electronic system. In summary, an interleaved DC-DC converter is an electronic device that efficiently converts continuous voltage from one level to another, using two synchronized conversion stages to improve energy efficiency and overall system performance. Thus, the usage of interleaved DC-DC converters in renewable energy applications offers certain advantages, namely better energy efficiency, managing energy source variability, integration of several renewable sources, accurate voltage regulation, reduced system size and weight, and increased reliability. These benefits contribute to a more efficient and reliable use of renewable energy in a variety of applications.

Advances in technology have led to a greater dependence on electrical products for daily needs. Achieving sustainable growth in society and the economy requires striking a balance between energy use and environmental conservation more than ever. This study combines a control strategy designed for DC-DC converters with a short conversion period. Changes made on the input immediately affect the output sides of converters, since there is no galvanic isolation between the input and output sides in a non-isolated converter architecture. The non-isolated converter topology has fewer components than the isolated one. However, non-isolated converters have a few minor issues that require attention, such as excessive duty cycle ratio, insufficient voltage gain, and additional circuitry for optimal operation. There are differences between each category of converter topology advantages and disadvantages. The application requirements are the basis for the decision. This study investigated some recently published studies on interleaved DC-DC converters. Previous studies examined the applications of conventional DC-DC converters in renewable energy applications, their limitations, and recent developments. However, they did not discuss the available control techniques for the operation of DC-DC converters [13]. An interleaved DC-DC converter converts a direct voltage from one level to another while reducing power losses and improving exit voltage stability.

The terminal voltages of renewable energy sources are usually low and change over time. Therefore, to offer reliable electrical energy, it is a standard procedure to interface with the DC bus using a bidirectional DC-DC converter with a high conversion ratio. The converter must lower the load current to ensure smooth power transfer, but this has the unexpected

consequence of restricting power capacity during conversion. This study also evaluates energy conversion between operating modes to offer a rapid energy conversion method.

II. THE BIDIRECTIONAL DC-DC INTERLEAVED CONVERTER OF PARALLEL CONNECTED MOSFET

The elevating converter is one converter type that is most frequently used to charge or discharge batteries due to its wide conversion range and user-friendliness [14-15]. This kind of converter allows the battery voltage levels to be adjusted to the DC bus voltage levels by maintaining a steady voltage and controlling the disruptions caused by connecting and disconnecting loads and energy sources. However, the battery receives a ripple in the switch's current, which leads to power losses and battery deterioration. When utilizing N phases made up of semiconductors and filter components, the output power can even get close to tenths of kilowatts. Reduced filter element dimensions arise from the operating frequency, becoming an integer multiple of the switching frequency. A straightforward solution for the application covered in this study is the typical interleaved bidirectional DC-DC converter. Figure 1 depicts a three-phase topology, with V1 and V2 standing for the DC link and the SC, respectively. Moreover, I1 and I2 are the corresponding currents across the SC and the DC link, whereas $P_I = V_I \times I_I$, and the same applies to the second phase power.

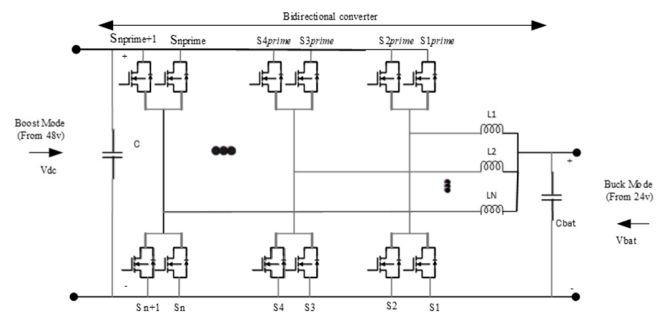


Fig. 1. Interleaved bidirectional DC-DC converter.

This study describes a qualitative and quantitative examination of the interleaved bidirectional DC-DC converter, which has four switches in each arm divided into lower-group and upper-group switches [16-19]. Figure 2 illustrates the three-phase interleaved bidirectional DC-DC converter employing n parallel-connected MOSFETs. Eight power switches ($S_{1...6}$ and $S_{1prime...6prime}$) and three inductors make up the power circuit (L1, L2, L3). The parasitic resistances, such as inductor resistances (RL1, RL2, RL3) and switch resistances, are considered to derive the non-ideal DC voltage gain (RS1). When the suggested converter runs in step-up mode (boost), where DH is the duty cycle of the gate signals S1-S6, energy passes from the battery to the DC bus side. The step-down mode (buck) transfers energy from the DC bus to the battery side, with DL being the duty cycle of the gate signals S_{1prime} - S_{6prime} . The main waveforms for the duty cycle range belonging to [0, 0.25] follow.

A. Boost Mode: First Stage [t_0-t_1]

S_1 , S_{4prime} , S_{3prime} , S_6 are turned on, and S_{1prime} , S_{2prime} , S_{5prime} , S_{6prime} , S_3 , S_4 , S_5 are turned off. Inductors L_1 and L_3 start storing energy from the battery through S_1 and S_6 at this point, and its current, $I_{L1}(t)$, $I_{L3}(t)$ increases linearly until t_1 . Since the voltages v_{bat} and $v_{bat} - v_{DC}$ are applied to L_1 , L_2 , and L_3 , respectively, current $I_{L2}(t)$ continues to fall linearly and the stored energy in L_2 is given to the load through S_{3prime} and S_{4prime} . The total of $I_{L1}(t)$ and $I_{L2}(t)$ equals the battery current $I_{bat}(t)$. Equations (1)-(3) can be used to calculate the instantaneous currents.

$$I_{L1}(t) = I_{L1}(t_0) + \left(v_{bat} - \frac{I_{bat}}{3} (R_{L1} + R_{S1}) \right) \cdot \frac{(t-t_0)}{L_1} \quad (1)$$

$$I_{L2}(t) = I_{L2}(t_0) + \left(v_{DC} + \frac{R_{L2} I_{bat}}{3} + \frac{R_{S4prime} I_{bat}}{8} - v_{bat} \right) \cdot \frac{(t-t_0)}{L_2} \quad (2)$$

$$I_{L3}(t) = I_{L3}(t_0) + \left(v_{bat} - \frac{I_{bat}}{3} (R_{L3} + R_{S6}) \right) \cdot \frac{(t-t_0)}{L_3} \quad (3)$$

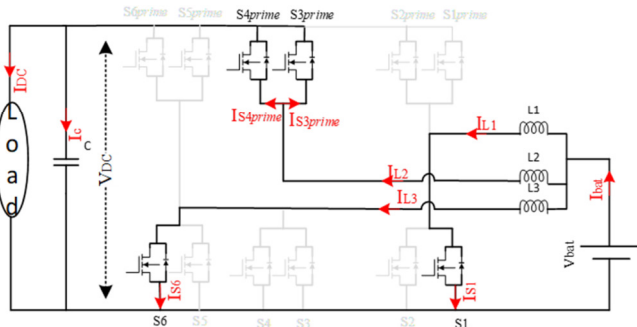


Fig. 2. Equivalent circuits of boost mode using three-phase interleaved converter.

B. Boost Mode: Second Stage [t_1-t_2]

S_1 , S_2 , S_3 , S_4 , S_5 , and S_6 are turned off, and S_{1prime} , S_{2prime} , S_{3prime} , S_{4prime} , S_{5prime} , and S_{6prime} are turned on. Through S_{1prime} , S_{2prime} , S_{3prime} , S_{4prime} , S_{5prime} , and S_{6prime} , both conductors now send energy to the load. Then, both conductors receive the voltage $v_{bat} - v_{DC}$. As a result, the instantaneous voltage equations for L_1 , L_2 , and L_3 are shown in (4)-(6), and the current across all inductors falls linearly.

$$v_{L1}(t) = v_{bat} - v_{DC} - I_{bat} \left(\frac{R_{L1}}{3} + \frac{R_{S2prime}}{8} \right) \quad (4)$$

$$v_{L2}(t) = v_{bat} - v_{DC} - I_{bat} \left(\frac{R_{L2}}{3} + \frac{R_{S3prime}}{8} \right) \quad (5)$$

$$v_{L3}(t) = v_{bat} - v_{DC} - I_{bat} \left(\frac{R_{L3}}{3} + \frac{R_{S6prime}}{8} \right) \quad (6)$$

Buck mode features eight steps in one switching time and two operation regions, just like the boost mode. The circuit is symmetric, so just the fourth period is examined. The main waveforms for the duty cycle range that belongs to $[0, 0.25]$ follow.

C. Buck Mode: First Stage [t_0-t_1]

S_3 , S_4 , S_{1prime} , and S_{5prime} are turned on, S_{3prime} , S_{4prime} , S_1 , S_2 , S_{2prime} , S_{6prime} , S_5 , and S_6 are turned off. As energy from the v_{DC}

is stored through the inductor L_1 , its current $I_{L1}(t)$ grows linearly until t_1 . The voltage $v_{DC} - v_{bat}$ and v_{bat} is delivered to L_1 , L_2 , and L_3 , and the current $I_{L2}(t)$ continues to fall linearly. As a result, the stored energy in L_2 is transferred to the load through S_3 and S_4 . Equations (7)-(9) can be used to calculate the instantaneous currents and voltages of L_1 , L_2 , and L_3 .

$$I_{L1}(t) = I_{L1}(t_0) + \left(v_{DC} - v_{bat} - I_{DC} (R_{L1} - R_{S1prime}) \right) \cdot \frac{(t-t_0)}{L_1} \quad (7)$$

$$I_{L2}(t) = I_{L2}(t_0) - \left(v_{bat} + (I_{bat} - I_{DC}) \cdot \left(R_{L2} + \frac{R_{S4}}{2} \right) \right) \cdot \frac{(t-t_0)}{L_2} \quad (8)$$

$$I_{L3}(t) = I_{L3}(t_0) + \left(v_{DC} - v_{bat} - I_{DC} (R_{L3} - R_{S5prime}) \right) \cdot \frac{(t-t_0)}{L_3} \quad (9)$$

D. Buck Mode: Second Stage [t_1-t_2]

S_3 , S_4 , S_1 , S_2 , S_5 , and S_6 are turned on, S_{3prime} , S_{4prime} , S_{2prime} , S_{1prime} , S_{5prime} , and S_{6prime} are turned off, as shown in Figure 3. Through S_3 , S_1 , S_2 , and S_4 , both conductors now deliver energy to the load. Then, both conductors receive the voltage $-v_{bat}$. As a result, the instantaneous current/voltage equations for L_1 , L_2 , and L_3 are shown in (10)-(12), and the current through all inductors falls linearly.

$$I_{L1}(t) = I_{L1}(t_1) - \left(v_{bat} + I_{bat} \left(\frac{R_{L1}}{3} + \frac{R_{S2}}{8} \right) \right) \cdot \frac{(t-t_1)}{L_1} \quad (10)$$

$$I_{L2}(t) = I_{L2}(t_1) - \left(v_{bat} + I_{bat} \left(\frac{R_{L2}}{3} + \frac{R_{S3}}{8} \right) \right) \cdot \frac{(t-t_1)}{L_2} \quad (11)$$

$$I_{L3}(t) = I_{L3}(t_1) - \left(v_{bat} + I_{bat} \left(\frac{R_{L3}}{3} + \frac{R_{S5}}{8} \right) \right) \cdot \frac{(t-t_1)}{L_3} \quad (12)$$

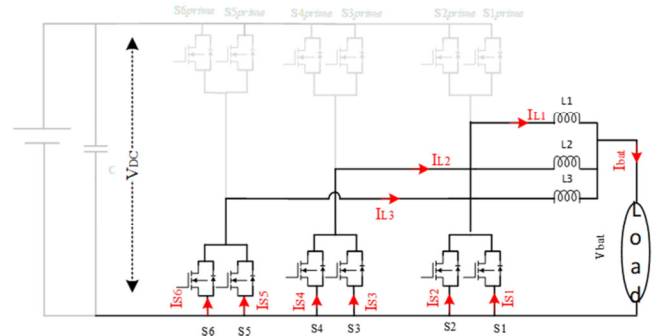


Fig. 3. Equivalent circuits of buck mode using the three-phase interleaved converter.

III. TECHNIQUE AND CONTROL STRATEGY

Only the functions of the variables that need to be controlled are employed in the system transfer functions that may be obtained from this state space. The control structure illustrated in Figure 4 is used to regulate the bus voltage and lower the battery's current ripple [19], where it is evident that measuring the three system status variables is required. Two PI controllers are used to control the charging and discharging of batteries: one generates reference current based on the mode of

operation (charging or discharging), and the other regulates the battery's current.

This case study includes two modes of operation:

- When the DC bus is connected and the battery's full voltage is reached at the control objective set point of the first PI closed loop, the charging mode switches to automated.
- Discharging mode: this mode is automatically triggered when the DC bus is disconnected, and the first PI closed loop's control purpose is to maintain a steady load voltage throughout the discharging process.

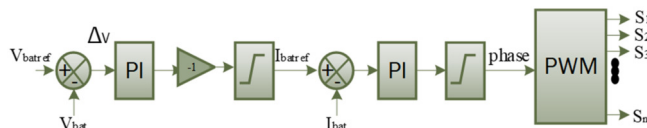


Fig. 4. Control system for the three-phase DC-DC converter.

IV. EXPERIMENTAL RESULTS

The measurements were performed using a battery-linked DC source to validate the functionality of PWM. The tests were configured with a fixed DC bus voltage $v_{DC} = 48$ V, a battery voltage $v_{bat} = 24$ V, $f_{s1} = 10$ kHz, and a nominal power of 500 W, as displayed in Table I. The simulation results are shown for the case of two-phase and three-phase DC-DC converters.

TABLE I. PARAMETERS OF THE EXPERIMENTAL SYSTEM

Parameter	Specification
Inductor	$57.6 \cdot 10^{-3}$ H
Battery capacitor	50 Ah
Capacitor	1000 mF

A resistive charge of 4.4Ω is connected as part of this operation, which simulates the converter during boost (discharge mode) on the low voltage side 24 V, 50 Ah battery, which is thought to be powered by the 48 V DC bus voltage. Comparably, the DC bus replacement resistive charge is connected to a 24 V, 50 Ah battery that is in charge of operation. The system is then simulated for buck operation mode, where the 24 V battery acts as a charge. Figures 5 and 6 demonstrate the comparison by the variation of the different battery initiation %SOC. A correct charging voltage is essential to prevent the battery from overloading or under-charging, which could damage the battery or reduce its capacity.

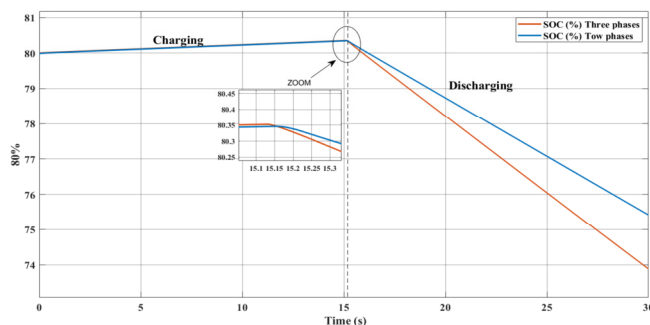


Fig. 5. State of charge of Li-ion batteries during charging and draining for an 80 percent SOC initiative.

On the other hand, the three-phase converter allows a more uniform distribution of current and reduces fluctuations. Using the latter can also reduce current harmonics and improve the quality and quantity of energy delivered to the battery. Thus, it offers better performance, energy efficiency, and high SOC to maintain the performance specifications. However, it should be noted that the choice of the three-phase interlaced converter depends on the specific system requirements and design constraints.

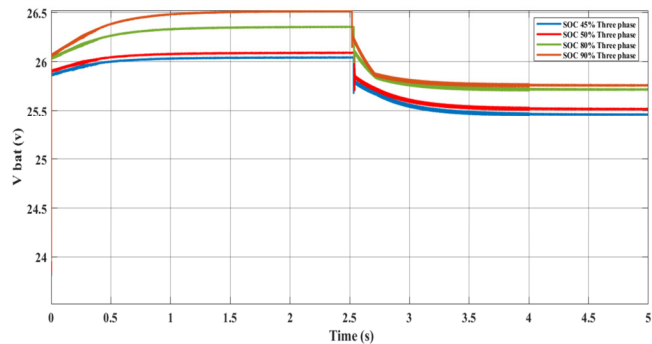


Fig. 6. Three-phase converter battery voltage.

It is crucial to examine some factors to compare the outcomes and functionality of the DC-DC converter interlaced with a parallel MOSFET in two and three phases. Similar benefits are provided by the two layouts, including fewer switching losses and current disturbances due to the distribution of load among the transistors. Nevertheless, these two strategies also differ greatly in important ways:

- The main difference lies in the number of phases used, as this difference can affect the current distribution and stability of the system.
- The number of phases can influence the distribution of current load between the MOSFET in parallel. In the case of the three-phase converter, this number is divided by three. This load distribution can have an impact on conduction losses and system efficiency.
- The number of phases can also influence battery state control. Different initial battery SOC values can affect system dynamics, regulation accuracy, and stability.

To obtain a more profound comprehension of the distinctions between these two scenarios, particular simulations and analyses must be carried out. Three phases provide a more evenly current distribution in the inductors of a three-phase interlaced DC-DC converter. A third of the total current flows through each phase, minimizing imbalances and oscillations amongst the inductors. In addition, proper balancing of the inductor currents helps to reduce the current ripples in the system. Current ripples can cause unwanted electromagnetic disturbances, additional energy losses, and degradation of the overall performance of the converter. It is important to note that balancing inductor currents also depends on the system design, component selection, and control strategy used. Proper design and accurate current regulation can help to improve the equilibration of inductor currents.

The three-phase converter can be marginally more efficient than the two-phase converter due to the ameliorated current distribution. More energy is transmitted to the battery to charge the SOC when the efficiency is higher, since less energy is lost as heat, which may cause the three-phase converter's battery SOC to increase. This leads to the second point, which is the response time: The three-phase converter can also offer a faster response time due to its three-phase switching configuration. The three-phase converter can react more quickly to load changes or battery SOC variations with three switching phases. This allows for more accurate control of the charge and discharge current, which can reduce the time required to reach a SOC target. Consequently, the three-phase converter can charge or discharge the battery faster, resulting in higher battery SOC values in a given time frame.

The employment of three sets of MOSFETs in parallel improves the load and power distribution among the phases in a three-phase interlaced DC-DC converter. Therefore, the system's current distribution improves, and switching losses are reduced. Additionally, the overall conversion efficiency of the three-phase converter might be higher. A more effective converter may extract energy from the power source and transmit it to the battery more effectively during charging or discharging. This indicates that the three-phase converter can charge the battery with less energy loss for the same amount of power supplied by the power source. By reducing energy losses, the three-phase converter can charge the battery more effectively, resulting in a higher SOC. This means that for the same amount of energy provided by the power source, the three-phase converter can store a greater amount of energy in the battery, leading to a higher SOC. It should be noted that additional variables, such as the MOSFETs threshold voltage, the battery control setup, the battery's internal resistance, etc., may also affect SOC. So, when analyzing three-phase converters, it is crucial to consider all of these details.

V. CONCLUSION

This study presented a new modulation (IPWM) on an interleaved bidirectional three-phase DC-DC converter topology to achieve greater efficiency and excellent current distribution between the switches. The inductor current ripple frequency was twice as high as the switching frequency, the battery current ripple frequency was four times higher than the switching frequency, and current ripple cancelation happened at 25% of the duty cycle. In summary, this study demonstrated that the proposed novel modulation of the bidirectional DC-DC converter could be used for both low- and high-power densities.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the approval and support of this study by Grant No. ENGA-2023-12-2236 from the Deanship of Scientific Research at Northern Border University, Arar, Saudi Arabia.

REFERENCES

- [1] A. D. Savio and J. A. Vimala, "Development of multiple plug-in electric vehicle mobile charging station using bidirectional converter," *International Journal of Power Electronics and Drive Systems (IJPEDS)*, vol. 11, no. 2, pp. 785–791, Jun. 2020, <https://doi.org/10.11591/ijpeds.v11.i2.pp785-791>.
- [2] N. Mutoh and T. Inoue, "A Control Method to Charge Series-Connected Ultraelectric Double-Layer Capacitors Suitable for Photovoltaic Generation Systems Combining MPPT Control Method," *IEEE Transactions on Industrial Electronics*, vol. 54, no. 1, pp. 374–383, Feb. 2007, <https://doi.org/10.1109/TIE.2006.885149>.
- [3] F. Musavi, M. Edington, and W. Eberle, "Wireless power transfer: A survey of EV battery charging technologies," in *2012 IEEE Energy Conversion Congress and Exposition (ECCE)*, Raleigh, NC, USA, Sep. 2012, pp. 1804–1810, <https://doi.org/10.1109/ECCE.2012.6342593>.
- [4] M. E. Bendib and A. Mekias, "Solar Panel and Wireless Power Transmission System as a Smart Grid for Electric Vehicles," *Engineering, Technology & Applied Science Research*, vol. 10, no. 3, pp. 5683–5688, Jun. 2020, <https://doi.org/10.48084/etasr.3473>.
- [5] I. W. Cox, "Electric vehicle traction control system and method," US9205758B2, Dec. 08, 2015.
- [6] D. G. Dorrell, M.-F. Hsieh, M. Popescu, L. Evans, D. A. Staton, and V. Grout, "A Review of the Design Issues and Techniques for Radial-Flux Brushless Surface and Internal Rare-Earth Permanent-Magnet Motors," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 9, pp. 3741–3757, Oct. 2010, <https://doi.org/10.1109/TIE.2010.2089940>.
- [7] S. A. Ahmad Tarusan, A. Jidin, M. L. Mohd Jamil, K. Abdul Karim, and T. Sutikno, "A review of direct torque control development in various multilevel inverter applications," *International Journal of Power Electronics and Drive Systems (IJPEDS)*, vol. 11, no. 3, pp. 1675–1688, Sep. 2020, <https://doi.org/10.11591/ijpeds.v11.i3.pp1675-1688>.
- [8] T. Ahmed, H. Kada, and A. Ahmed, "New DTC strategy of multi-machines single-inverter systems for electric vehicle traction applications," *International Journal of Power Electronics and Drive Systems (IJPEDS)*, vol. 11, no. 2, pp. 641–650, Jun. 2020, <https://doi.org/10.11591/ijpeds.v11.i2.pp641-650>.
- [9] D. Casadei, F. Profumo, G. Serra, and A. Tani, "FOC and DTC: two viable schemes for induction motors torque control," *IEEE Transactions on Power Electronics*, vol. 17, no. 5, pp. 779–787, Sep. 2002, <https://doi.org/10.1109/TPEL.2002.802183>.
- [10] L. Xi, X. Zhang, C. Sun, Z. Wang, X. Hou, and J. Zhang, "Intelligent Energy Management Control for Extended Range Electric Vehicles Based on Dynamic Programming and Neural Network," *Energies*, vol. 10, no. 11, Nov. 2017, Art. no. 1871, <https://doi.org/10.3390/en10111871>.
- [11] N. Priyadarshi, M. S. Bhaskar, P. Sanjeevikumar, F. Azam, and B. Khan, "High-power DC-DC converter with proposed HSFNA MPPT for photovoltaic based ultra-fast charging system of electric vehicles," *IET Renewable Power Generation*, Jun. 2022, <https://doi.org/10.1049/rpg2.12513>.
- [12] Y. Xing, E. W. M. Ma, K. L. Tsui, and M. Pecht, "Battery Management Systems in Electric and Hybrid Vehicles," *Energies*, vol. 4, no. 11, pp. 1840–1857, Nov. 2011, <https://doi.org/10.3390/en4111840>.
- [13] K. Fahem, D. E. Chariag, and L. Sbita, "On-board bidirectional battery chargers topologies for plug-in hybrid electric vehicles," in *2017 International Conference on Green Energy Conversion Systems (GECS)*, Hammamet, Tunisia, Mar. 2017, pp. 1–6, <https://doi.org/10.1109/GECS.2017.8066189>.
- [14] Y. Bian, Y. Zheng, W. Ren, S. E. Li, J. Wang, and K. Li, "Reducing time headway for platooning of connected vehicles via V2V communication," *Transportation Research Part C: Emerging Technologies*, vol. 102, pp. 87–105, May 2019, <https://doi.org/10.1016/j.trc.2019.03.002>.
- [15] A. Flah and C. Mahmoudi, "Design and analysis of a novel power management approach, applied on a connected vehicle as V2V, V2B/I, and V2N," *International Journal of Energy Research*, vol. 43, no. 13, pp. 6869–6889, 2019, <https://doi.org/10.1002/er.4701>.
- [16] X. Li *et al.*, "A Unified Control of Super-capacitor System Based on Bi-directional DC-DC Converter for Power Smoothing in DC Microgrid," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 3, pp. 938–949, 2023, <https://doi.org/10.35833/MPCE.2021.000549>.

-
- [17] K. B. Vikhyath and N. A. Prasad, "Combined Osprey-Chimp Optimization for Cluster Based Routing in Wireless Sensor Networks: Improved DeepMaxout for Node Energy Prediction," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12314–12319, Dec. 2023, <https://doi.org/10.48084/etasr.6542>.
- [18] N. Ding, K. Prasad, and T. t. Lie, "The electric vehicle: a review," *International Journal of Electric and Hybrid Vehicles*, vol. 9, no. 1, pp. 49–66, Jan. 2017, <https://doi.org/10.1504/IJEHV.2017.082816>.
- [19] Q. Yan, B. Zhang, and M. Kezunovic, "Optimization of electric vehicle movement for efficient energy consumption," in *2014 North American Power Symposium (NAPS)*, Pullman, WA, USA, Sep. 2014, pp. 1–6, <https://doi.org/10.1109/NAPS.2014.6965467>.

Statistical Analysis of the Factors influencing the In Situ U-Value of Walls

Smita Rashmi

Department of Architecture and Planning, National Institute of Technology Patna, India
smatar.phd20.ar@nitp.ac.in (corresponding author)

Ravish Kumar

Department of Architecture and Planning, National Institute of Technology Patna, India
ravish@nitp.ac.in

Received: 13 January 2024 | Revised: 2 February 2024 | Accepted: 7 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6904>

ABSTRACT

Building thermal performance testing requires in situ measurement techniques that are well supported and validated by simulation with statistics to improve the accuracy of the results. Local on-site performance of building components is different from the theoretical one, influenced by factors affecting the building's thermal conditions. The current paper reviews the factors influencing the measured U-value results in the heat flux method based on quantitative findings of other studies through regression and correlation statistics. The findings regarding the current status of knowledge are limited to in situ methods without detailed insights of response time, sensitivity analysis, and thermal boundary conditions in the local context. Regression analysis between wall characteristics, time duration, temperature difference, and the measured U-value shows a very strong and statistically significant impact of these variables on the accuracy of the measured U-value of low transmittance walls. The R^2 value indicates that three variables can collectively explain 91% of the variance in the measured U-value. There is a linear correlation between the wall characteristics and the measured U-value and a non-linear correlation between the time duration, temperature difference, and the measured U-value. Future work will focus on developing a measurement framework that considers time-dependent variables, dynamic weather, and uncertainty with high accuracy for different boundary conditions.

Keywords-thermal measurement; in situ measurement; U-value; heat flux method; regression analysis

I. INTRODUCTION

"If you cannot measure it, you cannot improve it" Lord Kelvin's famous quote applies to building energy use [1]. In situ, testing and dynamic data analysis are prerequisites for quantifying the actual performance of buildings and verifying the mathematical equations that describe it. The dataset obtained through in situ testing will also help validate the building simulation results. In situ testing is more complex in solution than in practice, and questions are raised regarding its accuracy and reliability. A high-quality method is required, starting with the test environment, sensors' quality and calibration, correct experimental setup, and data analysis. Thermal transmittance value (U-value) measurement and accuracy are crucial for the evaluation of the energy performance. Very few empirical studies have been conducted to identify the factors that influence the accuracy of numerical data. Management of zero-energy buildings requires monitoring techniques that include the thermal performance of the building. Despite some real-time investigations, the data collection concerning the most influential parameters (U values, time duration requirements, and temperature differences) still needs to be answered. If taken care of, data

utilization for determining the most significant parameters will provide almost accurate real-time U-values. About 91% of the variance in the measured U-value can be attributed to temperature difference, time duration, and wall characteristics. Near-zero buildings will require a higher temperature difference of $11^{\circ}\text{C} < \Delta T < 15^{\circ}\text{C}$ with measurements in the transient state and a short time duration of 24 to 72 hr. A double wall with a low U-value and a low-temperature difference of $< 11^{\circ}\text{C}$ demands a minimum time duration of 96–168 hr to reach the convergence value. This study will contribute to the field of performance testing and will widen the scope of taking up real-time building measurements.

Several green building rating systems, including GRIHA and LEED, do not ensure great performance [2] and satisfactory indoor air quality [3]. A significant gap was observed between the predicted and the actual performance of non-residential and residential buildings [4]. This performance gap persists in India [5]. No standardized protocol has gained international importance in assessing buildings in the use category. About 40% to 45% of the total heat load is due to the materials used and the design of the envelope, which is again governed by the U-value [6]. U-value also regulates the energy

consumption and thermal comfort inside the building. The conductive heat flow was reduced due to the evaporation process with natural stones as a building material, which increased the thermal comfort inside with a reduction in energy consumption [7]. Natural convection heat transfer rates increase in an open enclosure with an air cavity with an aspect ratio of 2 [8].

II. BACKGROUND

Gaps have been found regarding real-time empirical studies required for experimentation, specifically in low and medium thermal transmittance values. Several studies have been conducted to compare in situ and theoretical measurement values, e.g. U-value measurements and calculations of the ceramic wall carried out in [9] utilizing the heat flux method (the most used method) showed 8.1% and 18.9% variations with different inside and outside temperatures. Measured and theoretical U-values for precast concrete construction demonstrated a variation ranging from 4% to 75% in [10]. The error percentage between measured and theoretical U-values ranged between 17% and 153% for complex wall compositions [11].

A. Heat Flux Method: A Static Procedure

This method has a worldwide application based on heat-flux sensors [12]. Data on heat flux and air temperature are the two fundamental requirements of this method. A long measurement time will converge to the value of the thermal resistance. This value is based on data measurements of the surface temperature and average heat fluxes. Two thermocouples were installed on each side and the outside wall. Also, a heat flux meter was installed. The heat flux sensors were selected based on the following:

- Expected ranges of heat flux and temperature.
- Mode of heat conduction, convection, and radiation transfer (specific boundary conditions).
- Special requirements of the environment (chemical and mechanical, as these conditions have adverse effects on sensors).
- Measurement of an output signal voltage.

B. Instrument Location

The installation of the heat flow sensor should be 1.5 m above the floor [13]. The location should be far from the cold bridges [14], and the sensor should be 1.3m from the radiators or fan coil units. To avoid convective effects, sensors that measure the internal and external air temperatures were placed 30 cm to 40 cm away from the vertical wall surface [15]. No air gaps should exist between the sensor and the surface, as they act like insulators. When the sensor remains fixed and is not in a moving condition during the experiment, the cables should not be forced to cause stresses and strain relief of the cable should be provided by a cable tie mount.

C. Transient Analysis – Excitation Pulse Method

The theory of response factors was adopted in [16], and became a principle for this method. This theory has been widely applied in building simulation software for the heat

transfer modeling process. Excitation (cooling/heating) is applied on only one side of the wall to change the surface's temperature and heat flux response is measured and converted into wall response factors on both sides. The response factors were calculated from the thermal and physical properties of the wall. In this method, the internal surface temperature of the wall can be controlled by linear heating or cooling, which generates a triangular profile of the surface temperature. The heat fluxes on both sides of the wall were measured, and the response factor was calculated from a mathematical equation.

A radiative heater for heating and a convection fan for cooling were used for excitation. A box protects the exterior surface and the sensors and a data logger records the data [16]. Generally, to determine the maximum surface temperature, a duration of 15 min and a temperature range of 70 °C to 90 °C were found appropriate [17].

III. METHODOLOGY

The standard four-stage identification, screening, eligibility, and inclusion process was considered for review. Extensive literature research was conducted involving Scopus (2013-2023) and Google Scholar databases using the keywords "accuracy, in-situ, measurements, U value." Systematically conducted searches were performed by writing the keywords utilizing the Boolean operator "AND." Analysis of the search results revealed 74 documents from various disciplines. The initial screening after reading the abstracts led to 36 documents, and after the final screening of the complete text, only 14 papers were included. Only six studies were entailed in the review after examining the required factors and quantitative data. Very little work has been done regarding quantitative real-time assessment and the determination of the impact of time duration, temperature difference, and wall characteristics on the accuracy of the U-value.

TABLE I. QUANTITATIVE DATA TAKEN FROM THE LITERATURE

Reference	Publication year
[18]	2017
[19]	2020
[20]	2018
[21]	2018
[22]	2018
[23]	2017
[24]	2019
[25]	2018
[26]	2016

A. Data for the Statistical Technique

Data were collected from the studies mentioned above with theoretical U-values of 0.27, 0.36, and 0.52 W/m²K, and wall characteristics of (i) double-skin facade with internal insulation but no air cavities (thickness 0.33 m), (ii) four-layer wall panel incorporating a galvanized steel structure (thickness 0.30 m), (iii) double-skin facade with a non-ventilated air cavity and internal insulation, finished with continuous covering (thickness 0.34 m), time duration of 24, 48, 72, 96, 120, 144, and 168 hr and temperature differences of $\Delta T < 11$ °C, 11 °C $< \Delta T < 15$ °C, and $\Delta T > 15$ °C.

Pearson correlation was performed to determine if there was a correlation between the measured and the theoretical U-values. A high positive correlation result was found between the measured and the theoretical U-value with $r(33) = 0.95$, $p \leq 0.001$.

Multiple linear regression analysis was performed to examine the influence of the variables (temperature difference, time duration, wall characteristics) on the measured U value of low transmittance walls. It was hypothesized that they would positively predict the accurate results of measured U value. The regression model revealed (Table II) that the variables temperature difference explained 92.66% of the variance.

B. Linear Regression Assumptions

1) Quantile-Quantile (Q-Q) Plot

The model did not demonstrate any multicollinearity, which is problematic if tolerance < 0.10 or VIF > 10 .

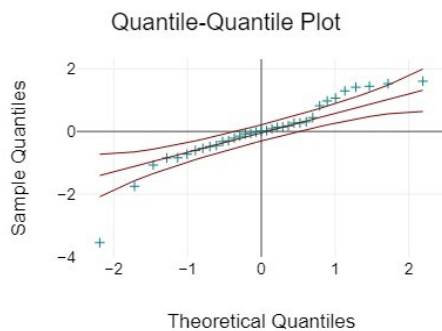


Fig. 1. Q-Q plot showing a normal distribution.

2) Multicollinearity

TABLE II. INDEPENDENT VARIABLE MODEL FOR TOLERANCE AND VIF

Model	Tolerance	VIF
Temperature difference $11^\circ\text{C} < \Delta T < 15^\circ\text{C}$	0.42	2.4
Temperature difference $\Delta T > 15^\circ\text{C}$	0.63	1.6
Time duration (hr)	1	1
Wall characteristics: The panel consisted of a four-layer wall incorporating a galvanized steel structure with a total thickness of 0.30 m.	0.42	2.4
Wall characteristics: Double-skin facade with a non-ventilated air cavity and internal insulation, finished with continuous covering, with a total thickness of 0.34 m.	0.31	3.2

3) Heteroskedasticity

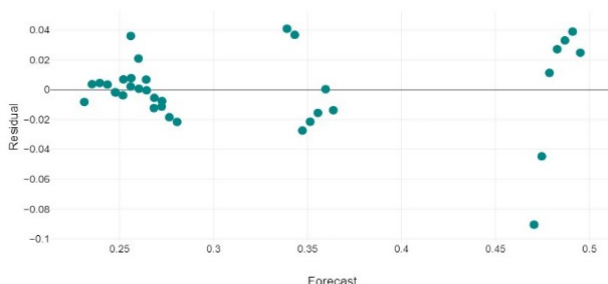


Fig. 2. Plot showing the assumption of heteroskedasticity.

The variance in the residuals was constant over the predicted values. The data did not show heteroscedasticity.

IV. RESULTS AND DISCUSSION

A. Model Summary

The model summary can be observed in Table III.

TABLE III. MODEL SUMMARY OF REGRESSION ANALYSIS.

R	R ²	Adjusted R ²	Standard error of the estimate
0.96	0.93	0.91	0.03

TABLE IV. ANOVA RESULTS

Model	df	F	p
Regression	5	73.18	<.001

Analysis of variance (ANOVA) was used to test whether this value significantly differed from zero (Table IV). It was found that the effect significantly varied from zero, with $F=73.18$, $p \leq 0.001$, $R^2 = 0.93$. The R-value (multiple correlation coefficient) showcases a strong correlation between the independent and the dependent variables. The greater the correlation, the better the regression model is. The coefficient of determination R^2 indicates that 93% of the variance of the dependent variable (measured U-value) can be explained by the three independent variables (temperature difference, wall characteristics, and time duration). The results denote that the three predictors can collectively account for 91% of the variance in the measured U value, with $p < 0.001$.

B. Cohens f^2

In Cohen's f^2 (Table V), the strength of the relationship of time duration, temperature difference, and wall characteristics is quite large with the measured U-value. As seen in Cohen's f^2 , the strength of the relationship between time duration, temperature difference, and wall characteristics is quite significant with the measured U-value.

TABLE V. COHENS F^2 STRENGTH OF THE RELATIONSHIP

Cohens f^2	
Temperature difference $11^\circ\text{C} < \Delta T < 15^\circ\text{C}$	3.44
Temperature difference $\Delta T > 15^\circ\text{C}$	3.47
Time Duration (hr)	3.47
Wall characteristics: The panel consisted of a four-layer wall incorporating a galvanized steel structure with a total thickness of 0.30 m.	3.47
Wall characteristics: Double-skin facade with a non-ventilated air cavity and internal insulation, finished with continuous covering, with a total thickness of 0.34 m.	3.47

C. Regression Coefficients

The following regression model (Table VI) was obtained:

- When all the independent variables are zero, the value of the measured U-value is 0.47.
- If the value of the variable temperature difference $11^\circ\text{C} < \Delta T < 15^\circ\text{C}$ changes by one unit, the measured U-value changes by 0.02.

- If the value of the variable temperature difference $\Delta T > 15$ °C changes by one unit, the value of the measured U-value changes by 0.02.
- If the value of the variable time duration changes by one unit, the value of the variable measured U-value does not change.
- If the value of the variable wall characteristics: Double-skin facade with a non-ventilated air cavity and internal insulation, finished with continuous covering, with a total

thickness of 0.34 m changes by one unit, the value of the variable measured U-value changes by -0.24.

- If the value of the variable wall characteristics - Double-skin facade with a non-ventilated air cavity and internal insulation, finished with continuous covering, having a total thickness of 0.34m changes by one unit, the value of the variable measured U value changes by -0.16.

TABLE VI. REGRESSION COEFFICIENTS

	Unstandardized coefficients	Standardized coefficients	Standard error	t	p	95% confidence interval for B	
Model	B	Beta	Standard error	t	p	Lower bound	Upper bound
(Constant)	0.47		0.01	33.28	<.001	0.44	0.5
Temperature difference $11 < \Delta T < 15$	0.02	0.13	0.01	1.67	0.105	-0.01	0.05
Temperature difference $\Delta T > 15$	0.02	0.07	0.01	1.13	0.267	-0.01	0.05
Time duration (hr)	0	0.09	0	1.76	0.089	0.0	0
Wall characteristics: The panel consisted of a four-layer wall incorporating a galvanized steel structure with a total thickness of 0.30 m.	-0.24	-1.26	0.01	-16.2	< 0.001	-0.27	-0.21
Wall characteristics: Double-skin facade with a non-ventilated air cavity and internal insulation, finished with continuous covering, with a total thickness of 0.34 m.	-0.16	-0.67	0.02	-7.47	< 0.001	-0.2	-0.11

D. Standardized Regression Coefficients

The standardized coefficient beta is independent of the measured variable and is always between -1 and 1. The more crucial the amount of beta, the greater the contribution of the respective independent variable is to explain the dependent variable's measured U-value. In this model, the variable wall characteristics are: The panel consisted of a four-layer wall incorporating a galvanized steel structure with a total thickness of 0.30 mm, which greatly influences the variable measured U-value.

E. p-Value

The calculated regression coefficients refer to the sample used for the regression analysis; therefore, it is interesting whether the individual coefficients only deviate from zero by chance or not. To test this, a null hypothesis is made for each coefficient equal to zero in the population. The standard error indicates the extent to which the respective coefficient will scatter on average when the regression analysis is calculated for a further sample. The test statistic t is calculated from the standard error and coefficient. The p-value for the coefficient of temperature difference $11^\circ\text{C} < \Delta T < 15^\circ\text{C}$ was 0.105. Thus, the p-value is more important than the significance level of 0.05, and the null hypothesis that the coefficient of temperature difference of $11^\circ\text{C} < \Delta T < 15^\circ\text{C}$ is zero if the population is maintained. Thus, it was assumed that the coefficient for the variable temperature difference of $11^\circ\text{C} < \Delta T < 15^\circ\text{C}$ in the population was not different from zero.

The p-value for the coefficient of temperature difference $\Delta T > 15^\circ\text{C}$ was 0.267. Therefore, the p-value was greater than the significance level of 0.05, and the null hypothesis that the coefficient of temperature difference $\Delta T > 15^\circ\text{C}$ was zero in the population was maintained. Consequently, it is thought that

the coefficient for the variable temperature difference $\Delta T > 15^\circ\text{C}$ in the population is not different from zero. The p-value for the coefficient of time duration (hr) is 0.089 and so the p-value is greater than the significance level of 0.05. As a result, it is presumed that the coefficient for the variable time duration (hr) in the population is not different from zero. The p-value for the coefficient of wall characteristics of a panel consisted of a four-layer wall incorporating a galvanized steel structure with a total thickness of 0.30 m is < 0.001 . Thus, the p-value is smaller than the significance level of 0.05, and the null hypothesis that the coefficient of this variable is zero in the population is rejected. It is therefore speculated that the coefficient for the variable wall characteristics of a panel consisted of a four-layer wall incorporating a galvanized steel structure with a total thickness of 0.30 m in the population is different from zero.

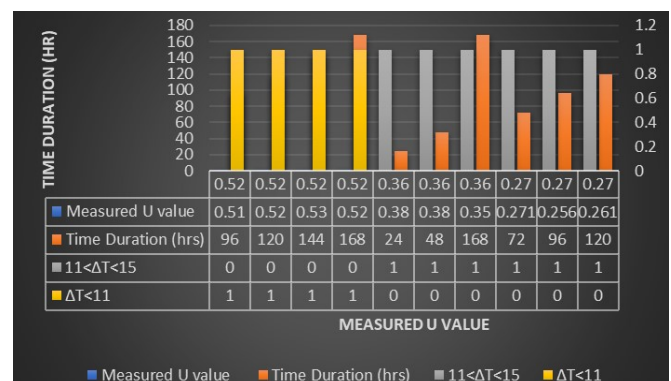


Fig. 3. Results showing the accuracy between theoretical and measured U-values of double skin facade and four-layer wall panel with air cavity and insulation in terms of time duration and temperature difference required for very low U-value walls.

The p-value for the coefficient of wall characteristics double-skin facade with a non-ventilated air cavity and internal insulation, finished with continuous covering, with a total thickness of 0.34 m is < 0.001 . Thus, the p-value is smaller than the significance level of 0.05, and the null hypothesis that the coefficient of this variable is zero in the population is rejected. It is consequently estimated that the coefficient for this variable in the population, is different from zero.

V. CONCLUSION

In situ measurements have become an inevitable solution due to many parameters, such as time, dynamic effects of climatic conditions, infiltration, exfiltration, and moisture. The expected in situ behavior of the envelope is quite different from its designed, simulated, and theoretical behavior. As per standard methods, in situ measurements are limited to certain boundary conditions. This study attempts to analyze the data taken from empirical studies to illustrate the influence of the range of temperature difference, test duration, and wall characteristics on the measured U-value for low transmittance walls through a statistical analysis. It was found that there is a strong relationship between the dependent and the three independent variables. More studies should be carried out in the future to evaluate the factors influencing the results and incorporate these factors for real-time assessments of different wall characteristics.

REFERENCES

- [1] J. Kuhle, "Comment: 'If you can't measure it, you can't improve it' (Lord Kelvin)," *Neurology*, vol. 87, no. 13, pp. 1335–1335, Sep. 2016, <https://doi.org/10.1212/WNL.00000000000003164>.
- [2] L. Sawyer, P. de Wilde, and S. Turpin-Brooks, "Energy performance and occupancy satisfaction: A comparison of two closely related buildings," *Facilities*, vol. 26, no. 13/14, pp. 542–551, Jan. 2008, <https://doi.org/10.1108/02632770810914299>.
- [3] K. W. Tham, P. Wargocki, and Y. F. Tan, "Indoor environmental quality, occupant perception, prevalence of sick building syndrome symptoms, and sick leave in a Green Mark Platinum-rated versus a non-Green Mark-rated building: A case study," *Science and Technology for the Built Environment*, vol. 21, no. 1, pp. 35–44, Jan. 2015, <https://doi.org/10.1080/10789669.2014.967164>.
- [4] B. Bordass, R. Cohen, M. Standeven, and A. Leaman, "Assessing building performance in use 2: technical performance of the Probe buildings," *Building Research & Information*, vol. 29, no. 2, pp. 103–113, Mar. 2001, <https://doi.org/10.1080/09613210010008027>.
- [5] A. M. Nii, A. Emmanuel, and A. Joshua, "Developing a Building Energy Efficiency Assessment Tool for Office Buildings in Ghana: Delphic Consultation Approach," *Energy Procedia*, vol. 111, pp. 629–638, Mar. 2017, <https://doi.org/10.1016/j.egypro.2017.03.225>.
- [6] A. S. Mahmoud, "Overview of Green Roof Technology as a Prospective Energy Preservation Technique in Arid Regions," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 8982–8989, Aug. 2022, <https://doi.org/10.48084/etasr.4919>.
- [7] T. Scussiato, W. H. Ito, J. Ramis, and P. I. B. de Queiroz, "A Numerical Model for Heat and Moisture Transfer in Porous Media of Building Envelopes," *Engineering, Technology & Applied Science Research*, vol. 12, no. 5, pp. 9239–9246, Oct. 2022, <https://doi.org/10.48084/etasr.5120>.
- [8] T. Chati, K. Rahmani, T. T. Naas, and A. Rouibah, "Moist Air Flow Analysis in an Open Enclosure. Part A: Parametric Study," *Engineering, Technology & Applied Science Research*, vol. 11, no. 5, pp. 7571–7577, Oct. 2021, <https://doi.org/10.48084/etasr.4344>.
- [9] G. Desogus, S. Mura, and R. Ricciu, "Comparing different approaches to in situ measurement of building components thermal resistance," *Energy and Buildings*, vol. 43, no. 10, pp. 2613–2620, Oct. 2011, <https://doi.org/10.1016/j.enbuild.2011.05.025>.
- [10] A. Ahmad, M. Maslehuddin, and L. M. Al-Hadhrani, "In situ measurement of thermal transmittance and thermal resistance of hollow reinforced precast concrete walls," *Energy and Buildings*, vol. 84, pp. 132–141, Dec. 2014, <https://doi.org/10.1016/j.enbuild.2014.07.048>.
- [11] L. Evangelisti, C. Guattari, P. Gori, and R. D. L. Vollaro, "In Situ Thermal Transmittance Measurements for Investigating Differences between Wall Models and Actual Building Performance," *Sustainability*, vol. 7, no. 8, pp. 10388–10398, Aug. 2015, <https://doi.org/10.3390/su70810388>.
- [12] L. Evangelisti, A. Scorza, R. De Lieto Vollaro, and S. A. Sciuto, "Comparison between Heat Flow Meter (HFM) and Thermometric (THM) Method for Building Wall Thermal Characterization: Latest Advances and Critical Review," *Sustainability*, vol. 14, no. 2, Jan. 2022, Art. no. 693, <https://doi.org/10.3390/su14020693>.
- [13] R. Albatici, A. M. Tonelli, and M. Chiogna, "A comprehensive experimental approach for the validation of quantitative infrared thermography in the evaluation of building thermal transmittance," *Applied Energy*, vol. 141, pp. 218–228, Mar. 2015, <https://doi.org/10.1016/j.apenergy.2014.12.035>.
- [14] L. Zalewski, S. Lassue, D. Rousse, and K. Boukhalfa, "Experimental and numerical characterization of thermal bridges in prefabricated building walls," *Energy Conversion and Management*, vol. 51, no. 12, pp. 2869–2877, Dec. 2010, <https://doi.org/10.1016/j.enconman.2010.06.026>.
- [15] H. Nowak and Ł. Nowak, "Non-Destructive Possibilities of Thermal Performance Evaluation of the External Walls," *Materials*, vol. 14, no. 23, Jan. 2021, Art. no. 7438, <https://doi.org/10.3390/ma14237438>.
- [16] G. P. Mitalas and D. G. Stephenson, "Room thermal response factors," *ASHRAE Transactions*, vol. 73, Jan. 1967, [Online]. Available: <https://www.osti.gov/biblio/5069807>.
- [17] A. Rasooli, "In-Situ Determination of Buildings' Thermo- Physical Characteristics: Method Development, Experimentation, and Computation," *A+BE | Architecture and the Built Environment*, no. 07, pp. 1–228, Jun. 2020, <https://doi.org/10.7480/abe.2020.07.4989>.
- [18] B. Tejedor, M. Casals, M. Gangoellés, and X. Roca, "Quantitative internal infrared thermography for determining in-situ thermal behaviour of façades," *Energy and Buildings*, vol. 151, pp. 187–197, Sep. 2017, <https://doi.org/10.1016/j.enbuild.2017.06.040>.
- [19] A. François, L. Ibos, V. Feuillet, and J. Meulemans, "Novel in situ measurement methods of the total heat transfer coefficient on building walls," *Energy and Buildings*, vol. 219, Jul. 2020, Art. no. 110004, <https://doi.org/10.1016/j.enbuild.2020.110004>.
- [20] K. Gaspar, M. Casals, and M. Gangoellés, "In situ measurement of façades with a low U-value: Avoiding deviations," *Energy and Buildings*, vol. 170, pp. 61–73, Jul. 2018, <https://doi.org/10.1016/j.enbuild.2018.04.012>.
- [21] K. Gaspar, M. Casals, and M. Gangoellés, "Review of criteria for determining HFM minimum test duration," *Energy and Buildings*, vol. 176, pp. 360–370, Oct. 2018, <https://doi.org/10.1016/j.enbuild.2018.07.049>.
- [22] B. Tejedor, M. Casals, and M. Gangoellés, "Assessing the influence of operating conditions and thermophysical properties on the accuracy of in-situ measured U-values using quantitative internal infrared thermography," *Energy and Buildings*, vol. 171, pp. 64–75, Jul. 2018, <https://doi.org/10.1016/j.enbuild.2018.04.011>.
- [23] X. Meng, T. Luo, Y. Gao, L. Zhang, Q. Shen, and E. Long, "A new simple method to measure wall thermal transmittance in situ and its adaptability analysis," *Applied Thermal Engineering*, vol. 122, pp. 747–757, Jul. 2017, <https://doi.org/10.1016/j.applthermaleng.2017.05.074>.
- [24] B. Tejedor, M. Casals, M. Macarulla, and A. Giretti, "U-value time series analyses: Evaluating the feasibility of in-situ short-lasting IRT tests for heavy multi-leaf walls," *Building and Environment*, vol. 159, Jul. 2019, Art. no. 106123, <https://doi.org/10.1016/j.buildenv.2019.05.001>.
- [25] A. Rodler, S. Guernouti, M. Musy, and J. Bouyer, "Thermal behaviour of a building in its environment: Modelling, experimentation, and

- comparison," *Energy and Buildings*, vol. 168, pp. 19–34, Jun. 2018, <https://doi.org/10.1016/j.enbuild.2018.03.008>.
- [26] A. Rasooli, L. Itard, and C. I. Ferreira, "A response factor-based method for the rapid *in-situ* determination of wall's thermal resistance in existing buildings," *Energy and Buildings*, vol. 119, pp. 51–61, May 2016, <https://doi.org/10.1016/j.enbuild.2016.03.009>.

Unveiling Shadows: Harnessing Artificial Intelligence for Insider Threat Detection

Erhan Yilmaz

Department of Computer Engineering, Ege University, Türkiye | Kron Technologies, Türkiye
erhan.yilmaz@itu.edu.tr (corresponding author)

Ozgu Can

Department of Computer Engineering, Ege University, Türkiye
ozgu.can@ege.edu.tr

Received: 15 January 2024 | Revised: 1 February 2024 | Accepted: 4 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6911>

ABSTRACT

Insider threats pose a significant risk to organizations, necessitating robust detection mechanisms to safeguard against potential damage. Traditional methods struggle to detect insider threats operating within authorized access. Therefore, the use of Artificial Intelligence (AI) techniques is essential. This study aimed to provide valuable insights for insider threat research by synthesizing advanced AI methodologies that offer promising avenues to enhance organizational cybersecurity defenses. For this purpose, this paper explores the intersection of AI and insider threat detection by acknowledging organizations' challenges in identifying and preventing malicious activities by insiders. In this context, the limitations of traditional methods are recognized, and AI techniques, including user behavior analytics, Natural Language Processing (NLP), Large Language Models (LLMs), and Graph-based approaches, are investigated as potential solutions to provide more effective detection mechanisms. For this purpose, this paper addresses challenges such as the scarcity of insider threat datasets, privacy concerns, and the evolving nature of employee behavior. This study contributes to the field by investigating the feasibility of AI techniques to detect insider threats and presents feasible approaches to strengthening organizational cybersecurity defenses against them. In addition, the paper outlines future research directions in the field by focusing on the importance of multimodal data analysis, human-centric approaches, privacy-preserving techniques, and explainable AI.

Keywords-cybersecurity; insider threats; artificial intelligence

I. INTRODUCTION

Insider threats present a considerable security risk, as malicious actors within an organization can heavily damage it through unauthorized access to sensitive information. Insider threats include current or former employees, contractors, and business partners who leverage their sanctioned access privileges to an organization's networks, systems, or data to intentionally compromise the confidentiality, integrity, or availability of the organization's information technology assets or the information itself. The gravity of the insider threat stems from the insider's knowledge of internal processes and systems, and their ability to intentionally misuse their authorized access to negatively affect the organization through data theft, sabotage, or fraud. Developing robust insider threat programs to detect and prevent such risks remains an imperative but challenging endeavor for most enterprises and institutions [1]. In [2], it was discovered that 27% of cybercrime incidents were suspected to be carried out by individuals within the organization, and 30% of those surveyed believed that insiders caused greater harm compared to external attackers. In [3],

internal fraudsters were identified as the main culprits in 29% of economic crime cases.

Detecting insider threats is challenging, as malicious insiders can operate within their authorized access permissions, and their actions may appear harmless when examined in isolation. Artificial Intelligence (AI) techniques offer potential solutions for more effective insider threat detection. AI can identify patterns and anomalies in user behavior that may indicate malicious intent before serious damage occurs [4]. An efficient insider threat detection program is imperative given the increasing number and severity of insider incidents. This study aims to address this primary concern. Even with AI's potential, most enterprises and institutions still find it difficult to put solid solutions into practice. This paper presents an overview of the current AI techniques utilized to detect insider threats, discusses the major challenges, and highlights potential research and development opportunities in the field to provide a comprehensive understanding. In addition, it explores cutting-edge approaches involving machine learning, natural language processing, knowledge graphs, and adversarial AI to identify insider threats. Furthermore, it addresses the main

challenges related to model interpretability, bias reduction, and privacy protection.

II. INSIDER THREAT

The insider threat describes the risk that someone who has authorized access to sensitive information or systems within an organization may misuse that access to compromise or harm it. This includes compromising the integrity, confidentiality, or availability of the organization's data, personnel, or physical premises. Since the threat comes from an insider, they already have some level of trust and access, making this type of threat especially risky [5]. Insider threats originate from individuals within the organization, such as regular staff, freelancers, interns, and other personnel linked to the company. These internal actors possess different degrees of trust and authority [6]. The total average cost of insider threat incidents was estimated to be \$15.4 million in 2022, up 37% over previous years [7]. The patterns shown in Figure 1 highlight the necessity of implementing robust insider threat mitigation programs.

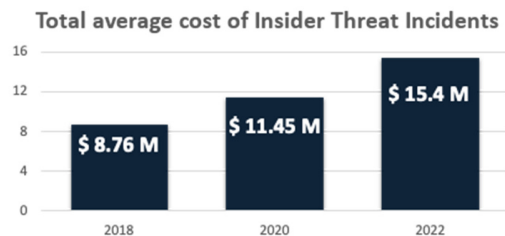


Fig. 1. Total average cost of insider threat incidents.

The Carnegie Mellon University-based CERT division has identified four distinct categories for malicious insider activity [8], which are shown in Figure 2:

- Information Technology (IT) Sabotage: This involves insiders employing IT resources to intentionally cause harm to either an organization or an individual.
- Intellectual Property (IP) Theft: Insiders exploit IT tools to steal intellectual property from the organization. This category encompasses instances of industrial espionage, which might also involve external actors.

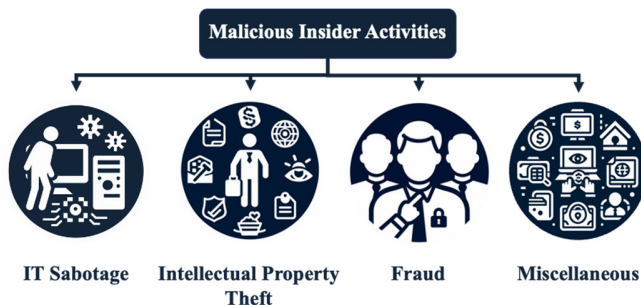


Fig. 2. Classification of malicious insider activities.

- Fraud: Insiders misuse IT capabilities to illicitly alter, append, or remove an organization's data (excluding programs or systems) with the intention of personal benefit. Additionally, this class encompasses instances where information theft leads to crimes such as identity theft or credit card fraud.
- Miscellaneous: This category encompasses cases where an insider's actions do not align with IP theft, fraud, or IT sabotage objectives.

The threat of insider security breaches presents a significant business vulnerability stemming from inadequate breach detection, delayed response, and inconsistent recovery methods. This has led businesses to allocate resources to security awareness training initiatives and implement a wide range of security protocols, procedures, and technological measures [9].

III. AI-BASED INSIDER THREAT DETECTION APPROACHES

Insider threat detection has become an important application area for AI. Various AI techniques have been explored to analyze user activities, communications, and patterns to identify potential insider threats. The following subsections explore aspects of AI-driven insider threat detection strategies, each providing targeted insights into the complex field of ensuring organizational security. The sophisticated methodologies and technological advancements discussed consist of user behavior analytics and anomaly detection, Natural Language Processing (NLP) techniques, the transformative influence of Large Language Models (LLMs), and the graph-based approaches employed to discern insider threats. These innovative AI methods analyze user behavior, textual data, language patterns, and diverse datasets, contributing to evolving cybersecurity defenses against insider threats.

A. User Behavior Analytics and Anomaly Detection

Anomalies are data patterns that deviate from normal behavior. They can be caused by various factors, such as fraud, hacking, terrorism, or system failures. Although the causes differ, anomalies share the trait of being noteworthy to analysts. Their relevance and interest in real-world situations make anomaly detection important [10-11]. Machine Learning (ML) can automatically create required models for anomaly detection based on the training data provided. This approach is motivated by the availability of necessary training data, which are easier to obtain than manually defining models. As attacks become more complex, ML enables the building and maintenance of anomaly detection systems with minimal human input [12].

User Behavior Analysis (UBA) is a cybersecurity tool that detects insider threats, targeted attacks, and financial fraud. UBA examines human behavior patterns and uses ML algorithms and statistical analysis to identify potential threats based on abnormal deviations from normal activities [13]. User behavior profiles utilize a user's historical activities within an organization to reflect their typical actions and psychological characteristics. Malicious insiders may disguise themselves as

legitimate users, quickly logging into security domains to conduct malicious activities, such as stealing files, before logging out again. Using user profiles helps detect anomalous behaviors that deviate from normal activities, signaling potential insider threats [14].

B. Natural Language Processing (NLP)-based Approaches

NLP techniques are well suited to analyze textual data, such as emails, chats, social media posts, and documents, to identify potential insider threats. NLP can extract key information from unstructured text data to detect signs of disgruntlement, malicious intent, or unauthorized sharing of sensitive information. In [15], a novel method was introduced to detect insider threats through anomaly detection, using NLP techniques to discern unusual patterns within textual data. The findings emphasized the effectiveness of this approach in uncovering insider threats that conventional methods struggle to identify.

C. Large Language Models (LLMs)

Recent advances in NLP have led to the development of LLMs with hundreds of billions of parameters or more, trained on massive textual datasets. These LLMs, based on transformer architecture, demonstrate impressive capabilities in language understanding and the completion of complex tasks through text generation [16]. Recent advances in LLMs, such as Bidirectional Encoder Representations from Transformers (BERT), Generative Pre-trained Transformer (GPT) 4, and Google's Language Model for Dialogue Applications (LaMDA), present new opportunities for insider threat detection through NLP. LLMs can identify unusual patterns in

the language used by insiders, such as the use of specific keywords or phrases associated with malicious intent. In [17], a pre-trained LLM was specifically developed to detect cybersecurity threats. This LLM could recognize 14 distinct attack categories, achieving an impressive accuracy rate of 98% across all identified attacks.

D. Graph-based Approaches

As technology has progressed, user data have become diverse and multidimensional, stemming from a range of sources including network activity, psychological elements, organizational dynamics, and employee conduct. These data exhibit distinct structural patterns. A graph-based method can be used to identify insider threats within this complex and heterogeneous dataset [18]. In [19], the challenge of insider threats within organizational contexts was addressed by proposing a method to detect insider behavior among employees by combining self-anomaly detection, which examines an employee's historical activities, with group anomaly detection, which compares an employee's behavior to that of their peers. This approach involved creating a contextual graph to provide relevant information for analysis. Experimental results demonstrated the algorithm's effectiveness in identifying insider instances and reducing false alarms when compared to using self-anomaly detection alone.

Table I presents a comparative analysis of insider threat detection, demonstrating how it benefits significantly from diverse AI-based approaches. Therefore, organizations might achieve significant improvements in the early detection and prevention of insider threats.

TABLE I. COMPARATIVE ANALYSIS OF APPROACHES FOR INSIDER THREAT DETECTION

Reference	Threat detection approach	Technique	Dataset	Accuracy metrics	Limitation
[20]	User behavior analytics for anomaly detection	LSTM autoencoder	CERT v4.2	True Positive Rate, False Positive Rate and Accuracy	Potential for missing important details when extracting features
[21]	User behavior analytics for anomaly detection	OCSVM, RNN, and Isolation Forest	Collected data from 4 employees of the organization	Precision, Recall, and Accuracy	Dataset with limited features
[22]	User behavior analytics for anomaly detection	Auto-encoder neural network, K-means clustering, and hidden Markov model	Credit card transactions by European cardholders	True Positive Rate, False Positive Rate	Limited dataset
[23]	NLP	Logistic regression and decision tree	Cyberbullying tweets	Accuracy	Need to investigate more methods for comparison
[15]	NLP	K-means and PCA	CERT r6.2	True Positive Rate	Lack of evaluation with real-world dataset
[17]	LLMs	SecurityLLM	Collected data from IoT devices	Accuracy, Recall, and F1-Score	Lack of different types of attacks
[24]	LLMs	Statistical analysis	Collected data from DC inside	Accuracy	Limited dataset
[25]	LLMs	Human-in-the-loop machine learning	Collected data from the communication of electrical devices	True Positive Rate, False Positive Rate, Precision and F1-Score	Low accuracy rate and lack of fine-tuning
[26]	Graph-Based Approach	Multi-edge weight relational graph neural network	CERT r6.2	Accuracy, Recall, Precision, and F1-Score	High delay in threat detection
[27]	Graph-Based Approach	Isolation Forest Algorithm	Collected data from individual users	Percentage of suspicious user	Limited features and dataset
[28]	Graph-Based Approach	Graph theoretic	Enron email dataset	Accuracy	Difficult to detect anomalies in dynamic data

IV. CHALLENGES AND LIMITATIONS OF DETECTING INSIDER THREATS

Detecting malicious insider threats poses significant challenges, as these events are often rare and difficult to model using only past data. Additionally, privacy concerns regarding employee monitoring and insufficient labeled data are critical issues for AI success. To train an AI model, large amounts of data are required. Thus, the model recognizes patterns, makes accurate predictions, and improves its performance. In this context, AI is used to model human behavior to identify insider threats. The challenges and limitations of detecting insider threats are as follows.

- **Insufficient Insider Threat Datasets:** Current research on insider threats struggles to validate and improve detection algorithms due to insufficient real-world organizational data. The lack of true insider threat data is a significant impediment to the development and evaluation of detection systems. Existing studies highlight that artificially generated datasets are not specifically designed to target insider threats. Furthermore, related research studies frequently use artificially generated datasets that are not suitable for insider threat scenarios. For example, some datasets lack malicious data or are out of date [3, 29].
- **Privacy Concerns Related to Employee Monitoring:** The solution proposed against insider threats should balance organizational security needs with employees' privacy rights and ethical concerns. For this purpose, more research is needed to develop monitoring strategies that are effective and respectful of employee autonomy and legal protections. The main problem with this concept arises from concerns about the employer's rationale for surveillance and its effects on morale and public perception [30].
- **Budget and Time Issues:** Supervised ML requires training data to build classification models for detecting insider threats. Existing approaches are mostly based on supervised learning that requires contextual user data and training procedures specific to insider threat detection. Although supervised ML methods are effective, they can be expensive and time-consuming [29].
- **Static Access Control Policy Rules:** Traditional access control systems have inherent vulnerabilities due to their reliance on predefined, inflexible policies and static attributes for making authorization decisions. These legacy systems fail to detect critical dynamic security events, such as changes in user behavior patterns, abuse of granted privileges, compromise of credentials through theft of access cards or passwords, and structural modifications of protected data [31].
- **Changing Employee Behavior:** Determining the threshold of deviant behavior that constitutes a malicious insider, or distinguishing normal from abnormal actions when a bad actor is already present, poses challenges. Employee routines fluctuate with shifting work demands and coverages. Therefore, abnormal behaviors can be mislabeled as malicious. Thus, careful management is necessary to avoid false accusations against employees,

reflecting the difficulties in building robust insider threat defenses [32].

- **Collusion Attacks:** In collusion attacks, two or more malicious insiders coordinate to compromise a system. Therefore, it is difficult to detect insider threats. The actions of each individual may seem benign on their own. However, insiders can gradually manipulate and steal data without raising red flags by avoiding large and abnormal actions of any user. To detect these attacks, monitoring must go beyond the tracking of individuals to identify collective suspicious patterns of activity on sensitive resources [33].

V. FUTURE DIRECTIONS AND RESEARCH OPPORTUNITIES

AI-based insider threat detection systems have made significant advances recently, and various fields warrant further research and exploration. Several key research avenues emerge to address the future directions of AI-based insider threat detection. First, building strong defenses is imperative due to the threat posed by adversarial attacks. To strengthen the resilience of AI systems against insider threats, adversarial attacks - which aim to circumvent or manipulate AI models - require concentrated efforts on strategies such as adversarial training, robust feature extraction, and anomaly detection in adversarial settings [34]. Improving the explainability of AI models used for insider threat detection is another important path. Researchers are urged to investigate methods that provide comprehensible justifications for the choices made by AI systems. Explainable AI models are expected to promote trust development, increase transparency, and enable productive cooperation between AI systems and human analysts [35]. Additionally, one promising direction is the integration of multimodal data analysis. To improve insider threat detection's precision and efficacy, multiple data modalities should be combined, including text, audio, and video. Future studies should focus on creating AI models that are adept at interpreting and combining data from various sources in a fluid and efficient manner. Insider threat detection through a thorough analysis of multimodal data is possible with techniques such as multimodal fusion, cross-modal learning, and deep learning architectures [36]. The development of privacy-preserving techniques becomes critical as the importance of privacy in insider threat identification increases. To effectively detect insider threats and protect employee privacy, research should focus on the analysis of cryptographic protocols, secure multiparty computation, and differential privacy. Ensuring that these techniques not only maximize efficiency and benefits but also unquestionably protect sensitive data integrity throughout the entire detection process is crucial [37-38]. Finally, human-centric methods that enable cooperation between AI systems and human analysts represent an important field for future study. It is crucial to investigate ways to incorporate contextual knowledge and human expertise into AI systems. Effective collaboration between human analysts and AI can be facilitated by using human-in-the-loop approaches, interactive visualization techniques, and decision support systems [39].

In summary, addressing these directions and research opportunities will be essential for future work in AI-based insider threat detection. Promising AI techniques that have been presented include LLMs, NLP, graph-based approaches, user behavior analytics, and anomaly detection. Continuous research and development will lead to strong and efficient insider threat detection systems, enhancing organizational security against changing cyber threats, despite obstacles such as limited datasets and privacy concerns.

VI. CONCLUSION

Insider threats pose significant risks to organizations. In this context, detecting and mitigating these threats is crucial for ensuring the confidentiality, integrity, and availability of information systems. This study presents an overview of the current AI techniques utilized for insider threat detection and highlights research challenges and opportunities in the field. AI techniques, such as user behavioral analytics and anomaly detection, NLP, LLMs, and graph-based approaches, have shown promise in identifying patterns and anomalies in user behavior that may indicate malicious intent. These techniques leverage ML algorithms, statistical analysis, and textual analysis to detect insider threats and potential indicators of malicious activities. However, several challenges and limitations must also be addressed to improve the effectiveness of insider threat detection. The main difficulty is the lack of sufficient datasets to detect insider threats. This deficiency hinders the development and evaluation of detection systems. In addition, privacy concerns related to employee monitoring and the need to balance organizational security needs with employee privacy rights and ethical considerations require further research. In this context, the development of monitoring strategies that respect privacy and legal protections is an essential requirement. Furthermore, the process of training accurate models to detect insider threats could be expensive and time-consuming, especially in supervised learning approaches that require contextual user data. Static access control policy rules and the changing behavior of employees pose additional challenges in accurately identifying insider threats and distinguishing them from normal actions.

Future directions of advanced AI-based insider threat research include the exploration of ensemble models, context-aware techniques, adversarial machine learning, and federated learning approaches. These approaches enable improving the accuracy and efficiency of insider threat detection systems.

In conclusion, insider threats present ongoing challenges for organizations, and the application of AI techniques holds promise in improving the early detection and prevention of such threats. Therefore, continued research and development in this field will contribute to the advancement of robust and effective insider threat detection systems, and ultimately improve the security posture of organizations against the evolving cybersecurity threats.

REFERENCES

- [1] J. R. C. Nurse *et al.*, "Understanding Insider Threat: A Framework for Characterising Attacks," in *2014 IEEE Security and Privacy Workshops*, San Jose, CA, USA, May 2014, pp. 214–228, <https://doi.org/10.1109/SPW.2014.38>.
- [2] "Cyber security breaches survey 2023," Department for Science, Innovation & Technology, London, UK. [Online]. Available: <https://www.gov.uk/government/statistics/cyber-security-breaches-survey-2023/cyber-security-breaches-survey-2023>.
- [3] I. Homoliak, F. Toffalini, J. Guarnizo, Y. Elovici, and M. Ochoa, "Insight Into Insiders and IT: A Survey of Insider Threat Taxonomies, Analysis, Modeling, and Countermeasures," *ACM Computing Surveys*, vol. 52, no. 2, Dec. 2019, <https://doi.org/10.1145/3303771>.
- [4] T. E. Senator *et al.*, "Detecting insider threats in a real corporate database of computer usage activity," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, Chicago, IL, USA, May 2013, pp. 1393–1401, <https://doi.org/10.1145/2487575.2488213>.
- [5] "Defining Insider Threats," CISA, <https://www.cisa.gov/topics/physical-security/insider-threat-mitigation/defining-insider-threats>.
- [6] "2023 Data Breach Investigations Report," Verizon, <https://www.verizon.com/business/resources/reports/dbir/>.
- [7] "2022 Cost of Insider Threats Global Report," Ponemon Institute, North Traverse City, MI, USA, 2022. [Online]. Available: <https://www.proofpoint.com/sites/default/files/threat-reports/pfpt-us-tr-the-cost-of-insider-threats-ponemon-report.pdf>.
- [8] M. Theis *et al.*, "Common Sense Guide to Mitigating Insider Threats, Sixth Edition," Carnegie Mellon University, report, Sep. 2020, <https://doi.org/10.1184/R1/12363665.v1>.
- [9] A. Georgiadou, S. Mouzakitis, and D. Askounis, "Detecting Insider Threat via a Cyber-Security Culture Framework," *Journal of Computer Information Systems*, vol. 62, no. 4, pp. 706–716, Jul. 2022, <https://doi.org/10.1080/08874417.2021.1903367>.
- [10] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, vol. 41, no. 3, Apr. 2009, <https://doi.org/10.1145/1541880.1541882>.
- [11] M. H. H. Khairi, S. H. S. Ariffin, N. M. A. Latiff, A. S. Abdullah, and M. K. Hassan, "A Review of Anomaly Detection Techniques and Distributed Denial of Service (DDoS) on Software Defined Network (SDN)," *Engineering, Technology & Applied Science Research*, vol. 8, no. 2, pp. 2724–2730, Apr. 2018, <https://doi.org/10.48084/etasr.1840>.
- [12] S. Omar, A. Ngadi, and H. H. Jebur, "Machine Learning Techniques for Anomaly Detection: An Overview," *International Journal of Computer Applications*, vol. 79, no. 2, pp. 33–41, Oct. 2013, <https://doi.org/10.5120/13715-1478>.
- [13] T. Akutota and S. Choudhury, "Big Data Security Challenges: An Overview and Application of User Behavior Analytics," *International Research Journal of Engineering and Technology*, vol. 4, no. 10, pp. 1544–1548, Oct. 2017.
- [14] X. Wang, Q. Tan, J. Shi, S. Su, and M. Wang, "Insider Threat Detection Using Characterizing User Behavior," in *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*, Guangzhou, China, Jun. 2018, pp. 476–482, <https://doi.org/10.1109/DSC.2018.00077>.
- [15] N. Garba, S. Rakshit, C. D. Mang, and N. R. Vajjhala, "An email content-based insider threat detection model using anomaly detection algorithms," in *Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2021*, Apr. 2021, <https://doi.org/10.2139/ssrn.3833744>.
- [16] W. X. Zhao *et al.*, "A Survey of Large Language Models." arXiv, Nov. 24, 2023, <https://doi.org/10.48550/arXiv.2303.18223>.
- [17] M. A. Ferrag *et al.*, "Revolutionizing Cyber Threat Detection with Large Language Models: A privacy-preserving BERT-based Lightweight Model for IoT/IIoT Devices." arXiv, Feb. 08, 2024, <https://doi.org/10.48550/arXiv.2306.14263>.
- [18] R. Nasir, M. Afzal, R. Latif, and W. Iqbal, "Behavioral Based Insider Threat Detection Using Deep Learning," *IEEE Access*, vol. 9, pp. 143266–143274, 2021, <https://doi.org/10.1109/ACCESS.2021.3118297>.
- [19] Pratibha, J. Wang, S. Aggarwal, F. Ji, and W. P. Tay, "Learning Correlation Graph and Anomalous Employee Behavior for Insider Threat Detection," in *2018 21st International Conference on Information Fusion (FUSION)*, Cambridge, UK, Jul. 2018, pp. 1–7, <https://doi.org/10.23919/ICIF.2018.8455358>.

- [20] B. Sharma, P. Pokharel, and B. Joshi, "User Behavior Analytics for Anomaly Detection Using LSTM Autoencoder - Insider Threat Detection," in *Proceedings of the 11th International Conference on Advances in Information Technology*, Bangkok, Thailand, Jul. 2020, pp. 1–9, <https://doi.org/10.1145/3406601.3406610>.
- [21] X. Xi *et al.*, "An Ensemble Approach for Detecting Anomalous User Behaviors," *International Journal of Software Engineering and Knowledge Engineering*, vol. 28, no. 11–12, pp. 1637–1656, Nov. 2018, <https://doi.org/10.1142/S0218194018400211>.
- [22] I. I. M. Abu Sulayman and A. Ouda, "User Modeling via Anomaly Detection Techniques for User Authentication," in *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, Vancouver, BC, Canada, Oct. 2019, pp. 0169–0176, <https://doi.org/10.1109/IEMCON.2019.8936183>.
- [23] T. Kanan, S. Hendawi, S. AlZu'bi, M. Elbes, and A. Mughaid, "Revolutionizing Cyberbullying Prevention: A Cutting-Edge Natural Language Processing-Based Approach," in *2023 International Conference on Information Technology (ICIT)*, Amman, Jordan, Aug. 2023, pp. 220–225, <https://doi.org/10.1109/ICIT58056.2023.10225847>.
- [24] T. Kwon and C. Kim, "Efficacy of Utilizing Large Language Models to Detect Public Threat Posted Online." arXiv, Dec. 29, 2023, <https://doi.org/10.48550/arXiv.2401.02974>.
- [25] A. Zaboli, S. L. Choi, T. J. Song, and J. Hong, "ChatGPT and other Large Language Models for Cybersecurity of Smart Grid Applications." arXiv, Nov. 09, 2023, <https://doi.org/10.48550/arXiv.2311.05462>.
- [26] J. Xiao, L. Yang, F. Zhong, X. Wang, H. Chen, and D. Li, "Robust Anomaly-Based Insider Threat Detection Using Graph Neural Network," *IEEE Transactions on Network and Service Management*, vol. 20, no. 3, pp. 3717–3733, Nov. 2022, <https://doi.org/10.1109/TNSM.2022.3222635>.
- [27] A. Gamachchi, L. Sun, and S. Boztas, "A Graph Based Framework for Malicious Insider Threat Detection." arXiv, Sep. 01, 2018, <https://doi.org/10.48550/arXiv.1809.00141>.
- [28] W. Eberle, J. Graves, and L. Holder, "Insider Threat Detection Using a Graph-Based Approach," *Journal of Applied Security Research*, vol. 6, no. 1, pp. 32–81, Dec. 2010, <https://doi.org/10.1080/19361610.2011.529413>.
- [29] M. N. Al-Mhiqani *et al.*, "A Review of Insider Threat Detection: Classification, Machine Learning Techniques, Datasets, Open Challenges, and Recommendations," *Applied Sciences*, vol. 10, no. 15, Jan. 2020, Art. no. 5208, <https://doi.org/10.3390/app10155208>.
- [30] F. L. Greitzer, "Insider Threats: It's the HUMAN, Stupid!," in *Proceedings of the Northwest Cybersecurity Symposium*, Richland, WA, USA, Dec. 2019, <https://doi.org/10.1145/3332448.3332458>.
- [31] M. Raissi-Dehkordi and D. Carr, "A multi-perspective approach to insider threat detection," in *2011 - MILCOM 2011 Military Communications Conference*, Baltimore, MD, USA, Nov. 2011, pp. 1164–1169, <https://doi.org/10.1109/MILCOM.2011.6127457>.
- [32] N. Saxena, E. Hayes, E. Bertino, P. Ojo, K. K. R. Choo, and P. Burnap, "Impact and Key Challenges of Insider Threats on Organizations and Critical Businesses," *Electronics*, vol. 9, no. 9, Sep. 2020, Art. no. 1460, <https://doi.org/10.3390/electronics9091460>.
- [33] I. A. Gheyas and A. E. Abdallah, "Detection and prediction of insider threats to cyber security: a systematic literature review and meta-analysis," *Big Data Analytics*, vol. 1, no. 1, Aug. 2016, Art. no. 6, <https://doi.org/10.1186/s41044-016-0006-0>.
- [34] N. Papernot and P. McDaniel, "Deep k-Nearest Neighbors: Towards Confidential, Interpretable and Robust Deep Learning." arXiv, Mar. 13, 2018, <https://doi.org/10.48550/arXiv.1803.04765>.
- [35] S. S. Taher, S. Y. Ameen, and J. A. Ahmed, "Advanced Fraud Detection in Blockchain Transactions: An Ensemble Learning and Explainable AI Approach," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12822–12830, Feb. 2024, <https://doi.org/10.48084/etasr.6641>.
- [36] T. Baltrušaitis, C. Ahuja, and L. P. Morency, "Multimodal Machine Learning: A Survey and Taxonomy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, Jan. 2018, <https://doi.org/10.1109/TPAMI.2018.2798607>.
- [37] M. Abadi *et al.*, "Deep Learning with Differential Privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, Vienna, Austria, Oct. 2016, pp. 308–318, <https://doi.org/10.1145/2976749.2978318>.
- [38] P. Geetha, C. Naikodi, and L. Suresh, "Optimized Deep Learning for Enhanced Trade-off in Differentially Private Learning," *Engineering, Technology & Applied Science Research*, vol. 11, no. 1, pp. 6745–6751, Feb. 2021, <https://doi.org/10.48084/etasr.4017>.
- [39] M. R. Endsley, "From Here to Autonomy: Lessons Learned From Human–Automation Research," *Human Factors*, vol. 59, no. 1, pp. 5–27, Feb. 2017, <https://doi.org/10.1177/0018720816681350>.

Enhancing Cognitive Radio WSN Communication through Cluster Head Selection Technique

Shraddha Panbude

Dr. Babasaheb Ambedkar Technological University, India
shraddha.panbude@gmail.com

Prachi Deshpande

Department of Computer Science & Engineering, Shreeyash College of Engineering & Technology, India
hod.cse@sycet.org

Brijesh Iyer

Dr. Babasaheb Ambedkar Technological University, India
brijeshiyer@dbatu.ac.in (corresponding author)

A. B. Nandgaonkar

Dr. Babasaheb Ambedkar Technological University, India
abnandgaonkar@dbatu.ac.in

Received: 23 December 2023 | Revised: 19 January 2024 | Accepted: 28 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6803>

ABSTRACT

The demand for frequency spectrum is increasing rapidly with the wide growth of wireless communications. Spectrum sensing issues present in Cognitive Radio Sensor Networks (CRSN) are detected dynamically using spectral sensing techniques, which also help to utilize frequency bands more effectively. The study proposes a novel Cosine Sand Cat Optimization (CSCO) protocol to address spectral sensing problems by selecting the optimal Cluster Head (CH) in a CRSN. The CRSN is simulated, and spectral allocation is performed using LeNet to extract signal components. Then, Primary User (PU) aware optimal CH selection is performed using the proposed CSCO by taking account of multi-objective fitness parameters. Finally, data communication is performed between nodes after CH selection using the CSCO protocol. The simulation results of CSCO were validated to determine its superiority concerning Secondary User (SU) density, and it attained residual energy, network lifetime, Packet Delivery Ratio (PDR), normalized throughput, and delay of 69.457 J, 77, 75.89%, 74.473, and 4.782ms, respectively.

Keywords-cosine sand cat optimization; sine cosine algorithm; sand cat swarm optimization; LeNet

I. INTRODUCTION

Cognitive Radio (CR) is an intelligent radio device that adapts itself to activities within its Radio Frequency (RF) environment [1]. Spectrum scarcity issues are effectively addressed by CR technologies, which also provide a solution to increase spectrum utilization by intelligently using temporarily unused spectrums [2]. Today, CR techniques have overcome the limitations encountered by conventional Wireless Sensor Networks (WSNs) to create Cognitive Radio Sensor Networks (CRSNs) [3-4]. In general, idle time slots or other resources are utilized by CR if they are not used by the Primary User (PU) or licensed users. CR uses frequency resources when the

connection is used by a Secondary User (SU) or an unlicensed user [5]. CR performs dynamic variation in operating parameters, spectrum sensing, and the estimation of vacant bands. In general, CR has two objectives: efficient utilization of spectrum resources and maintaining permanent and reliable communication [5]. Several methods have been proposed for the effective utilization of unused spectrum [6]. In general, spectrum sensing is performed by a few of the received signal features to effectively distinguish the signal from noise. PU signals are accurately detected by various spectral sensing approaches. The approaches used for spectrum sensing are categorized into blind and non-blind techniques. Blind spectrum sensing approaches do not utilize information on PU

signals, such as pulse shape, carrier frequency, modulation, etc. Wireless links are generally affected by different channel impairment issues, such as receiver impairments, multipath fading, and shadowing. These imperiments may cause loss of unpredictable signals resulting in spectrum sensing degradation. Moreover, hidden node issues affect spectrum sensing based on a single cognitive user and create shadowing issues. Thus, it becomes difficult for an SU to detect PU. Cooperative Spectrum Sensing (CSS) has been introduced to provide solutions in challenging scenarios by increasing the reliability of spectrum sensing [7]. CRSN performance has been improved by grouping adjacent nodes into clusters using cluster routing protocols, inter-cluster relay, and intra-cluster aggregation using multi-hop data delivery [8-9]. Thus, the clustering routing protocol introduced for CRSNs has become a trending research topic. CRSN assumes perfect spectrum sensing while designing cluster routing protocols. However, this assumption is suitable only for designing simple routing protocols that do not consider the original perceptual performance of CRSN nodes [9].

CRSNs also perform intracluster data transmission that applies to clustering protocols. This involves communication between clusters and inter-cluster data transmission and helps to increase scalability by reducing energy consumption and promoting communication between the sink and the cluster. The predetermined CHs present in intercluster communication help to transmit data to the sink. In CRSN, CHs are selected mostly among available sensor nodes that are highly vulnerable to energy depletion issues. This is due to the burden caused while performing functions related to CR, such as channel assignment and sensing, data forwarding, and data aggregation. In general, CH experiences high energy consumption while performing single-hop communication. The lifetime of CRSN is increased by performing multi-hop routing. However, the residual energy of CHs near the sink severely decreases due to frequent traffic relaying [10]. Therefore, the energy consumption of the end users is considered the critical parameter of CRSNs. In addition, artificial intelligence techniques, such as machine learning and deep learning approaches, are employed to provide solutions to CH-related problems in CRSNs. Several studies have been conducted based on the use of neural networks to design and optimize wireless communication systems [11-12].

Currently, several clustering algorithms are employed to form a fully connected topology of the network. However, the algorithms only paid minimum attention to problems based on ensuring maximum energy efficiency. CRSN is introduced by integrating CR capacity into sensor networks [13] and helps reduce high contention delays, eliminate collisions, and deploy multiple overlaid sensor networks. This aspect motivated the present study, which selected the optimal CH in a CRSN using the CSCO protocol. The CRSN was initially simulated and was further allowed for spectrum allocation using LeNet by extracting various signal components, such as energy, test statics, eigenstatistics, wavelet transform, and matched filter. Then, optimal PU-aware CH selection was performed on the allocated spectrum using CSCO by taking into account different multiobjective fitness parameters, such as trust factor, energy, delay, and distance. Later, in the data communication

phase, the selected CH carried out data communication by transferring data packets to other nodes.

II. THE PROPOSED CSCO PROTOCOL FOR OPTIMAL CH SELECTION IN CRSN

This study presents the CSCO protocol for the selection of optimal CH in CRSN. Initially, the CRSN simulation is performed, followed by the allocation of the spectrum. Spectrum allocation is performed by extracting signal components using LeNet [14]. The different signal components, such as energy, eigenstatistics, test statics, matched filter, and wavelet transform, are extracted. The allocated spectrum is allowed for PU-aware optimal CH selection, which is performed using the CSCO protocol by considering multiobjective fitness parameters, such as trust factor, energy, delay, and distance. CSCO was designed by integrating the SCA [15] and SCSO [16] algorithmic approaches. Finally, the selected CH performs data communication by transferring data packets from one node to another. Figure 1 shows the diagrammatic view of the CSCO protocol used for optimal CH selection in CRSN.

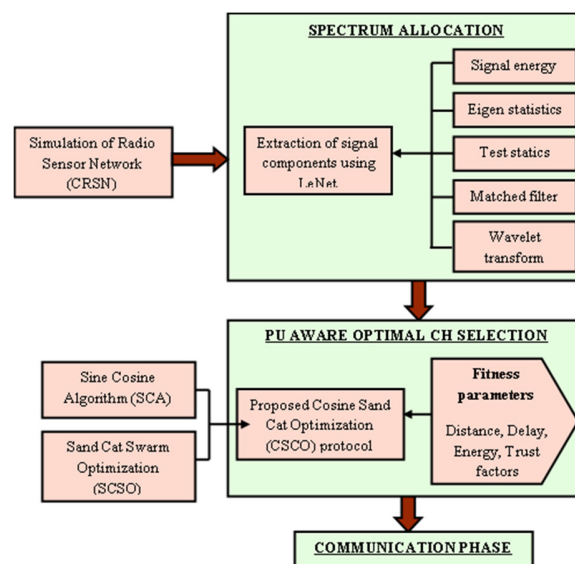


Fig. 1. Schematic diagram of CSCO protocol for optimal CH selection in CRSN.

A. Test Statistic

The test statistic [17] was used to observe null and alternative hypotheses in binary hypothesis testing problems with the presence and absence of PU.

B. LeNet Architecture

LeNet [14], a Convolutional Neural Network (CNN) gradient-based learning method, was applied for the spectrum allocation in CRSN. The extracted signal components are fed into the input layer and the output from the final layer is received. The LeNet comprises deep layers that help to accurately allocate spectrum in less execution time. It consists of convolutional, pooling, fully connected, and softmax layers. The total parameters used for training are effectively reduced by utilizing fully connected layers in LeNet.

C. CH Selection Using the CSCO Protocol

The optimal CH is selected using the proposed CSCO protocol, which is designed by integrating the SCA [15] and SCSO [16] techniques. SCSO is a metaheuristic algorithm designed to consider the natural survival behavior of sand cats. In general, sand cats live in stony and sandy deserts. The low-frequency noises are easily heard by sand cats through their extraordinary hearing sense. SCSO is designed based on two distinct characteristics of a sand cat, such as foraging and attacking the prey. Extraordinary features are used by sand cats to easily locate their prey. The SCSO effectively controls transitions in a balanced manner, which also effectively determines optimal solutions using few parameters in the exploitation and exploration phases. Similarly, SCA is a population-based optimization algorithm designed by considering the effects of sine and cosine. In this algorithm, the solution is repositioned in a cyclic pattern of sine and cosine functions. The SCA effectively converges to the global optimum and identifies promising regions of the search space. The SCA is incorporated with SCSO to exploit promising regions of the search space and promote the best approximation of local optimum. The mathematical modeling of CSCO is described below.

1) Phase 1: Population Initialization

The solution of each cat is randomly initiated, given by:

$$Y = [y_1, y_2, y_3, \dots, y_e] \quad (1)$$

The upper and lower boundaries are considered to locate Y , and a candidate matrix is generated by considering the population of sand cats and based on the problem size ($\theta_p \times \theta_e$).

2) Phase 2: Fitness Computation

Equation (16) is used for the computation of the fitness function during the selection of the optimal CH after the solution is initiated.

3) Phase 3: In search of Prey

Sand cats consider the emission of low-frequency noise to execute prey search mechanisms. The sensitivity ranges of sand cats are considered to start from 2 kHz to 0 during prey search, where the sensitivity range is determined by:

$$\vec{S}_A = Z_h - \left(\frac{2 \times Z_h \times Iter_p}{Iter_{Max}} \right) \quad (2)$$

where \vec{S}_A denotes the general sensitivity range, Z_h represents the hearing characteristics of a sand cat, the present iteration is signified as $Iter_p$, and the maximum iteration is indicated as $Iter_{Max}$. The final parameter is initialized to control the transition between the exploration and exploitation phase, which is given by:

$$\vec{\tau} = 2 \times \vec{S}_A \times R[0,1] - \vec{S}_A \quad (3)$$

where $\vec{\tau}$ represents the parameter vector initialized for transitions control and $R[0, 1]$ signifies a random number set to $[0, 1]$. The sensitivity range of each sand cat is expressed as:

$$\vec{S} = \vec{S}_A \times R[0,1] \quad (4)$$

The best-candidate position is considered by each sand cat to update its position \vec{X}_p , sensitivity range \vec{S} , as well as the present best position \vec{X}_B . Then, the best prey position is determined by sand cats as:

$$\vec{X}(f+1) = \vec{S} \cdot (\vec{X}_B(f) - R[0,1] \cdot \vec{X}_p(f)) \quad (5)$$

where \vec{X}_B signifies the best-candidate position of the sand, \vec{S} denotes sensitivity range, and the present position of the sand cat is indicated by \vec{X}_p . Let us consider, $\vec{X}(f+1) = X(f+1)$, $\vec{X}_B(f) = X_B(f)$, $\vec{S} = S$, and $\vec{X}_p(f) = X(f)$. Thus, (5) becomes:

$$X(f+1) = S \cdot (X_B(f) - R[0,1] \cdot X(f)) \quad (6)$$

SCA [15] is incorporated with SCSO [16] for the exploitation of promising regions of the search space and to promote the best approximation of local optimum. From SCA:

$$X(f+1) = X(f) + y_1 * \cos(y_2) * |y_3 u_D - X(f)| \quad (7)$$

where y_1 , y_2 , and y_3 are random numbers, u_D is the position of the destination in the D^{th} dimension, and $X(f)$ signifies the position of the present solution. Assuming $u_D > X(f)$, (7) becomes:

$$X(f) = \frac{X(f+1) - y_1 * \cos(y_2) * y_3 u_D}{(1 - y_1 \cos(y_2))} \quad (8)$$

Substituting (8) into (6) gives:

$$X(f+1) = S \cdot \left(X_B(f) - R[0,1] \cdot \left[\frac{X(f+1) - y_1 * \cos(y_2) * y_3 u_D}{(1 - y_1 \cos(y_2))} \right] \right) \quad (9)$$

Thus, the updated equation of CSCO is given by:

$$X(f+1) = \frac{1}{(1 - y_1 \cos(y_2) + S \cdot R[0,1])} [S \cdot X_B(f) (1 - y_1 \cos(y_2)) + S \cdot R[0,1] \cdot y_1 * \cos(y_2) * y_3 u_D] \quad (10)$$

4) Phase 4: Attacking the Prey

Based on its hearing ability, the sand cat effectively determines the prey, and the distance between its present and best positions is determined. The random position ensures that the position of the cat is near to the prey and is expressed as:

$$\vec{X}_R = |R \cdot \vec{X}_B(f) - \vec{X}_p(f)| \quad (11)$$

where \vec{X}_R symbolizes the random position of the sand cat. Moreover, the local optimum trap is avoided and the direction of movement is identified using a random angle when the sensitivity range of the sand cat is supposed to be a circle. Thus, the modified prey position based on the direction of movement is given by:

$$\vec{X}(f+1) = \vec{X}_B(f) - R \cdot \vec{X}_R \cdot \cos(\theta) \quad (12)$$

5) Phase 5: Solution Feasibility Check

The reevaluation of fitness is performed using (4) to identify the optimal solution to select optimal CH in CRSN. If any new solutions are identified to be more efficient than the current one, the solution can be replaced. Thus, the CSCO protocol effectively selects the optimal CH in the PU-aware CRSN system from the allocated spectrum by considering multi-objective fitness parameters.

III. RESULTS AND DISCUSSION

The NS2 simulator was used to implement the CSCO protocol for optimal CH selection in a CRSN. Figure 2 shows the simulation results obtained by the CSCO protocol during the selection of optimal CH. Figure 2 shows the simulation results obtained for 0.002 s and 6.028 s. In Figure 2(a), the algorithm does not declare a clear CH, while Figure 2(b) shows the CH selected with the black triangular spot.

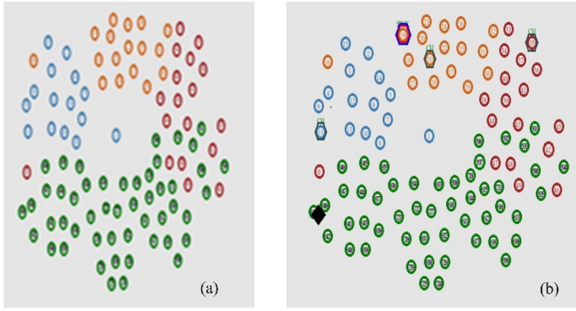


Fig. 2. Experimental simulation results of CSCO: (a) 0.002 s, (b) 6.028 s.

A. Evaluation Parameters

The following parameters, concerning SU density, were used to evaluate the performance of CSCO in CH selection.

Residual energy is the remaining energy presented in the nodes after the transmission of data packets in CRSN, and it is calculated by:

$$\text{Residual energy} = 1 - W_R(K) \quad (13)$$

where $W_R(K)$ denotes consumed energy.

Network lifetime is computed by considering the death of the sensor node, showing the ability of the model to prolong the network functioning during data transmission.

Throughput is the total data packets sent to the destination by the nodes at the stipulated period, expressed as:

$$\text{Throughput} = \frac{\eta}{t} \quad (14)$$

where t denotes the time duration taken by the nodes and η indicates the total number of node counts.

PDR is the proportion of total packets delivered to the total packets transmitted to the destination, formulated as:

$$\text{PDR} = \frac{\rho}{\sigma} \quad (15)$$

where ρ represents the total packets delivered, and σ indicates the total packets passed to the destination.

Delay is the time taken by the data packets to reach the destination, and is determined using (5).

B. Comparative Analysis

The effectiveness of the CSCO protocol in CH selection was validated by comparing its performance with that of traditional protocols, namely ISSMCRP [9], ESAUC [10], sensing-after prediction scheme [18], and Dynamic Fuzzy-based PU aware Clustering (DFPC) [19]. Comparative validation of the designed CSCO protocol in CH selection was performed for 1000 rounds. Figure 3 shows the analysis of the CSCO for the optimal selection of a CH with 1000 rounds. Figure 3(a) shows the analysis of the CH selection protocols using residual energy, where CSCO recorded maximum residual energy of 60.983J for an SU density of 200. The residual energy measured by existing models, such as ESAUC, the sensing-after prediction scheme, ISSMCRP, and DFPC was 50.297 J, 51.831 J, 53.187 J, and 56.765 J, respectively.

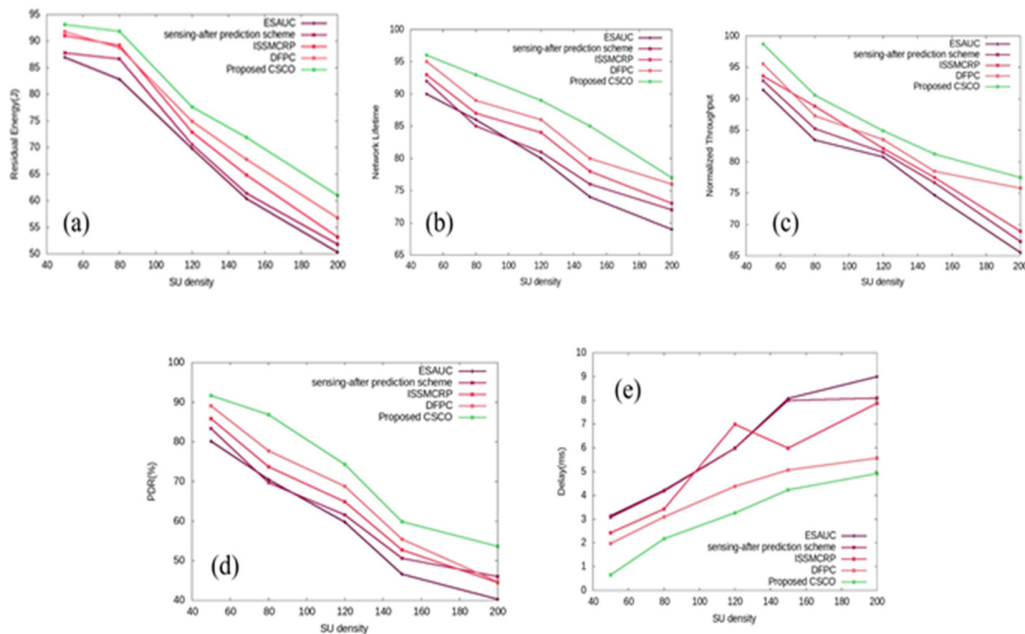


Fig. 3. Analysis of CSCO with 1000 rounds based on: (a) Residual energy, (b) network lifetime, (c) normalized throughput, (d) PDR, and (e) delay.

Figure 3(b) presents the validation using the network lifetime of the CH selection protocols. For an SU density of 200, the network lifetime obtained by ESAUC was 69, for the sensing-after prediction scheme was 72, for the ISSMCRP was 73, for the DFPC was 76, and for the proposed CSCO protocol was 77. Figure 3(c) shows an analysis of various CH selection models used in CRSN using normalized throughput. For an SU density of 200, the CSCO protocol recorded a maximum normalized throughput of 77.438, and the prevailing techniques, such as ESAUC, sensing-after prediction scheme, ISSMCRP, and DFPC measured a normalized throughput of 65.529, 67.255, 68.914, and 75.748, respectively. Figure 3(d) shows the PDR obtained by the various CH selection methods. For an SU density of 200, the maximum PDR of 53.64% was recorded by the proposed CSCO protocol, and PDRs of 40.27, 46, 44.475, and 44.764% were obtained by ESAUC, sensing-after prediction scheme, ISSMCRP, and DFPC, respectively. Figure 3(e) shows the validation of the performance of different CH selection protocols using delay. For an SU density of 200, the delays recorded by the existing models were 9 ms by ESAUC, 8.089 ms for the sensing-after prediction scheme, 7.874 ms for ISSMCRP, and 5.563 ms for DFPC. The proposed CSCO protocol obtained a minimum delay of 4.910 ms.

IV. CONCLUSIONS AND FUTURE WORK

This paper presented the CSCO algorithmic model for optimal CH selection to effectively perform CH-based communication in CRSNs by reducing the energy consumption of the network. The experimental results obtained showed that CSCO outperformed previous approaches. The CSCO achieved superior performance compared to previous techniques, having a residual energy of 69.457 J, a network lifetime of 77, a normalized throughput of 74.473, a PDR of 75.894%, and a delay of 4.782 ms. In the future, the performance of the proposed CSCO algorithm can be studied in various real-world applications. The use of various technologies to schedule subclusters to increase the energy efficiency of spectrum sensing in CRSN can also be studied.

REFERENCES

- [1] O. P. Awe, D. A. Babatunde, S. Lambrotharan, and B. AsSadhan, "Second order Kalman filtering channel estimation and machine learning methods for spectrum sensing in cognitive radio networks," *Wireless Networks*, vol. 27, no. 5, pp. 3273–3286, Jul. 2021, <https://doi.org/10.1007/s11276-021-02627-w>.
- [2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, Feb. 2005, <https://doi.org/10.1109/JSAC.2004.839380>.
- [3] H. Luo, Z. Huang, and T. Zhu, "A Survey on Spectrum Utilization in Wireless Sensor Networks," *Journal of Sensors*, vol. 2015, Mar. 2015, Art. no. e624610, <https://doi.org/10.1155/2015/624610>.
- [4] O. B. Akan, O. B. Karli, and O. Ergul, "Cognitive radio sensor networks," *IEEE Network*, vol. 23, no. 4, pp. 34–40, Aug. 2009, <https://doi.org/10.1109/MNET.2009.5191144>.
- [5] Y. Cui, X. jun Jing, S. Sun, X. Wang, D. Cheng, and H. Huang, "Deep learning based primary user classification in Cognitive Radios," in *2015 15th International Symposium on Communications and Information Technologies (ISCIT)*, Jul. 2015, pp. 165–168, <https://doi.org/10.1109/ISCIT.2015.7458333>.
- [6] M. Höyhty et al., "Spectrum Occupancy Measurements: A Survey and Use of Interference Maps," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2386–2414, 2016, <https://doi.org/10.1109/COMST.2016.2559525>.
- [7] M. K. Giri and S. Majumder, "Eigenvalue-based cooperative spectrum sensing using kernel fuzzy c-means clustering," *Digital Signal Processing*, vol. 111, Apr. 2021, Art. no. 102996, <https://doi.org/10.1016/j.dsp.2021.102996>.
- [8] K. B. Vikhyath and N. A. Prasad, "Combined Osprey-Chimp Optimization for Cluster Based Routing in Wireless Sensor Networks: Improved DeepMaxout for Node Energy Prediction," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12314–12319, Dec. 2023, <https://doi.org/10.48084/etasr.6542>.
- [9] A. Rajab, "Genetic Algorithm-Based Multi-Hop Routing to Improve the Lifetime of Wireless Sensor Networks," *Engineering, Technology & Applied Science Research*, vol. 11, no. 6, pp. 7770–7775, Dec. 2021, <https://doi.org/10.48084/etasr.4484>.
- [10] T. Stephan, F. Al-Turjman, S. J. K., and B. Balusamy, "Energy and spectrum aware unequal clustering with deep learning based primary user classification in cognitive radio sensor networks," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 11, pp. 3261–3294, Nov. 2021, <https://doi.org/10.1007/s13042-020-01154-y>.
- [11] G. K. Walia, M. Kumar, and S. S. Gill, "AI-Empowered Fog/Edge Resource Management for IoT Applications: A Comprehensive Review, Research Challenges and Future Perspectives," *IEEE Communications Surveys & Tutorials*, 2023, <https://doi.org/10.1109/COMST.2023.3338015>.
- [12] M. Zhang et al., "Exploiting Deep Learning for Secure Transmission in an Underlay Cognitive Radio Network," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 1, pp. 726–741, Jan. 2021, <https://doi.org/10.1109/TVT.2021.3050104>.
- [13] Z. Chen and W. Yue, "Differential Space-time Block Coding Based Cooperative Spectrum Sensing over Fading Environments in Cognitive Radio Sensor Networks," *Journal of Information*, vol. 9, no. 15, pp. 4599–4606, 2012.
- [14] G. Wei, G. Li, J. Zhao, and A. He, "Development of a LeNet-5 Gas Identification CNN Structure for Electronic Noses," *Sensors*, vol. 19, no. 1, Jan. 2019, Art. no. 217, <https://doi.org/10.3390/s19010217>.
- [15] S. Mirjalili, "SCA: A Sine Cosine Algorithm for solving optimization problems," *Knowledge-Based Systems*, vol. 96, pp. 120–133, Mar. 2016, <https://doi.org/10.1016/j.knsys.2015.12.022>.
- [16] A. Seyyedabbasi and F. Kiani, "Sand Cat swarm optimization: a nature-inspired algorithm to solve global optimization problems," *Engineering with Computers*, vol. 39, no. 4, pp. 2627–2651, Aug. 2023, <https://doi.org/10.1007/s00366-022-01604-x>.
- [17] A. Patel, H. Ram, A. K. Jagannatham, and P. K. Varshney, "Robust Cooperative Spectrum Sensing for MIMO Cognitive Radio Networks Under CSI Uncertainty," *IEEE Transactions on Signal Processing*, vol. 66, no. 1, pp. 18–33, Jan. 2018, <https://doi.org/10.1109/TSP.2017.2759084>.
- [18] P. Kumar, N. Chauhan, M. Kumar, and L. K. Awasthi, "Clustering based opportunistic traffic offloading technique for device-to-device communication," *International Journal of System Assurance Engineering and Management*, vol. 14, no. 3, pp. 827–839, Jul. 2023, <https://doi.org/10.1007/s13198-021-01136-5>.
- [19] S. Panbude, B. Iyer, A. B. Nandgaonkar, and P. S. Deshpande, "DFPC: Dynamic Fuzzy-based Primary User Aware clustering for Cognitive Radio Wireless Sensor Networks," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12058–12067, Dec. 2023, <https://doi.org/10.48084/etasr.6279>.

Investigation of Maximum Mid-Span Displacement and Reaction Forces in Fiber-reinforced Concrete Beams subjected to Impact

Zena Ahmed Mahmoud

Civil Engineering Department, University of Anbar, Iraq
zen21e1009@uoanbar.edu.iq

Muhannad Haqi Aldosary

Dam and Water Resources Department, University of Anbar, Iraq
muhannad_dosary@uoanbar.edu.iq

Abdulkader Ismail Al-Hadithi

Civil Engineering Department, University of Anbar, Iraq
abdulkader.alhadithi@uoanbar.edu.iq

Received: 10 November 2023 | Revised: 11 December 2023 | Accepted: 11 December 2023

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6606>

ABSTRACT

Self-Compacting Fiber-Reinforced Concrete (SCFRC) is a specialized type of concrete that combines the properties of Self-Compacting Concrete (SCC) with the addition of fibers for reinforcement. SCFRC is designed to have excellent flowability and self-leveling characteristics while providing enhanced tensile strength, ductility, and crack resistance. This paper presents a discussion on the topic of SCFRC and the impact load behavior of SCFRC beams reinforced with Waste Plastic Fibers (WPFs). A comparison with reinforced concrete beams without fibers is also conducted. This study aims to predict the maximum mid-span displacement and the maximum reaction force of the fiber concrete beams under impact load. Twelve beams that represent the total adopted parameters were tested under impact loading. The beams were divided into three main groups according to the longitudinal steel ratio. The steel ratio was varied by using steel bars of 10, 8, and 6 mm diameter, with PET waste fibers with different volume ratios $V_f\%$ of 0, 0.5, 0.75, and 1%. The results showed that the use of beams is reinforced with ρ_{max} , $\rho_{max} > \rho > \rho_{min}$ and ρ_{min} having reduced maximum deflection by 24.23%, 35.9%, and 46.28%, respectively, when using WPFs with a volumetric value of 1%. This paper also covers work steps, model details, and the tests that were carried out on the specimens, which were made from materials available in local markets.

Keywords-*self-compacting reinforced concrete; waste plastic fibers; impact load*

I. INTRODUCTION

The behavior of reinforced concrete under impact is an active area of research that continues to be explored. While significant progress has been made in understanding the response of concrete structures to static loads, the dynamic behavior of reinforced concrete subjected to impact loads is still a challenging topic [1]. Examples include reinforced concrete structures designed to resist accidental loading scenarios, such as falling rock impact, vehicle or ship collisions with buildings, bridges, or offshore facilities; and structures that are used in high-threat or high-hazard applications, such as military fortification structures or nuclear facilities. Several factors contribute to the complexity of studying the impact behavior in reinforced concrete [2]. These factors include the heterogeneous nature of concrete, the presence of reinforcement, the dynamic response of different materials, and

the variability in impact loading conditions. Additionally, the behavior can vary depending on factors like material composition, mix design, reinforcement details, and specimen geometry [3]. Research and experimentation have been carried out to assess the reaction of continuous steel-reinforced, Ultra-High-Performance Fiber-Reinforced Concrete (UHPFRC) beams under low-velocity loading. In order to thoroughly assess the behavior of UHPFRC beams subjected to low-velocity loads, these investigations combine experimental testing, numerical simulations, and analytical modeling techniques [4]. The inertia force mostly resists the impact force due to the impact loading's brief duration. Owing to the dynamic equilibrium that is created as a result, the RC member's bending moment and internal force are more complex [5–7]. The most commonly used design codes for structures that specialize in impact loading, like ACI 349-13 [8]

and UFC 340-02 [9], are only applicable to special structures like nuclear power plants and military installations, and their primary concerns are focused on high-velocity impact loading, like aircraft impact and explosion. This suggests that the application of these rules to general social infrastructure, such as bridges and buildings, is restricted because these structures are more likely to be exposed to low-velocity impact loading than high-velocity impact loading, including rock falls, cars, and ship accidents. In contrast to high-velocity impact phenomena, where local damage is the primary indicator of structural failure, when structures are subjected to low-velocity impact loads, global reactions, such as displacement and rotation, are the prominent markers of structural failure. However, there is currently no precise technology available to assess and control the overall behavior of concrete members under low-velocity impact loads [10].

The second most frequently dumped plastic material, behind polythene, is polythene terephthalate (PET) [11, 12]. PET, one of the most produced and commonly used plastics in the world, is utilized to package soft drinks, drinking water, food, and other consumer items [13, 14]. Plastic waste poses several environmental challenges. Plastics are not easily biodegradable and can persist in the environment for hundreds of years, leading to long-term pollution. Improper disposal of plastic waste, such as littering or improper landfilling, can result in plastic debris ending up in the environment. Plastic waste also can clog drainage systems, leading to flooding in urban areas, or contribute to the formation of microplastics, which are small particles that can accumulate in ecosystems and have detrimental effects on organisms. Plastic waste recycling in construction, i.e. the use of shredded plastic waste as aggregates or plastic fibers in the production of concrete is a well-known method of recycling. This approach aims to reduce the environmental impact of plastic waste while incorporating its beneficial properties into construction materials. However, it is important to note that the exact proportions and types of plastic used in concrete mixtures can vary depending on the desired properties and structure requirements. SCFRC finds applications in a wide range of construction projects, including bridges, tunnels, high-rise buildings, precast elements, and infrastructure projects. It is commonly used in structural elements, like beams, columns, slabs, and walls, where improved crack control and enhanced durability are critical. Overall, recycling plastic waste in construction, particularly in the production of SCFRC, can contribute to sustainable building practices, reduce the environmental impact of plastic waste, and promote the efficient use of resources [15, 16].

This study investigated the effect of the ratio of plastic fibers incorporated in SCFRC mix design. The behavior of concrete beams under impact load was studied while varying this ratio. The main goals of this study are to improve the performance of SCFRC mix designs and create design standards for structures using SCFRC beams under low-velocity loading conditions. This study offers important insights on how to improve the behavior of SCFRC beams, making them better suited for structural applications where low-velocity loads are an important factor. SCC provides numerous benefits in contrast to conventional concrete. These advantages consist of enhanced construction quality, expedited

construction process, cost reduction, improved quality of in-place concrete, even in challenging casting conditions, and decreased occurrences of accidents, noise, and vibrations. Additionally, it also facilitates the achievement of a superior surface quality [17]. As far as is known, no previous research has been conducted on self-compacting reinforced concrete beams containing plastic waste fibers under impact loads, representing a considerable gap in the literature. This research study significantly advances the comprehension of the structural behavior of SCC beams upon the inclusion of plastic fibers, with a specific focus on deflection and reaction characteristics.

II. EXPERIMENTAL WORK

A. Materials

Ordinary Portland Cement (Al-Mass), which met Iraqi requirements (IQS No.5/ 2019) [18], was the cement utilized in this project (see Table I). Additionally, coarse aggregates (with a particle size of up to 10 mm) and fine aggregates, which comply with the Iraqi standard specification (I.Q.S.) No.45/84 (see Table II) [19] were utilized. Tap water was used for mixing and curing. The additives employed in this research included EPSILONE I 21 superplasticizer and High Range Water Reducing/Superplasticizer, confirming to ASTM C 494 Type G and F [20]. Silica Fume (SF) Type Mega Add MS(D) with a specific gravity of 2.2 was used in the SCC mixes. The United Arab Emirates (UAE) produces SF, according to ASTM C1240 [21]. Waste Plastic Fiber (WPFs) with 3 mm width and 30 mm length that were produced by a paper shredded machine (See Table III) were also utilized.

TABLE I. PHYSICAL CEMENT COMPOSITION PROPERTIES

Physical properties	Result	Limit No.5/2019
Fineness using the Blain method (m ² /kg)	361	≥ 250
Time of initial setting (min)	160	≥ 45
Time of Final setting (hr)	260	≤ 10
Soundness (mm)	1	≤ 10
Autoclave %	0.1	≤ 0.8%
The compressive strength of mortar		
2 days (MPa)	20	≥ 10
28 days (MPa)	37	≥ 32.5

TABLE II. COARSE AND FINE AGGREGATES SIEVE ANALYSIS

Sieve Size (mm)	Passing %	Iraqi specification No. 45/1984 limit
Coarse aggregates		
20	100	100
14	100	90-100
10	82.9	50-85
5	8	0-10
Fine aggregates		
10	100	100
4.75	97	90-100
2.36	81	75-100
1.18	68	55-90
0.6	43	35-55
0.3	9	8-30
0.15	2	0-10

TABLE III. DIMENSIONS AND PHYSICAL PROPERTIES OF WPF

Property	Description
Aspect Ratio	28.025
Tensile Strength (MPa)	105
Density (kg/m ³)	1.1
Water absorption	0.00

B. Mix Design

The specifics of the experimental mixtures are presented in Table IV including the amount of coarse aggregates, plasticizers, and fillers according to the limitations indicated in the [22].

TABLE IV. MIX PROPORTIONS FOR SCC (kg/m³).

Mix	Fiber content%	Cement	Sand	Coarse aggregates	Water	SF	Superplasticizers
MO*	0	475	950	700	169	50	13
MA*	0.5	475	950	700	169	50	13
MB*	0.75	475	950	700	169	50	13
MC*	1	475	950	700	169	50	13

*MO (mix class 0 with 0% fiber content), MA (mix class A with 0.5% fiber content), MB (mix class B with 0.75% fiber content), *MC (Mix class C with 1% fiber content).

C. Testing Procedure

1) Compressive Strength and Static Modulus of Elasticity Tests

Three standard cylindrical specimens with diameter and height measurements of 150 and 300 mm, respectively, were used for compressive strength tests at a rate of 5.30 KN/s, subjected to ASTM C39-05 [23] and for static modulus of elasticity tests subjected to the ASTM C469-02. [24] It is possible to determine Young's modulus in a compressive test machine at a loading rate of 5.3 kN/s by encircling the sample with a 150 mm diameter compressor-meter ring and measuring the longitudinal axis of the sample's deformation through compression via a dial gauge with an accuracy of 0.002 mm. The value of elastic modulus can be established by:

$$E_c = \frac{S_2 - S_1}{\epsilon_2 - 0.00005} \quad (1)$$

where E_c is the static modulus of elasticity (MPa), S is the stress corresponding to ϵ (MPa), ϵ_1 is the longitudinal strain of 0.00005, and S_2 is the 40% of the ultimate stress.

2) Splitting Tensile Strength Test

This examination was conducted in accordance with ASTM C496-11 [25]. After 28 days of curing, it was carried out on 150×300 cylindrical samples. The tests were conducted in the same compression strength device at a rate of 0.94 kN/s.

D. Testing Program for Reinforced Concrete Beams

The beams were designed with shear capacity greater than flexural capacity, according to ACI 318-19code [26]. The dimensions of the proposed model were 1500 mm length, 100

mm width, and 150 mm depth. Since 12 wooden molds were created for casting the aforementioned beams, the molds were designed with separate parts to allow for easy editing of the models after solidification. Rebars with diameters of 6, 8, 10 mm were used in the tension zone and two with diameters of 6 mm were used in the compression zone. Shear reinforcement (stirrups) with a diameter of 6 mm were utilized. All sides of the net cover were 20 mm thick. Figures 1-2 depict the specimen's geometry and cross-section details.



Fig. 1. Geometry of the specimen and cross-section data.

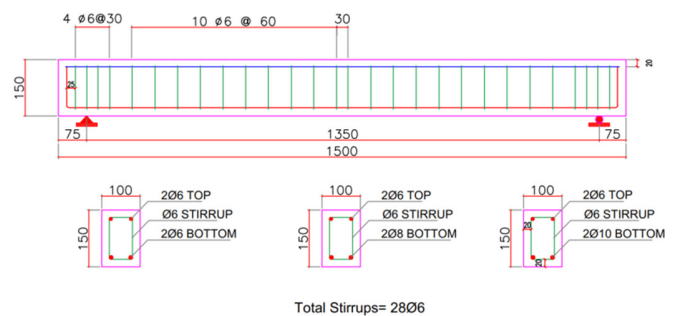


Fig. 2. Details of reinforced concrete beams.

E. Impact Loading Test

The tests were carried out at the University of Technology's Laboratory. The same procedure was followed to test all specimens. The beam was supported by the testing rig and the released dropped mass was led by gravity without any other external force. The specimens were placed in the testing frame with their finished faces up. The falling mass was dropped once, and the deflection was measured. The following steps summarize the test procedure:

- All of the equipment and sensors were properly positioned.
- The beam was securely fastened to the rigid supporting frame.
- The two load cells are located on the bottom face of the beam.
- The laser sensor was positioned beneath the center of the beam.
- The projectile was held in place by a cable.
- The cable was withdrawn and left to allow for a free fall.
- The impacted falls vertically to the sample midspan using a steel tube.

Figure 3 depicts the stages of beam testing, the setup, and the support system for the impact test. The reinforced concrete beam was precisely positioned in the center of the testing frame. A circular-section tube was used as a vertical guide for the falling mass, ensuring a mid-span impact. The impact body weight was constant at 24.450 kg. The impact body was shaped like a cylinder and was made up of two parts: a solid steel ball with a diameter of 100 mm and a weight of 4.450 kg at the

bottom to deliver the hammer's impact energy to the sample and a steel cylinder with a height of 330 mm and a diameter of 100 mm at the top to deliver the hammer's impact energy to the sample. The height distance from the impact body's face to the beam surface was constant at 2.1 m for all samples. The total applied dropped mass was equal to the combined mass of the solid steel ball and steel cylinder.

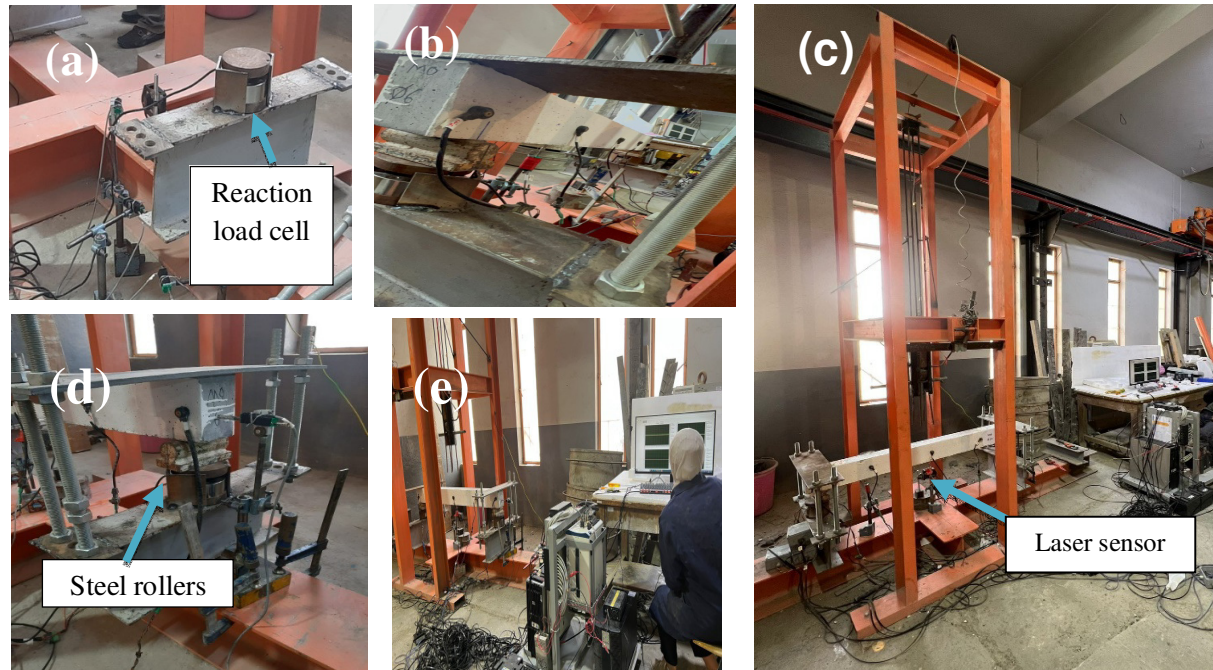


Fig. 3. Instruments used for impact testing.

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. Compressive Strength

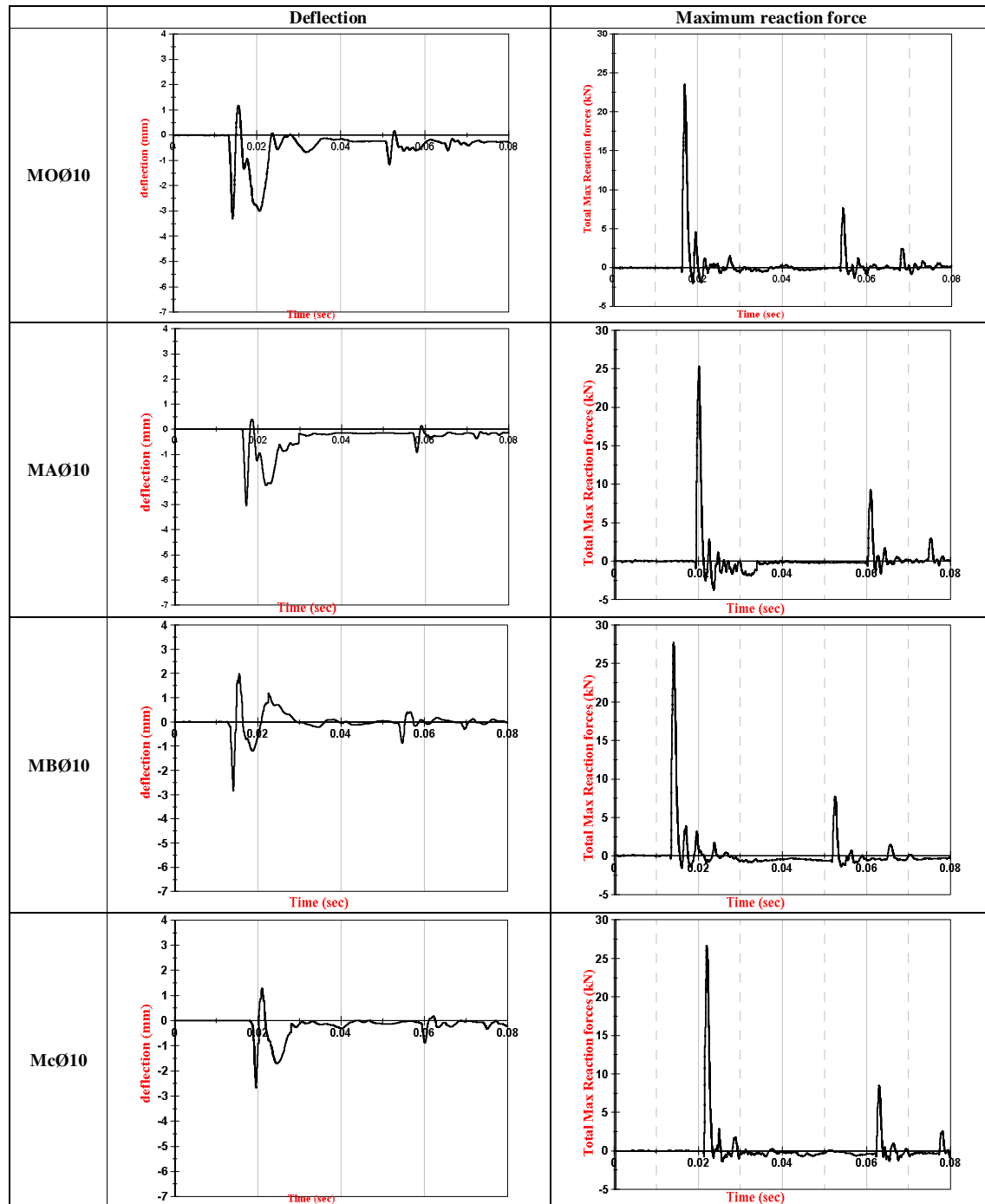
Table V shows the impact of the increase in the PET ratio in the concrete mixture on compressive strength. The experiment was carried out on samples at the age of 28 days with different PET ratios. The results show that the increase in the ratio of fiber increases the compressive strength up to 0.75%. This improvement can be attributed to the ability of WPFs to lessen these tensions, bear the stresses, and bridge the cracks. When microscopic cracks develop in the bonding material (matrix), the fibers in the area around the cracks work to hinder the development of the cracks and prevent them from spreading. This causes the cracks to move in a zigzag pattern, requiring more energy to keep spreading and high pressures to collapse with different fiber content [27-29]. Then any increase in the proportion results in a decrease in compressive strength [30, 31]. This decrease was caused by the increase in WPF volume ratios, since it led to an uneven distribution of fibers in the concrete mix, which in turn led to fiber gathering and clumping together. Air gaps were build up beneath the PF as a consequence of the cement paste's declining homogeneity and adhesion to fiber surfaces [32].

B. Static Modulus of Elasticity Test

Adding WPFs to concrete mixture can have a significant impact on the modulus of elasticity of concrete. Plastic fibers, when incorporated into a concrete mixture, can enhance the mechanical properties of the resulting material. The addition of plastic fibers can increase the tensile strength and ductility of concrete, as well as improve its resistance to cracking and shrinkage. Consequently, this can lead to a raise in the modulus of elasticity. [33]. Table V displays the influence of the increase in the PET ratio in the concrete mixture on the modulus of elasticity. The results also indicate that the raise in the ratio of fiber increases the modulus of elasticity up to 0.75%. The bond strength between the fibers and the concrete mix has been enhanced by the close proximity and the short gaps between the fibers, the fibers obstruction of the cracks' growth and bridging, and the resistance to stresses. The modulus of elasticity begins to decline but still remains higher than that of the reference mix. This decline was caused by the fibers' erratic distribution, which made the fibers to aggregate, weaken the link between the fibers and the mix, and increase the air gaps beneath the fibers. In addition to the fact that plastic fibers have a lower modulus of elasticity than concrete at advanced ages, adding too much fiber content to concrete weakens the structure and lowers its modulus of elasticity.

TABLE V. PERCENTAGE INCREASE IN COMPRESSION STRENGTH, MODULUS OF ELASTICITY AND SPLITTING STRENGTH COMPARED TO PET RATIO

Mix	Fiber content%	Compressive strength at 28 days (MPa)	Increment (%)	Modulus of elasticity (GPa)	Increment (%)	Splitting strength f_t (MPa)	Increment (%)
MO	0	48.8	0	31.4832	0	4.806	0
MA	0.5	50.6	3.688	32.4131	2.9536	5.485	14.128
MB	0.75	54.2	11.065	33.0524	4.9842	5.908	22.929
MC	1	52.7	7.991	32.7211	3.9319	5.722	19.059

Fig. 4. Deflection time history and total maximum reaction force for the first group (ρ_{max}).

C. Splitting Strength

Table V shows the results of splitting tests for tensile strength for all concrete mixtures. The addition of WPFs to concrete mixes resulted in an increase in the compared tensile strength [34]. Adding fibers at a volume ratio of 0.5–0.75% demonstrated the highest increase in tensile strength. However, when the volumetric ratios of fibers exceeded 1%, the splitting tensile strength started to decrease due to the irregular fiber distribution, fiber agglomeration, and increased air voids. Despite the decrease, the fiber-containing concrete still had a higher tensile strength than the reference mix. In the reference mix, the failure mode after the initiation of cracks was the splitting of the cylinder into two parts. However, in the fiber-containing concrete, there was no sudden separation into two parts. Instead, the concrete cylinder exhibited resistance, and the cracks continued to develop and spread. This behavior persisted even under increasing tensile strength due to overloading until the ultimate failure was reached. Figure 5 provides a visual representation of this phenomenon.



Fig. 5. Failure mode of concrete cylinders under splitting strength.

D. Impact Loading Testing

1) Deflection Time-History

Figures 4, 6, 7 illustrate the representative time history responses of deflection. The shapes of the time histories of deflection were similar in reinforced concrete samples with the same diameter of reinforcement. It can be concluded that when the reinforcement ratio remained constant, increasing the proportion of plastic fibers clearly reduced both maximum deflection and residual deflection values. This means that the addition of these fibers increased the ductility of concrete beams made of SCC and containing WPFs. The factors mentioned above apply to the role of WPFs in enhancing ductility, they are uniformly distributed within the structure of the concrete mixture, augmenting homogeneity and reducing the amount of voids therein while also making the concrete body more cohesive and hard. When microcracks start to emerge within the matrix, WPFs attempt to stop their spread in the surrounding region by regulating their growth. Therefore, the cracking route becomes winded, requiring more energy to continue. When beams are subjected to impact load, fibers aid in minimizing deformations and excessive bending. Consequently, the elasticity of SCC containing WPFs was improved by the fiber addition.

2) Time Histories of Total Maximum Reaction Force

Figures 4, 6, 7 represent the time histories of total maximum reaction force. The shapes of the time histories of

total maximum reaction force were similar in reinforced concrete samples with the same diameter of reinforcement. These findings demonstrate SCC's total maximum reaction force, and for beams containing WP fibers with different volumetric percentages (0.5%, 0.75%, and 1%). The reasons explained above are also related to the role of WPFs in increasing the total reaction forces due to the qualities of sizable compressive strength and sufficient tensile strength, taking into account that both tensile and compressive strength are very important in concrete. The outcomes of testing impact loads showed that adding WPFs to concrete increases impact resistance, and this can be attributed to the ability of WPFs to absorb the energy resulting from impact loads, transforming concrete to a more ductile material. In addition, fibers reduce the crack width and crack bridging leading to large energy absorption concrete. The addition of WPFs with different volumetric ratios in SCCs has a positive effect as it raises impact resistance and thus raises the total maximum reaction force for all mixes concrete containing fibers. After the volumetric ratio of 0.75%, the total maximum reaction force began to decrease but it was still higher than that of the reference mix. The concrete exhibits ductile rather than brittle behavior as the outcome of the inclusion of WPFs.

3) Effect of Test Variables on the Impact Behaviors

Table VI portrays the impact responses with respect to the test variables. The values of total maximum reaction force, maximum deflection, and residual deflection are shown. Table VI confirms that the impact responses were affected by the experimental variables. The deflection and the remaining deflection values are inversely proportional to the PET ratio in the concrete mix and rebar diameters. After the result comparison, the following were noted:

- The first collection of SCCs beams reinforced with (ρ_{max}) had variable WPFs percentages of (0, 0.5, 0.75, and 1%) under impact loading. The findings reveal that the use of WPFs has a positive impact on every beam in this group. Using such fibers had an improvement in SCC beam behavior. The deflection value at the stage of failure was reduced when compared to reference mix SCC beams. It can be established that when employing 1% WPFs, the amount of the maximum deflection during the first impact was 24.23% less than that of the beams made by the reference mix. The highest increase in the value of the total maximum reaction force according to the reference SCCs beam was 18% and was achieved when using 0.75% WPFs. The significant effects of combining longitudinal steel reinforcement and WPFs become more apparent in this group when compared to the other groups.

The second collection of SCCs beams was reinforced with $\rho_{max} > \rho > \rho_{min}$. The observations that can be inferred are the same as those reached above. The greatest reduction in the value of maximum deflection during the first impact, compared to the reference SCCs beam, was equal to 35.9%. This reduction was achieved by incorporating WPFs with a volumetric ratio of 1%. The maximum increment in the value of total maximum reaction force according to the reference RC beam was 21.77%. This was achieved when employing WPFs with a volumetric ratio equal to 0.75%.

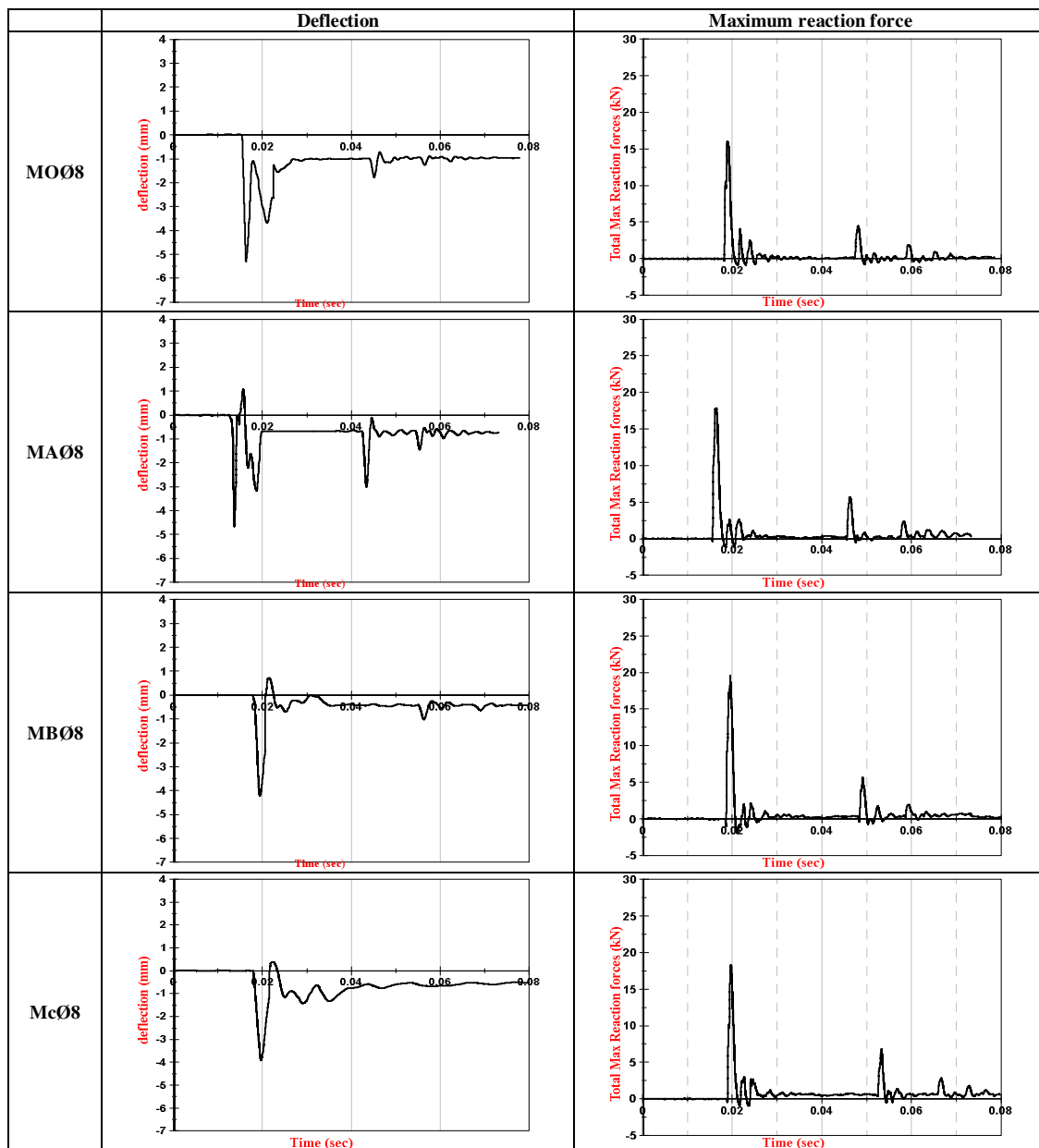


Fig. 6. Deflection time history and total maximum reaction force for the first group ($\rho_{max} > \rho_{min}$).

TABLE VI. IMPACT RESPONSES OF TOTAL MAXIMUM REACTION FORCE, MAXIMUM DEFLECTION, AND RESIDUAL DEFLECTION

Group	Beam	WPF %	Total Maximum reaction force (KN)	Increment (%)	Maximum deflection (mm)	Decrease (%)	Residual deflection (mm)
G1 Ø10 (ρ_{max})	MOØ10	0%	23.5	0	3.317	0	0.25
	MAØ10	0.5%	25.28	7.57	3.04	8.97	0.198
	MBØ10	0.75%	27.73	18	2.837	16.9	0.062
	MCØ10	1%	26.67	13.49	2.67	24.23	0.091
G2 Ø8 (ρ)	MOØ8	0%	16.03	0	5.3	0	0.82
	MAØ8	0.5%	17.84	11.29	4.67	13.5	0.77
	MBØ8	0.75%	19.52	21.77	4.21	25.89	0.49
	MCØ8	1%	18.29	14.09	3.9	35.9	0.65
G3 Ø6 (ρ_{min})	MOØ6	0%	13.7	0	6.89	0	3.09
	MAØ6	0.5%	16.2	18.24	5.68	21.3	2.99
	MBØ6	0.75%	17.889	30.57	5.14	34.05	2.9
	MCØ6	1%	16.67	21.67	4.71	46.28	2.97

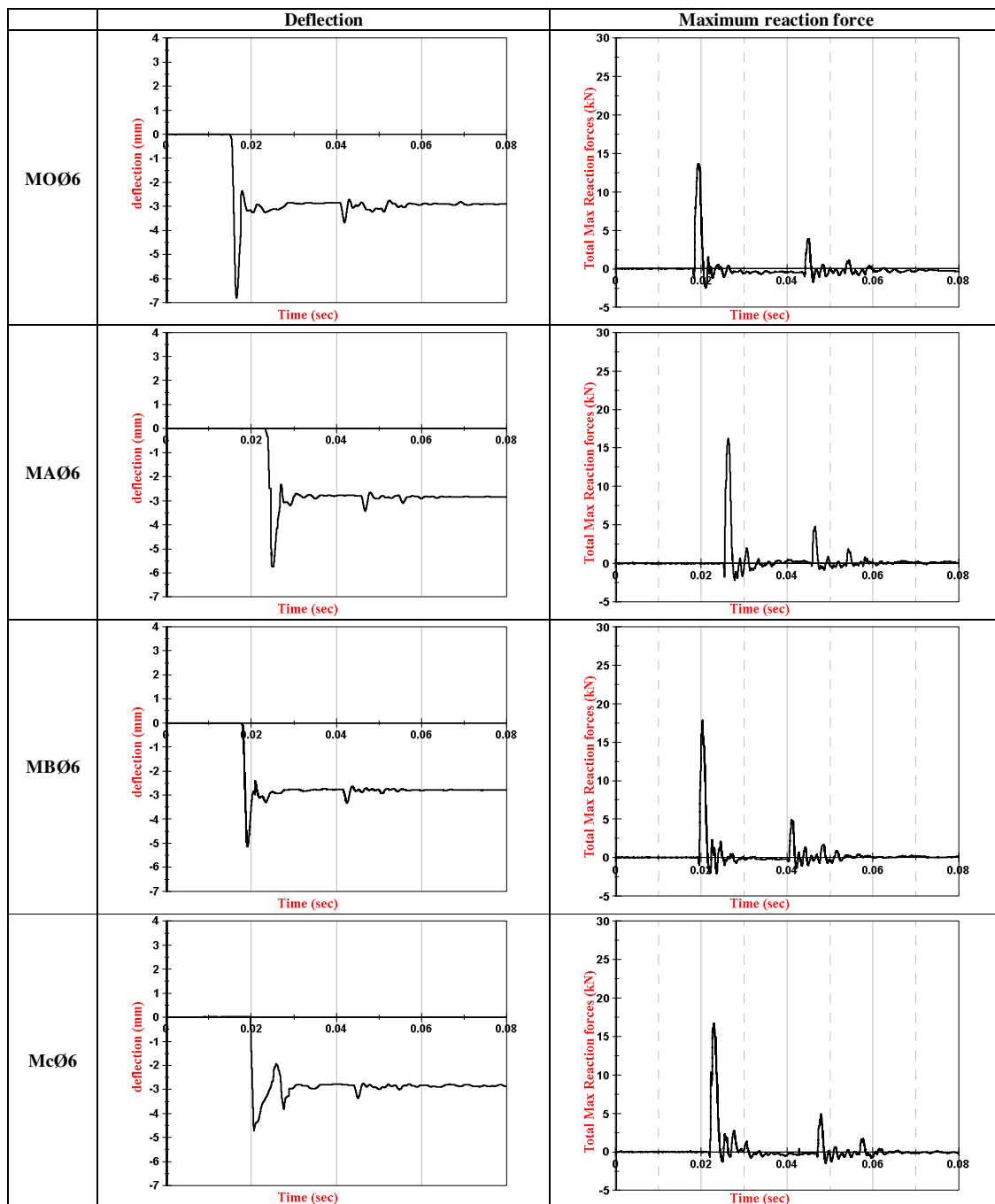


Fig. 7. Deflection time history and total maximum reaction force for first group (ρ_{min}).

- The third group of SCCs beams was reinforced with the minimum area of steel (ρ_{min}). The combined effect of longitudinal steel reinforcement and WPFs in this group appears to be lower than in the other groups, due to the minimum amount of steel utilized for the reinforcement of the SCCs beams in this group. However, the addition of WPFs had a significant influence on reducing the maximum deflection. The highest reduction achieved in the value of maximum deflection during the first impact was equivalent to 46.28% compared to the beam made with the reference mix. This reduction was achieved when using WPFs with a volumetric value of 1%. The maximum increment in the value of total maximum reaction force in comparison with the reference SCCs beam was 30.57%, achieved when using WPFs with a volumetric ratio equal to 0.75%.
- It can be concluded that, when the reinforcement ratio is held constant, increasing the proportion of plastic fibers has a noticeable impact on the maximum deflection and residual deflection values, reducing both. This indicates that

the addition of fibers augmented the ductility of concrete beams made of SCC and containing WPFs.

- When waste plastic fibers and steel reinforcement were combined in reinforced concrete beams subjected to impact loading, their synergistic effect can lead to improved performance. Generally, they are able to bridge the cracks and redistribute the stresses in the beam body to carry more load with a smaller deflection at the same load level [35].

IV. CONCLUSION

The current research study significantly advances the comprehension of the structural behavior of Self-Compacting Concrete (SCC) beams after the inclusion of plastic fibers, with a specific focus on deflection and reaction characteristics. In practice, plastic waste fibers can be added to reduce the deflection and reaction of SCC beams, especially when there is a pressing need to eliminate the deflection without increasing beam's depth. Based on the findings of the experimental investigation, the following conclusions can be derived:

Adding 0.75% plastic waste fibers to concrete mixtures has a number of benefits, including increased compressive strength up to 11.065%, increased modulus of elasticity up to 4.984%, increased splitting strength up to 22.929% and increased maximum reaction force. Beyond this addition percentage, compressive strength, modulus of elasticity, splitting strength, and maximum reaction force are decreased. Furthermore, maximum deflection and residual deflection are decreased, and the significant effects of combining longitudinal steel reinforcement and plastic waste fibers become more apparent in the first group (ρ_{max}). When increasing the reinforcement ratio, better performance was seen under impact loading circumstances. As the reinforcement ratio increased, there was a discernible decrease in the SCFRC beams' maximum and residual deflections during the initial impact event. Additionally, the behavior was better when the reinforcement ratio was larger, especially when it came to deflection recovery.

Overall, the incorporation of plastic waste fibers in a concrete mixture can enhance its mechanical properties, increase its cracking resistance, and improve the distribution of forces within the material. The specific benefits and performance improvements depend on the ratio of plastic waste fibers. The optimum ratio of 0.75% is recommended.

REFERENCES

- [1] T. D. Hrynyk and F. J. Vecchio, "Behavior of Steel Fiber-Reinforced Concrete Slabs under Impact Load," *Structural Journal*, vol. 111, no. 5, pp. 1213–1224, Sep. 2014, <https://doi.org/10.14359/51686923>.
- [2] J. M. Adam, M. Buitrago, E. Bertolesi, J. Sagaseta, and J. J. Moragues, "Dynamic performance of a real-scale reinforced concrete building test under a corner-column failure scenario," *Engineering Structures*, vol. 210, May 2020, Art. no. 110414, <https://doi.org/10.1016/j.engstruct.2020.110414>.
- [3] S. D. Adhikary, B. Li, and K. Fujikake, "Residual resistance of impact-damaged reinforced concrete beams," *Magazine of Concrete Research*, vol. 67, no. 7, pp. 364–378, Apr. 2015, <https://doi.org/10.1680/mac.14.00312>.
- [4] D.-Y. Yoo, N. Banthia, S.-W. Kim, and Y.-S. Yoon, "Response of ultra-high-performance fiber-reinforced concrete beams with continuous steel reinforcement subjected to low-velocity impact loading," *Composite Structures*, vol. 126, pp. 233–245, Aug. 2015, <https://doi.org/10.1016/j.compstruct.2015.02.058>.
- [5] S. Saatci and F. J. Vecchio, "Effects of Shear Mechanisms on Impact Behavior of Reinforced Concrete Beams," *Structural Journal*, vol. 106, no. 1, pp. 78–86, Jan. 2009, <https://doi.org/10.14359/56286>.
- [6] D. M. Cotsovos, N. D. Stathopoulos, and C. A. Zeris, "Behavior of RC Beams Subjected to High Rates of Concentrated Loading," *Journal of Structural Engineering*, vol. 134, no. 12, pp. 1839–1851, Dec. 2008, [https://doi.org/10.1061/\(ASCE\)0733-9445\(2008\)134:12\(1839\)](https://doi.org/10.1061/(ASCE)0733-9445(2008)134:12(1839)).
- [7] D. M. Cotsovos, "A simplified approach for assessing the load-carrying capacity of reinforced concrete beams under concentrated load applied at high rates," *International Journal of Impact Engineering*, vol. 37, no. 8, pp. 907–917, Aug. 2010, <https://doi.org/10.1016/j.ijimpeng.2010.01.005>.
- [8] *ACI 349M-13(2013), Code Requirements for Nuclear Safety-Related Concrete Structures and Commentary*. Farmington Hills, MI, USA: ACI Concrete, 2013.
- [9] *UFC 3-340-02 Structures to Resist the Effects of Accidental Explosions*. Washington DC, USA: UFC, 2014.
- [10] J. Abd and I. K. Ahmed, "The Effect of Low Velocity Impact Loading on Self-Compacting Concrete Reinforced with Carbon Fiber Reinforced Polymers," *Engineering, Technology & Applied Science Research*, vol. 11, no. 5, pp. 7689–7694, Oct. 2021, <https://doi.org/10.48084/etasr.4419>.
- [11] D. de Mello, S. H. Pezzin, and S. C. Amico, "The effect of post-consumer PET particles on the performance of flexible polyurethane foams," *Polymer Testing*, vol. 28, no. 7, pp. 702–708, Oct. 2009, <https://doi.org/10.1016/j.polymertesting.2009.05.014>.
- [12] D. Foti, "Use of recycled waste pet bottles fibers for the reinforcement of concrete," *Composite Structures*, vol. 96, pp. 396–404, Feb. 2013, <https://doi.org/10.1016/j.compstruct.2012.09.019>.
- [13] M. Frigione, "Recycling of PET bottles as fine aggregate in concrete," *Waste Management*, vol. 30, no. 6, pp. 1101–1106, Jun. 2010, <https://doi.org/10.1016/j.wasman.2010.01.030>.
- [14] F. A. Al-Fahdawi, A. I. Al-Hadithi, and J. A. Al-Asafi, "The Mechanical Properties of Ferrocement Mortar with Waste Plastic Fibers at Elevated Temperatures," *Engineering, Technology & Applied Science Research*, vol. 12, no. 5, pp. 9347–9350, Oct. 2022, <https://doi.org/10.48084/etasr.5209>.
- [15] J. Santos, A. Pham, P. Stasinopoulos, and F. Giustozzi, "Recycling waste plastics in roads: A life-cycle assessment study using primary data," *Science of The Total Environment*, vol. 751, Jan. 2021, Art. no. 141842, <https://doi.org/10.1016/j.scitotenv.2020.141842>.
- [16] A. Rahimi and J. M. García, "Chemical recycling of waste plastics for new materials production," *Nature Reviews Chemistry*, vol. 1, no. 6, pp. 1–11, Jun. 2017, <https://doi.org/10.1038/s41570-017-0046>.
- [17] N. Memon *et al.*, "A Review on Self Compacting Concrete with Cementitious Materials and Fibers," *Engineering, Technology and Applied Science Research*, vol. 8, no. 3, pp. 2969–2974, Jun. 2018, <https://doi.org/10.48084/etasr.2006>.
- [18] *I.Q.S. No. 5/2019 Specification, Portland Cement*. Baghdad, Iraq: Central Organization for Standardization & Quality Control (COSQC), 2019.
- [19] *I.Q.S. No. 45 Aggregate from Natural Sources for Concrete and Construction*. Baghdad, Iraq: Central Organization for Standardization & Quality Control (COSQC), 1984.
- [20] *ASTM C494/C494M-08(2008), Standard Specification for Chemical Admixtures for Concrete*. West Conshohocken, PA, USA: ASTM International, 2008.
- [21] *ASTM C1240-15(2015), Standard Specification for Silica Fume Used in Cementitious Mixtures*. West Conshohocken, PA, USA: ASTM International, 2015.
- [22] *Specification and Guidelines for Self-Compacting Concrete*. Farnham, UK: EFNARC, 2002.
- [23] *ASTM C39/C39M-05(2005), Standard Test Method for Compressive Strength of Cylindrical Concrete Specimens*. West Conshohocken, PA, USA: ASTM International, 2005.

- [24] ASTM C469-02(2002), *Standard Test Method for Static Modulus of Elasticity and Poisson's Ratio of Concrete in Compression*. West Conshohocken, PA, USA: ASTM International, 2002.
- [25] ASTM C496/C496M-11(2011), *Standard Test Method for Splitting Tensile Strength of Cylindrical Concrete Specimens*. West Conshohocken, PA, USA: ASTM International, 2011.
- [26] ACI 318-19(2019), *Building Code Requirements for Structural Concrete and Commentary*. Farmington Hills, MI, USA: ACI Concrete, 2019.
- [27] J. Barros, E. Pereira, and S. Santos, "Lightweight Panels of Steel Fiber-Reinforced Self-Compacting Concrete," *Journal of Materials in Civil Engineering*, vol. 19, no. 4, pp. 295–304, Apr. 2007, [https://doi.org/10.1061/\(ASCE\)0899-1561\(2007\)19:4\(295\)](https://doi.org/10.1061/(ASCE)0899-1561(2007)19:4(295)).
- [28] A. Al-Hadithi, "Some properties of concrete using waste plastic fiber with a very small percentages," in *The First International Engineering Sciences Conference of Aleppo University*, Aleppo, Syria, Nov. 2008, pp. 1–8.
- [29] T. Ochi, S. Okubo, and K. Fukui, "Development of recycled PET fiber and its application as concrete-reinforcing fiber," *Cement and Concrete Composites*, vol. 29, no. 6, pp. 448–455, Jul. 2007, <https://doi.org/10.1016/j.cemconcomp.2007.02.002>.
- [30] R. Nibudey, P. B. Nagarnaik, D. Parbat, and A. Pande, "Strength and fracture properties of post consumed waste plastic fiber reinforced concrete," *International Journal of Civil, Structural, Environmental and Infrastructure Engineering Research and Development*, vol. 3, no. 2, pp. 9–16, Jun. 2013.
- [31] R. Nibudey, P. B. Nagarnaik, D. Parbat, and A. Pande, "Strengths Prediction of Plastic fiber Reinforced concrete (M30)," *International Journal of Engineering Research and Applications*, vol. 3, no. 1, pp. 1818–1825, Jan. 2013.
- [32] R. Sharma and P. P. Bansal, "Use of different forms of waste plastic in concrete – a review," *Journal of Cleaner Production*, vol. 112, pp. 473–482, Jan. 2016, <https://doi.org/10.1016/j.jclepro.2015.08.042>.
- [33] A. I. A.- Hadithi and M. A. Abbas, "The Effects of adding Waste Plastic Fibers on the Mechanical Properties and Shear Strength of Reinforced Concrete Beams," *Iraqi Journal of Civil Engineering*, vol. 12, no. 1, pp. 110–124, 2018.
- [34] Y. Ghernouti, B. Rabehi, T. Bouziani, H. Ghezraoui, and A. Makhloufi, "Fresh and hardened properties of self-compacting concrete containing plastic bag waste fibers (WFSCC)," *Construction and Building Materials*, vol. 82, pp. 89–100, May 2015, <https://doi.org/10.1016/j.conbuildmat.2015.02.059>.
- [35] A. Abed, A. Al-Hadithi, and A. S. Mohammed, "The effects of adding waste plastic fibers on some properties of roller compacted concrete," *MATEC Web of Conferences*, vol. 162, 2018, Art. no. 02008, <https://doi.org/10.1051/mateconf/201816202008>.

A Framework for Efficient and Accurate Automated CLO and PLO Assessment

Hafedh Mahmoud Zayani

Department of Electrical Engineering, College of Engineering, Northern Border University, Saudi Arabia
hafedh.zayani@nbu.edu.sa (corresponding author)

Walid Abdelfattah

Department of Mathematics, College of Arts and Science, Northern Border University, Saudi Arabia
walid.abdelfattah@nbu.edu.sa

Rahma Sellami

Department of Computer Science, Applied College, Northern Border University, Saudi Arabia
rahma.ali@nbu.edu.sa

Jihane Ben Slimane

Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
jehan.saleh@nbu.edu.sa

Amani Kachoukh

Department of Information Systems, Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
amani.khasookh@nbu.edu.sa

Received: 31 December 2023 | Revised: 19 January 2024 | Accepted: 23 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6846>

ABSTRACT

Accurate and efficient learning outcome assessment is crucial for ensuring high-quality education, but traditional methods can be time-consuming, error-prone, and inconsistent. We developed a novel Excel Macro-enabled framework for automating the evaluation of Course Learning Outcomes (CLOs) and Program Learning Outcomes (PLOs) in higher education. The framework consists of two Excel Macro-enabled workbooks. The course section workbook guides instructors through the assessment process, automatically calculates CLO achievement levels, and generates reports for the coordinators and the Head of Department (HoD). The course-level workbook aggregates data from all course sections and calculates CLO and PLO achievement levels relative to the course. Proven successful in three FCIT (Faculty of Computer and Information Technology) programs at NBU (Northern Border University), the framework demonstrably reduces assessment time and errors, improves consistency, and facilitates data-driven program improvement, making it a valuable tool for enhancing program quality.

Keywords-Course Learning Outcomes (CLOs); Program Learning Outcomes (PLOs); assessment automation; Excel macro-enabled workbooks; higher education; accreditation

I. INTRODUCTION

Ensuring that students acquire the intended knowledge, skills, and values is paramount in the dynamic world of higher education. This necessitates effective assessment of Course Learning Outcomes (CLOs) and Program Learning Outcomes (PLOs), which capture the intended knowledge, skills, and values students should acquire by the end of a course or program, respectively. However, traditional assessment

methods often fall short, proving to be time-consuming, prone to errors, and lacking consistency in application across different courses or programs [1, 2], leading to inaccurate evaluations, hinder program improvement efforts, and ultimately impede student success. A critical need arises for innovative solutions that streamline and enhance learning outcome assessment, in [3-5]. This paper introduces a novel automation framework designed to address these challenges. Built on readily accessible Excel Macro-enabled workbooks, the framework

simplifies and automates many of the assessment tasks within the direct assessment methods, empowering both instructors and coordinators to:

- Reduce the time and effort invested in assessment: Manual calculations and report generation are replaced with automated processes, freeing up valuable time for educators to focus on teaching and student interaction [6].
- Minimize errors and inconsistencies: Standardized calculations and automated data aggregation ensure accuracy and consistency across assessments, providing reliable data for program evaluation and improvement.
- Facilitate data-driven decision making: Comprehensive reports generated by the framework provide rich insights into CLO and PLO achievement levels, enabling informed decisions about curriculum development, instructional strategies, and resource allocation.

The framework's efficacy is demonstrated through its successful implementation in the Faculty of Computer and Information Technology (FCIT) at Northern Border University (NBU). Used by three programs and lauded by accreditation agencies like NCAAA [7] and ABET in [8], it stands as a testament to the framework's potential to revolutionize learning outcome assessment in higher education institutions. This paper delves deeper into the framework's design, implementation, and impact. We explore the functionalities of the two Excel workbooks, one for course sections and another for course-level assessment. We showcase how the framework automates calculations, generates reports, and aggregates data from multiple sections to provide comprehensive insights into program-level PLOs achievement. Furthermore, we present a detailed case study of the framework's implementation in the FCIT, highlighting its demonstrable effects on reducing assessment time, minimizing errors, and facilitating data-driven program improvement. Finally, we discuss the broader implications of the framework for higher education institutions seeking to streamline and enhance their CLO and PLO assessment practices, particularly those adhering to similar national qualification frameworks. This framework lays the way for a future in which learning outcome assessment is not a burden, but a strong instrument for driving continuous improvements in student learning and program quality.

II. CLO AND PLO ASSESSMENT AUTOMATION FRAMEWORK

Ensuring students achieve high-quality learning outcomes is paramount in today's rapidly evolving higher education landscape, where inadequate learning can hinder career prospects and personal growth [9, 10]. Central to this goal is the effective assessment of CLOs and PLOs. However, traditional assessment methods often fall short, proving to be time-consuming, prone to errors, and lacking consistency. This hinders accurate evaluation, limits data-driven decision making, and poses challenges for accreditation requirements. To address these challenges, we present the novel CLO and PLO Assessment Automation Framework. This framework leverages readily available Excel Macro-enabled workbooks to

streamline and automate many of the assessment tasks, empowering both instructors and coordinators to:

- Reduce the time and effort invested in assessment, freeing up valuable resources for teaching and student interaction.
- Minimize errors and inconsistencies through standardized calculations and data aggregation, ensuring reliable data for program evaluation and improvement.
- Facilitate data-driven decision making by providing rich insights into CLO and PLO achievement levels, enabling informed decisions about curriculum development, instructional strategies, and resource allocation.
- Simplify accreditation compliance by generating reports tailored to accreditation agency requirements, minimizing data collection and reporting burdens.

This framework operates through two key components: Course Section Assessment Workbook and Course Assessment Automation Workbook.

Fig. 1. Course section assessment workbook home page.

A. Course Section Assessment Workbook

This workbook guides instructors through the assessment process for each individual course section. It automates calculations, generates reports, and facilitates evidence upload, empowering instructors to focus on delivering quality instruction and providing insightful feedback to students. This instructor-facing workbook simplifies and automates the assessment of CLOs within individual sections. Key features include:

- Automated Calculations: Manual calculations are implemented through pre-defined weighting schemes and rubrics.
- Individual Student Reports: Detailed reports are generated for each student, summarizing performance and providing feedback.
- Cloud-based Evidence Upload: Securely upload evidence of assessments via integrated cloud links.
- Marks Summary for HoD Approval: Streamline mark submission with a dedicated report for Head of Department (HoD) approval.


The Course Section Assessment Workbook includes hidden (protected to be visible by the instructor) sheets and visible sheets can be displayed or accessed using links from the home page as shown in Figure 1. Hidden sheets act like a database used to store the registration data, the study plan, and the CLO-PLO alignment:

- Program study plan: Provides a comprehensive overview of all courses within the program, including details such as course names, credit hours, prerequisites, and levels.
- Program PLO-SO alignment: Maps the program's PLOs to external frameworks such as ABET's Student Outcomes (SOs), ensuring alignment with accreditation requirements and industry standards.
- Course-PLO alignment: Specifies the alignment of each course with specific PLOs, indicating the degree to which each course contributes to the achievement of program-level outcomes.
- CLO-PLO Alignment: Delineates the alignment between individual CLOs and the corresponding PLOs, ensuring coherence between course-level and program-level learning objectives.
- Course Sections (updated each term): Maintains a list of course sections offered in the current term, including section IDs, instructors, and timetables, ensuring up-to-date information for assessment.
- Student Registrations (updated each term): Records student enrollment in each course section, providing accurate student lists for assessment activities.
- Coordinators (updated each term): Stores contact information for course coordinators, facilitating communication and collaboration in assessment processes.
- Process Sheet (parameters, data, calculations): Contains various parameters, filters, and calculations used within the workbook, enabling automated processes and data manipulation.

Accessible Sheets allowing section instructor to specify the assessment plan of the course, upload marks of each student in each assessment tools relative to CLOs and upload evidence relative to each assessment tools, and obtain CLOs and PLOs, SO attainment and the marks summary to be shared with head of the department:


- Section information (main application interface): Serves as the primary interface for instructors, providing access to other sheets and functionalities within the workbook.
- Student list (automatically generated): Displays the list of students enrolled in the selected section, extracted from the hidden "Student registrations" sheet, ensuring accurate student information.
- CLO-PLO alignment (automatically generated): Presents the alignment of CLOs with PLOs for the specific course, providing a clear overview of learning objectives and their contribution to program outcomes.

- Assessment Plan (assessment tool coordination table): Facilitates coordination among instructors by aligning assessment tools and weighting schemes across different course sections, promoting consistency in assessment practices.
- Assessment Tools sheets (Quizzes, Assignments, Exams, etc.) with cloud links to gather evidence: Provide dedicated spaces for instructors to enter assessment scores for various assessment methods, along with cloud links for secure upload of assessment evidence, streamlining evidence collection and storage.




Covered CLOs & Allocated Marks	1.1	1.2	2.1	2.2	2.3	AMs Weight
	17	27	19	18	19	100
Assessment Methods (AMs)						
Lab			4	3	3	10
Project						0
Quizzes	2	2	2	2	2	10
Assignments & G. Discussions			3	3	4	10
Participation						0
MidTerm Exam	15	15				30
Final Exam		10	10	10	10	40

Fig. 2. Course assessment plan sheet.



CLOs Assessment						PLUS SOs Assessment												
CLOs	Assig Marks	Nbr Std	Nbr CLO	CLOs Achiev. %	CLOs Achiev.	NCMAA							ABET					
						PLOs	K1	S1	S1	S3	S4	V1	V2	V3				
1.1	17	16		88.9%	A	PLOs Achiev. %	92.3%	87.0%										
1.2	27	17		94.4%	A													
2.1	19	14		77.8%	A	SOs	SO1	SO2	SO3	SO4	SO5	SO6	SO7					
2.2	18	16		88.9%	A	SOs Achiev. %	87.5%											
2.3	19	17		94.4%	A													

Fig. 3. CLO, PLO, and SO assessment.



Marks Details
CPIT405
Internet Application

جامعة الحدود الشمالية
NORTHERN BORDER UNIVERSITY
كلية الهندسة والعلوم
College of Engineering and Sciences

Course Title: Internet Application
Course Code: CPIT405
Section ID: 02450

Term: Sem. 1-1443/1442
Progra: Information Technology
Branch: Boys

Instructor: هالة بنو منصور, م.ع.م.

Assessment Methods -> Allocated Marks ->		Participation	Quizzes	Assignments	Lab	Project	MidTerm	Total_50	FinalExam	Total_100	Extra Mark	Banner	Grade
Student IDs	Student Names	5	5	10	10	5	10	60	40	100	0		
1	2E+08 رعد الرحمن عبد الحامد العنود	4.00	5.00	7.75	7.75	0.00	16.75	28.68	28.75	67.35	0.00	68	D+
2	2E+08 عبدالرحمن طهون الوائلي العنود										0.00	DN	DN
3	2E+08 تارن عبد الله العنود	1.50	5.00	9.53	8.60	0.00	17.43	41.51	29.68	71.51	0.00	72	C
4	2E+08 فواز عبد الله العنود	3.00	5.00	8.53	7.60	0.00	21.55	52.25	21.25	73.52	0.00	74	C
5	2E+08 فهد عبد الله العنود										0.00	W	W
6	2E+08 رعد صلاح عبد الله العنود	3.00	4.50	7.00	8.60	0.00	12.42	34.52	19.65	56.17	0.00	57	F

Grade Distribution

STATISTICAL INFORMATION

Grade	A+	A	B+	B	C+	C	D+	D	F
Nbr of Std	0	0	0	0	0	2	1	0	1
Percentage	0.0%	0.0%	0.0%	0.0%	0.0%	50.0%	25.0%	0.0%	25.0%

Status: Starting Course: Completing Course

Status	Starting Course	Completing Course
DN Drop	1	1
	7	4

Fig. 4. Course section marks summary sheet.

- CLO Attainment (automatically generated): Summarizes the attainment levels of CLOs for each student, calculated based on entered scores and pre-defined weighting

schemes, providing a clear overview of student performance at the course level.

- PLO-SO Attainment (automatically generated): Aggregates CLO attainment data to generate PLO and SO attainment levels, demonstrating student achievement at the program level relative to the section.
- Marks Summary (for HoD approval): Presents a comprehensive summary of student marks for review and approval by the Head of Department (HoD), streamlining mark submission and approval processes.

Covered CLOs & Allocated Marks	1.1	1.2	2.1	2.2	Total
Assigned CLOs Marks	10	10	10	10	40
1 202400091	10.00	10.00	10.00	10.00	40.00
2 202400103	10.00	10.00	10.00	10.00	40.00
3 202400199	10.00	10.00	2.00	10.00	32.00
4 202400215	10.00	2.00	10.00	2.00	24.00
5 202400241	10.00	10.00	10.00	8.00	38.00
6 202401040	10.00	10.00	8.00	10.00	36.00
7 202401448	10.00	6.00	10.00	10.00	36.00
8 202402001	5.00	10.00	10.00	10.00	35.00
9 202402116	10.00	1.00	10.00	10.00	31.00
10 202402729	10.00	10.00	10.00	10.00	40.00
11 202402817	10.00	10.00	10.00	10.00	40.00
12 202402851	10.00	10.00	8.00	10.00	38.00
13 202403027	5.00	6.00	10.00	10.00	31.00
14 202403590	10.00	10.00	10.00	10.00	40.00
15 202403616	7.00	10.00	10.00	10.00	37.00
16 202403619	10.00	10.00	10.00	10.00	40.00
17 202403620	10.00	10.00	7.00	10.00	37.00
18 202403703	10.00	10.00	10.00	10.00	40.00

Fig. 5. Uploading assessment data for an assessment tool sheet.

B. Course Assessment Automation Workbook

This workbook acts as a central hub for coordinators, aggregating data from all course sections and providing comprehensive insights into program-level PLO achievement. It offers automated data analysis, report generation, and facilitates data sharing with stakeholders, simplifying program evaluation and accreditation processes. This workbook aggregates data from all sections to provide comprehensive program-level insights relative to the course:

- Effortless Data Aggregation: Seamlessly imports data from individual section files, minimizing manual entry and errors.
- Rich Reports and Visualizations: Generates reports and charts showcasing CLO and PLO achievement across sections and student cohorts.
- Accreditation Compliance: Simplifies data collection and reporting for accreditation agencies.
- Centralized Data Management: Maintains an organized repository of course information, alignment matrices, and assessment data for future reference.

The Course Assessment Automation Workbook has the following key features:

- Data Aggregation: Seamlessly imports data from multiple section files, eliminating manual entry and ensuring consistency across sections.
- Automated Report Generation: Automatically generates informative reports, including:

- Course information sheet (Figure 6): Summarizes key course details and statistics.
- CLO-PLO alignment sheet: Visualizes the alignment of CLOs with PLOs for the course.
- Grade distribution sheet: Displays the distribution of student grades across sections.
- CLO Assessment sheet: Summarizes CLO attainment levels across sections.
- PLO Assessment sheet: Presents PLO attainment levels for the course.

- Centralized data management: Serves as a repository for course information, alignment matrices, and assessment data, facilitating program-level analysis and decision-making.
- Compliance support: Generates XLSX reports (without macros) for seamless sharing with accreditation agencies, streamlining compliance efforts.

COURSE CLOS PLOS SOS ASSESSMENT

جامعة الحدود الشمالية
NORTHERN BORDER UNIVERSITY
كلية العلوم والتكنولوجيا
College of Science and Technology

✓ Date: 5/19/2022 ✓ Academic Year: 1443/1442 ✓ Semester: الفصل الدراسي الأول

✓ Course Code: CPT405 ✓ Course Name: Internet Application ✓ Ar. Name: تطبيقات الشبكة الحاسوبية

✓ Program: Information Technology ✓ Department: Information Technology ✓ Level: 9

✓ Credit Hr: 3 ✓ Contact Hr: 4 ✓ Nbr. Std. (Male: 4, Female: 5, Total: 9)

✓ Coordinator: Dr. Hafeedh Zayani ✓ Nbr. Section (Male: 1, Female: 1, Total: 2)

Generate CLOs-PLOs-ASS for Course File

Fig. 6. Course assessment workbook home page.

Course-level CLO and PLO assessment automation is achieved through a custom VBA script function named GenCLOsPLOsAssCourse() (Figure 7). This function simplifies data aggregation by prompting the user to select multiple section-level assessment workbooks. It then iterates through each workbook, extracting relevant data such as section ID, student count, grade distributions, CLO and PLO mapping, and CLOs with assigned marks. The extracted data are then populated into corresponding worksheets within a single course-level assessment workbook. Notably, the function dynamically handles branches (male/female) and department information, utilizing it to select the appropriate worksheet for CLO-PLO mapping and tracking student and section counts accordingly. Overall, this function streamlines the generation of comprehensive course-level assessments by consolidating data from multiple sections.

The framework seamlessly integrates instructor and coordinator tasks, streamlining assessment processes from the individual section level to program-wide analysis. Instructors leverage the Course Section Assessment Workbook to manage student information, enter scores, visualize CLO and PLO

attainment, upload evidence, and generate reports for approval. Coordinators, in turn, utilize the Course Assessment Automation Workbook to effortlessly import section files, visualize course-level statistics and grade distributions, analyze CLO and PLO achievement across sections, and generate comprehensive reports for quality assurance purposes.

```
Function GenerateCLOsPLOSAssCourse() As String
    ' User selects section assessment workbooks
    Dim fd As FileDialog
    fd.InitialFileName = ActiveWorkbook.Path
    fd.AllowMultiSelect = True
    If fd.Show = -1 Then

        ' Process each selected workbook
        For i = 1 To fd.SelectedItems.Count
            Dim wbSection As Workbook
            Set wbSection = Workbooks.Open(fd.SelectedItems(i), UpdateLinks:=0)

            ' Extract relevant data and populate course workbook
            ' ... (code omitted for brevity) ...

            wbSection.Close
        Next i

        ' Determine department and return value
        ' ... (code omitted for brevity) ...

    End If

    ' Activate course workbook and return department
    wbCourse.Activate
    GenerateCLOsPLOSAssCourse = dep
End Function

Done:
    Exit Function

STP:
    MsgBox "Error: " & Err.Description, vbCritical, "Importing Sec. Ass. Data"
End Function
```

Fig. 7. Generation function.

III. IMPLEMENTATION AND CASE STUDY

The implementation of the CLO and PLO Assessment Automation Framework at FCIT necessitated meticulous integration with the current procedures. The initial phase entailed engaging in collaboration with teachers and coordinators to comprehend their assessment methodologies and requirements. As a result, the Assessment Plan sheet was modified to incorporate the use of current assessment tools and weighting systems, ensuring a seamless transition without causing any disruption to established procedures.

Challenges were inevitable. The initial concerns encompassed a range of technical anxieties and apprehensions over the division of effort. To tackle these issues, extensive training sessions were carried out, emphasizing the intuitive interface and the potential for significant time savings. In addition, open communication channels were established to handle technical issues and give continuous assistance. This was necessary because the framework is built on the NBU licensed Office 365, and some instructors are reluctant to upgrade to the required one. The framework's effectiveness was validated through its application in three diverse programs: Information Systems (IS), Computer Sciences (CS), and Information Technology (IT). Here's a glimpse into the transformative impact:

- Reduced Assessment Time:
 - Instructors from 3 departments reported a 50% reduction in time spent on calculations and report generation.

- All course coordinators noted a 30% decrease in workload associated with data aggregation and analysis.
- Enhanced Accuracy:
 - Automated calculations eliminated human error, leading to greater confidence in data integrity.
 - Standardized formulas across sections ensured consistency in attainment calculations.
- Data-Driven Decision Making:
 - PLO-SO attainment insights in IT helped identify areas for curriculum revision, aligning program outcomes with industry expectations.
 - Real-time CLO attainment data in IS facilitated targeted interventions for struggling students, boosting their performance.
 - CS program coordinators utilized grade distribution reports to identify sections requiring additional resources or faculty development.

The achievements observed in each of these programs provide strong evidence of the potential of the framework. In addition to enhancing efficiency and precision, it facilitates data-driven decision making, which fosters ongoing program enhancement and guarantees that graduates possess the necessary competencies and understanding to prosper in the contemporary society.

IV. COMPARISON WITH EXISTING WORKS

To demonstrate the distinctiveness and value of our framework, we present a comparison with existing solutions for CLO and PLO assessment. This comparison highlights the unique features and advantages our framework offers, positioning it as a valuable tool for higher education institutions. While existing commercial learning management systems like Blackboard Learn [11] and Desire2Learn [12] offer assessment tools and reporting features, they lack the tailored focus on CLO and PLO alignment and reporting found in our framework. Open-source solutions like ELMS LN [13] and OAT [14], while flexible, require technical expertise and don't directly address CLOs and PLOs. Research like Al-Naki et al.'s [12] explores CLO-PLO mapping automation using NLP, but lacks broader assessment features. Others [4, 6], propose models and frameworks for program improvement through assessment data, but lack the specific software implementation offered by our Excel-based solution. Our framework's strengths lie in its ease of use, specific focus on CLO and PLO, data-driven decision-making capabilities, and customizability, making it a unique and valuable tool for higher education institutions.

The comparison presented in Table I underscores the unique advantages of our CLO and PLO assessment automation framework. Its specific focus on CLOs and PLOs, user-friendly Excel-based interface, comprehensive assessment automation features, data-driven decision-making capabilities, and customizable design make it a compelling choice for higher education institutions seeking to streamline and enhance

their assessment processes. By effectively addressing the complexities of CLO and PLO assessment, our framework empowers institutions to make data-informed decisions that

drive continuous program improvement and ensure student success.

TABLE I. COMPARISON TABLE

Feature	Proposed	[11]	[12]	[13]	[14]	[15]	[4]	[6]
Focus on CLOs and PLOs	Yes	No	No	No	No	Yes (mapping only)	No	No
Assessment automation	Yes	Partially	Partially	Requires additional tools	No	No	No	No
Data-driven decision making	Yes	Yes	Yes	Yes	Limited	No	Yes (model only)	No
Ease of use	Very high	Medium	Medium	High technical expertise needed	High technical expertise needed	Low	Medium	Medium
Customizability	High	Medium	Medium	High	High	No	Medium	Medium
Cost	Free	Paid	Paid	Free	Free	Research paper	Research paper	Research paper

V. CONCLUSION

The proposed CLO and PLO Assessment Automation Framework demonstrates the significant influence that technology may have on the fields of education. The advantages of the proposed framework are indisputable: instructors and coordinators experience a decrease in burden, assessment accuracy and consistency are enhanced, and above all, there is a notable transition towards curriculum development guided by data and student learning. These benefits are especially significant in academic institutions, which prioritize the correlation between program outcomes and student achievement.

The framework's potential transcends the confines of FCIT. Its adaptability and modular design make it a compelling solution for other higher education institutions, as evidenced by its recent adoption by the Electrical Engineering (EE) program at NBU for term 2, AY 2023/2024. Disseminating this framework, with appropriate customization and context-specific adjustments, can foster a broader culture of assessment excellence. This is particularly relevant within institutions like NBU, which emphasize program outcome-student achievement alignment. However, the framework's flexibility extends beyond specific frameworks, making it a valuable tool for fostering a more comprehensive approach to assessment and program improvement in diverse educational contexts.

While the framework demonstrates significant strengths in streamlining CLO and PLO assessment, it also presents some limitations that require ongoing attention. To ensure the framework's correctness and effectiveness, it is necessary to do periodical modifications at the start of each term. The QAAC faces a substantial burden since they are required to manually collect and update data from Banner, including information on available sections and registrations. Additionally, they must align CLOs and PLOs and prepare data for the workbooks. The manual approach is both time-consuming and susceptible to errors. To mitigate this load and guarantee seamless adoption in the future, other options could be considered:

- Creating automatic data import tools: By optimizing the process of retrieving data from Banner, we can greatly reduce the need for manual labor and mitigate the potential for errors. Collaboration with IT teams is necessary to

develop interfaces that efficiently send pertinent data to the workbooks.

- Integrating pre-filled templates: Offer pre-configured templates for aligning CLOs and PLOs, minimizing the requirement for human data input and conserving significant time.
- Discovering the capabilities of cloud-based data storage: Transferring data storage to the cloud could streamline access and communication among instructors and coordinators, obviating the necessity for manual file transfers and guaranteeing data availability across departments.

Through the implementation of these solutions, their attention can be directed towards the analysis of data and the provision of vital insights to enhance student learning and the efficacy of the program. This will guarantee the ongoing success of the framework and decrease the administrative workload linked to quarterly upgrades.

Essentially, the CLO and PLO Assessment Automation Framework serves as a catalyst for a fundamental change in our assessment strategy and enhances educational quality, going beyond being only a technological solution. By acknowledging its capacities and exploring further possibilities for expansion, we may collectively cultivate an educational setting in which data not only improves the assessment process, but also the entire procedure of nurturing confident and successful graduates.

ACKNOWLEDGMENT

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA, for funding this research work through the project number NBU-FFR-2024-1563-01.

REFERENCES

- [1] A. A. Glatthorn, "Outcome-Based Education: Reform and the Curriculum Process," *Journal of Curriculum and Supervision*, vol. 8, no. 4, pp. 354–364, 1993.
- [2] T. R. Guskey, "Defining the Differences between Outcome-Based Education and Mastery Learning," *School Administrator*, vol. 51, no. 8, pp. 34–37, 1994.
- [3] A. Alyasin, R. Nasser, M. El Hajj, and H. Harb, "Assessing Learning Outcomes in Higher Education: From Practice to Systematization," *TEM*

- Journal, pp. 1593–1604, Aug. 2023, <https://doi.org/10.18421/TEM123-41>.
- [4] M. A. Al-Shargabi, "An Integrated Decision Support Model For Enhancing Continuous Improvement Of Academic Programs," *Engineering, Technology & Applied Science Research*, vol. 9, no. 5, pp. 4835–4841, Oct. 2019, <https://doi.org/10.48084/etasr.3079>.
- [5] N. Shahmohammadi, "The Impact of Total Quality Management on the Effectiveness of Educational Programs of Karaj First Grade High Schools," *Engineering, Technology & Applied Science Research*, vol. 8, no. 1, pp. 2433–2437, Feb. 2018, <https://doi.org/10.48084/etasr.1674>.
- [6] E. A. M. Jibreel, N. M. A. Zakari, and J. A. S. Smida, "Develop assessment framework inspiring-Saudi Arabia s Vision 2030 for health program in higher education: an experience from Al Maarefa University, Riyadh, Saudi Arabia," *International Journal of Medicine in Developing Countries*, vol. 4, no. 8, pp. 1233–1233, Jul. 2020, <https://doi.org/10.24911/IJMD.51-1592585802>.
- [7] "National Commission for Academic Accreditation and Assessment," *Eucation and Training Evaluation Commission*. <https://beta.etc.gov.sa:3443>.
- [8] "Home - ABET." <https://www.abet.org/>.
- [9] A. B. Altamimi, "Big Data in Education: Students at Risk as a Case Study," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11705–11714, Oct. 2023, <https://doi.org/10.48084/etasr.6190>.
- [10] H. Zaghloul, "Assessment of CLOs, PLOs, ILOs, SLOs: An academic programme for educational theatre as a model," *The Education and science journal*, vol. 25, no. 6, pp. 99–125, Jun. 2023, <https://doi.org/10.17853/1994-5639-2023-6-99-125>.
- [11] "Educational Technology Services | Blackboard | North America." <https://www.blackboard.com/>.
- [12] "D2L | Online Learning Platform | Creators of the Brightspace LMS Software," *D2L*, Feb. 14, 2022. <https://www.d2l.com/>.
- [13] "ELMS Learning Network." <https://www.elmsln.org/>.
- [14] "Open-Assistant/oasst-model-eval." Open-Assistant, Jan. 04, 2024, [Online]. Available: <https://github.com/Open-Assistant/oasst-model-eval>.
- [15] N. Zaki, S. Turaev, K. Shuaib, A. Krishnan, and E. Mohamed, "Automating the mapping of course learning outcomes to program learning outcomes using natural language processing for accurate educational program evaluation," *Education and Information Technologies*, vol. 28, no. 12, pp. 16723–16742, Dec. 2023, <https://doi.org/10.1007/s10639-023-11877-4>.

Effect of Fire Exposure on the Properties of Self-Compacting Concrete reinforced by Glass Fibers

Rawaa K. Aboud

Department of Civil Engineering, University of Baghdad, Iraq
rawaa.khalid@coeng.uobaghdad.edu.iq (corresponding author)

Hadeel K. Awad

Department of Civil Engineering, University of Baghdad, Iraq
hadeel.kalid@coeng.uobaghdad.edu.iq

Shatha D. Mohammed

Department of Civil Engineering, University of Baghdad, Iraq
shatha.dh@coeng.uobaghdad.edu.iq

Received: 17 January 2024 | Revised: 1 February 2024 | Accepted: 13 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6924>

ABSTRACT

The optimal design of any structural elements requires examining all environmental risks, emergency accidents, and standard load cases. Exposure to fire is one of the most common safety threats. Nowadays wide developments are achieved in the field of concrete technology, therefore, experimental and theoretical investigations should be performed on the characteristics of such developed materials under different loading conditions. This study investigates the impact of fire exposure on the mechanical characteristics of self-compacting concrete, specifically compressive and tensile strength, modulus of elasticity, and stress-strain relation. The adopted fire exposure consisted of six steady-state temperatures (300, 400, 500, 600, 700, and 800°C) for one hour and a sudden cooling method. Four glass fiber volume fractions were adopted: 0, 0.5, 1, and 1.5%. The glass fiber volume fractions considered (0.5-1.5%) improved the mechanical properties investigated. Two states were detected for the effect of fire exposure. The effect of fire exposure was inversely proportional to fiber content in burning temperatures of 300-700°C, while the reduction in mechanical properties of 1.5% fiber content was greater than those of 0.5 and 1% when the temperature increased to 800°C. Furthermore, the addition of glass fiber changed the brittle mode stress-strain relation to semi-ductile for the non-burned and burned up to 600°C specimens, whereas a brittle behavior was detected when the temperature increased above 600°C. In general, a similar effect was noticed for all the glass fiber ratios considered regarding the slope of the stress-strain linear stage compared to the non-burned specimens, which was more salient when the burning temperature increased.

Keywords-self compacting concrete; glass fiber; flame; sudden cooling; modulus of elasticity; compressive strength; tensile strength

I. INTRODUCTION

Concrete is a congenital construction material. Although its tensile strength and brittleness are low [1-3], it is widely used in construction since its constitutive materials are available, workable, inexpensive, and have high strength. Currently, Self-Compacting Concrete (SCC) is classified as an advanced and widely used concrete due to its high flow ability, strength, low noise pollution, labor expense, and construction period, and it is used in important and difficult building structures [4-5]. High-rise buildings and underground structures are spreading due to continuous development in infrastructure manufacturing. These structures are more vulnerable to fire due to gas use,

electrical appliances, and in some countries to terrorism. Many studies investigated the impact of fire and high-temperature exposure on the ultimate strength and durability of concrete. Fire flames and high temperatures also significantly affect spalling, stiffness, and strength properties [6]. Some studies observed that the residual compressive strength of SCC decreased when heated to an elevated temperature of up to 650°C [7-8]. Several studies showed that strength reduction is lower at burning temperatures up to 400°C [9-10].

In [11], SCC in reinforced concrete beams influenced by repeated loads and exposed to different fire-flame burning temperatures (200, 300, 400, and 500°C) for 30 minutes had a

decrease in maximum ultimate load capacity (16, 23, 54, and 71%) in the case of sudden cooling conditions and a lower reduction in gradual cooling (8, 14, 36, and 64%). This variation between the two cooling methods was reduced with an augmentation in the burning temperature. In [7], the unconsumed compressive strength, tensile strength, and modulus of elasticity of SCC under the impact of different high temperatures and cooling methods were evaluated. A non-significant increase in compressive strength was observed at 150°C for both cooling cases, sudden and gradual. As the temperature rose to 300°C, the mechanical properties were being continuously decreased, and the residual compressive strength dropped by 22.26%. Regarding the tensile strength and modulus of elasticity, the reductions were approximately 50% at temperatures of 450-600°C.

Several types of fiber were considered in [12-14] to improve the mechanical properties of fire-exposed concrete. In [13], burning Reactive Powder Concrete (RPC) samples at temperatures of 300, 400, and 500°C were gradually and suddenly cooled. Three volume fraction percentages of hooked steel fiber (0, 1, and 1.5%) were considered. Increasing the burning temperature reduced the compressive strength of the concrete without fiber. In the case of 1% fiber content, there was a smaller reduction at 400°C burning temperature, especially for gradual cooling. Sudden cooling was more effective than gradual in the decrease in compressive strength. In [14], three cooling methods (sudden, gradual, and foam) were examined to investigate the influence of burning temperatures (300, 400, and 500°C) on RPC reinforced with micro steel fiber (0, 1.0, and 1.5%). The results showed that the addition of 1.5% fiber at 400°C with gradual cooling improved the concrete properties by 1.6%. Sudden cooling had a negative effect at different temperatures, especially when the burning temperature exceeded 400°C. The results also demonstrated that the worst behavior of concrete properties was observed for samples without fiber content at 500°C, reaching 42.65 and 33.5% decrease in gradual and sudden cooling, respectively, whereas there was an improvement in concrete properties with fiber content of 1.0 and 1.5% to 6.67, 3.13%, 3.14 and 0.57% for gradual and foam cooling, respectively.

This study aimed to investigate the effect of E-Glass fiber volume fraction on some fresh and hard properties of SCC and examine the influence of glass fiber reinforced SCC exposed to different burning temperatures (300, 400, 500, 600, 700, and 800°C) for one hour on different mechanical properties, such as compressive strength, splitting tensile strength, modulus of elasticity, and stress-strain diagram.

II. THE EXPERIMENTAL PROGRAM

A. Materials

This study used Tasloga Ordinary Portland cement (CEM I 42.5 N) in all SCC mixes, which follows [15] and [16]. Table I shows its physical and chemical characteristics. Crushed coarse aggregate of grade 5-20 mm and natural fine aggregate classified as zone 2, according to the requirements of the modified Iraqi specification No. 45/1984 and ASTM C33/C33M [17-18], were used, as illustrated in Tables II and III. This study also used 2% silica fume as a partial

replacement for cement, which conforms to [19]. Densified silica fume of Chinese origin was supplied from "Sika Fume - HR, Concrete Additive", as displayed in Figure 1, and Table IV presents its chemical analysis and physical properties. To control the internal cracks that may occur when concrete is exposed to fire flame, different volume fraction (vf) percentages of E-Glass chopped fiber (GF) were adopted, 0.5, 1, and 1.5%, as exhibited in Figure 2. Table VI depicts the properties of the E-Glass fiber.

TABLE I. CEMENT CHEMICAL COMPOSITION AND PHYSICAL PROPERTIES

Physical properties			
Test	Test results	Limit of [15] type CEM I 42.5N	Limit of [16] Type I
Specific surface area (Blaine method), (m ² /kg)	320	≥ 250	Air permeability test ≥260 [20]
Setting time (Vicat's method)			
Initial setting time: (hrs:min)	0:57	≥ 45 min	≥ 45 min
Final setting time: (hrs:min)	4:40	≤ 10 hrs.	≤ 6 hrs 15 min
Compressive strength (MPa)			
2-day	17.6	≥ 10 MPa	----
3-day	22.8	----	≥ 12 MPa
7-day	31.6	----	≥ 19.0 MPa
28-day	43.2	≥ 42.5 MPa	---
Soundness (Autoclave method)	0.14 %	≤ 0.8 %	≤ 0.8 %
Chemical composition			
Compound	Content %	Limit of [15]	[16] Type I
Lime (CaO)	62.8	----	----
Silica (SiO ₂)	20.57	----	----
Alumina (Al ₂ O ₃)	5.56	----	----
Iron Oxide (Fe ₂ O ₃)	3.32	----	----
Magnesia (MgO)	2.91	Max. 5.0	Max. 6.0
Sulfate (SO ₃)	2.21	2.8 if C ₃ A > 3.5 ≤ 2.5 if C ₃ A ≤ 3.5	if C ₃ A > 8.0 Max. 3.5
Insoluble residue (I.R)	0.77	Max 1.5	Max 1.5
Loss of ignition (L.O.I)	1.94	Max 4.0	Max 3.0
Main compounds of OPC			
Tri calcium silicate (C ₃ S)		50.79	
Di calcium silicate (C ₂ S)		20.74	
Tri calcium aluminate (C ₃ A)		9.12	
Tetra calcium aluminate ferrite (C ₄ AF)		10.10	

TABLE II. PHYSICAL AND CHEMICAL PROPERTIES OF COARSE AGGREGATE

Sieve Size (mm)	% Passing by weight	Percentage of materials by mass passing sieves	
		[17]	[18]
Sieve Size (mm)		(5-20) mm	4.75-19.0 mm
37.5	100.0	100	---
25	---	---	100
20 or 19	97.9	95-100	90-100
14	100.0	Not limited	---
10 or 9.5	---	30-60	20-55
5 or 4.75	---	0-10	0-10
2.36	38.1	Not limited	0-5
SO ₃ content %	0.075	Max. 0.1	---

TABLE III. PHYSICAL AND CHEMICAL PROPERTIES OF FINE AGGREGATE

Sieve size (mm)	% passing by weight	Percentage of materials by mass passing sieves	
		[17]	[18]
10 or 9.5	100.0	100	100
4.75	97.6	90 –100	95-100
2.36	87.5	75 – 100	80-100
1.18	76.7	55 – 90	50-85
0.60	57.1	35 – 59	25-60
0.30	23.9	8 – 30	5-30
0.15	5.0	0 – 10	0-10
Finer than 0.075 mm	3.8	Max. 5	Max. 5
Clay lumps and friable particles	0.37	Max. 1	2-10
SO ₃ content (%)	0.43	Max. 0.5	---

TABLE IV. CHEMICAL AND PHYSICAL PROPERTIES OF SILICA FUME

Chemical composition		
Oxides	Content %	[19]
SiO ₂	92.95	Min. 85%
Fe ₂ O ₃	1.22	----
Al ₂ O ₃	0.34	----
CaO	0.54	----
MgO	0.90	----
SO ₃	0.4	----
L.O.I	3.4	Max. 6%
Moisture content	1.1	Max. 3%
Physical properties		
Property	Test result	[19]
Color	Dark grey	----
Specific surface area (m ² /kg)	20000	Min. 15000
Retaining on sieve 45 μ (%)	3.7%	Max. 10
Strength activity index (%)	113	Min. 105

TABLE V. PROPERTIES OF SUPER-PLASTICIZER (GLINEUM 51)*

Property	Details
Form	Viscous liquid
Color	Light brown
Relative density	1.1 @ 20 °C
PH	6.6
Viscosity	128 \pm 30 cps @ 20°C
Transport	Not classified as dangerous

*Properties of super-plasticizers according to the manufacture datasheet

TABLE VI. GLASS FIBER PROPERTIES*

Property	Details
Relative Density	2540 kg/m ³
Average Length	30 mm
Diameter	0.6 mm
Modulus of Elasticity	80x10 ³ MPa
Tensile Strength	1200 MPa
Aspect Ratio (l/d)	50
Softening point	840 °C
Strength degradation point	350 °C

*According to the conformity certificate



Fig. 1. E-Glass chopped fiber.



Fig. 2. Silica fume.

B. Tested Specimens

1) Self-Compacting Concrete (SCC) Mix Design

The SSC mix was designed according to [21]. Several trial mixes were created to reach a compressive strength of 45 MPa with 0.34 w/b ratio. E-Glass chopped strand glass fiber was used throughout the experimental process with an aspect ratio of 50 and volume fraction (0.5, 1.0, and 1.5%), as shown in Table VII. The fresh tests for SCC mixes were based on [21] and test methods, as portrayed in Figure 3, and were all within its standards, as illustrated in Table VIII.

TABLE VII. DETAILS OF SCC MIXTURES

Mix Type	Mix Proportion (kg/m ³)					SP (% by cement weight)	Silica fume (kg/m ³)
	Water	Cement	Sand	Gravel	GF		
SR	154	445	886	942	0	1.3	9
SG0.5	154	445	886	942	12.7	1.3	9
SG1	154	445	886	942	25.4	1.3	9
SG1.5	154	445	886	942	38.1	1.3	9

TABLE VIII. SCC FRESH TEST RESULTS ACCORDING [21]

Characteristic	Tests method	Test result	[21]	
			Class	Limits
Flow/filling ability	Slump flow	735 mm	SF2	660-750
Viscosity/flow ability	T ₅₀₀	2.5 sec	VS2/VF2	> 2
Passing ability	L-box	0.94	PA2	\geq 0.8 with three bars
Segregation resistance	Sieve segregation	19.6 %	SR1	\leq 20



Fig. 3. Slump flow test.

C. Experimental Program

The SCC mixes were cast into molds of 150×150×150 mm to test compressive strength according to [22]. A splitting tensile strength test 150×300 mm cylinder mold was cast according to [23]. The static modulus of elasticity test for all mixes was carried out following [24]. This test was performed on a cylindrical specimen with dimensions 150×300 mm. A dial gauge with an accuracy of 0.001 was used to measure the vertical displacement, as displayed in Figure 4. The stress-strain curve was used to determine the modulus of elasticity, according to:

$$E_c = S2 - S1 / E2 - 0.00005 \quad (1)$$

The fire exposure test procedure was performed in a U-shaped steel furnace of dimensions 3000×1350×500 mm. The considered steady-state temperatures were 300, 400, 500, 600, 700, and 800°C for one hour, which was measured after the target temperature was reached. An ATP DT-612 thermocouple device and a wire sensor type K were used to monitor the temperature, as depicted in Figure 5. The temperature rise inside the furnace was according to [25]. Sudden cooling was performed for all the specimens immediately after the burning test was completed, and then they were tested.

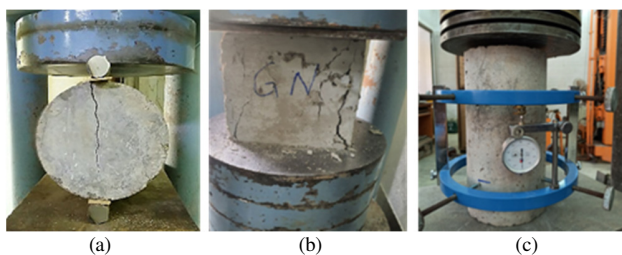


Fig. 4. Tests of SCC specimens: (a) splitting tensile strength test, (b) compressive strength test, and (c) static modulus of elasticity test.

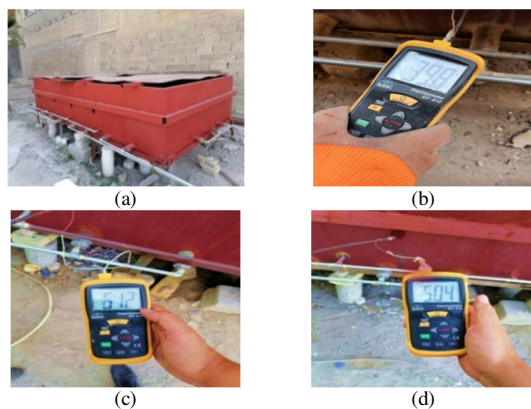


Fig. 5. (a) Burning furnace, (b) digital thermometer at 400°C, (c) digital thermometer at 500°C, (d) digital thermometer at 600°C.

III. TEST RESULTS AND DISCUSSION

A. Fresh Properties of SCC

Table IX demonstrates the results of fresh properties. Slump flow diameter, L-Box test, and Segregation Index (SI) percentage ranged between 652-733 mm, 0.96-0.82, and 18.4-13.1%, respectively. The highest slump flow diameter, L-Box,

and SI were achieved for the mixture without fiber (SR). The addition of fiber decreased the fresh test results compared to SR. The reduction was directly proportional to the fiber content, as for the SG1.5 mix it reached 11.05, 14.58, and 28.8% slump flow diameter, L-Box, and SI, respectively. The addition of fiber increases flow resistance and decreases flow ability as an outcome of increased interlocking and friction between fiber and aggregate, as presented in [26]. The T500 test was performed to evaluate the viscosity of the SCC mixes. The T500 test results ranged from 2.4 to 4.5 s. The findings indicate that the T500 test results for the mixes containing fiber were longer than the mixes without. In general, adding fiber produces concrete with more cohesion and interlocking, leading to a slow flow of the SCC mixture, as it obstructs the flow within the constrained region, as shown in [27].

TABLE IX. FRESH PROPERTIES OF SCC MIXES

Mixes	Slump flow (mm)	T500 (s)	L-box	Segregation Index % (SI)
SR	735	2.5	0.94	19.6
SG0.5	692	3.3	0.89	18.7
SG1	675	3.9	0.85	17.6
SG1.5	662	4.6	0.81	15.5

B. Mechanical Properties of SCC

1) Compressive Strength

Compressive strength is one of the most essential concrete characteristics. Figure 6 and Table X illustrate the compressive strength results at 56 days, including 28 days of water curing and 28 days under laboratory conditions, after exposure to fire flame at 300, 400, 500, 600, 700, and 800°C. The results reveal that compressive strength was directly proportional to the glass fiber content and inversely proportional to the burning temperatures for all fiber volume fractions. In the case of the SR mix, the reduction in compressive strength reached 30.3, 37.9, and 53.4% at 500, 600, and 700°C, respectively, whereas there was a difference in decrease values as the fiber content increased, being 29.5, 27.2, and 23.1% at 500°C, 36.9, 33.4, and 28.9% at 600°C, and 50.2, 47.4 and 44.6% at 700°C for 0.5, 1.0, and 1.5%, accordingly. When the specimens were burned, many changes occurred in the properties of the concrete, resulting in a considerable reduction in compressive strength. When the temperature increases above 300°C, calcium hydroxide in cement begins to dehydrate, releasing more water vapor and significantly lowering the compressive strength of concrete. The deterioration of the interfacial bond between the aggregate and the cement paste at high temperatures is responsible for the reduction in the compressive strength of the concrete [28]. This can be related to the improvement in the mechanical bond of concrete, where the fiber can delay the creation of microcracks and stop their spread to some extent [29]. This behavior changed at 800°C, causing a slight modification with 1 and 1.5% fiber content, reaching a reduction of 63.3 and 65.2%, respectively. The specific drop is attributed to the softening of the glass fiber at this temperature, i.e., a reduction in the glass fiber characteristics. Therefore, the tensile and compressive strength properties were attenuated since porosity and microcracks increased.

TABLE X. COMPRESSIVE STRENGTH TEST RESULTS

Burning temperature (°C)	Compressive strength (MPa)			
	SR	SG0.5	SG1.0	SG1.5
Normal temp.	45.23	45.78	46.89	48.4
300	38.89	40.08	41.13	43.16
400	35.39	36.55	38.42	40.29
500	31.52	32.28	34.15	37.23
600	28.07	28.9	31.23	34.41
700	21.09	22.78	24.65	26.84
800	14.12	17.78	17.23	16.84

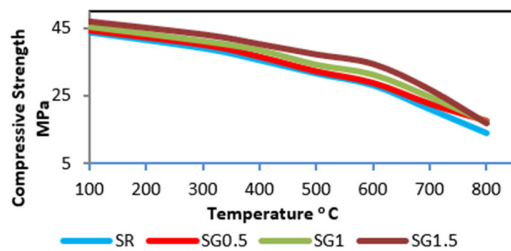


Fig. 6. Effect of burning temperature on the compressive strength of SCC.

2) Modulus of Elasticity

Table XI and Figure 7 exhibit the results of the elasticity modulus, indicating a behavior similar to the compressive strength test results. Glass fiber improved the static modulus of elasticity, as the optimum modification was found in the samples with 1.5% fiber content, reaching an 11.8% increase at ambient temperature. The glass fiber modulus (80×103 MPa) enhanced the modulus of elasticity of SCC. As the burning temperature increased, the deterioration of the elastic modulus for the same fiber content increased. Figure 8 demonstrates that the reduction reached 17.2, 25.7, and 36.8% at 600, 700, and 800°C, respectively. In addition, this deterioration increased proportionally as the burning temperature rose for all fiber volume fractions. At a burning temperature of 800°C, these trends changed since the glass fiber softening point was achieved, so samples with 1.5% fiber by volume had the maximum reduction of 42.4%.

TABLE XI. MODULUS OF ELASTICITY TEST RESULTS

Burning temperature (°C)	Modulus of elasticity (MPa)			
	SR	SG0.5	SG1.0	SG1.5
Normal temp.	28219	29390.5	30732.6	31566.2
300	26167	27564.5	28910.1	29833.9
400	24961.7	26368.3	28108.4	29102.3
500	23557.5	24839.7	26520.6	27602.3
600	22230.8	23557.2	25448.9	26613.9
700	19269.8	21026.9	22832.7	23738.4
800	16847.3	18693.2	18417.2	18164.1

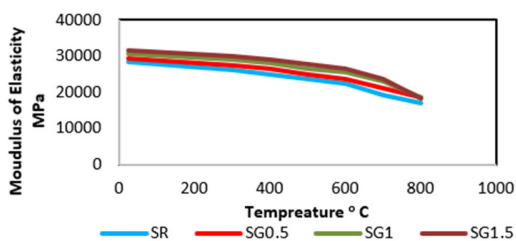


Fig. 7. Effect of burning temperature on the modulus of elasticity of SCC.

3) Splitting Tensile Strength

Table XII and Figure 8 present splitting tensile strength results. Glass fiber had a direct effect on splitting tensile strength, as the maximum reduction was 21.6 % for the samples with 1.5% content. This may be attributed to the fiber mechanism in preventing crack propagation and improving matrix bonding. The splitting tensile strength was inversely proportional to the burning temperature. After exposure to fire flame, it was observed that glass fiber limited splitting tensile strength reduction since it eliminates the damage, spalling, and cracking of SCC specimens. These reductions were 23.6, 21.4, and 22.7% at 500°C, 34.7, 27.6, and 25.6% at 600°C, and 39, 38.3, and 35.2% at 700°C for 0.5, 1, and 1.5% volume fractions of glass fiber, respectively. At higher temperatures the evaporation of water from concrete pores increases, resulting in an increase in shrinkage, which leads to a decrease in splitting tensile strength. Also, the worst case of splitting tensile strength reduction was achieved at 800°C with a fiber content of 1.5%, reaching 78.6%.

TABLE XII. SPLITTING TENSILE STRENGTH TEST RESULTS

Burning Temperature (°C)	Splitting Tensile Strength MPa			
	SR	SG0.5	SG1.0	SG1.5
Normal temp.	4.8	5.12	5.6	5.84
300	4.12	4.70	4.9	5.43
400	3.82	4.26	4.6	4.82
500	3.36	3.91	4.4	4.51
600	3.08	3.34	4.05	4.34
700	2.78	3.12	3.45	3.78
800	1.03	1.65	1.54	1.25

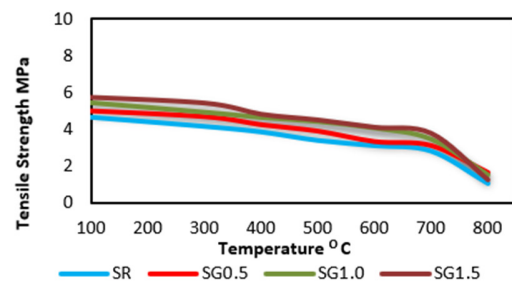


Fig. 8. Effect of burning temperature on the splitting tensile strength of SCC mixes.

4) Stress-Strain Relation

The specific action of stress-strain with the presence of glass fiber changed from brittle to semi-ductile behavior, as fiber contributes to concrete cracks bridging and preventing sudden failure. This behavior was present and clear at a temperature of 25°C for all samples containing glass fiber, as displayed in Figure 9. The particular behavior was lost when the burning temperatures increased and the glass fiber content changed. When the burning temperature reached 600°C, this behavior began to return to a brittle stage. The stress-strain stress relation at 25°C showed that fiber enhanced specimen rigidity, even with 0.5% fiber content. As the fiber content increased, the improvement percentages converged, i.e., there was no significant difference between the fiber volume fractions considered. The glass fiber effect was more pronounced as the burning temperature increased, since for the

samples containing 1.5% glass fiber content, the decrease in the E was less than that of samples containing 0.5 or 1%. As exhibited in Figures 10-15, the curves began to diverge from each other.

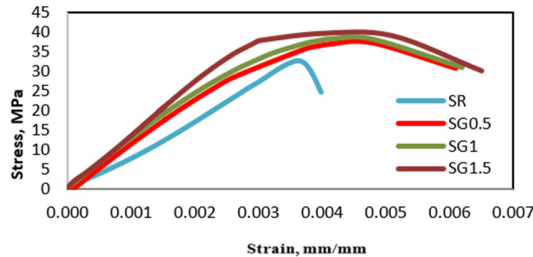


Fig. 9. Stress-strain at ambient temperature.

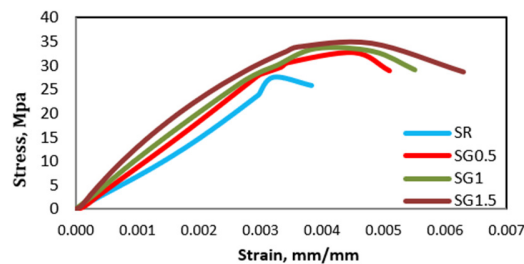


Fig. 10. Stress-strain at 300°C.

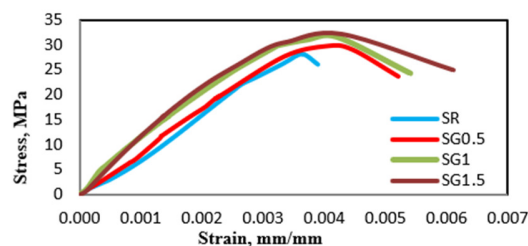


Fig. 11. Stress-strain at 400°C.

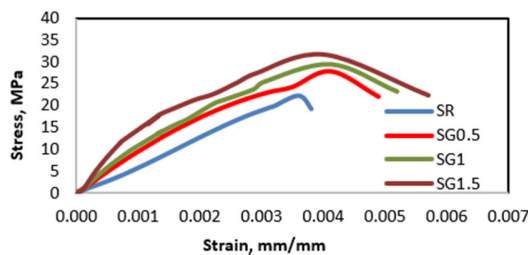


Fig. 12. Stress-strain at 500°C.

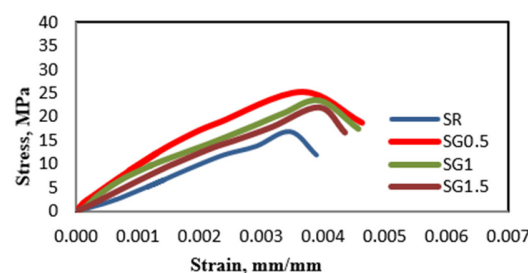


Fig. 13. Stress-strain at 600°C.

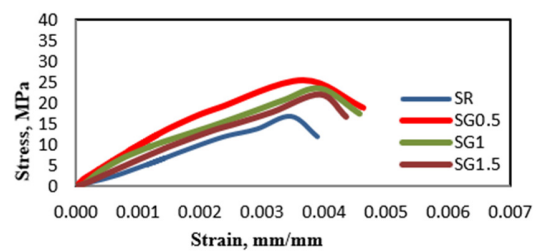


Fig. 14. Stress-strain at 700°C.

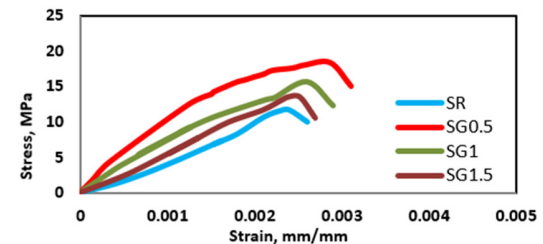


Fig. 15. Stress-strain at 800°C.

IV. CONCLUSIONS

This study investigated the effect of glass fiber content on the mechanical properties of SCC exposed to fire, concluding the following:

- Adding glass fiber deteriorated the fresh properties of SCC, i.e. slump, L-Box, and segregation. Furthermore, there was an increase in T500 compared to the reference mix. However, they were still within the EFNARC standard requirements.
- The glass fiber volume fractions considered (0.5-1.5%) improved compressive strength, tensile strength, and modulus of elasticity for the non-fire-exposed specimens.
- The reduction in mechanical properties due to fire exposure was inversely proportional to the fiber volume fraction for burning temperatures of 300-700°C.
- Increasing the burning temperature to 800°C changed the context of the glass fiber effect. The percentage of decrease in mechanical properties of samples with 1.5% fiber content was greater than those of 0.5, and 1%.
- Adding glass fiber changed the brittle trend of concrete stress-strain relation to semi-ductile for the non-burned and burned up to 600°C specimens, while the behavior was more brittle as the temperature increased more than 600°C.
- A similar effect was noticed for the considered glass fiber ratios regarding the slope of the stress-strain linear stage concerning the non-burned specimens.
- Increasing the glass fiber volume fraction was more significant in terms of linear stage behavior as the burning temperature increased, i.e., the stress-strain curves diverged more as the burning temperature rose.
- For the fire exposure of structural elements, it is recommended to limit the glass fiber volume fraction to

avoid reducing strength as the burning temperature approaches the glass fiber softening point.

REFERENCES

- [1] P. Rashiddadash, A. A. Ramezaniapour, and M. Mahdikhani, "Experimental investigation on flexural toughness of hybrid fiber reinforced concrete (HFRC) containing metakaolin and pumice," *Construction and Building Materials*, vol. 51, pp. 313–320, Jan. 2014, <https://doi.org/10.1016/j.conbuildmat.2013.10.087>.
- [2] S. Lohar, T. W. Cheng, D. Nicolaides, I. Lohar, D. Panias, and K. Sakkas, "Valorisation of glass wastes for the development of geopolymer composites – Durability, thermal and microstructural properties: A review," *Construction and Building Materials*, vol. 222, pp. 673–687, Oct. 2019, <https://doi.org/10.1016/j.conbuildmat.2019.06.169>.
- [3] S. Lohar, T. W. Cheng, D. Nicolaides, I. Lohar, D. Panias, and K. Sakkas, "Valorisation of glass waste for development of Geopolymer composites – Mechanical properties and rheological characteristics: A review," *Construction and Building Materials*, vol. 220, pp. 547–564, Sep. 2019, <https://doi.org/10.1016/j.conbuildmat.2019.06.041>.
- [4] N. A. Memon, M. A. Memon, N. A. Lakho, F. A. Memon, M. A. Keerio, and A. N. Memon, "A Review on Self Compacting Concrete with Cementitious Materials and Fibers," *Engineering, Technology & Applied Science Research*, vol. 8, no. 3, pp. 2969–2974, Jun. 2018, <https://doi.org/10.48084/etasr.2006>.
- [5] J. Abd and I. K. Ahmed, "The Effect of Low Velocity Impact Loading on Self-Compacting Concrete Reinforced with Carbon Fiber Reinforced Polymers," *Engineering, Technology & Applied Science Research*, vol. 11, no. 5, pp. 7689–7694, Oct. 2021, <https://doi.org/10.48084/etasr.4419>.
- [6] A. H. Abdullah and S. D. Mohammed, "The Fire Effect on the Performance of Reinforced Concrete Beams with Partial Replacement of Coarse Aggregates by Expanded Clay Aggregates," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12220–12225, Dec. 2023, <https://doi.org/10.48084/etasr.6412>.
- [7] S. Paul, M. H. Rashid, and M. A. Rahman, "Effect of Elevated Temperature on Residual Strength of Self Compacted Concrete," *Journal of Engineering Science*, vol. 11, no. 2, pp. 107–115, Dec. 2020, <https://doi.org/10.3329/jes.v11i2.50902>.
- [8] B. Demirel and O. Keleştemur, "Effect of elevated temperature on the mechanical properties of concrete produced with finely ground pumice and silica fume," *Fire Safety Journal*, vol. 45, no. 6, pp. 385–391, Nov. 2010, <https://doi.org/10.1016/j.firesaf.2010.08.002>.
- [9] N. Anand, G. Arulraj, and C. Aravindhan, "Stress-Strain Behaviour of Normal Compacting and Self Compacting Concrete Under Elevated Temperatures," *Journal of Structural Fire Engineering*, vol. 5, no. 1, pp. 63–76, Jan. 2014, <https://doi.org/10.1260/2040-2317.5.1.63>.
- [10] A. A. Hammadi, A. F. Izzat, and J. A. Farhan, "Effect of Fire Flame (High Temperature) on the Self Compacted Concrete (SCC) One Way Slabs," *Journal of Engineering*, vol. 18, no. 10, 2012.
- [11] S. D. Mohammed and N. M. Fawzi, "Fire Flame Influence on the Behavior of reinforced Concrete Beams Affected by Repeated Load," *Journal of Engineering*, vol. 22, no. 9, pp. 206–223, 2016.
- [12] Z. H. Dakhel and S. D. Mohammed, "Castellated Beams with Fiber-Reinforced Lightweight Concrete Deck Slab as a Modified Choice for Composite Steel-Concrete Beams Affected by Harmonic Load," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 8809–8816, Aug. 2022, <https://doi.org/10.48084/etasr.4987>.
- [13] R. K. Aboud, H. K. Awad, and S. D. Mohammed, "Fire flame effect on the compressive strength of reactive powder concrete using different methods of cooling," *IOP Conference Series: Materials Science and Engineering*, vol. 518, no. 2, Feb. 2019, Art. no. 022029, <https://doi.org/10.1088/1757-899X/518/2/022029>.
- [14] H. K. Awad, "Influence of Cooling Methods on the Behavior of Reactive Powder Concrete Exposed to Fire Flame Effect," *Fibers*, vol. 8, no. 3, Mar. 2020, Art. no. 19, <https://doi.org/10.3390/fib8030019>.
- [15] "Iraqi Standard No. 5: Portland Cement." The Central Organization for Standardization and Quality Control, Iraq, 2019.
- [16] "ASTM C150/C150M-21: Standard Specification for Portland Cement." American Society for Testing and Materials, 2021.
- [17] "Iraqi Specification No. 45: Aggregate from Natural Sources for Concrete and Construction." 1983.
- [18] "ASTM C33/C33M-18: Standard Specification for Concrete Aggregates." American Society for Testing and Materials, 2018.
- [19] "ASTM C1240-20: Standard Specification for Silica Fume Used in Cementitious Mixtures." American Society for Testing and Materials, 2020.
- [20] "ASTM C204-18e1: Standard Test Methods for Fineness of Hydraulic Cement by Air-Permeability Apparatus." American Society for Testing and Materials, 2018.
- [21] "The European Guidelines for Self-Compacting Concrete," European Federation of Concrete Admixture Associations | European Federation of Specialist Construction Chemicals and Concrete Systems, May 2005.
- [22] "BS EN 12390-3:2019 - TC: Testing hardened concrete - Compressive strength of test specimens." British Standards Institution, 2019.
- [23] "ASTM C496/C496M-15: Standard Test Method for Splitting Tensile Strength of Cylindrical Concrete Specimens." American Society for Testing and Materials, 2017.
- [24] "ASTM C469/C469M-15: Standard Test Method for Static Modulus of Elasticity and Poisson's Ratio of Concrete in Compression." American Society for Testing and Materials, 2015.
- [25] "ASTM E119-20: Standard Test Methods for Fire Tests of Building Construction and Materials." American Society for Testing and Materials, 2020.
- [26] A. Al-Ameeri, "The Effect of Steel Fiber on Some Mechanical Properties of Self Compacting Concrete," *American Journal of Civil Engineering*, vol. 1, no. 3, pp. 102–110, 2013, <https://doi.org/10.11648/j.ajce.20130103.14>.
- [27] S. L. Hake, S. S. Shinde, P. K. Bhandari, P. R. Awasarmal, and B. D. Kanawade, "Effect of Glass Fibers on Self Compacting Concrete," *E3S Web of Conferences*, vol. 170, 2020, Art. no. 06018, <https://doi.org/10.1051/e3sconf/202017006018>.
- [28] D. B. Kulkarni and S. N. Patil, "Comparative Study of Effect of Sustained High Temperature on strength Properties of Self Compacting Concrete and Ordinary Conventional Concrete," *International Journal of Engineering and Technology*, vol. 3, no. 2, pp. 106–118, 2011.
- [29] E. T. Dawood and M. Ramli, "Mechanical properties of high strength flowing concrete with hybrid fibers," *Construction and Building Materials*, vol. 28, no. 1, pp. 193–200, Mar. 2012, <https://doi.org/10.1016/j.conbuildmat.2011.08.057>.

Design and Development of a Wheelchair Prototype

Van-Tinh Nguyen

School of Mechanical Engineering, Hanoi University of Science and Technology, Vietnam
tinh.nguyenvan@hust.edu.vn

Tran Thanh Tung

Faculty of Mechanics and Automation, VNU University of Engineering and Technology, Vietnam
tranthanh tung@vnu.edu.vn (corresponding author)

Received: 1 January 2024 | Revised: 2 February 2024 | Accepted: 11 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6851>

ABSTRACT

The rate of people with disabilities in Vietnam is about 7% out of a total of 98 million. Wheelchairs are popular assistive devices for disabled people and are often found in many places around the world, including Vietnam. This study proposes a novel design of a prototype electric wheelchair to support people with disabilities in Vietnam. The electric wheelchair model was successfully simulated and manufactured, fully meeting the proposed technical requirements. The proposed model works well and is suitable for the shape and physical strength of Vietnamese people.

Keywords-disabilities; wheelchair; prototype; design; manufacturing

I. INTRODUCTION

People with disabilities in Vietnam account for about 7% of the total of 98 million. The disability rate tends to increase with age, whereas women account for 58% and men for 42% [1-2]. These people have difficulty moving from one place to another. Wheelchairs are popular assistive devices for disabled people and can be found in many places around the world, including Vietnam. A conventional wheelchair is operated manually by the person sitting on it, sometimes making the user feel stressed and tired. In many cases, even though there are wheelchairs to assist disabled people, they still need the help of others to move. This dependency makes people with disabilities feel helpless and depressed. Therefore, electronic wheelchairs have been developed, which is a type of motor-controlled wheelchair that allows users to move more easily to the desired place without much effort.

In [3], an electric wheelchair was proposed, which was equipped with mecanum wheels, allowing it to move in any desired direction and speed using a joystick controller. In [4], an electric wheelchair equipped with a joystick controller was presented to serve as a personal mobility device for people with disabilities. In [5], a mechanical design for a wheelchair was presented that was capable of transitioning between sitting, lying, and standing positions. The upright position helps patients access elevated areas and participate in physical activity. In [6], the durability of the materials used in standing electric wheelchairs was examined. This study focused on evaluating the strength of the materials when subjected to different patient weights, ranging from 40 to 90 kg. In [7], a design for an automated multiposition dynamic wheelchair was

proposed, which could transport quadriplegic patients by modifying the structure of a manual wheelchair. In [8], a new wheelchair was introduced, which incorporated a body posture adjustment feature. This innovative design aimed to meet the specific needs of older people and individuals with disabilities. In [9], a method was proposed to detect and classify a step and subsequently autonomously climb it safely. In [10], a design optimization study of non-pneumatic castors for wheelchairs was carried out using Finite Element Analysis (FEA). This study compared the mechanical properties of these castors with those of a pneumatic castor with similar characteristics. In [11], a scientific examination of ten distinct ramps was carried out to assess the structures that directly impact the sense of comfort or discomfort experienced by wheelchair users with assistance, as well as the coefficients of friction of various flooring surfaces. In [12], a detailed examination of an automated control system was presented, which allowed a robotic assistant to push a wheelchair upstairs. In [13], SmartWheels was proposed, which is a method for identifying urban characteristics by analyzing data from inertial sensors that are generated by wheelchair movements. In [14], the design and development of a 4-wheel driven omni wheelchair was presented, which was specifically designed for indoor navigation and aimed to minimize wheel slippage and vibration. In [15], a design process for a new electric wheelchair controller was proposed, which allowed both wheelchair users and their caregivers to have control access.

Due to advancements in science and technology, machines equipped with integrated electrical control systems have become increasingly familiar [16-23]. This study aims to establish a systematic approach to develop and produce a

wheelchair prototype that is both cost-effective and equipped with comprehensive support features that help people with disabilities. The electric wheelchair model was successfully simulated and a prototype was manufactured, fully meeting the proposed technical requirements. This machine worked well and was suitable for the shape and physical strength of the Vietnamese.

II. METHODOLOGY

A. Design Object

The vehicle's parameters for wheelchair design are heavily influenced by the user's physical characteristics. The fundamental variables consist of width, length, height, and allowed load. The average height of Vietnamese is approximately 1.65 meters, while their average weight is approximately 70 kg. The width of the chair ranges from 380 to 450 mm, and the width from the backrest to the knee measures approximately 400-500 mm. The selected dimensions were width×length = 380×480 mm. The height varies between 900-1000 mm. The proposed electric wheelchair was designed with dimensions of approximately 1000×700×920 mm, considering the average physical condition and height of Vietnamese people. Additionally, the vehicle weight is less than 30 kg. Figure 1 illustrates the configuration of the wheelchair prototype, with the control box positioned on the right side, which is typically the dominant-hand side for most people.

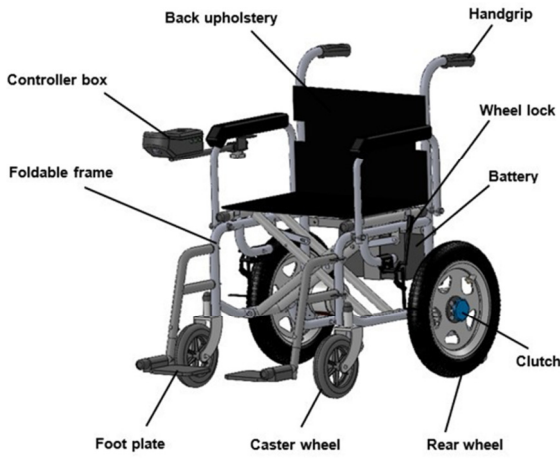


Fig. 1. The mechanical structure of the electric wheelchair prototype.

B. Mechanical Design of Wheelchair

Figure 2 illustrates the force system exerted on the wheelchair frame. The force balance equations is described in (1):

$$\vec{F} + \vec{F}_{ms} + \vec{F}_{qt} + \vec{R} + \vec{G} + \vec{N} = 0 \quad (1)$$

$$\begin{cases} F - F_{ms} - F_{qt} = 0 \\ R + G - N = 0 \end{cases} \quad (2)$$

$$N = R + G \quad (3)$$

$$F = f * N + \frac{R+G}{g} * a = (f * g + a) * \frac{R+G}{g} \quad (4)$$

where a is the acceleration of the vehicle when the speed changes from $v_o = 0$ km/h to $v = 6$ km/h.

$$a = \frac{v - v_o}{t} = \frac{5}{6} \text{ (m/s}^2\text{)} \quad (5)$$

$$F = \left(0,005 * 9.8 + \frac{5}{6}\right) * \frac{(70+30)*9.8}{9.8} = 150 \text{ (N)} \quad (6)$$

Using two motors for propulsion results in the distribution of torque on the shaft of each motor:

$$T = \frac{150}{2} * \frac{16*2.54}{2.100} = 15 \text{ (Nm)} \quad (7)$$

The power of the motor is:

$$P = \frac{T * n}{9.55} \quad (8)$$

and the motor shaft revolutions are:

$$n = \frac{1000 * v}{\pi * d} = \frac{1000 * 6}{\pi * 16 * 25.4 * 0,06} = 78.3 \text{ rpm} \quad (9)$$

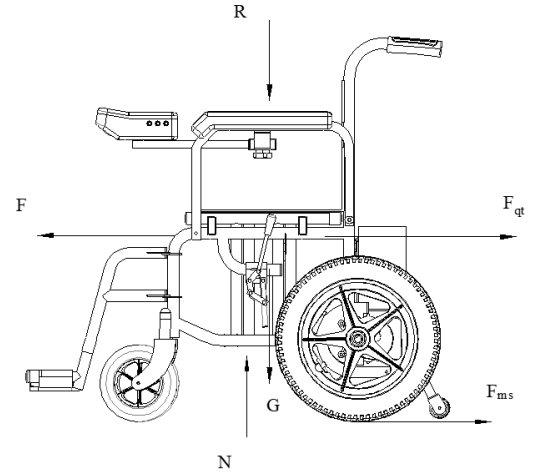


Fig. 2. Forces acting on the wheelchair.

C. Finite Element Analysis

Simulation analyses are necessary to verify that the prototype satisfies the required technical specifications before fabrication [24-28]. The numerical simulation process was carried out using SolidWorks software, as shown in Figure 3. The frame material is stainless steel SUS 304 which has an elastic modulus of 1.9×10^5 Mpa, a Poisson ratio of 0.29, and a yield strength of 206.807 Mpa. The external force acting on the frame is the weight, which is set for the most common simulation, equal to 70 kg. This force is transferred to two blue rods, as shown in Figure 3. In addition, the supporting points are set at the bottom of the frame. Figures 4 and 5 show the simulation and test results for strength and displacement. Stress analysis indicates that the frame provides complete durability assurance, with the highest stress measured at 23 Mpa, far below the allowable limit of 250 Mpa. The displacement analysis indicates that the maximum displacement measured is 0.122 mm, indicating an insignificant magnitude of displacement. The results indicate that the wheelchair structure fully satisfies the technical specifications.

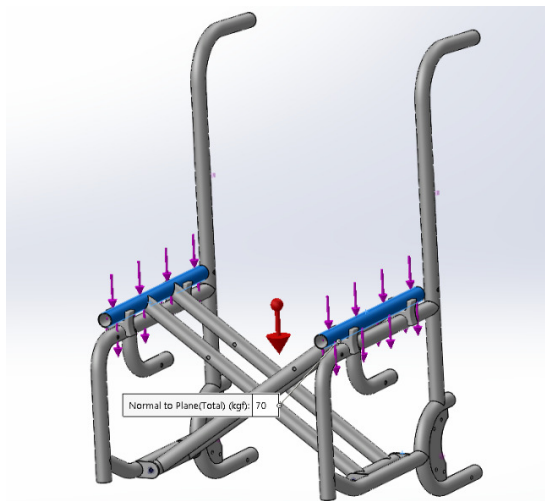


Fig. 3. Simulation in Solidworks.

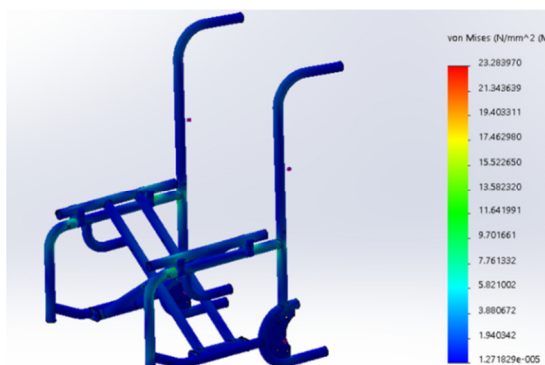


Fig. 4. Stress result.

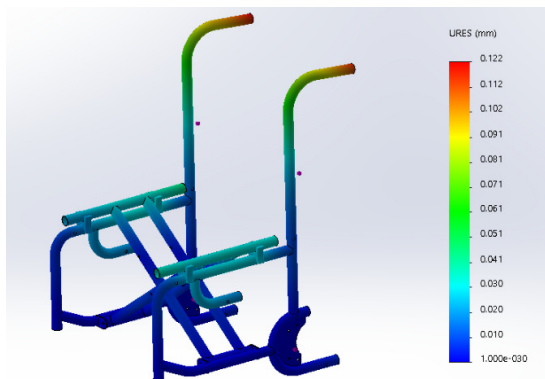


Fig. 5. Displacement result.

III. RESULTS AND DISCUSSION

Figure 6 illustrates the prototype wheelchair that was successfully manufactured. The prototype's dimensions were 1000x670x930 mm (lengthxwidthxheight). When folded, its length was reduced to 330 mm. The wheelchair has a speed range of 2 to 6 km/h, a maximum travel distance of 20 km, and can climb slopes with a maximum inclination of 35°. Furthermore, the wheelchair is equipped with an integrated GPS position tracker.



Fig. 6. A prototype of the electric wheelchair.



Fig. 7. Testing of the wheelchair.

Three men in good health, weighing between 55 and 70kg, were recruited to evaluate the electric wheelchair prototype. Figure 7 shows a momentary representation of the wheelchair control process. The results indicate that the mechanical movements exhibit a high degree of fluidity, allowing the vehicle to navigate with smooth motion and effortless control operations.

IV. CONCLUSION AND FUTURE WORK

This study presented the design and development of a prototype electric wheelchair to assist disabled people. The

electric wheelchair model was successfully simulated and manufactured, fully meeting the proposed technical requirements. The machine works well and is suitable for the shape and physical strength of Vietnamese people. Future work could be done to improve the vehicle's ability to overcome obstacles and optimize the lifting and lowering ability of the handrail. In addition, it is possible to integrate more advanced techniques into electric wheelchairs, such as the ability to control the vehicle via Bluetooth and add electronic brakes.

REFERENCES

- [1] "Situational Analysis Report on the Rights of Persons with Disabilities in Viet Nam," UN Partnership on the Rights of Persons with Disabilities, Mar. 2023. [Online]. Available: <https://www.undp.org/vietnam/publications/situational-analysis-report-rights-persons-disabilities-viet-nam>.
- [2] "Report on Improving Employment Opportunities for Persons with Disabilities in Viet Nam," UN Partnership on the Rights of Persons with Disabilities, Dec. 2020. [Online]. Available: <https://www.undp.org/vietnam/publications/report-improving-employment-opportunities-persons-disabilities-viet-nam>.
- [3] N. Thongpance and P. Chotikunnan, "Design and Construction of Electric Wheelchair with Mecanum Wheel," *Journal of Robotics and Control (JRC)*, vol. 4, no. 1, pp. 71–82, Mar. 2023, <https://doi.org/10.18196/jrc.v4i1.17095>.
- [4] H. R. Fajrin, T. A. Zain, and M. Irfan, "Design of Electric Wheelchair with Joystick Controller as Personal Mobility for Disabled Person," in *Proceedings of the 1st International Conference on Electronics, Biomedical Engineering, and Health Informatics*, Surabaya, Indonesia, 2021, pp. 151–159, https://doi.org/10.1007/978-981-33-6926-9_14.
- [5] D. Wang and H. Yu, "Mechanical Design of a Wheelchair with Multi-Posture Characteristics," presented at the 2017 2nd International Conference on Electrical, Automation and Mechanical Engineering (EAME 2017), Apr. 2017, <https://doi.org/10.2991/eame-17.2017.44>.
- [6] R. H. Putra, A. G. W. Rahman, E. S. Ningrum, and D. S. Purnomo, "Design and stress analysis on electric standing wheelchair," in *2017 International Electronics Symposium on Engineering Technology and Applications (IES-ETA)*, Sep. 2017, pp. 112–117, <https://doi.org/10.1109/ELECSYM.2017.8240388>.
- [7] L. A. Aguilar-Pérez, J. C. Paredes-Rojas, J. I. Sanchez-Cruz, J. A. Leal-Naranjo, A. Oropeza-Osornio, and C. R. Torres-SanMiguel, "Design of an Automated Multiposition Dynamic Wheelchair," *Sensors*, vol. 21, no. 22, 2021, <https://doi.org/10.3390/s21227533>.
- [8] J. Rui and Q. Gao, "Design and Analysis of A Multifunctional Wheelchair," *IOP Conference Series: Materials Science and Engineering*, vol. 538, no. 1, Feb. 2019, Art. no. 012045, <https://doi.org/10.1088/1757-899X/538/1/012045>.
- [9] A. Botta, R. Bellincioni, and G. Quaglia, "Autonomous detection and ascent of a step for an electric wheelchair," *Mechatronics*, vol. 86, Oct. 2022, Art. no. 102838, <https://doi.org/10.1016/j.mechatronics.2022.102838>.
- [10] I. Ali, M. Z. U. Rahman, A. Hussain, A. Ghaffar, and S. Zafar, "A finite element analysis based design of a non-pneumatic wheel chair castor," *Journal of Engineering Research*, Aug. 2023, <https://doi.org/10.1016/j.jer.2023.08.013>.
- [11] V. Tatano and R. Revellini, "An alternative system to improve accessibility for wheelchair users: The stepped ramp," *Applied Ergonomics*, vol. 108, Apr. 2023, Art. no. 103938, <https://doi.org/10.1016/j.apergo.2022.103938>.
- [12] H. Ikeda, T. Toyama, D. Maki, K. Sato, and E. Nakano, "Cooperative step-climbing strategy using an autonomous wheelchair and a robot," *Robotics and Autonomous Systems*, vol. 135, Jan. 2021, Art. no. 103670, <https://doi.org/10.1016/j.robot.2020.103670>.
- [13] S. Mascetti, G. Civitarese, O. El Malak, and C. Bettini, "SmartWheels: Detecting urban features for wheelchair users' navigation," *Pervasive and Mobile Computing*, vol. 62, Feb. 2020, Art. no. 101115, <https://doi.org/10.1016/j.pmcj.2020.101115>.
- [14] A. S. Kundu, O. Mazumder, P. K. Lenka, and S. Bhaumik, "Design and Performance Evaluation of 4 Wheeled Omni Wheelchair with Reduced Slip and Vibration," *Procedia Computer Science*, vol. 105, pp. 289–295, Jan. 2017, <https://doi.org/10.1016/j.procs.2017.01.224>.
- [15] K. Ma and Z. Qi, "A Human-Centered Design of Electric Wheelchair Controller With Dual Control Access for Both Drivers of Disabled People and Caregiver," *Journal of Computing and Information Science in Engineering*, vol. 17, no. 3, Feb. 2017, Art. no. 031005, <https://doi.org/10.1115/1.4034742>.
- [16] J. Iqbal, "Modern Control Laws for an Articulated Robotic Arm: Modeling and Simulation," *Engineering, Technology & Applied Science Research*, vol. 9, no. 2, pp. 4057–4061, Apr. 2019, <https://doi.org/10.48084/etasr.2598>.
- [17] S. Husnain and R. Abdulkader, "Fractional Order Modeling and Control of an Articulated Robotic Arm," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12026–12032, Dec. 2023, <https://doi.org/10.48084/etasr.6270>.
- [18] T. T. K. Ly, N. T. Thanh, H. Thien, and T. Nguyen, "A Neural Network Controller Design for the Mecanum Wheel Mobile Robot," *Engineering, Technology & Applied Science Research*, vol. 13, no. 2, pp. 10541–10547, Apr. 2023, <https://doi.org/10.48084/etasr.5761>.
- [19] N.-K. Nguyen, T. T. Pham, and V.-T. Nguyen, "A Novel Design of Automatic Longan Seed Removing Machine," in *Proceedings of the 8th International Conference on Mechanical, Automotive and Materials Engineering*, Hanoi, Vietnam, 2023, pp. 59–65, https://doi.org/10.1007/978-981-99-3672-4_5.
- [20] T. T. Tung, N. X. Quynh, and T. V. Minh, "Development and Implementation of a Mini CNC Milling Machine," *Acta Marisensis. Seria Technologica*, vol. 18, no. 2, pp. 24–28, Dec. 2021.
- [21] T. V. Minh and T. T. Tung, "A Prototype of Automated Chayote Peeling Machine," *Annual Journal of Technical University of Varna*, vol. 6, no. 2, pp. 83–93, Dec. 2022, <https://doi.org/10.29114/ajtuv.vol6.iss2.285>.
- [22] N.-K. Nguyen and V.-T. Nguyen, "A Prototype of Lotus Fiber Extracting Machine," in *Advances in Engineering Research and Application*, 2023, pp. 65–73, https://doi.org/10.1007/978-3-031-22200-9_7.
- [23] D. H. Diep, B. T. Phong, N. X. Quynh, and T. T. Tung, "Design and Fabricate a Prototype Smart Helmet," *TNU Journal of Science and Technology*, vol. 228, no. 10, pp. 118–126, May 2023, <https://doi.org/10.34238/tnu-jst.7632>.
- [24] T. T. Tung, N. Van Tinh, D. T. Phuong Thao, and T. V. Minh, "Development of a prototype 6 degree of freedom robot arm," *Results in Engineering*, vol. 18, Jun. 2023, Art. no. 101049, <https://doi.org/10.1016/j.rineng.2023.101049>.
- [25] T. V. Minh, V. C. Thiet, and T. T. Tung, "Design of an automated plastic bag packaging machine," *Acta Technologica*, vol. 9, no. 1, pp. 35–38, Mar. 2023, <https://doi.org/10.22306/atec.v9i1.166>.
- [26] E. Gardie Damtie and H. Dubale, "Finite element analysis of crank and slotted lever quick return mechanism for shaper machine application," *Results in Materials*, vol. 14, Jun. 2022, Art. no. 100288, <https://doi.org/10.1016/j.rinma.2022.100288>.
- [27] T. T. Tung, L. T. H. Ngoc, N. X. Quynh, and T. V. Minh, "Design and Analysis of a Paper Shredder Machine," *Acta Technologica*, vol. 7, no. 3, pp. 93–97, Sep. 2021, <https://doi.org/10.22306/atec.v7i3.117>.
- [28] T. T. Tung, L. T. H. Ngoc, N. X. Quynh, and T. V. Minh, "Design and Analysis of a Paper Shredder Machine," *Acta Technologica*, vol. 7, no. 3, pp. 93–97, Sep. 2021, <https://doi.org/10.22306/atec.v7i3.117>.
- [29] D. Apparao and M. V. Jagannadha Raju, "Design and analysis of spur gear manufactured by DMLS process," *Materials Today: Proceedings*, vol. 46, pp. 149–153, Jan. 2021, <https://doi.org/10.1016/j.matpr.2020.07.078>.

Structural Behavior of Concrete One-Way Slab with Mixed Reinforcement of Steel and Glass Fiber Polymer Bars under Fire Exposure

Mohammed R. Rasheed

Department of Civil Engineering, Engineering College, University of Baghdad, Iraq
mohammed.zaed2101m@coeng.uobaghdad.edu.iq (corresponding author)

Shatha D. Mohammed

Department of Civil Engineering, Engineering College, University of Baghdad, Iraq
shatha.dh@coeng.uobaghdad.edu.iq

Received: 21 December 2023 | Revised: 3 February 2024 | Accepted: 11 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6795>

ABSTRACT

Steel Reinforced Concrete (RC) frequently faces durability problems. In certain areas, Glass Fiber-Reinforced Polymer (GFRP) rebars are considered a non-corrodible substitute for steel reinforcement. Elevated temperatures have a significant impact on the mechanical characteristics and the adhesiveness of GFRP rebars to concrete, particularly when the polymeric matrix's glass transition temperature is approached or surpassed. Three simply supported reinforced concrete slabs were considered in the experimental program. Each specimen had identical dimensions of 1500×540×120 mm. For the fire resistance requirements, a 45 mm clear concrete cover and an exception of a 200 mm unexposed (cool) anchor zone at the ends were considered. The GFRP replacement ratio was 0, 20, and 40%. The burning procedure involved fire exposure for an hour with a steady-state temperature of 500 °C in accordance with ASTM E-119 regarding the temperature time elevation and a sudden cooling condition. The optimal concrete cover was detected by testing a fire-exposed small model reinforced by GFRP bars of varying concrete cover. The specimen was tested under static intense loads. The reference slab and the slab with a replacement percentage of 20% failed due to flexural failure, whereas the slab with a replacement percentage of 40% failed due to shear failure. The influence of the GFRP replacement ratio was extended to include toughness and ultimate load. A replacement percent of 20% increased them by 18.30, and 2.62%, respectively, while a replacement percent of 40% decreased them by 28.16, and 3.13%, accordingly. It was also shown that the location of replacing the GFRP and 200 mm of unexposed (cold) installation area at the ends with a 45 mm concrete cover has a significant impact. The more the GFRP is located in the middle, away from the ends, the better the fire resistance is.

Keywords-reinforced concrete slabs; fire resistance; Glass Fiber-Reinforced Polymer (GFRP); concrete cover

I. INTRODUCTION

Over the past ten years, Fiber-Reinforced Polymer (FRP) materials have attracted increasing attention in civil engineering research. The original applications for FRPs were in the chemical processing, automotive, and aerospace industries. High strength-to-weight ratios and corrosion resistance are just two of their advantages, which make them desirable for usage in a wide range of civil engineering applications. There are many FRP field applications due to advancements in knowledge and FRP technology, and the rise in their utilization. These developments have also decreased material cost and increased structural designers' trust. FRPs show significant potential across various applications, with a notable impact on internal concrete reinforcement. In this role, FRPs can effectively substitute traditional steel reinforcement, providing a solution to combat electrochemical corrosion.

Electrochemical corrosion has substantially contributed to the ongoing global infrastructure crisis and the deterioration observed in reinforced concrete structures over the last 50 years [1-6]. The performance of FRP-reinforced concrete structures has been the subject of numerous studies [7-15]. However, research indicates that FRP-reinforced concrete slabs usually fracture abruptly and unevenly. The combination of FRP with steel bars, specifically by placing Glass Fiber-Reinforced Polymer (GFRP) at the ends of a structure, emerges as an effective method to prevent corrosion of stainless steel within the construction. This integration acts as a preventive measure, strategically inhibiting the corrosion of stainless-steel components in the structure. The benefits of both FRP and steel bars enhance the overall performance of the slab. Many researchers have explored the performance of concrete structures reinforced with this combination [16-18].

As its name implies, FRPs are composed of tiny fibers embedded in a polymer matrix. Epoxies, polyesters, or vinyl esters, which are typically carbon (graphite), glass, or aramid (Kevlar), are commonly employed as matrices for the fibers used in civil engineering. The mechanical and thermal properties of FRPs are greatly influenced by temperature, and even a slight rise in temperature (between 100 and 200 °C) can result in a discernible deterioration in the FRPs' mechanical qualities [1]. FRP-reinforced parts' fire endurance can be affected by a number of factors, including the type of reinforcement used, the type of aggregate utilized, and the thickness of the concrete cover. With different FRP materials exhibiting varying degrees of mechanical and bond property degradation at high temperatures, the kind of reinforcement is a crucial factor to take into account [19]. The type of aggregate and the concrete cover have a significant impact on the heat transfer to the reinforcement, which in turn influences the temperatures in the exposed and anchoring zones of the bars during a fire [19].

Before the most recent edition of ACI 440.1R (ACI 2015), the use of FRP reinforcing bars was not advised in circumstances where maintaining structural integrity in the face of fire was critical [19]. Experimental research, however, has raised expectations for GFRP bars' fire performance [20-21]. CSA S806 [22] recommends being cautious and using a thick concrete cover (e.g. 45 mm for 1 hr of fire resistance) to prevent GFRP bars from deteriorating during a fire.

Authors in [22] delved into assessing the fire resistance of concrete slabs reinforced with FRP. Their research encompassed numerical parametric analysis and a comprehensive review of the existing literature in this domain. The investigation specifically employed numerical finite difference analysis to scrutinize the time-temperature response of a concrete slab internally reinforced with FRP under fire exposure. Upon thorough examination of the limited available literature, the authors concluded that, when reaching a critical temperature of 250°C in the internal FRP reinforcement, certain effects or behaviors were observed. The numerical approach graded fire endurance and represented the worst-case scenario for GFRP. The study discovered that thicker concrete covers, carbonate aggregate concrete, and FRP-reinforced concrete slabs have higher fire endurance than slabs reinforced with conventional reinforcing steel. It was also found that material property data for FRPs at elevated temperatures are required, and structural fire endurance tests are necessary to validate the numerical models. In [24], the impact of various reinforcement techniques on the firing behavior of RC beams under service load was analytically investigated. When a beam's resistance could no longer sustain the service load, the beam failed. A steel beam shattered after 100 min of fire, but a Carbon Fiber-Reinforced Polymer (CFRP) beam withstood the heat for 60 min. However, after being exposed to fire for 40 minutes, the GFRP-RC and AFRP beams crumbled.

Authors in [21, 25] conducted an experimental research, which showed that fire-exposed GFRP-RC slab bond failure can be avoided by providing sufficient anchorage of bars with extended unexposed (cool) portions. Although previous experimental research raised concerns regarding fire safety and

GFRP bar application in concrete, these findings may reduce bond failure in fire-exposed GFRP-RC slabs. Two GFRP RC slabs containing carbonate and siliceous coarse aggregates were tested in [26]. With temperatures 10% lower than the slab with siliceous aggregates, the slab with carbonate aggregates fared better in the fire. The authors conducted numerical parametric investigations on slabs, including those two aggregates as well as expanded shale aggregates as a result of their findings. The enlarged shale aggregate slabs demonstrated a fire resistance of 85 min with a 50 mm suggested cover, followed by siliceous (65 min) and carbonate (72 min) slabs. GFRP-RC slabs with cover thicknesses of 25 and 38 mm were tested in a fire scenario without any load in [27]. As predicted, after 4 hr of fire exposure, the bottom temperatures of the longitudinal rebars varied significantly, with a lower slab having 100 °C higher temperature than the higher slab.

The literature study presented above emphasizes a few crucial findings: (i) the thickness of the concrete cover significantly impacts GFRP-RC elements' thermal response, (ii) concrete elements containing carbonate aggregates fare slightly better in fires than those made with siliceous aggregates, (iii) overall slab thickness does not appear to significantly affect fire behavior, (iv) while comparative studies are limited, CFRP-RC elements generally perform better than GFRP-RC elements, though the results are not very clear, (v) the bond between FRP reinforcement and concrete greatly affects fire behavior, bent rebars at the far end of the anchorage can increase fire resistance, and (vi) lap splices in FRP rebars significantly affect the fire resistance of FRP-RC beams.

According to the outcomes of the literature review concerning the structural behavior of GFRP-reinforced concrete elements under fire exposure, it was concluded that the support location would be damaged, and the fire effect would cause local failure. The current analysis aims to offer valuable insights into the slab post-fire behavior, especially with mixed reinforcement while considering different GFRP replacement ratios of 0, 20, and 40%, defining an optimized combination of a 45 mm thick concrete cover and a 200 mm unexposed (cold) fixation area at the end to improve the structural behavior. Examining the test results includes assessing critical aspects, such as ultimate load capacity, load-deflection behavior, load-strain relationships, and the material's toughness.

II. EXPERIMENTAL PROGRAM

A. Materials

The casting of all samples and control specimens involved the utilization of Ordinary Portland Cement (CEM I 42.5R) obtained from the Mass brand in Iraq. An evaluation of the cement's adherence to the Iraqi Specification No. 5/2019 [28] was conducted. The chosen sand was classified as belonging to Zone 2. The siliceous aggregates utilized in the study consisted of crushed gravel with a maximum particle size of 12 mm. Both fine and coarse aggregates exhibited conformity with the Iraqi Specification (IQS) No. 45/1993 [29]. The details about the combinations can be found in Table I. The specifics of the utilized GFRP and steel bars are presented in Table II.

TABLE I. DETAILS OF CONCRETE MIX PROPORTIONS.

Cement (kg/m ³)	Gravel (kg/m ³)	Sand (kg/m ³)	Water (l/m ³)	Optima 100 (kg/m ³)	Silica fume (kg/m ³)	Specimen strength f_{cu} (MPa)
470	945	827	170	6.2	20	54

TABLE II. THE PROPERTIES OF REINFORCEMENT BARS.

Bars	Nominal diameter (mm)	Tensile strength (MPa)	Modulus of elasticity E, (MPa)	Elongation (%)
GFRP	10	1207	48280	2.5
Steel	10	437	200000	11
Steel	8	434	200000	12

B. Concrete Cover Checking

Before the specimens were prepared, a 400 × 210 × 110 mm miniature model with a compressive strength of $f_{cu} = 40$ MPa was cast and reinforced by four GFRP bars, the mix proportion is displayed in Table III. As discerned in Figure 1, the model was cast on a variety of concrete covers. The model was coated with water-soaked canvases that were sprayed every day. The miniature model was then subjected, in compliance with ASTM E-119, to fire at a steady-state temperature of 500 °C for 1 hr (Figure 2). When the fire exposure ended, rapid cooling technique (water spraying) was used to chill the model. After that, as observed in Figure 3, the model was broken and the GFRP bars were removed. The GFRP bars on the sides were seen to be burned, and it was discovered that the middle cover, which was 45 mm from the bottom, best prevented burning.

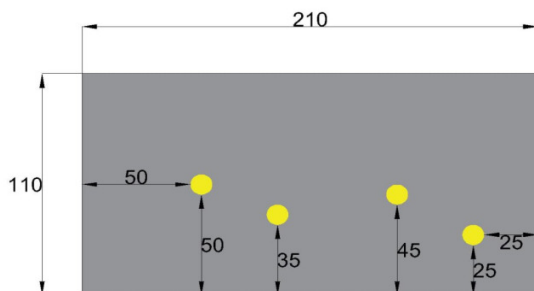


Fig. 1. Dimensions of a miniature model (mm).

TABLE III. MIXTURE SPECIFICATIONS FOR THE ADOPTED COMPRESSIVE STRENGTH MODEL

Water (kg/m ³)	Cement (kg/m ³)	Sand (kg/m ³)	Gravel (kg/m ³)	f_{cu} (MPa)
200	400	755	1002	40



Fig. 2. Casting and burning of the miniature model.

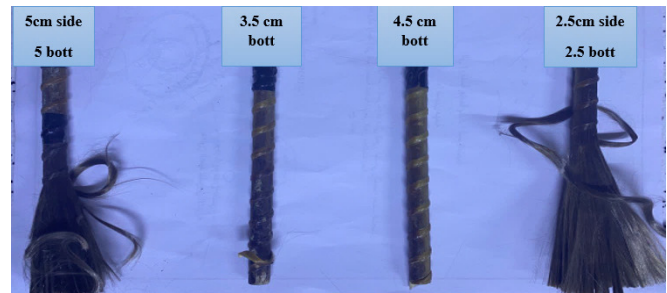


Fig. 3. Condition of the GFRP bars after exposure to fire.

C. Tested Specimens

The experimental program included testing three identical concrete slabs with the same geometric layout and concrete compressive strength. Each slab was 1500 mm, 550 mm, and 120 mm in length, width, and depth, respectively. Additional details regarding the test specimens are portrayed in Figure 4 and Table IV.

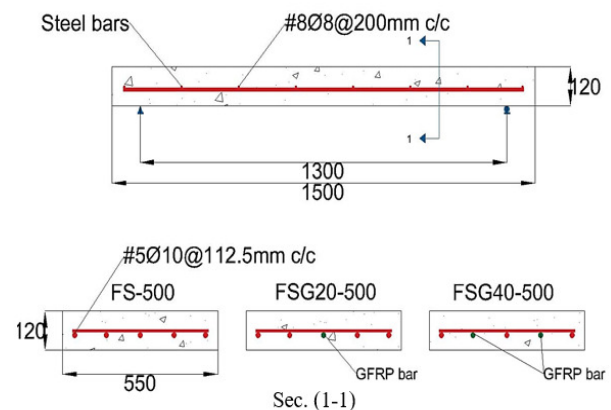


Fig. 4. Dimensions and reinforcement details of the specimens.

TABLE IV. CHARACTERISTICS OF THE TESTED SPECIMENS.

Slab designation	f'_c MPa	Temperature (°C)	GFRP ratio (%)
FS-500	43	500	0
FSG20-500	43	500	20
FSG40-500	43	500	40

D. Burning and Cooling

As shown in Figures 5 and 6, the specimens were set inside a gas furnace and were exposed to a direct flame for 1 hr while keeping the temperature steady at 500±10 °C. A digital thermometer reader fitted with sensor wire type K was used to track the temperature of the specimen and furnace region, according to ASTM E-119 [30]. The specimens were sprayed with water (sudden cooling) to lower their temperature after the burning procedure.

D. Testing Procedure

Two hemispherical supports, intended to offer straightforward support, were attached to the slabs. The length of the slabs under examination and the measurement of the distance between the supports were 1300 mm. After that, a 100-ton hydraulic jack was used to apply the load. A steel

loading, I-section girder equally distributed the load, producing two equal forces divided by a 400 mm span. Several measurements were made during the experiment, including the strain experienced by the concrete surface, the applied load amplitude, and the vertical deflection at the center of the slab. Every stage of the loading procedure involved recording of these measurements. Each phase of the load process entailed a 2.5 kN increase until failure occurred. The crack formation was meticulously monitored and recorded following each step. See Figures 7-9 for a more thorough explanation of the experimental setup.



Fig. 5. Setup of the furnace and specimens.

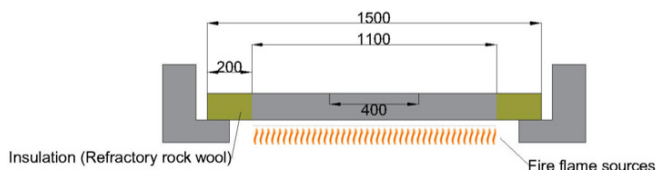


Fig. 6. Fire test setup (mm).

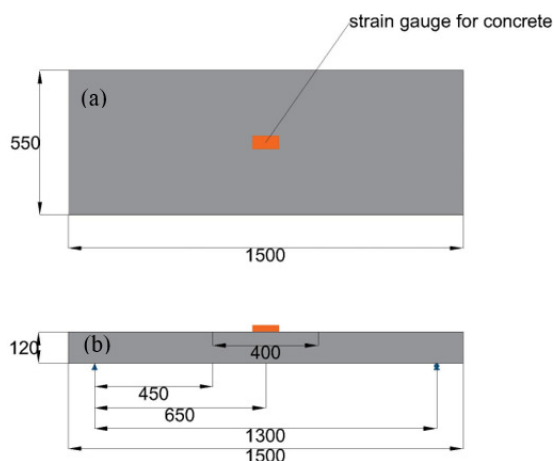


Fig. 7. Locations of the strain gauge: (a) Top face, (b) side face (mm).

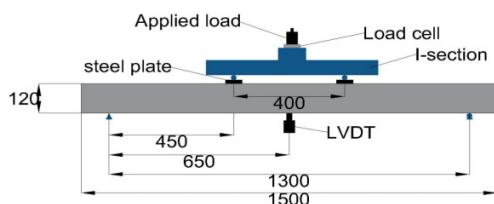


Fig. 8. Details of the test setup (mm).

III. TEST RESULTS AND DISCUSSION

An application of a two-point load was utilized to test three reinforced concrete slabs with different ratios of GFRP bars in place of steel bars as depicted in Figure 9. The load was increased steadily until the point of failure. The discussion comprised four categories to enhance the comprehension of the slab's structural behavior. These are:

- Crack pattern, failure load, and mode of failure.
- Load-deflection behavior.
- Load-strain relation.
- Toughness.

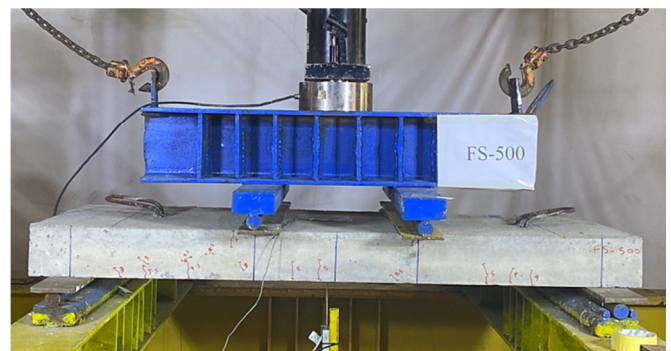


Fig. 9. Test setup.

A. Cracks Pattern, Failure Load, and Mode of Failure

The results for the burned specimens revealed that, after the firing and cooling process, cracks began to occur on the slabs' surfaces. Flexural cracks were also developed as an outcome of the fire in the slabs' sides and bottom. Once the burning and cooling procedure were completed, the specimens were put through a two-line static load test. As exposed in Table V, the variation in the ultimate load for the specimen FSG20-500 was +2.62% compared to the reference slab, whereas it was decreased by 3.13% for the specimen FSG40-500. The reduction in strength and mechanical properties of concrete material prevents the flexural reinforcement from reaching its maximum ultimate strength. Consequently, no significant influence was observed regarding the effect of glass fiber bars.

The slabs FS-500 and FSG20-500 experienced flexural failure while the slab (FSG40-500) experienced shear failure. First, additional cracks that emerged from fire fractures in the slab's bottom face were created. Flexural cracks spread along the bottom surface of the reference slab FS-500 and the slab with a replacement percent of 20% (FSG20-500), in a direction parallel to the support direction and the original crack as the stress increased. Cracks began to appear at the failure stage, ultimately spreading to the sides of the slab and the compression chord. The cracks were located in the central third of the slab. None was observed in the vicinity of the supports. The final specimens' breaking pattern is spotted in Figures 10 and 11. These figures unequivocally demonstrate that no fractures were observed in the area of support zones and that flexural cracks are roughly parallel to one another.

TABLE V. LOAD CAPACITY AND ULTIMATE DEFLECTION OF THE BURNED SPECIMENS

Specimen	Ultimate deflection (mm)	Ultimate load (kN)	% Variation of ultimate load (with respect to FS-500)
FS-500	49.445	69.3	---
FSG20-500	56.5708	71.12	+2.62
FSG40-500	41.3141	67.13	-3.13

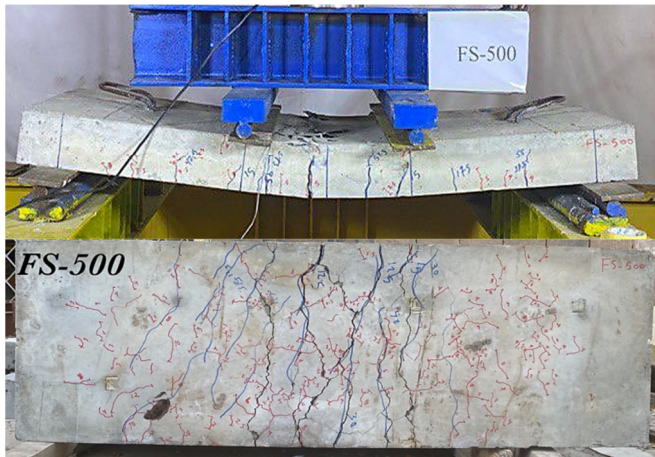


Fig. 10. Crack pattern of the specimen FS-500. Top: side face, bottom: bottom face.

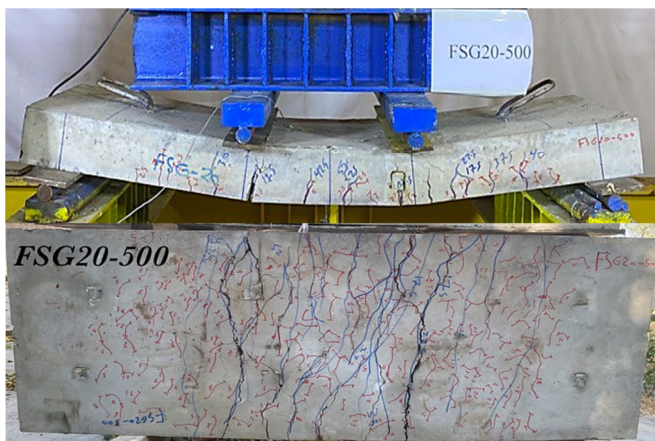


Fig. 11. Crack pattern of the specimen FSG20-500. Top: side face, bottom: bottom face.

In the slab FSG40-500, in which the substituted ratio was 40%, the pattern of the cracks was completely different. As the applied load increased, more flexural fracture development and propagation were observed on the slab's bottom surface. The supports' orientation and the initial fracture were parallel to one other. Subsequently, diagonal shear cracks started to propagate along the slab's sides. When the applied load was gradually increased, the shear fractures in the slab seemed to get larger and travel toward the loading location. As soon as the diagonal shear cracks were generated by the debonding of the GFRP bars, the slab collapsed. Due to the 40% replacement percent (2 from 5 bars) that increased the slab's flexural resistance, shear failure mode rather than flexure mode occurred.

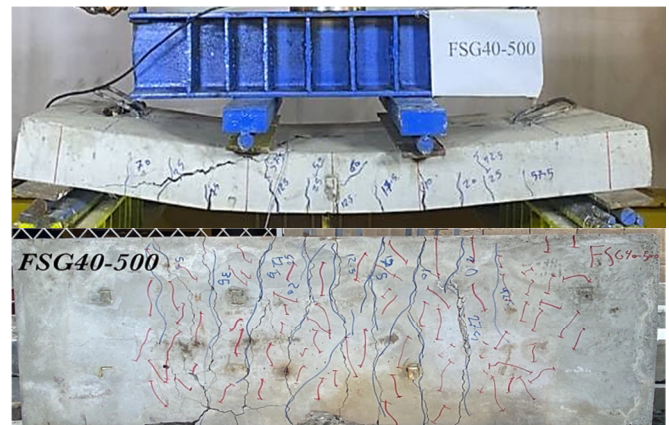


Fig. 12. Crack pattern of the specimen FSG40-500. Top: side face, bottom: bottom face.

B. Load-Deflection Behavior

Every load increment during the test procedure involved measuring the deflection at the slab center. The service and ultimate load phases were examined for sample deflection. It is generally assumed that the service load accounts for around 70–75% of the total load [31]. In the current experiment, the service load for each specimen was determined to be the 70% of the ultimate load. The specimens' ultimate loads were calculated based on the maximum bearing load, as illustrated in Table V. Throughout the incremental loading process, several distinct phases were generally observed. In the elastic zone, the deflection first increased gradually and steadily. When cracks began to develop and spread, the deflection rate quickened and accelerated. The deflection curve's slope then started to decline as this pattern persisted until the tension stress in the steel reinforcement reached its yield point. The test is declared to be terminated when the deflection keeps rising without the applied load increasing in proportion. Figure 13 exhibits how the GFRP bar percentage adjustments affect the mid-span load-deflection characteristics. FS-500 serves as the control sample, and the outcomes of the specimens FSG20-500 and FSG40-500 are compared with it. The load-deflection curves evidents that the three slabs' degrees of stiffness vary inside the elastic zone. Table VI showcases the specimens' rigidity (stiffness) at the elastic zone. The results indicate that the stiffness decreased by about 5.18 and 21.29% for a replacement percent of 20 and 40%, respectively.

TABLE VI. STIFFNESS OF THE TESTED SPECIMENS

Specimen	Load (kN)	Deflection at 20 kN load	Stiffness $K=P/\Delta$ (kN/mm)	Stiffness decrease (%)
FS-500	20	3.7	5.40	---
FSG20-500	20	3.9	5.12	5.18
FSG40-500	20	4.7	4.25	21.29

TABLE VII. MID-SPAN DEFLECTIONS OF THE TESTED SAMPLES AT SERVICE AND ULTIMATE LOADS

Specimen	Deflection at service load (mm)	% Increase in deflection at service load	Deflection at ultimate load (mm)	% Change in deflection at ultimate load
FS-500	9.3	Ref.	49.44	Ref.
FSG20-500	11.3	21.50	56.57	+14.4
FSG40-500	14.2	52.68	41.31	-16.45

The discrepancy is attributed to the different moduli of elasticity of the steel and GFRP, with the GFRP having a lower modulus of elasticity than the steel, given that all the other parameters of the specimens studied are compatible. In comparison to the control slab, this causes the load-deflection curve for the GFRP-reinforced slab to have a lower initial slope. These results validate the conclusions obtained in [32].

The trend of the load-deflection comprised two stages. The behavior before the steel reinforcement yield point constitutes the first stage. At this stage, the replacement of 20% demonstrated a behavior slightly similar to that of the reference slab, with an insignificant decrease in stiffness. There was an increase in the deflection, which led to a decrease in stiffness. As for the case of 40% replacement, the deflection increased prominently in the elastic region, and this led to a significant reduction in stiffness and deterioration of the concrete in the elastic region. This difference between the modulus of elasticity of steel and GFRP caused this jump in the structural behavior, as the GFRP has lower modulus of elasticity than steel. In the second stage, which starts after the steel yielding, it can be noticed that the behavior with 20% replacement was similar to that of the reference slab, as the deflection began to increase significantly with a slight increase in the load, i.e. the flexural resistance decreased and led to the flexure failure. In the case of 40% replacement after the point of steel yielding, the GFRP bar had a greater role in increasing load, and the load began to increase with a slight augmentation in deflection. The specimen began to tend toward the brittle (linear) behavior, which is compatible with the GFRP load-strain behavior. This means the flexural resistance increased and led to shear failure.

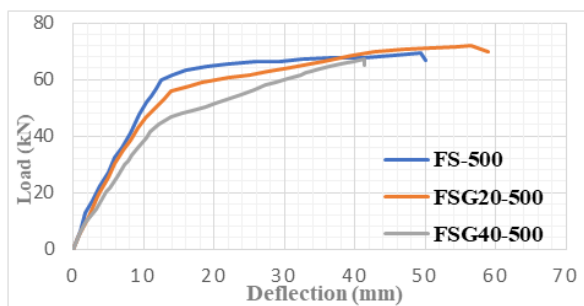


Fig. 13. Load-deflection behavior at mid-span.

C. Load-Strain Relation.

The concrete top surface's strain-load relations were measured at the mid-span. Figure 14 illustrates the effect of increasing the GFRP replacement percentage on the load-strain relations of the top concrete surface at the mid-span. In the elastic zone before the yield of steel reinforcement, the concrete's compressive strain was 710, 802, and 920 micro-strains for specimens FS-500, FSG20-500, and FSG40-500, respectively. This indicates that because steel and GFRP have different elastic moduli, the strain in the elastic area was directly proportional to the replacement ratio. The deflection in the elastic zone increases as the elastic modulus decreases, and the strain in the concrete in the elastic region increases as the deflection increases. The elastic region experienced an increase in strain and a significant decrease in stiffness at 40% material

replacement. Premature shear failure occurred as a result of the concrete's degradation in the elastic zone. The FS-500 specimen had a compressive strain of 4125 micro-strains, the FSG20-500 specimen had a micro-strain of 3362, and the FSG40-500 specimen had a micro-strain of 3220 at the ultimate load.

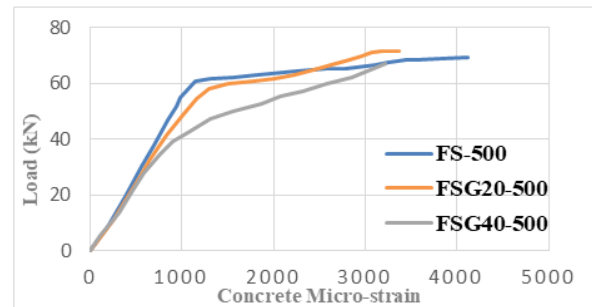


Fig. 14. Load-strain curves for concrete top fiber at the mid-span.

D. Toughness

The integral of the load-deflection curves for the slabs was computed to calculate the flexural toughness, also cited as total energy. Flexural toughness refers to a material's total ability to absorb energy. One important property in the realm of concrete structures is the ability of a loaded structure to store energy. The highest load magnitude and the deflection seen at the point of failure both have an impact on the absorbed energy, which is measured by the area under the load-deflection curve. The total energy of the slabs that were tested is displayed in Table VIII. It is evident that increased load capacity and deflection modified the toughness by 18.30% when 20% of the steel was replaced. Due to the lower deflection and load capacity, the toughness decreased by 28.16% when the replacement ratio was 40%

TABLE VIII. TOUGHNESS OF TESTED SLABS

Specimen	Toughness at ultimate load (kN.mm)	% variation of toughness at ultimate load	Toughness at failure load (kN.mm)	% variation of toughness at failure load
FS-500	2859.72	---	2897.55	---
FSG20-500	3254.28	+13.83	3427.92	+18.30
FSG40-500	1906.84	-27.21	1957.65	-28.16

IV. CONCLUSIONS

In this paper, a set of post-fire experiments was carried out on three reinforced concrete slabs to evaluate their structural performance. The collected data were subsequently analyzed, considering a variety of hybrid reinforcement ratios and placements. The main conclusions of this study are:

- The use of a fire protection process and optimal concrete cover prevent supporting local failure in fire-exposed specimens that contain mixed steel and GFRP reinforcement.
- The location of replaced GFRP bars plays a crucial role. Optimal positioning is essential for enhanced fire resistance. Specifically, placing GFRP more centrally, and

away from the ends, has a positive impact, resulting in superior fire resistance performance.

- To resist a fire of 500 °C for 1 hr, the utilization of 200 mm of unexposed (cold) installation surface at the ends with a 45 mm concrete cover was adequate.
- Increasing the GFRP replacement ratio up to 40% decreased the ultimate load. Consequently, no significant influence was observed regarding the effect of glass fiber, so it is recommended to use high compressive strength when using mixed reinforcement.
- The percentage of 20% replacement of steel reinforcement by GFRP bars did not change the failure mode. It was observed that the majority of the fractures were found in the middle third of the slab. In contrast, no cracks were found near the supports, suggesting flexural failure mode. Moreover, increasing the GFRP replacement ratio up to 40% changed the failure mode to shear failure.
- Since steel has a higher modulus of elasticity than GFRP, the stiffness in the elastic zone for mixed reinforced specimens decreases as the replacement percentage increases.
- A significant effect was absorbed regarding the compressive concrete strain as a mixed reinforcement of steel and GFRP was considered. The effect was directly proportional with the GFRP replacement ratio.
- It is evident that a 20% replacement of GFRP modified the toughness, whereas increasing the replacement ratio to 40% reduced it.

REFERENCES

- [1] L. A. Bisby, "Fire behaviour of fibre-reinforced polymer (FRP) reinforced or confined concrete," Ph.D. dissertation, Queen's University, Kingston, CA, USA, 2003.
- [2] T. H. Ibrahim, I. A. S. Alshaarabaf, A. A. Allawi, N. K. Oukaili, A. El-Zohairy, and A. I. Said, "Theoretical Analysis of Composite RC Beams with Pultruded GFRP Beams subjected to Impact Loading," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12097–12107, Dec. 2023, <https://doi.org/10.48084/etasr.6424>.
- [3] A. Jalil and A. H. Al-Zuhairi, "Behavior of Post-Tensioned Concrete Girders Subject to Partially Strand Damage and Strengthened by NSM-CFRP Composites," *Civil Engineering Journal*, vol. 8, no. 7, pp. 1507–1521, Jul. 2022, <https://doi.org/10.28991/CEJ-2022-08-07-013>.
- [4] M. A. E. Zareef, "An Experimental and Numerical Analysis of the Flexural Performance of Lightweight Concrete Beams reinforced with GFRP Bars," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10776–10780, Jun. 2023, <https://doi.org/10.48084/etasr.5871>.
- [5] B. Abdulkareem and A. F. Izzet, "Serviceability of Post-fire RC Rafters with Openings of Different Sizes and Shapes," *Journal of Engineering*, vol. 28, no. 1, pp. 19–32, Jan. 2022, <https://doi.org/10.31026/j.eng.2022.01.02>.
- [6] H. Q. Abbas and A. H. Al-Zuhairi, "Use of EB-CFRP to Improve Flexural Capacity of Unbonded Post-Tensioned Concrete Members Exposed to Partially Damaged Strands," *Civil Engineering Journal*, vol. 8, no. 6, pp. 1288–1303, Jun. 2022, <https://doi.org/10.28991/CEJ-2022-08-06-014>.
- [7] B. Saikia, P. Kumar, J. Thomas, K. S. N. Rao, and A. Ramaswamy, "Strength and serviceability performance of beams reinforced with GFRP bars in flexure," *Construction and Building Materials*, vol. 21, no. 8, pp. 1709–1719, Aug. 2007, <https://doi.org/10.1016/j.conbuildmat.2006.05.021>.
- [8] Z. Wang, X. Liang, Y. Wang, and T. Zhai, "Experimental and theoretical investigations on the flexural behavior of RC slabs with steel-PVA hybrid fiber reinforced cementitious composite (HFRCC) permanent formwork," *Case Studies in Construction Materials*, vol. 17, Dec. 2022, Art. no. e01432, <https://doi.org/10.1016/j.cscm.2022.e01432>.
- [9] W. Qu, X. Zhang, and H. Huang, "Flexural Behavior of Concrete Beams Reinforced with Hybrid (GFRP and Steel) Bars," *Journal of Composites for Construction*, vol. 13, no. 5, pp. 350–359, Oct. 2009, [https://doi.org/10.1061/\(ASCE\)CC.1943-5614.0000035](https://doi.org/10.1061/(ASCE)CC.1943-5614.0000035).
- [10] C. Barris, L. Torres, J. Comas, and C. Miàs, "Cracking and deflections in GFRP RC beams: An experimental study," *Composites Part B: Engineering*, vol. 55, pp. 580–590, Dec. 2013, <https://doi.org/10.1016/j.compositesb.2013.07.019>.
- [11] A. I. Said and O. M. Abbas, "Serviceability behavior of High Strength Concrete I-beams reinforced with Carbon Fiber Reinforced Polymer bars," *Journal of Engineering*, vol. 19, no. 11, pp. 1515–1530, Nov. 2013, <https://doi.org/10.31026/j.eng.2013.11.10>.
- [12] S. I. Ali and A. A. Allawi, "Effect of Web Stiffeners on The Flexural Behavior of Composite GFRP- Concrete Beam Under Impact Load," *Journal of Engineering*, vol. 27, no. 3, pp. 76–92, Feb. 2021, <https://doi.org/10.31026/j.eng.2021.03.06>.
- [13] A. H. A. Al-Ahmed and M. H. M. Al-Jburi, "Behavior of Reinforced Concrete Deep Beams Strengthened with Carbon Fiber Reinforced Polymer Strips," *Journal of Engineering*, vol. 22, no. 8, pp. 37–53, Aug. 2016, <https://doi.org/10.31026/j.eng.2016.08.03>.
- [14] M. Abas Golham and A. H. A. Al-Ahmed, "Behavior of GFRP reinforced concrete slabs with openings strengthened by CFRP strips," *Results in Engineering*, vol. 18, Jun. 2023, Art. no. 101033, <https://doi.org/10.1016/j.rineng.2023.101033>.
- [15] J. Abd and I. K. Ahmed, "The Effect of Low Velocity Impact Loading on Self-Compacting Concrete Reinforced with Carbon Fiber Reinforced Polymers," *Engineering, Technology & Applied Science Research*, vol. 11, no. 5, pp. 7689–7694, Oct. 2021, <https://doi.org/10.48084/etasr.4419>.
- [16] K. Protchenko, K. Mlodzik, M. Urbanski, E. Szmigiera, and A. Garbacz, "Numerical estimation of concrete beams reinforced with FRP bars," *MATEC Web of Conferences*, vol. 86, 2016, Art. no. 02011, <https://doi.org/10.1051/mateconf/20168602011>.
- [17] H. Y. Leung and R. V. Balendran, "Flexural behaviour of concrete beams internally reinforced with GFRP rods and steel rebars," *Structural Survey*, vol. 21, no. 4, pp. 146–157, Jan. 2003, <https://doi.org/10.1108/02630800310507159>.
- [18] R. A. Hawileh, "Finite element modeling of reinforced concrete beams with a hybrid combination of steel and aramid reinforcement," *Materials & Design (1980-2015)*, vol. 65, pp. 831–839, Jan. 2015, <https://doi.org/10.1016/j.matdes.2014.10.004>.
- [19] *ACI 440.1R-15(2015), Guide for the Design and Construction of Structural Concrete Reinforced with Fiber-Reinforced Polymer (FRP) Bars*. Farmington Hills, MI, USA: ACI Concrete, 2015.
- [20] L. A. Bisby and V. K. R. Kodur, "Evaluating the fire endurance of concrete slabs reinforced with FRP bars: Considerations for a holistic approach," *Composites Part B: Engineering*, vol. 38, no. 5, pp. 547–558, Jul. 2007, <https://doi.org/10.1016/j.compositesb.2006.07.013>.
- [21] E. Nigro, G. Cefarelli, A. Bilotta, G. Manfredi, and E. Cosenza, "Fire resistance of concrete slabs reinforced with FRP bars. Part I: Experimental investigations on the mechanical behavior," *Composites Part B: Engineering*, vol. 42, no. 6, pp. 1739–1750, Sep. 2011, <https://doi.org/10.1016/j.compositesb.2011.02.025>.
- [22] *CSA - S806(2012), Design and construction of building structures with fibre-reinforced polymers*. Toronto, ON, Canada: Canadian Standards Association, 2012.
- [23] V. K. R. Kodur and D. Baingo, "Fire Resistance of FRP Reinforced Concrete Slabs," National Research Council of Canada, Ottawa, ON, Canada, Internal Report No. 758, 1998.

- [24] M. Saafi, "Effect of fire on FRP reinforced concrete members," *Composite Structures*, vol. 58, no. 1, pp. 11–20, Oct. 2002, [https://doi.org/10.1016/S0263-8223\(02\)00045-4](https://doi.org/10.1016/S0263-8223(02)00045-4).
- [25] H. Hajiloo, M. F. Green, M. Noel, N. Benichou, and M. Sultan, "Fire tests on full-scale FRP reinforced concrete slabs," *Composite Structures*, vol. 179, pp. 705–719, Nov. 2017, <https://doi.org/10.1016/j.compstruct.2017.07.060>.
- [26] V. K. R. Kondur and L. Bisby, "Evaluation of fire endurance of concrete slabs reinforced with FRP bars," *Journal of Structural Engineering*, vol. 131, no. 1, pp. 34–43, 2005.
- [27] V. K. R. Kodur, L. A. Bisby, and S. H.-C. Foo, "Thermal Behavior of Fire-Exposed Concrete Slabs Reinforced with Fiber-Reinforced Polymer Bars," *Structural Journal*, vol. 102, no. 6, pp. 799–807, Nov. 2005, <https://doi.org/10.14359/14787>.
- [28] *I.Q.S. No. 5/2019 Specification, Portland Cement*. Baghdad, Iraq: Central Organization for Standardization & Quality Control (COSQC), 2019.
- [29] *I.Q.S. No. 45 Aggregate from Natural Sources for Concrete and Construction*. Baghdad, Iraq: Central Organization for Standardization & Quality Control (COSQC), 1984.
- [30] *ASTM E119-22(2022), Standard Test Methods for Fire Tests of Building Construction and Materials*. West Conshohocken, PA, USA: ASTM International, 2022.
- [31] K. H. Tan and H. Zhao, "Strengthening of Openings in One-Way Reinforced-Concrete Slabs Using Carbon Fiber-Reinforced Polymer Systems," *Journal of Composites for Construction*, vol. 8, no. 5, pp. 393–402, Oct. 2004, [https://doi.org/10.1061/\(ASCE\)1090-0268\(2004\)8:5\(393\)](https://doi.org/10.1061/(ASCE)1090-0268(2004)8:5(393)).
- [32] I. F. Kara, A. F. Ashour, and M. A. Koroglu, "Flexural behavior of hybrid FRP/steel reinforced concrete beams," *Composite Structures*, vol. 129, pp. 111–121, Oct. 2015, <https://doi.org/10.1016/j.compstruct.2015.03.073>.

Wind Resource Evaluation in Libya: A Comparative Study of Ten Numerical Methods for the Estimation of Weibull Parameters using Multiple Datasets

Youssef Kassem

Department of Mechanical Engineering, Energy, Environment, and Water Research Center, Science, Technology, Engineering Education Application and Research Center, Near East University, Cyprus
yousseuf.kassem@neu.edu.tr (corresponding author)

Huseyin Camur

Department of Mechanical Engineering, Science, Technology, Engineering Education Application and Research Center, Near East University, Cyprus
huseyin.camur@neu.edu.tr

Almonsef Alhadi Salem Mosbah

High and Intermediate Institute of Agricultural Technology, Libya
almonsef4400@gmail.com

Received: 9 January 2024 | Revised: 4 February 2024 | Accepted: 7 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6882>

ABSTRACT

This study examines Libya's pursuit of sustainable wind energy solutions, using nine sites with mast measurements before the 2011 civil war and six gridded datasets, including CFSR, ERA5, EAR5-Ag, MERRA2, EAR5-Land, and TerraClimate. Employing the Weibull distribution function with ten methods, the empirical method of Justus proved to be optimal for calculating Weibull parameters across datasets. Al Bayda and Darnah exhibit substantial wind power potential (116.80-123.00 W/m²) based on MERRA2 data, making them ideal for large-scale wind turbine deployment. Furthermore, the results showed that wind power density was estimated below 100 W/m² for all selected locations according to CFSR, ERA5, EAR5-Ag, EAR5-Land, and TerraClimate. This study emphasizes the need for new mast measurements to refine dataset selection, which is crucial for accurate assessments and large wind farm planning. Consequently, this study provides key insights into optimizing wind energy utilization in diverse Libyan regions, addressing both the potential and the challenges in sustainable energy development.

Keywords-Libya; wind energy potential; Weibull distribution; numerical estimation method; measured data; gridded datasets

I. INTRODUCTION

Energy is an essential catalyst for promoting economic growth and facilitating industrialization [1]. The global energy supply chain mainly relies on fossil fuels and plays an indispensable role in meeting global energy needs [2]. Therefore, renewable energy can be considered as an alternative solution to the energy crisis due to the limited fossil fuel reserves and their environmental consequences [3]. Renewable sources such as wind energy are rapidly developing and becoming more economically competitive [4]. Wind energy is a promising alternative that has enormous global capacity and is poised to transform the energy landscape, offering abundant and sustainable power for the future [5-6].

Wind energy's first advantage is its global availability. Certain countries with favorable geography have higher wind capabilities [7]. The second advantage is the high energy output achievable with commercial wind turbines, now reaching multi-megawatt capacities [8]. In general, understanding local wind characteristics is crucial for efficient planning and construction of wind power plants [9]. Therefore, the initial stage of assessing wind power potential includes utilizing an anemometer to measure wind speed data at the specific location of the intended wind power plant [10-11]. However, in certain cases, direct wind speed measurement may be challenging due to factors, namely budget constraints, logistical difficulties, or lack of available monitoring infrastructure. In such cases,

gridded datasets emerge as an alternative data source to evaluate wind speed [12]. Gridded datasets are spatial datasets that provide information on various meteorological parameters across a grid of geographic points [13]. These datasets are often generated by integrating observational data, satellite imagery, and numerical models [12-13]. Moreover, choosing suitable gridded datasets for wind speed assessments is a challenge. Accuracy depends on input data quality, spatial resolution, and interpolation methods. Inaccuracies can compromise the success of wind power projects, highlighting the need for careful consideration and validation. Several studies have investigated the challenges and opportunities associated with the use of gridded datasets to assess wind power prospects [14-19]. For example, in [14], it was demonstrated that the ERA5 wind speed data were closely aligned with ground measurements, indicating higher accuracy and reliability than MERRA. In [16], the ERA wind data were validated as valuable for accurately estimating the wind potential. In [17], the importance of higher-resolution wind speed data was emphasized for local and global site prioritization. In [18], the ERA5 data were proved to be reliable for offshore and onshore locations. In [19], a stronger agreement was observed for the ERA5 data with onshore sites, compared to MERRA2.

As a continuation of [20-25], this study aimed to identify a suitable gridded dataset, from CFSR, ERA5, EAR5-Ag, MERRA2, EAR5-Land, and TerraClimate, for evaluating the wind energy potential in Libya and to find the best numerical method (Least Squares Regression Method (LSRM), Energy Pattern Factor Method (EPFM), Empirical Method of Lysen (EML), Empirical Method of Justus (EMJ), Method of Moments (MM), Maximum Likelihood Method (MLM), Modified Maximum Likelihood Method (MMLM), Mean Standard Deviation Method (MSDM), Moment Iteration Method (MIM), and Empirical Method of Mabchour (EMM)) to determine the parameters of the Weibull distribution function, which is used to estimate the wind energy potential at a given location. A review of previous studies [22], NASA datasets, and wind speed measurements collected before 2011 were employed for the evaluation of wind energy in Libya. As far as is known, this is the first study to establish a methodology for the decision-making process in the initial stages of a wind farm investment in Libya engaging data collected from multiple sources.

II. METHODS

The main objective of this study is to estimate the wind energy potential of nine selected sites in Libya using various datasets. Figure 1 illustrates a flowchart of the overall research methodology adopted.

A. Measured Wind Speed Data

The most reliable wind resource data are collected from weather station masts, utilizing an anemometer to measure the wind speed. For this study, mean monthly wind data from nine stations were collected for the period 1990-2010. In recent years, the Civil War in Libya resulted in a scarcity of the instruments required for wind speed measurement. Detailed geographic information for these selected stations can be found in Table I.

TABLE I. DETAILS OF THE SELECTED LOCATIONS

Location	Latitude [°N]	Longitude [°E]	Altitude [m]	Period
Tripoli	32.892	13.173	81	1981-2010
Nalut	31.874	10.979	568	1981-2010
Espiaa	32.537	13.177	73	1993-2009
Al bayda	32.754	21.757	626	2000-2009
Benghazi	32.129	20.082	2	2000-2009
Al-kufrah	24.199	23.293	394	2000-2009
Misratah	32.375	15.090	8	1996-2010
Sabha	27.033	14.432	426	1995-2010
Darnah	32.766	22.624	48	2000-2009

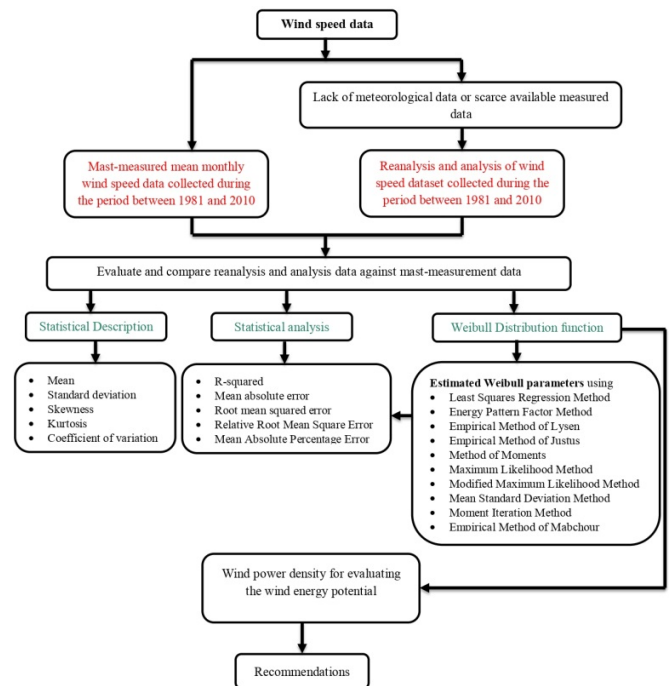


Fig. 1. Research methodology.

B. Reanalysis and Analysis of Wind Speed Dataset

In regions lacking measurement tools, it is vital to assess Satellite Products (SPs) thoroughly before using them for meteorological data [22]. This ensures the reliability and accuracy of service providers in specific areas. Dataset analysis and reanalysis involve creating comprehensive meteorological datasets, applying advanced data assimilation techniques, and combining various data sources, such as weather stations and satellite data [26-27]. These datasets are critical for assessing wind energy potential [12, 22]. This study chose six SPs based on high spatial resolution, coverage range, and availability periods (Table II). The performance of these datasets was evaluated using statistical metrics, like the coefficient of determination (R^2), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Relative Root Mean Square Error (RRMSE), and Mean Absolute Percentage Error (MAPE), as outlined in (1)-(5).

$$R^2 = 1 - \frac{\sum_{i=1}^n (a_{a,i} - a_{p,i})^2}{\sum_{i=1}^n (a_{p,i} - a_{a,ave})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (a_{a,i} - a_{p,i})^2} \quad (2)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |a_{a,i} - a_{p,i}| \quad (3)$$

$$RRMSE = \sqrt{\frac{\frac{1}{n} \sum_{i=1}^n (a_{a,i} - a_{p,i})^2}{\frac{1}{n} \sum_{i=1}^n (a_{p,i})^2}} \quad (4)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{a_{a,i} - a_{p,i}}{a_{a,i}} \right| \times 100 \quad (5)$$

where n is the number of data, $a_{p,i}$ is the predicted value, $a_{a,i}$ is the actual value, $a_{a,ave}$ is the average actual value, and i is the number of input variables.

TABLE II. INFORMATION OF USED SATELLITE PRODUCTS

Product	Description	Resolution	Period
TerraClimate	Global gridded dataset of meteorological and water balance for global terrestrial surfaces	0.042°×0.042°	1958-present
ERA5	Fifth-generation reanalysis product of the European Centre for Medium-Range Weather Forecasts	0.05°/1 d	1979-present
ERA5-Land	ERA5-Land has been produced by replaying the land component of the ECMWF ERA5 climate reanalysis	0.125°×0.125°	1963-present
ERA5-Ag	Agriculture-specific dataset of the ECMWF ERA5	0.1°×0.1°	1979-present
MERRA-2	Second-generation Modern-ERA Retrospective Analysis for Research and Applications	0.5°×0.625°	1981-present
CFSR	NCEP (NOAA NWS National Centers for Environmental Prediction) Climate Forecast System Reanalysis dataset	1/5°	1979-present

TABLE III. METHODS USED FOR CALCULATING THE PARAMETERS OF 2P-W

Method	Weibull distribution parameters
LSRM	$k = \frac{n \sum_{i=1}^n \ln(v) \times \ln[-\ln\{1 - F(v)\}]}{n \sum_{i=1}^n \ln(v^2) - \{\sum_{i=1}^n \ln(v)\}^2}$; $c = \left(\frac{1}{n} \sum_{i=1}^n v_i^k \right)^{1/k}$
EPFM	$k = 1 + \frac{3.69}{EPF^2}$; $EPF = \frac{1}{\bar{v}^3} \left(\sum_{i=1}^n \frac{v_i^3}{N} \right)$; $c = \frac{\bar{v}}{\Gamma(1 + \frac{1}{k})}$; $\bar{v} = \frac{1}{n} \sum_{i=1}^n v_i$
EML	$k = \left(\frac{\sigma}{\bar{v}} \right)^{-1.086}$; $c = \bar{v} \left(0.568 + \frac{0.433}{k} \right)$; $\bar{v} = \frac{1}{n} \sum_{i=1}^n v_i$
EMJ	$k = \left(\frac{\sigma}{\bar{v}} \right)^{-1.086}$; $c = \frac{\bar{v}}{\Gamma(1 + \frac{1}{k})}$; $\bar{v} = \frac{1}{n} \sum_{i=1}^n v_i$
MM	$k = \left(\frac{0.9874 \bar{v}}{\sigma} \right)^{1.0983}$; $c = \frac{\bar{v}}{\Gamma(1 + \frac{1}{k})}$; $\sigma = c \left[\Gamma(1 + \frac{2}{k}) - \Gamma^2(1 + \frac{1}{k}) \right]^{0.5}$; $\bar{v}^2 = \frac{\left\{ \Gamma(1 + \frac{1}{k}) \right\}^2}{\left[\Gamma(1 + \frac{2}{k}) - \Gamma^2(1 + \frac{1}{k}) \right]^{0.5}}$
MLM	$k = \left(\frac{\sum_{i=1}^n v_i^k \ln(v_i) - \sum_{i=1}^n \ln(v_i)}{\sum_{i=1}^n v_i^k} \right)^{-1}$; $c = \left(\frac{1}{n} \sum_{i=1}^n v_i^k \right)^{1/k}$
MMLM	$k = \left(\frac{\sum_{i=1}^n v_i^k \ln(v_i) f(v_i) - \sum_{i=1}^n \ln(v_i) f(v_i)}{\sum_{i=1}^n v_i^k f(v_i)} \right)^{-1}$; $c = \left(\frac{1}{f(v \geq 0)} \sum_{i=1}^n v_i^k \right)^{1/k}$
MSDM	$k = \left(\frac{\sigma}{\bar{v}} \right)^{-1.086}$; $c = \frac{\bar{v}}{\Gamma(1 + \frac{1}{k})}$
MIM	$\frac{\bar{v}^2}{\sigma^2} = \frac{\left\{ \Gamma(1 + \frac{1}{k}) \right\}^2}{\left[\Gamma(1 + \frac{2}{k}) - \Gamma^2(1 + \frac{1}{k}) \right]^{0.5}}$; $\bar{v} = \frac{1}{n} \sum_{i=1}^n v_i$; $\sigma = c \left[\Gamma(1 + \frac{2}{k}) - \Gamma^2(1 + \frac{1}{k}) \right]^{0.5}$
EMM	$k = 1 + [0.483 \times (v_i - 2)]^{0.51}$; $c = \frac{\bar{v}}{\Gamma(1 + \frac{1}{k})}$

C. Wind Speed Distribution

Wind speed analysis is vital for designing and operating wind energy systems. The two-parameter Weibull (2p-W) distribution function is widely utilized to assess wind speed data, wind potential, and energy generation in a specific area [28]. It describes wind speed variation through the probability density function $f(v)$ and cumulative distribution function $F(v)$ [28-29].

$$f(v) = \frac{k}{c} \left(\frac{v}{c} \right)^{k-1} \exp \left[- \left(\frac{v}{c} \right)^k \right] \quad (6)$$

$$F(v) = 1 - \exp \left[- \left(\frac{v}{c} \right)^k \right] \quad (7)$$

where v represents the wind speed, and c and k are the scale and shape parameters of the Weibull distribution function.

D. Estimation of Weibull Parameters

Various methods have been proposed to estimate the c and k Weibull parameters [30-33]. In [30], energy pattern factor, mean standard deviation, and maximum likelihood were used to estimate the Weibull parameters by analyzing wind speed data in Iran. In [31], the empirical approach and wasp algorithm were employed for data in Pakistan, noting limited effectiveness in certain locations. In [32], the maximum likelihood and the modified maximum likelihood methods were found suitable to determine the Weibull parameters in Brazilian stations. In [33], the modified maximum likelihood method was the best for representing wind data in Northern Pakistan. This study utilized 10 numerical methods to estimate the Weibull distribution parameters, displayed in Table III.

E. Wind Potential Estimation

Wind Power Density (WPD) is one of the most important indicators for designing a wind farm. WPD is used to assess the potential of wind resources at a particular location and can be determined by the Weibull two-parameter method [34]:

$$\frac{P}{A} = \frac{1}{2} \rho c^3 \Gamma\left(\frac{k+3}{k}\right) \quad (8)$$

where P represents the wind power, ρ is air density ($\rho = 1.23 \text{ kg/m}^3$), \bar{v} is the mean wind speed in m/s, c is the scale parameter of the Weibull distribution function in m/s, and k is the shape parameter of the Weibull distribution function. The wind energy potential is categorized according to the average WPD as exhibited in Table IV.

TABLE IV. WIND ENERGY POTENTIAL CLASSIFICATION

Number	Power class	P (W/m ²) at 10 m	P (W/m ²) at 30 m
1	Poor	≤100	≤160
2	Marginal	≤150	≤240
3	Moderate	≤200	≤320
4	Good	≤250	≤400
5	Excellent	≤300	≤480
6	Excellent	≤400	≤640
7	Excellent	≤1000	≤1600

III. RESULTS

A. Statistical Description of the Wind Speed Data Based on the Period of Mast Measurement Data

Figure 2 summarizes the statistical details of the average monthly wind speed for all locations. It includes mean, Standard Deviation (SD), Coefficient of Variation (CV), Minimum ($Min.$), Maximum ($Max.$), Kurtosis (K), and Skewness (S).

1) Mast Measurements

The average monthly wind speeds range from 2.12 m/s (Espiaa) to 6.13 m/s (Benghazi). In general, the mean and SD values indicate consistent wind behavior. CV values range from 6.75% to 15.36%. A CV of 6.75% suggests a low variation around the mean, while 15.36% implies a slightly higher variation. The minimum wind speeds were 5.14 m/s in Darnah and 1.75 m/s in Espiaa, while the maximum was 6.91 m/s in Benghazi. Most locations showcase negatively skewed distributions, implying left-skewed data. K values range from -1.83 to 2.82, indicating differences in data flatness. A negative K suggests a slightly flatter distribution, while a positive K indicates heavier tails than normal distribution. Wind speeds vary between locations, but they demonstrate consistency with relatively low variation, left-skewed distributions, and varying degrees of flatness in their data distributions.

2) CFSR

The mean wind speed varied between 2.48 m/s in Al Bayda and 4.63 m/s in Misratah, indicating consistency in wind patterns. CV values ranged from 7.82% to 11.94%, showing relatively low to moderate variation around the mean. The dataset exhibited a minimum monthly wind speed of 2.08 m/s in Darnah and a maximum of 6.91 m/s in Misratah. The S values for Benghazi, Darnah, Espiaa, Misratah, and Nalut were

negative, implying left-skewed distributions. Al Bayda, Al-Kufrah, Sabha, and Tripoli had right-skewed distributions. K values, ranging from -1.96 to 0.21, indicated varying degrees of flatness in the data distributions.

3) EAR5

The mean wind speed varied from 2.161 m/s (Espiaa) to 4.69 m/s (Benghazi). In general, the mean and SD values disclose consistent wind behavior. CV values ranged from 7.79% to 19.68%. The minimum monthly wind speed was recorded in Al Bayda (2.02 m/s) while the maximum was 5.78 m/s in Benghazi. The S values were positive for all locations except Nalut, indicating right-skewed distributions. The K values, ranging from -1.95 to 1.14, revealed differences in the flatness of the data distributions.

4) EAR5-Land

The mean wind speed varied between 2.68 m/s (Al Bayda) and 4.47 m/s (Misratah). In general, the mean and SD values suggest a high level of consistency in wind behavior. CV values, ranging from 9.47% to 19.89%, indicate moderate variability around the mean. A CV of 9.47% suggests relatively low variation, while 19.89% indicates slightly higher variability. Remarkably, minimum monthly wind speeds of 2.10 m/s and 3.68 m/s were observed in Al Bayda and Misratah, respectively, while the maximum monthly wind speed of 5.40 m/s was also recorded in Misratah. S values for most locations are positive, demonstrating right-skewed distributions. Additionally, K values ranged from -1.76 to 0.74.

5) EAR5-Ag

The mean wind speeds ranged from 3.73 m/s in Al-Kufrah to 4.60 m/s in Benghazi. CV values ranged from 9.40% to 12.80%. Furthermore, Al-Kufrah had a minimum monthly wind speed of 3.044 m/s. On the other hand, Benghazi recorded the highest monthly wind speed of 5.20 m/s. The S values for most locations were negative, indicating left-skewed distributions. K values ranged from -2.11 to 0.50.

6) MERRA2

The MERRA2 dataset shows wind speeds ranging from 3.96 m/s (Al-Kufrah) to 5.75 m/s (Darnah), indicating consistent wind behavior. The CV values, ranging from 6.20 to 17.31%, suggest moderate variability around the mean. Al-Kufrah recorded a minimum monthly wind speed of 3.30 m/s, while Darnah had the highest monthly wind speed of 7.10 m/s. Most locations display right-skewed distributions based on positive S values. K values ranging from -1.42 to -0.07 indicate varying degrees of distribution flatness, with negative values suggesting slightly flatter distributions.

7) TerraClimate

The TerraClimate dataset illustrates wind speeds ranging from 3.23 m/s (Espiaa) to 5.15 m/s (Darnah). The mean and SD values indicate consistent wind behavior. CV values, ranging from 6.60% to 12.06%, suggest moderately low variability around the mean. Darnah and Espiaa had the minimum and maximum monthly wind speeds. Most locations exhibited right-skewed distributions with positive S values. K values, from -1.59 to -0.07, represent the data's flatness.

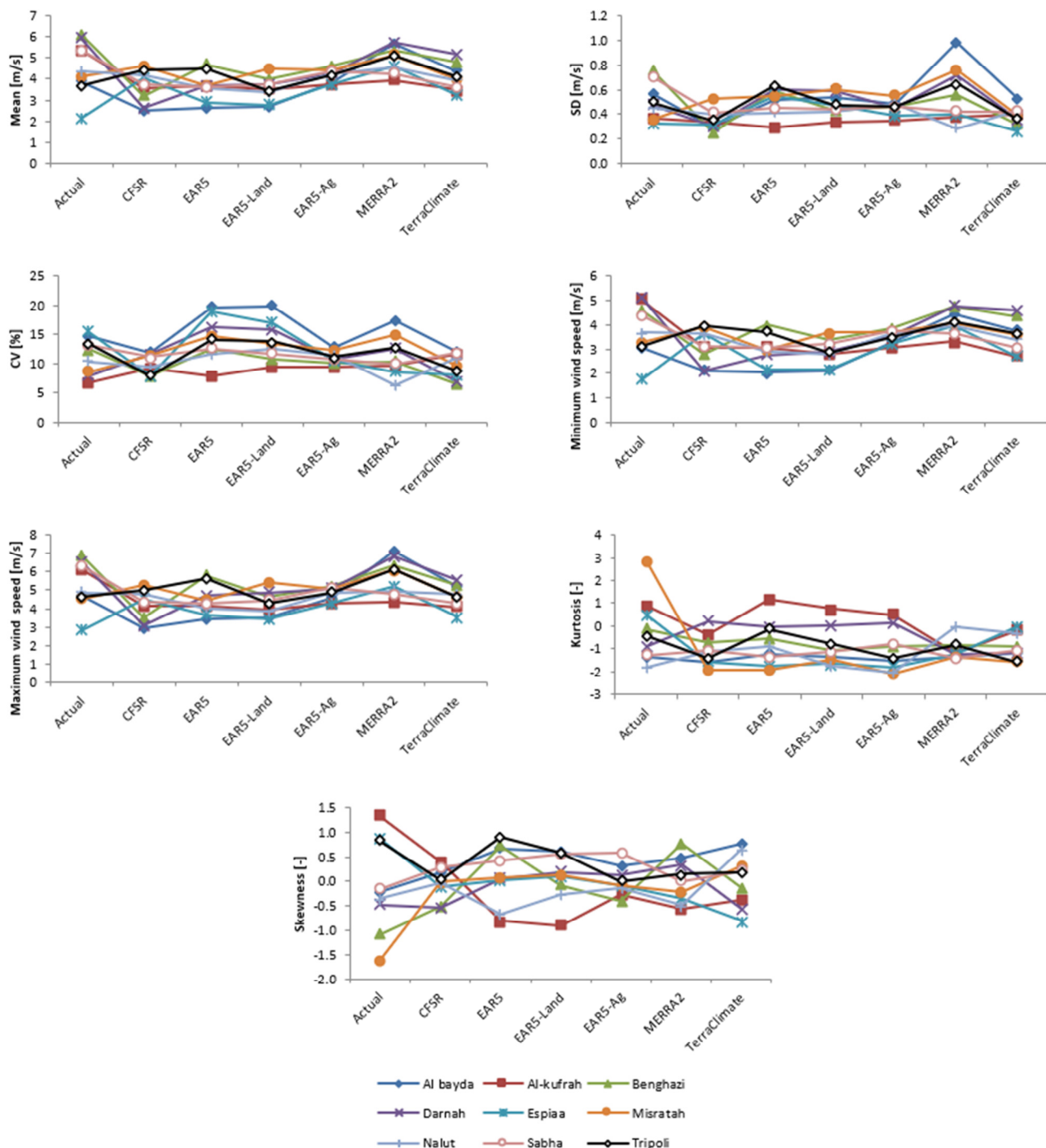


Fig. 2. Statistical description of average monthly wind speed values. Mean, SD, max. speed, kurtosis, and skewness.

B. Best Satellite Products Based on the Period of Mast Measurement Data

Figure 3 compares the measured and estimated data collected from SPs. This study used various statistical metrics to find the best SP for evaluating wind energy potential. The evaluation of the SPs' performance was based on R^2 , which measures the degree of the linear relationship between

observed and modeled values. The highest value of R^2 was found to be 0.9344 for the EARS-Ag dataset in Nalut, while the minimum value of 0.00016 was obtained for the TerraClimate dataset in Misratah. Generally, R^2 is a statistical measure that falls within the 0 to 1 range and signifies the proportion of variance in observed data explained by a model. It is important to understand that higher R^2 values do not automatically indicate the superiority of one dataset over another.

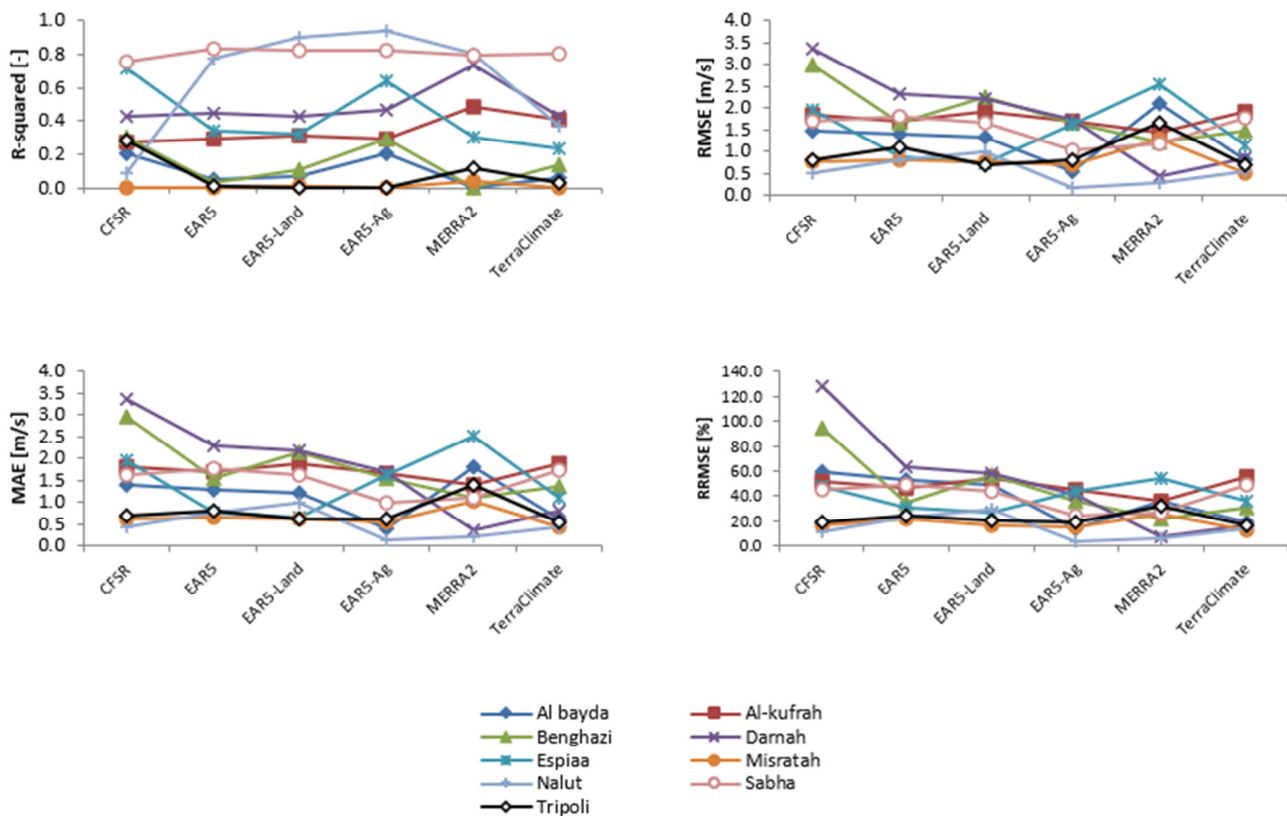


Fig. 3. Statistical description of average monthly wind speed values. R^2 , RMSE, MAE, and RMSE (%).

Table V presents the SP ranking for evaluating the wind potential in the selected regions based on $RMSE$, MAE , and $RRMSE$. In general, $RMSE$ is used to evaluate the accuracy of the estimated data, as it measures the average magnitude of the differences between the observed and estimated values. A zero $RMSE$ signifies a perfect match between the observed and predicted data, signifying that the model predictions are precise. However, as $RMSE$ increases, it denotes a progressively poorer match between the observed and predicted values, implying that the model's accuracy decreases. EAR5-Ag exhibited superior performance compared to other datasets for Al Bayda, Nalut, and Sabha. Furthermore, MERRA2 showed better performance compared to other datasets in Al-Kufrah, Benghazi, Darnah, Misratah, and Tripoli. EAR5-Land exhibited better performance compared to other datasets in Espiaa. In addition, $RRMSE$ and $MAPE$ were used to evaluate the SPs' performance. All SPs ranged above 30%, particularly for Al-Kufrah and Espiaa, implying that their performance can be considered poor. Based on $MAPE$, CFSR has a high $MAPE$ compared to other SPs in Darnah, showing that the performance can be considered inaccurate. In Espiaa, the $MAPE$ for TerraClimate, EAR5-Ag, CFSR, and MERRA2 was notably higher than 50% compared to other SPs. According to the performance rate of these products, the latter can be characterized as inaccurate estimates.

C. Weibull Distribution Function and Wind Power Density

The parameters of the Weibull distribution were calculated using ten methods. Additionally, the WPD was computed to

evaluate the wind potential for each selected location. Table VI lists the ranking of methods employed to estimate the WPD for evaluating the wind potential based on the percentage absolute error. Figure 3 depicts the WPD value for each location depending on the best approach. Based on the actual data, Benghazi and Darnah recorded the highest WPD values (141.52-148.26 W/m^2 and 129.09-131.15 W/m^2 , respectively). It can be concluded that the wind power potential in both Benghazi and Darnah is quite promising, making these locations suitable for harnessing wind energy through large-scale wind turbines. Focusing on the MERRA2 dataset, the highest WPD values were observed in Al Bayda and Darnah, with WPD from 122.22 to 120.58 W/m^2 and from 116.80 to 123.00 W/m^2 , respectively. These findings strongly suggest that both Al Bayda and Darnah possess significant wind power potential, making them excellent candidates for utilizing large-scale wind turbines to harness wind energy effectively. In contrast to the previously mentioned data, other datasets reveal that the WPD values in the region are less than 100 W/m^2 , which can be categorized as poor. These low values signify that wind resources may not be suitable for the deployment of high-capacity wind turbines. However, there is still potential for utilizing wind energy in the regions using small-scale wind turbines.

TABLE V. SP RANKING PER AREA AND METRIC

Location	The rank of SPs based on RMSE and MAE					
	1	2	3	4	5	6
Al Bayda	EAR5-Ag	TerraClimate	EAR5-Land	EAR5	CFSR	MERRA2
Al-Kufrah	MERRA2	EAR5-Ag	EAR5	CFSR	EAR5-Land	TerraClimate
Benghazi	MERRA2	TerraClimate	EAR5-Ag	EAR5	EAR5-Land	CFSR
Darnah	MERRA2	TerraClimate	EAR5-Ag	EAR5-Land	EAR5	CFSR
Espiaa	EAR5-Land	EAR5	TerraClimate	EAR5-Ag	CFSR	MERRA2
Misratah	TerraClimate	EAR5-Ag	CFSR	EAR5-Land	EAR5	MERRA2
Nalut	EAR5-Ag	MERRA2	CFSR	TerraClimate	EAR5	EAR5-Land
Sabha	EAR5-Ag	MERRA2	EAR5-Land	CFSR	TerraClimate	EAR5
Tripoli	TerraClimate	EAR5-Ag	EAR5-Land	CFSR	EAR5	MERRA2
Location	Rank of SPs based on RRMSE					
	1	2	3	4	5	6
Al bayda	EAR5-Ag	TerraClimate	EAR5-Land	EAR5	CFSR	MERRA2
Al-kufrah	MERRA2	EAR5-Ag	EAR5	CFSR	EAR5-Land	TerraClimate
Benghazi	MERRA2	TerraClimate	EAR5-Ag	EAR5	EAR5-Land	CFSR
Darnah	MERRA2	TerraClimate	EAR5-Ag	EAR5-Land	EAR5	CFSR
Espiaa	EAR5-Land	EAR5	TerraClimate	EAR5-Ag	CFSR	MERRA2
Misratah	TerraClimate	EAR5-Ag	CFSR	EAR5-Land	EAR5	MERRA2
Nalut	EAR5-Ag	MERRA2	CFSR	TerraClimate	EAR5	EAR5-Land
Sabha	EAR5-Ag	MERRA2	EAR5-Land	CFSR	TerraClimate	EAR5
Tripoli	TerraClimate	CFSR	EAR5-Ag	EAR5-Land	EAR5	MERRA2
Location	Rank of SPs based on MAPE					
	1	2	3	4	5	6
Al bayda	EAR5-Ag	TerraClimate	EAR5-Land	EAR5	CFSR	MERRA2
Al-kufrah	MERRA2	EAR5-Ag	EAR5	CFSR	EAR5-Land	TerraClimate
Benghazi	MERRA2	TerraClimate	EAR5-Ag	EAR5	EAR5-Land	CFSR
Darnah	MERRA2	TerraClimate	EAR5-Ag	EAR5-Land	EAR5	CFSR
Espiaa	EAR5-Land	EAR5	TerraClimate	EAR5-Ag	CFSR	MERRA2
Misratah	TerraClimate	EAR5-Ag	CFSR	EAR5-Land	EAR5	MERRA2
Nalut	EAR5-Ag	MERRA2	CFSR	TerraClimate	EAR5	EAR5-Land
Sabha	EAR5-Ag	MERRA2	CFSR	TerraClimate	EAR5	EAR5-Land
Tripoli	TerraClimate	EAR5-Land	EAR5-Ag	CFSR	EAR5	MERRA2

TABLE VI. THE RANK OF THE METHODS USED FOR ESTIMATING WPD

Location	Dataset	Rank									
		1	2	3	4	5	6	7	8	9	10
Al Bayda	Measured and SPs	EMJ	EPFM	LSRM	MMLM	MM	MLM	EML	EMM	MSDM	MIM
Al-kufrah	Measured and SPs	EMJ	EPFM	LSRM	MM	MMLM	EMM	MSDM	MLM	EML	MIM
Benghazi	Measured and SPs	EMJ	EPFM	MMLM	LSRM	MM	MLM	EML	EMM	MSDM	MIM
Darnah	Measured and SPs	EMJ	LSRM	EPFM	MMLM	MM	MLM	EMM	EML	MSDM	MIM
Espiaa	Measured and SPs	EMJ	EPFM	LSRM	MMLM	MM	EMM	MSDM	MLM	MIM	EML
Misratah	Measured	EMJ	EML	EPFM	LSRM	MMLM	MM	MLM	EMM	MSDM	MIM
	SPs	EMJ	EPFM	LSRM	MMLM	MM	MLM	EMM	EML	MSDM	MIM
Nalut	Measured and SPs	EMJ	EPFM	LSRM	MMLM	MM	MLM	EMM	EML	MSDM	MIM
Sabha	Measured and SPs	EMJ	EPFM	LSRM	MMLM	MM	MLM	EML	EMM	MSDM	MIM
Tripoli	Measured and SPs	EMJ	EPFM	LSRM	MMLM	MM	EML	EMM	MLM	MSDM	MIM

IV. DISCUSSION AND CONCLUSIONS

Understanding the volatile potential of wind energy is vital for advising Libyan policymakers and investors. This understanding helps in the selection of ideal sites and methods to measure wind speed, which influences decision-making in wind energy initiatives. The evaluation of wind speed variation is crucial during the feasibility assessment of wind power systems. The use of satellite data and reanalysis are necessary due to limited in situ measurements. Based on the findings, the annual wind speed at the selected sites is above 2 m/s at an altitude of 10 m, and the average monthly wind speed lies within the range of 2.12-6.13 m/s. Additionally, the mean monthly wind speed data for Benghazi are 6.13 m/s, 4.69 m/s, and 4.60 m/s based on mast measurements, EAR5 and EAR5-

Ag, respectively, fort Darnah are 5.75 and 5.15 m/s focusing on MERRA2 and TerraClimate, accordingly, and for Misratah are 4.47 and 4.63 m/s contingent on EAR5-Land and CFSR, correspondingly, which were higher compared to other locations. The results demonstrate that MERRA-2, EAR5-Land, EAR5-Ag, and TerraClimate performed the best, while CFSR and EAR5 were the weakest in terms of average monthly wind speed data. Therefore, it can be concluded that the selection of an SP and reanalysis of the data to evaluate wind energy potential depend on the specific location [12, 22]. In [22], it was displayed that CFSR and ERA5-Land were the most suitable for evaluating wind resources. According to [12], spatial resolution plays a crucial role in both satellite data and reanalysis, significantly influencing the level of detail. Higher-

resolution data offer a more comprehensive understanding of Earth's features, capturing subtle variations. Additionally, accuracy can be influenced by processing algorithms, introducing potential inconsistencies in the data.

This study also investigated the wind energy potential in different locations in Libya engaging different data sources and the Weibull distribution function. Ten different numerical methods were utilized to estimate the Weibull distribution parameters. The accuracy of these methods was assessed using different statistical analysis techniques. Subsequently, WPD was calculated based on the estimated parameters, considering the results obtained from these methods. The outcomes showed that the EMJ and the EPFM were the most effective approaches to analyze the mean monthly data series at selected locations. The former also exhibited that Benghazi and Darnah had the highest values of WPD. Previous studies support this finding,

as in [20], where the maximum annual wind speed was recorded in Benghazi followed by Darnah, in [35] the highest mean wind speeds were obtained in Darnah, Misurata, and Tobruk, and in [36], the highest annual wind speed was recorded in Benghazi followed by Darnah.

This study introduces a wind energy roadmap to attract investors in clean energy for sustainable development in Libya, address energy problems, and meet domestic demands. Wind energy aligns with the goals of sustainable development, reducing emissions. Despite initial costs, benefits include energy security and reduced dependence on fossil fuels, which help sustainable development. Furthermore, this study identified the three most important grid datasets for assessing the country's wind potential, improving the accuracy of the assessments for investors and policymakers.

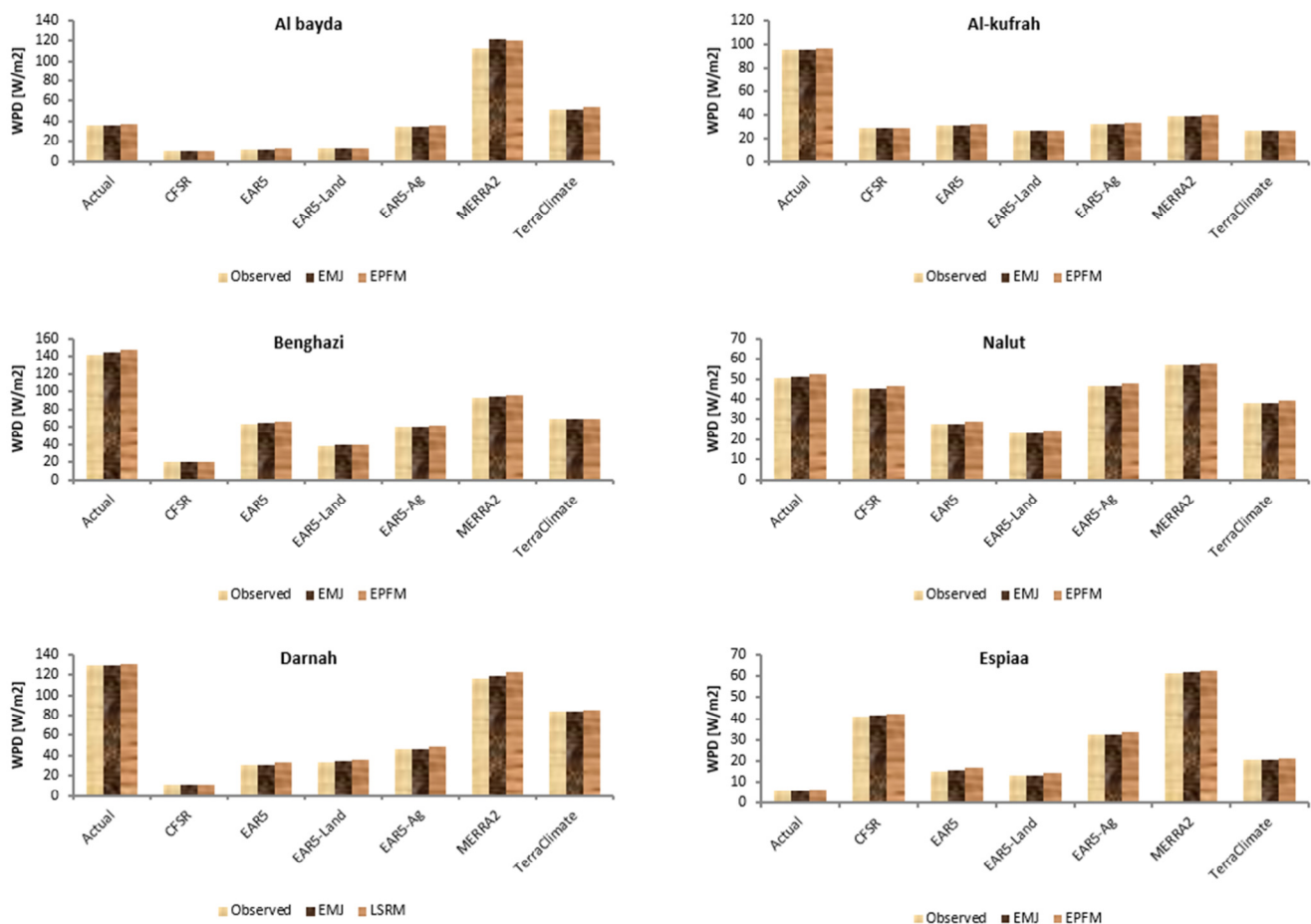


Fig. 4. Annual WPD based on multiple data sources.

V. FUTURE WORK

Future studies could evaluate the accuracy of satellite data and selected reanalysis by comparing daily or hourly measured data. This comparative analysis can help to gain insight into the limitations and uncertainties associated with the data, leading to a better understanding of their reliability and applicability to assess the wind potential in Libya. This study showed that

Benghazi, Darnah, and Misuratah are suitable for installing wind farms in the future. Therefore, future studies can perform techno-economic analysis of wind turbines with various specifications using mathematical modeling and/or simulation tools. In addition, future research should focus on examining the economic data on the performance of wind farms in different locations in Libya.

REFERENCES

- [1] W. Fang, Z. Liu, and A. R. Surya Putra, "Role of research and development in green economic growth through renewable energy development: Empirical evidence from South Asia," *Renewable Energy*, vol. 194, pp. 1142–1152, Jul. 2022, <https://doi.org/10.1016/j.renene.2022.04.125>.
- [2] O. Aslanturk and G. Kiprizh, "The Role of Renewable Energy in Ensuring Energy Security of Supply and Reducing Energy-Related Import," *International Journal of Energy Economics and Policy*, vol. 10, no. 2, pp. 354–359, 2020.
- [3] F. Chien, H. W. Kamran, G. Albashar, and W. Iqbal, "Dynamic planning, conversion, and management strategy of different renewable energy sources: A Sustainable Solution for Severe Energy Crises in Emerging Economies," *International Journal of Hydrogen Energy*, vol. 46, no. 11, pp. 7745–7758, Feb. 2021, <https://doi.org/10.1016/j.ijhydene.2020.12.004>.
- [4] S. Potrč, L. Čuček, M. Martin, and Z. Kravanja, "Sustainable renewable energy supply networks optimization – The gradual transition to a renewable energy system within the European Union by 2050," *Renewable and Sustainable Energy Reviews*, vol. 146, Aug. 2021, Art. no. 111186, <https://doi.org/10.1016/j.rser.2021.111186>.
- [5] Y. Kassem, H. Gokcekus, H. Camur, and A. H. A. Abdelnaby, "Wind Power Generation Scenarios in Lebanon," *Engineering, Technology & Applied Science Research*, vol. 12, no. 6, pp. 9551–9559, Dec. 2022, <https://doi.org/10.48084/etasr.5258>.
- [6] F. Elmahmoudi, O. E. K. Abra, A. Raihani, O. Serrar, and L. Bahatti, "Elaboration of a Wind Energy Potential Map in Morocco using GIS and Analytic Hierarchy Process," *Engineering, Technology & Applied Science Research*, vol. 10, no. 4, pp. 6068–6075, Aug. 2020, <https://doi.org/10.48084/etasr.3692>.
- [7] G. Bandoc, R. Prăvălie, C. Patriche, and M. Degeratu, "Spatial assessment of wind power potential at global scale. A geographical approach," *Journal of Cleaner Production*, vol. 200, pp. 1065–1086, Nov. 2018, <https://doi.org/10.1016/j.jclepro.2018.07.288>.
- [8] Y. Kassem, H. Gokcekus, and A. M. S. Essayah, "Wind Power Potential Assessment at Different Locations in Lebanon: Best-Fit Probability Distribution Model and Techno-Economic Feasibility," *Engineering, Technology & Applied Science Research*, vol. 13, no. 2, pp. 10578–10587, Apr. 2023, <https://doi.org/10.48084/etasr.5686>.
- [9] L. Bilir, M. İmir, Y. Devrim, and A. Albostan, "An investigation on wind energy potential and small scale wind turbine performance at İncek region – Ankara, Turkey," *Energy Conversion and Management*, vol. 103, pp. 910–923, Oct. 2015, <https://doi.org/10.1016/j.enconman.2015.07.017>.
- [10] M. Brower, *Wind Resource Assessment: A Practical Guide to Developing a Wind Project*. Hoboken, NJ, USA: John Wiley & Sons, 2012.
- [11] E. Chavero-Navarrete, M. Trejo-Perea, J. C. Jáuregui-Correa, R. V. Carrillo-Serrano, G. Ronquillo-Lomeli, and J. G. Ríos-Moreno, "Hierarchical Pitch Control for Small Wind Turbines Based on Fuzzy Logic and Anticipated Wind Speed Measurement," *Applied Sciences*, vol. 10, no. 13, 2020, <https://doi.org/10.3390/app10134592>.
- [12] Y. Kassem, H. Gökçekuş, and M. H. A. Abdalla, "Wind energy resource assessment based on the use of multiple satellite data for sustainable energy production in Sudan," *Environment, Development and Sustainability*, Nov. 2023, <https://doi.org/10.1007/s10668-023-04221-w>.
- [13] N. Salvação, A. Bentamy, and C. Guedes Soares, "Developing a new wind dataset by blending satellite data and WRF model wind predictions," *Renewable Energy*, vol. 198, pp. 283–295, Oct. 2022, <https://doi.org/10.1016/j.renene.2022.07.049>.
- [14] A. Piasecki, J. Jurasz, and A. Kies, "Measurements and reanalysis data on wind speed and solar irradiation from energy generation perspectives at several locations in Poland," *SN Applied Sciences*, vol. 1, no. 8, Jul. 2019, Art. no. 865, <https://doi.org/10.1007/s42452-019-0897-2>.
- [15] A. Khatibi and S. Krauter, "Validation and Performance of Satellite Meteorological Dataset MERRA-2 for Solar and Wind Applications," *Energies*, vol. 14, no. 4, Jan. 2021, Art. no. 882, <https://doi.org/10.3390/en14040882>.
- [16] C. Satyanarayana Gubbala, V. B. R. Dodla, and S. Desamsetti, "Assessment of wind energy potential over India using high-resolution global reanalysis data," *Journal of Earth System Science*, vol. 130, no. 2, Apr. 2021, Art. no. 64, <https://doi.org/10.1007/s12040-021-01557-7>.
- [17] C. Jung and D. Schindler, "On the influence of wind speed model resolution on the global technical wind energy potential," *Renewable and Sustainable Energy Reviews*, vol. 156, Mar. 2022, Art. no. 112001, <https://doi.org/10.1016/j.rser.2021.112001>.
- [18] G. Gualtieri, "Analysing the uncertainties of reanalysis data used for wind resource assessment: A critical review," *Renewable and Sustainable Energy Reviews*, vol. 167, Oct. 2022, Art. no. 112741, <https://doi.org/10.1016/j.rser.2022.112741>.
- [19] V. Yildirim, E. Rusu, and F. Onea, "Wind Energy Assessments in the Northern Romanian Coastal Environment Based on 20 Years of Data Coming from Different Sources," *Sustainability*, vol. 14, no. 7, 2022, <https://doi.org/10.3390/su14074249>.
- [20] Y. Kassem, H. Çamur, and R. A. F. Aateg, "Exploring Solar and Wind Energy as a Power Generation Source for Solving the Electricity Crisis in Libya," *Energies*, vol. 13, no. 14, 2020, <https://doi.org/10.3390/en13143708>.
- [21] Y. Kassem, H. Camur, and A. A. S. Mosbah, "Feasibility Analysis of the Wind Energy Potential in Libya using the RETScreen Expert," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11277–11289, Aug. 2023, <https://doi.org/10.48084/etasr.6007>.
- [22] H. S. A. Lagili, A. Kiraz, Y. Kassem, and H. Gökçekuş, "Wind and Solar Energy for Sustainable Energy Production for Family Farms in Coastal Agricultural Regions of Libya Using Measured and Multiple Satellite Datasets," *Energies*, vol. 16, no. 18, 2023, <https://doi.org/10.3390/en16186725>.
- [23] Y. Kassem, H. Çamur, and M. A. AbuGharara, "Assessment of Wind Energy Potential for Selecting Small-Scale Wind Turbines in Low Wind Locations in Libya: A Comparative Study," *International Journal of Engineering Research and Technology*, vol. 12, no. 6, pp. 820–836, 2019.
- [24] Y. Kassem, H. Gökçekuş, and R. A. Faraj, "Evaluation of the Wind Energy Potential in Libya's Eastern Mediterranean Coast Area Using Weibull Distribution Function," *International Journal of Applied Engineering Research*, vol. 14, no. 10, pp. 2483–2491, 2019.
- [25] Y. Kassem, H. Gökçekuş, and M. A. AbuGharara, "An Investigation on Wind Energy Potential in Nalut, Libya, using Weibull Distribution," *International Journal of Applied Engineering Research*, vol. 14, no. 10, pp. 2474–2482, 2019.
- [26] A. Dogra, J. Thakur, and A. Tandon, "Do satellite-based products suffice for rainfall observations over data-sparse complex terrains? Evidence from the North-Western Himalayas," *Remote Sensing of Environment*, vol. 299, Dec. 2023, Art. no. 113855, <https://doi.org/10.1016/j.rse.2023.113855>.
- [27] G. R. C. Essou, F. Brissette, and P. Lucas-Picher, "Impacts of combining reanalyses and weather station data on the accuracy of discharge modelling," *Journal of Hydrology*, vol. 545, pp. 120–131, Feb. 2017, <https://doi.org/10.1016/j.jhydrol.2016.12.021>.
- [28] Y. Kassem, H. Çamur, S. Abughinda, and A. Sefik, "Wind Energy Potential Assessment in Selected Regions in Northern Cyprus Based on Weibull Distribution Function," *Journal of Engineering and Applied Sciences*, vol. 15, pp. 128–140, Oct. 2019, <https://doi.org/10.36478/jeasci.2020.128.140>.
- [29] Y. Kassem, H. Gökçekuş, and W. Janbein, "Predictive model and assessment of the potential for wind and solar power in Rayak region, Lebanon," *Modeling Earth Systems and Environment*, vol. 7, no. 3, pp. 1475–1502, Sep. 2021, <https://doi.org/10.1007/s40808-020-00866-y>.
- [30] H. Teimourian, M. Abubakar, M. Yildiz, and A. Teimourian, "A Comparative Study on Wind Energy Assessment Distribution Models: A Case Study on Weibull Distribution," *Energies*, vol. 15, no. 15, 2022, <https://doi.org/10.3390/en15155684>.
- [31] I. Hussain, A. Haider, Z. Ullah, M. Russo, G. M. Casolino, and B. Azeem, "Comparative Analysis of Eight Numerical Methods Using Weibull Distribution to Estimate Wind Power Density for Coastal Areas in Pakistan," *Energies*, vol. 16, no. 3, 2023, <https://doi.org/10.3390/en16031515>.

- [32] J. A. Guarienti, A. Kaufmann Almeida, A. Menegati Neto, A. R. de Oliveira Ferreira, J. P. Ottonelli, and I. Kaufmann de Almeida, "Performance analysis of numerical methods for determining Weibull distribution parameters applied to wind speed in Mato Grosso do Sul, Brazil," *Sustainable Energy Technologies and Assessments*, vol. 42, Dec. 2020, Art. no. 100854, <https://doi.org/10.1016/j.seta.2020.100854>.
- [33] M. Khalid Saeed, A. Salam, A. U. Rehman, and M. Abid Saeed, "Comparison of six different methods of Weibull distribution for wind power assessment: A case study for a site in the Northern region of Pakistan," *Sustainable Energy Technologies and Assessments*, vol. 36, Dec. 2019, Art. no. 100541, <https://doi.org/10.1016/j.seta.2019.100541>.
- [34] B. Bilal, K. H. Adjallah, K. Yetilmezsoy, M. Bahramian, and E. Kıyan, "Determination of wind potential characteristics and techno-economic feasibility analysis of wind turbines for Northwest Africa," *Energy*, vol. 218, Mar. 2021, Art. no. 119558, <https://doi.org/10.1016/j.energy.2020.119558>.
- [35] A. M. Jary, M. Elmnifi, Z. Said, L. J. Habeeb, and H. Moria, "Potential wind energy in the cities of the Libyan coast, a feasibility study," *Journal of Mechanical Engineering Research and Developments*, vol. 44, no. 7, pp. 236–252, 2021.
- [36] F. Ahwide and S. Bousheha, "Estimation of Wind Characteristics and Energy Yield at Different Towns in Libya," *International Journal of Energy and Power Engineering*, vol. 11, no. 5, pp. 611–617, Jun. 2017.

Fostering Sustainability through the Integration of Renewable Energy in an Agricultural Hydroponic Greenhouse

Aymen Lachheb

Laboratory Smart Electricity & ICT, SEICT, LR18ES44, National Engineering School of Carthage, University of Carthage, Tunisia
aymen.lachheb@enicarthage.rnu.tn (corresponding author)

Rym Marouani

Laboratory ATSSEE, Department of Physics, Faculty of Sciences de Tunis, University Tunis El Manar, Campus Universitaire Farhat Hached, Tunisia
rym.marouani@fst.utm.tn

Chabakata Mahamat

Universite de Guyane, UMR Espace-Dev, Cayenne, French Guiana, France
chabakata.mahamat@univ-guyane.fr

Safa Skouri

Centre de Recherches et des Technologies de l'Energie, Technopole de Borj-Cedria, Tunisia
skouri_safa@yahoo.fr

Salwa Bouadila

Centre de Recherches et des Technologies de l'Energie, Technopole de Borj-Cedria, Tunisia
salwa.bouadila@crten.rnrt.tn

Received: 21 January 2024 | Revised: 11 February 2024 | Accepted: 13 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6939>

ABSTRACT

This research explores the feasibility of integrating renewable energy sources, such as solar and wind, to power a hydroponic greenhouse. In this way, the latter's energy autonomy is ensured. The study begins by evaluating the annual electricity consumption of the examined system. A renewable energy system capable of meeting its energy requirements throughout the year is also designed. The main objective is to assess the efficiency of two types of renewable energy sources, namely photovoltaic panels and wind turbines, and to improve their integration within the agricultural chamber by implementing a model simulation. Two scenarios were examined: the first one represents a photovoltaic power plant with storage, connected to the grid, while the second scenario presents a wind power plant connected to the grid. This numerical analysis is supplemented by a one-year experimental study of a photovoltaic installation connected to the network with storage, which in turn is connected to the experimental device. To handle energy within the renewable energy greenhouse, an energy management system was developed based on a fuzzy logic controller. This system aims to maintain energy balance and ensure continuous power supply. The energy management system optimizes energy flow to minimize consumption, reduce grid dependence, and improve overall system efficiency, resulting in cost savings and certain environmental benefits.

Keywords-hydroponic greenhouse; PV installation; wind energy; energy management system; fuzzy logic control

I. INTRODUCTION

A notable increase in fossil fuel consumption has been detected across various sectors over the past few years, with

Africa emerging as a growing influence in global energy trends, contributing to this boost in demand [1]. Recognizing the environmental challenges posed by the extended use of

fossil fuels, the region is highly expected to shift to renewable energy sources as a strategic solution to address this problem. This shift reflects a broader global effort to promote sustainable and environmentally friendly energy practices, highlighting the need for a transition to cleaner alternatives in the face of growing energy demand [2]. According to the International Renewable Energy Agency (IREA), there is a compelling projection that renewable energy will account for more than 60% of the world's total energy consumption by 2050, signifying a substantial move towards sustainable energy practices. This transition is already evident in the increasing contribution of solar and wind sources to electricity production, with a projected 17% growth compared to last year's 16%. In alignment with this global trend, the Tunisian Solar Plan has set ambitious targets for the year 2030. The plan aims to significantly enhance the country's renewable energy capacity by adding 3815 MW of installed power through various technologies. The distribution of this capacity includes 1510 MW for photovoltaic (PV) panels, 1755 MW for wind power generation, 450 MW for Concentrated Solar Power (CSP), and an additional 100 MW from power plants utilizing biomass resources. These targets underscore a comprehensive strategy to diversify the energy mix and embrace a more sustainable approach to meet growing energy demands while reducing dependence on fossil fuels. The concerted efforts outlined in both global and regional initiatives indicate a commitment to a cleaner, more resilient energy future.

Agrivoltaic systems, which combine PV energy production with agricultural activities on the same land, offer promising synergies for both sectors, potentially mitigating competition between energy and food production [3-6]. Authors in [7] combined PV panels, a hydrogen producer, and a ground source geothermal heat pump to heat a small greenhouse. The system offered an overall efficiency equal to 11%. Authors in [8] presented a simulation model analyzing solar radiation distribution within Mediterranean agrivoltaic greenhouses, a valuable tool for optimizing such systems. A hybrid system was proposed in [9] to analyze and predict zero energy potential in real-time air adaptation of direct PV systems for air conditioning of three regions in China. In [10], a system that incorporates both wind and heating is proposed, directly converting wind energy into thermal energy through water agitation heating in the greenhouse. In recent years, researchers have shown a strong interest in wind-resistant greenhouse designs [11-15]. A great amount of attention is also drawn in exploring methods for conditioning soilless technologies. The use of renewable energy in greenhouses demonstrates a great potential to reduce energy demand, operating costs, and CO₂ emissions [3, 5]. In many cases, significant amounts of energy are required to ensure the ideal growing conditions under greenhouses, posing major challenges for the greenhouse sector to eliminate energy demands while simultaneously improving the quality and quantity of the harvest [6]. Consequently, there is a crucial need to accelerate the further development and implementation of clean energy technologies to meet the challenges of the energy crisis and climate change [16].

This paper presents an improved approach to optimizing energy distribution in hydroponic greenhouses, aiming to

enhance efficiency and sustainability. It introduces a control strategy encompassing two scenarios, designed to streamline energy distribution from various sources to successfully meet the greenhouse's electrical demands. The study begins with an overview of the components constituting the electrical energy system of a hydroponic greenhouse, followed by a detailed analysis to determine its annual energy demand. Subsequently, simulation assessments are performed to evaluate the integration of PV and wind systems into the intensified cropping system, guaranteeing that the latter aligns with the greenhouse's electrical requirements. An experimental investigation is also carried out to analyze the annual electricity consumption within the greenhouse.

II. ELECTRIC POWER REQUIRED IN HYDROPONIC GREENHOUSE

The hydroponic greenhouse is located on an experimental platform in Borj Cedria, Tunisia. Borj Cedria is situated on the northeastern coast of Tunisia in North Africa, approximately 25 km southeast of Tunis, at coordinates 36°41'34" N latitude and 10°24'39" longitude. Recognized for its coastal setting, Borj Cedria is a town within the greater Tunis metropolitan area. The experimental platform is equipped with a meteorological station enabling the measurement of various climatic parameters (maximum, minimum, and average) at 10-minute intervals. These parameters include wind speed, ambient temperature, and global solar irradiation. The northern region of Tunisia enjoys a Mediterranean climate, characterized by mild winters and hot and dry summers. The studied hydroponic greenhouse entails all the elements that guarantee an optimal environment for cultivation. Control of air temperature, airflow, water temperature, water circulation, and lighting are all incorporated into the design and can be fixed or vary depending on the needs of the vegetation. Table I illustrates the actuators installed in the hydroponic greenhouse. The experimental hydroponic greenhouse involves a comprehensive array of electrical equipment, encompassing ventilation and climate control devices, thermal conditioning systems (heating and cooling), lighting, and irrigation systems (Figure 1).



Fig. 1. Experimental hydroponic greenhouse.

TABLE I. ACTUATORS USED IN THE HYDROPONIC GREENHOUSE

Equipment	Water pump	Irrigation controller	Aerator	Speed variator	Fan type 1	Fan type 2	Centrifugal pump	Lamp	Acquisition chain
Units	1	1	2	1	2	1	1	8	1
Power (W)	80	10	60	900	250	50	370	16	20

The electrical energy consumption in a hydroponic greenhouse fluctuates throughout the year due to seasonal variations, operational needs, and vegetation requirements. For various operations (irrigation, ventilation, lighting, and thermal conditioning), the consumption varies between 0.75 and 2 kWh, with consumption being highest during peak load times. The consumption of the irrigation system shows a fluctuating pattern throughout the year, ranging between 25 and 165 Wh. As with lighting, the electricity consumption of the irrigation system is small compared to the total consumption of the greenhouse and does not exceed 0.16 kWh. The air conditioning systems remain the main electricity consumers.

III. ARCHITECTURE AND DESCRIPTION OF THE HYBRID POWER SYSTEM

Utilizing renewable energy sources to supply power hydroponic greenhouses represents a pivotal stride in the direction of sustainable agriculture. The effective integration and connection of these system components is imperative in order to guarantee a dependable and sustainable operation.

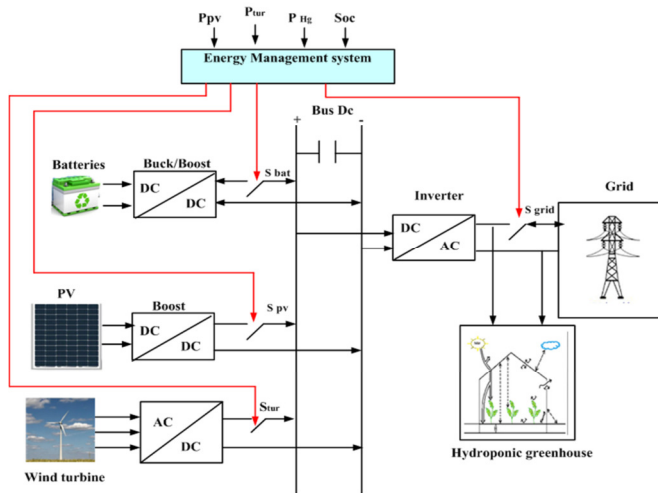


Fig. 2. Synoptic hydroponic greenhouse.

This research seeks to create a hybrid energy system connected to the grid and featuring an integrated battery storage system. The proposed design facilitates smooth grid connection/disconnection, effectively managing energy surplus or deficits. Incorporating this PV-battery system into a grid-connected hydroponic greenhouse can substantially decrease the reliance on grid energy, leading to environmental benefits, such as reduced CO₂ emissions and reduced bimonthly energy expenses. However, initial investment costs and dependence on weather conditions are potential challenges that require further consideration. Figure 2 outlines the proposed hybrid system architecture, with arrows indicating energy flow directions. This architecture serves as the foundation for this study. The

proposed system comprises several components, including a renewable energy source (PV array or wind turbine generator) with a unidirectional DC/DC converter for Maximum Power Point Tracking (MPPT), a battery bank with a bidirectional DC/DC converter for managing the energy transfer between the battery and the DC bus, and a DC/AC inverter that facilitates energy exchange between the DC bus and the grid (see Figure 9). The setup ensures that the hybrid system can effectively deliver the necessary energy to fulfill the demands of the hydroponic greenhouse. The research delineates two scenarios:

- **Scenario 1:** PV/Battery/grid-connected power plant.
- **Scenario 2:** Wind turbine/ grid connected power plant.

A. PV-Battery -Grid-connected Power Plant Modeling

1) PV Power Plant Modeling

PV energy is widely recognized as the leading renewable power source. It holds a prominent position among renewable energy resources due to its eco-friendliness, abundance, and recyclability. In essence, a PV cell can be likened to an electrical current generator shunted with a diode. This diode is created through the formation of a p-n junction. To delve into the physical phenomena occurring within the cell, the model embodies two intrinsic resistors, R_s and R_p , in series and parallel arrangements [17-18]. Equation (1) describes the current (I_{pv}) generated by a PV module, which consists of several PV cells either connected in series or in parallel [1]. The current generated depends on the number of cells connected in series N_s and parallel N_p .

$$I_{pv} = N_p I_{pv} - N_p I_s \left[\exp\left(\frac{V_{pv} + R_s I_{pv}}{nKT N_s} - 1\right) \right] - N_p Q \left[\frac{V_{pv} + R_s I_{pv}}{N_s R_p} \right] \quad (1)$$

where I_{pv} is the current generated by the PV panel, I_s is the saturation current of the diode, V_{pv} is the voltage generated by the panel, R_s is the cell series resistance, R_p is the cell parallel resistance, Q is the electron charge, n is the ideality diode factor (between 1 and 2), K is the Boltzmann constant, and T is the junction temperature. In this work, the FSM-300 W PV module was used. To ensure that the simulations provide accurate results and effectively visualize the performance of the PV power plant in relation to the available solar radiation, calculations were performed to determine the power generated. Power equation (2) can be expressed as a function of solar irradiation and ambient temperature [18]:

$$P_{pv}(t) = P_{pv(rated)} \left[1 + K(T_{amb} + (0.0256I(t)) - T_{ref}) \right] \frac{I(t)}{I_{ref}} \quad (2)$$

where $P_{pv}(t)$ is the power generated by the PV panel at time t , $P_{pv(rated)}$ is the rated power of the PV panel, $I(t)$ is the solar irradiation at time t measured in W/m^2 , I_{ref} is the reference solar irradiation ($1000 W/m^2$), T_{amb} is the ambient temperature and T_{ref} is the reference temperature (at $25^\circ C$).

This scenario involves the study of a PV grid-connected power plant, where the proposed PV system is utilized to meet the energy demands of the hydroponic greenhouse. The hydroponic greenhouse energy requirements are primarily fulfilled by the PV system. If the power requirement is higher, additional energy is imported from the grid. The DC-DC converter is used to efficiently convert the DC voltage output from the PV modules to the voltage level required for the grid-connected inverter. The DC-DC converter boosts or bucks the voltage level depending on the specific needs of the grid-connected inverter and reassures that the MPPT algorithm is optimized for the operating conditions of the PV array. The DC-DC converter also improves the overall competence of the PV system by reducing power losses that can occur due to voltage mismatches between the PV panels and the grid-connected inverter.

MPPT is a technique to ensure that the PV system is operating at maximum efficiency and produces the maximum possible amount of energy under all operating conditions [19]. An MPPT algorithm is utilized to adjust the duty cycle (D) of the DC/DC converter. The duty cycle is the ratio of the on-time of a switch to the period of the waveform. By adjusting the duty cycle, the DC/DC converter voltage can be varied to track the Maximum Power Point (MPP) under all operating conditions. This signal is generated by a controller that monitors the output power of the PV array and compares it to the maximum power point. The controller then adjusts the duty cycle of the DC/DC converter to match the maximum power point.

The Perturb and Observe (P&O) algorithm is one of the most widely adopted techniques for achieving MPPT in PV systems. The P&O algorithm works by continually perturbing the operating point of the PV system and observing the resulting change in output power. In particular, the algorithm perturbs the duty cycle of the DC/DC converter by a small amount, typically 0.1%, and observes the resulting change in output power. Based on this observation, the algorithm adjusts the operating point of the PV system in the direction of power increase until it reaches the MPP of the system. The flowchart of the MPPT algorithm is provided in Figure 3.

2) Modeling of the Storage System

Energy Storage Systems (ESSs) play an important role in reducing the differences between energy production and consumption, which can be particularly difficult to manage when energy supplies rely on intermittent renewable sources, like wind and solar energy [20-22]. Batteries are among the most advanced electronic devices used to improve the robustness of smart electrical microgrids and store electrical energy. They are particularly valuable in addressing the inherent variability of renewable energy sources. In the field of

battery modeling, numerous electrochemical models have been presented in the literature. These include models, such as the internal resistance battery model, the RC network battery model, and the Randles circuit. Among them, the Shepherd model is considered one of the most accurate models with corresponding electrical characteristics.

$$V_{bat} = E - R_{in} i_{bat} \quad (3)$$

$$E = E_0 - K \frac{Q}{Q - it} + Ae^{-Bit} \quad (4)$$

$$SOC(\%) = 100 \cdot (1 - \frac{Q_d}{C_{Bat}}) = 100 \cdot (1 - \frac{i_{Bat} \cdot t}{C_{Bat}}) \quad (5)$$

where E , V_{Bat} , i_{Bat} , Q_d , and C_{Bat} are the battery electromotive force, the battery voltage, the battery current, ampere-hours stored, and the internal capacity, respectively.

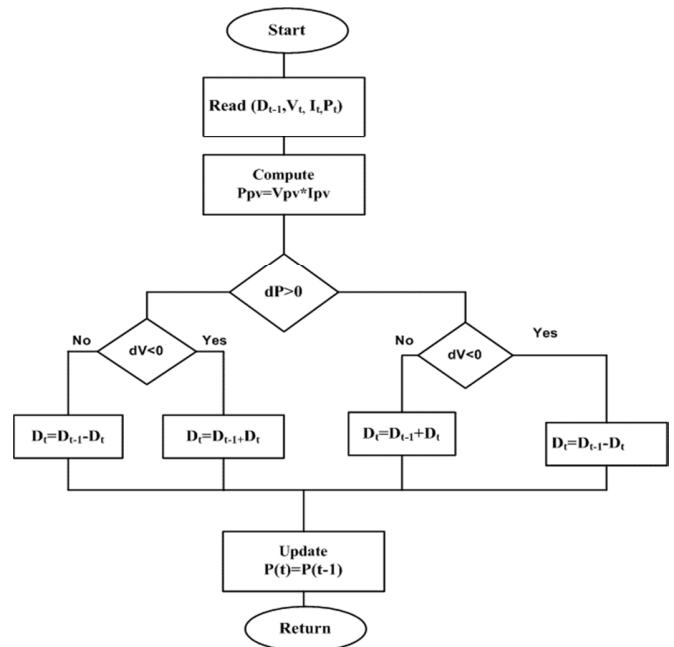


Fig. 3. Flowchart of the MPPT algorithm.

The Battery State Of Charge (SOC), representing the battery's stored electricity, is a crucial parameter for effective control. The proposed supervisory system must monitor the battery's SOC to make informed decisions based on its current state and the required power demand. Within the battery, the ampere-hours (Q_d) stored over a time interval (t) are related to the nominal battery capacity (C_{bat}) and the charging current (i_{Bat}). The hydroponic greenhouse is equipped with a lead-acid battery bank. Lead acid batteries are a prevalent choice in both off-grid and grid-connected PV systems. They are valued for their reliability, widespread global production, and cost-effectiveness in comparison to alternative PV storage technologies, like the lithium batteries. Additionally, lead-acid batteries offer several advantages, including low maintenance requirements, robust high current handling, suitability for deep

discharge conditions, favorable power density, and a wide operating temperature range. The battery bank consists of 16 cells with each cell having a capacity of 500 Ah and a nominal voltage of 24 V. Therefore, the total voltage of the battery bank is 48 V with a capacity of around 22 kWh. The battery is controlled through a bidirectional converter and a PI controller.

B. Energy Management based on Fuzzy Logic Control

The Fuzzy Logic Controller has emerged as one of the most effective tools in the scientific realm. This approach is distinguished by its straightforwardness and adaptability, drawing from theoretical advancements and human expert insights regarding the energy management of a hybrid system to power hydroponic greenhouses. These insights are employed to formulate fuzzy rules in the "if-then" linguistic format, all aimed at optimizing energy performance (Figure 4).

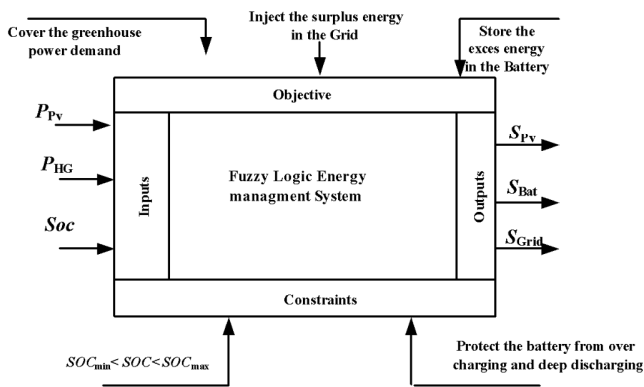


Fig. 4. Energy management system.

The fuzzy logic management algorithm consist of 3 inputs (P_{PV} , P_{HG} , SOC) and 3 outputs (S_{pv} , S_{Bat} , S_{grid}), the management algorithm comprises three operating modes:

- Mode 1: The required energy of the load is covered through the PV installation.
- Mode 2: The required energy of the load is covered through batteries.
- Mode 3: The required energy of the load is covered through the grid.

To use the battery safely, two limits of the state of charge were defined: $SOC_{min}=20\%$ and $SOC_{max}=80\%$. The management strategy should primarily use the power generated by the PV to satisfy the greenhouse load demand. Figure 5 illustrates the implementation of the proposed fuzzy logic controller engaging the Mamdani fuzzy inference system. This controller serves as the central decision-making hub for supplying the hydroponic greenhouse and charging batteries, ensuring they align with the specified conditions and operational requirements.

The principle of the fuzzy logic controller is based on generating three control signals S_{pv} , S_{Bat} , and S_{grid} starting from three inputs: PH Power (P_{pv}), greenhouse power demand (P_{HG}), and the battery state of charge SOC . S_{pv} is the control signal of

the switch of the PV generators, S_{Bat} is the control signal of the battery switch, and S_{grid} is the control signal of the grid switch. Table II displays the rule table of a fuzzy controller where the inputs of the matrix are fuzzy sets of SOC , P_{pv} and P_{HG} . The output of this table is the state of the three switches S_{pv} , S_{Bat} and S_{grid} .

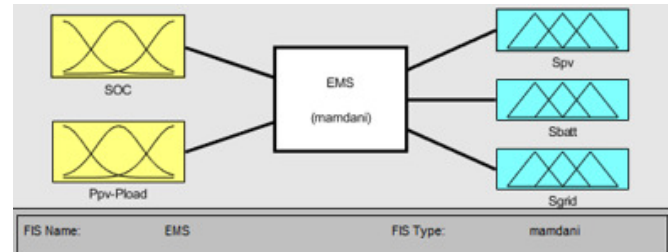


Fig. 5. EMS based on fuzzy logic controller.

TABLE II. DECISIONS MADE BY THE FUZZY LOGIC CONTROLLER.

G	$P_{pv}-P_{HG}$	SOC	S_{pv}	S_{Bat}	S_{Grid}	Decision
H	P	L	on	on	0	PVs supply HG and charge the battery
H	P	H	on	off	on	PVs supply HG and inject surplus energy to the grid
L	N	M	off	off	off	Battery supplies the HG
L	N	L	off	off	on	Grid supplies the HG

C. Simulation Results of the PV Power Plant

To verify the accuracy and effectiveness of the modeling and control techniques applied to the PV system that powers a hydroponic greenhouse, simulations were conducted using MATLAB software for both daily and annual periods.

1) Simulation Results of the PV Power Plant for One Day

Figure 6 illustrates the power output of the PV system, including the power supplied to the grid and the power demand of the hydroponic greenhouse. It exhibits the amount of electricity generated by the PV system during the simulation period. This information is crucial to understand the energy yield of the PV system and evaluate its performance. During the simulation period, it was observed that the power output of the PV array was zero between 00:00 Am and 05:00 Am and the grid supplied electricity during this period to meet the load demand. From the 5th hour, the PV system started to generate enough energy to cover the load demand. The grid reacted dynamically to the fluctuations in the load and the power of the PV system, thus ensuring that the entire energy requirement of the system was covered in the event of a production failure of the PV system. The excess energy from the PV system was fed into the grid. The simulation results presented in Figure 6 provide valuable insight into the performance and behavior of the PV system powering the hydroponic greenhouse. They demonstrate that the PV array can effectively meet the energy needs of the greenhouse and that the grid can be used as a backup power source when the PV array output fluctuates. The findings can be used to fine-tune the control strategy and improve system efficiency and reliability.

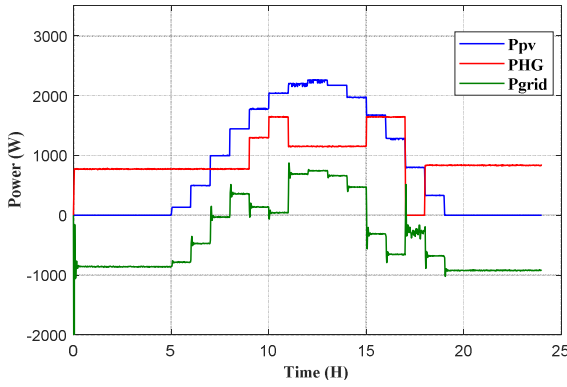


Fig. 6. Daily power generation from the PV system, grid supply, and hydroponic greenhouse power demand.

2) Simulation Results of the PV Power Plant for One Year

MATLAB was utilized to run simulations over a period of one year to experimentally validate the accuracy and efficacy of the modeling for the PV power plant. Figure 7 portrays the monthly electricity production of the PV system, the power supplied to the grid, and the power required by the hydroponic greenhouse. These values are strongly dependent on the solar radiation. In January, the PV system generated an average of 235 kWh. In May, the expected production was 358 kWh, indicating an increase of 52%. The highest production takes place in July with 370 kWh, constituting an increase of almost 58% compared to January. Figure 8 represents the percentage of PV system power generation for hydroponic greenhouse power demand and grid injection.

In the winter season, the power demand of the HG presented a percentage varying between 30% and 40% of the total electric energy generated by the PV system. During the summer season, the energy injected into the grid presented an average value of 85%. Thus, the designed PV system is able to meet the total hydroponic greenhouse electricity demand without considering energy storage systems.

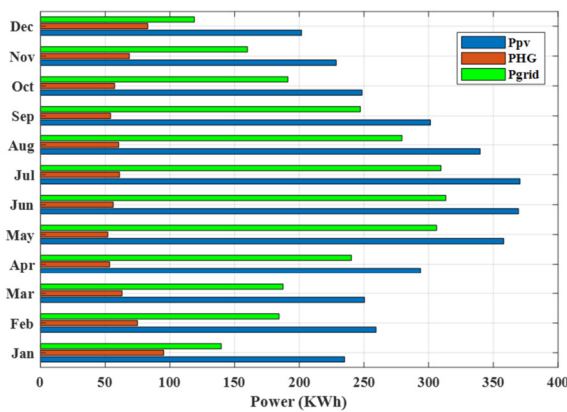


Fig. 7. Annual power generation from the PV system, grid supply, and hydroponic greenhouse power demand.

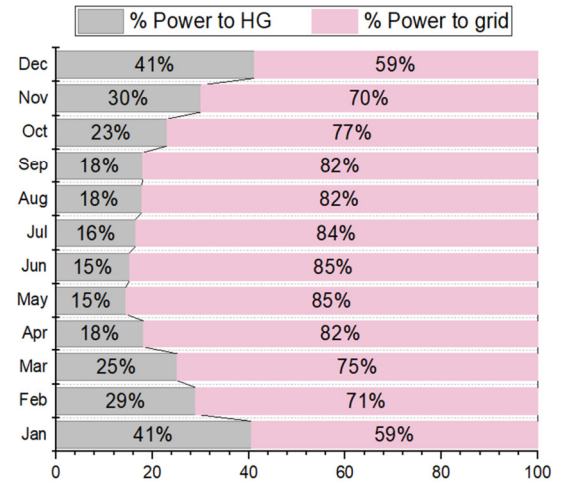


Fig. 8. Percentage of the PV system power generation for hydroponic greenhouse power demand and grid injection.

D. Wind Turbine Grid Connected Power Plant

1) Modeling

In this case, the aim is to create a wind turbine system for a hydroponic greenhouse that is connected to the electricity grid. To achieve this, the system is designed based on input data for the simulation, such as wind speed and direction, greenhouse location and orientation, energy demand profile and grid connection requirements. The main objective is to determine the ideal wind turbine specifications that will ensure consistent and efficient power generation for the greenhouse. This is accomplished by capturing the kinetic energy of the wind with aerodynamic blades, which in turn rotate the Permanent Magnet Synchronous Generator (PMSG) to produce electricity. The fundamental equation describing the mechanical power of the wind turbine is [23]:

$$P_{WT} = \frac{1}{2} \rho \pi R^2 V_{WT}^3 C_p(\lambda, \beta) \quad (6)$$

where P_{WT} is the mechanical power, ρ is the air density, R is the blade-radius, V_{WT} is the wind speed, and $C_p(\lambda, \beta)$ is the aerodynamic power coefficient which can be expressed as:

$$C_p(\lambda, \beta) = C_1 \left(C_2 \frac{1}{\lambda} - C_3 \beta - C_4 \right) e^{\frac{C_5}{\lambda}} \quad (7)$$

$$\frac{1}{\lambda_i} = \frac{1}{\lambda + 0.08\beta} - \frac{0.035}{\beta^3 + 1}$$

where λ is the tip speed ratio, β is the blade pitch angle, $C_1 = 0.5$, $C_2 = 116$, $C_3 = 0.4$, $C_4 = 5$, and $C_5 = 21$.

When β is maintained constant, we can see that the power coefficient C_p has only one maximum value C_{p_max} , which corresponds to the optimal value of C_p [6]. Therefore, C_{p_max} can be expressed as:

$$\lambda = \frac{\omega_m R}{V_{WT}} \quad (8)$$

where ω_m is the mechanical angular velocity of the rotor.

The electrical properties of the PMSG can be expressed mathematically by [7]:

$$\begin{cases} u_d = R_s i_d + \frac{d(L_d i_d + \psi_m)}{dt} - \omega_e L_q i_q \\ u_q = R_s i_q + \frac{d(L_q i_q)}{dt} + \omega_e (L_d i_d + \psi_m) \end{cases} \quad (9)$$

where d and q are the synchronous rotating reference frame, R_s is the stator winding-resistance, ω_e represents the electric pulsation, ψ_m is the generator rotor flux, i_d and i_q are the stator currents expressed in the $d-q$ reference frame, (L_d , L_q) and (u_d , u_q) are respectively the stator inductances and the stator voltages expressed in the $d-q$ reference frame.

The PMSG electromagnetic torque (T_e) can be calculated using the electromagnetic power (P_e) formula [25]:

$$P_e = \omega_m T_e = \frac{3}{2} \omega_e [(L_d i_d + \psi_m) i_q - L_q i_q i_d] \quad (10)$$

The active power and reactive power of a system can be calculated by [9]:

$$\begin{cases} P = u_d i_d + u_q i_q \\ Q = u_d i_q - u_q i_d \end{cases} \quad (11)$$

2) Simulation Results

Figure 12 depicts the MATLAB/Simulink schematic blocks of the wind turbine model. Wind turbine parameters are taken into consideration as inputs to the program. Such parameters are the wind speed and inclination angle, which are defined as variables or as constants. An AC/DC rectifier mainly converts the three-phase output into a single-phase output system. The main consumer is located between the wind power system and the connected grid, so that it can switch between them if necessary to supply the hydroponic greenhouse. The design configuration of this wind power system is shown below. The physical parameters of the proposed WTS are given in Table III.

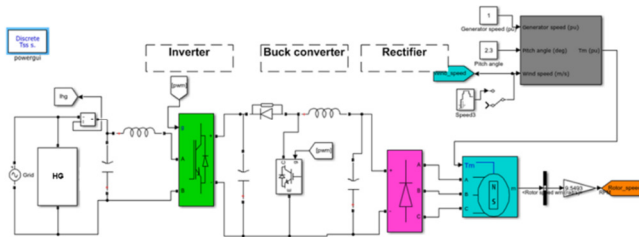


FIG. 9. Simulink assembly of the wind power system.

TABLE III. CHARACTERISTICS OF THE WIND TURBINE SYSTEM

Stator winding resistance of the PMSG (Ω)	0.18
Number of pole pairs	2
Moment of inertia (Kg/m^2)	0.0062
Armature inductance (H)	0.0835

One of the parameters characterizing the different models of wind turbines is the speed of rotation, which depends on wind speed. To compute the power of the wind turbine system, the wind speed is determined. In this study a wind turbine with a vertical axis of 1 kW is chosen. The recorded values for the minimum wind speed fluctuated between 0 and 4 m/s. The maximum value reached 10 m/s at noon. The average wind speed in north-west Tunisia varies between 0 and 6 m/s. Figure 10 demonstrates the daily electricity generation from the wind system, the grid supply, and the hydroponic greenhouse electricity demand. The presented numerical results disclose the electrical behavior of the wind system, which supplies the hydroponic greenhouse with an average value of about 500 W. They reveal that the wind generator can cover 50% of the energy needs of the greenhouse. The remainder of the electricity requirement is covered by the electricity grid. Compared to the first scenario, it becomes clear that the demand for electrical energy is covered more by PV than by wind turbines. The current generated at the output of the AC/DC/AC converter of the 1 kW vertical axis wind turbine is displayed in Figure 11. The current of the wind turbine only stabilizes when the wind speed is constant, which explains the fluctuations shown. Eventually, the wind speed stabilizes at the maximum value of 9 m/s, giving a maximum current of 11 A. The obtained simulation results indicate that the current has a sinusoidal shape, and the maximum amplitude varies according to the power requested by the hydroponic greenhouse load.

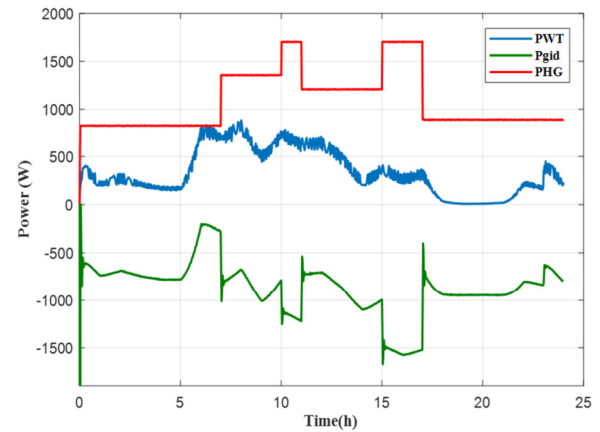


Fig. 10. Daily power generation from wind system, grid supply, and hydroponic greenhouse power demand.

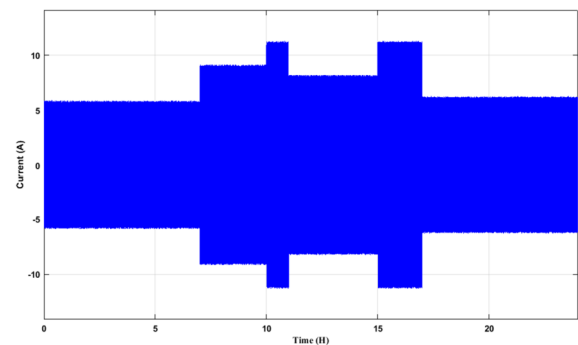


FIG. 11. Current generated by the wind grid-connected power plant.

The three-phase voltage generated by the proposed generator undergoes a conversion process that transforms it into single-phase voltage. This transformation is achieved through a rectifier and a single-phase inverter setup. The resulting output is maintained as AC voltage, with a maximum amplitude of 320 V. This voltage can supply the hydroponic greenhouse or be injected into the grid as per the requirements of the application. Figure 12 illustrates the relationship between the PMSG speed and the Wind Turbine System (WTS) electromagnetic torque. The rotor speed of the PMSG exhibits random fluctuations due to the variability in wind speed. Consequently, these fluctuations in the PMSG rotor speed result in corresponding changes in the electromagnetic torque and the power output of the generator.

The power created by a wind turbine as a function of turbine speed is described by its power curve at different wind speeds, as can be noticed in Figure 13. The generated turbine power is a polynomial curve. It depends on the parameters of wind speed and turbine speed. There are two significant turbine speeds: a maximum speed and a minimum speed at which the turbine produces the least power. The optimal speed is recorded when the turbine produces the maximum power, but this speed is variable according to the wind speed. For example, at a speed of 10 m/s, the turbine generates an optimal power of 1.2 kW.

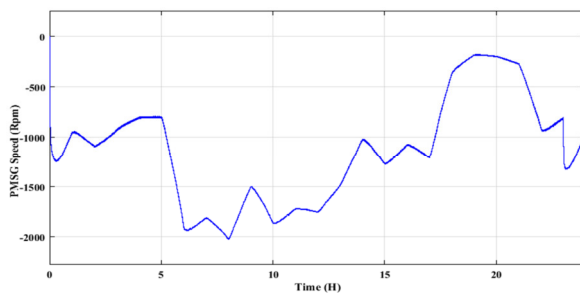


Fig. 12. Speed of the PMSG.

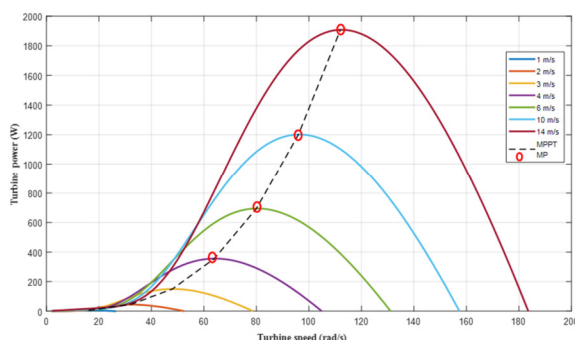


Fig. 13. Turbine power generated vs turbine speed for different wind speed values.

IV. IMPLEMENTATION AND EXPERIMENTAL VALIDATION OF THE PV - STORAGE SYSTEM

To design and optimize the performance of a PV system to supply a hydroponic greenhouse, various factors need to be considered during the design process. These include the

available space for the solar panels, the orientation and angle of inclination of the panels, the location and climate of the site and the expected solar irradiation. Based on these factors, the solar panels, inverters, batteries, and cabling can be properly sized to ensure maximum efficiency and effectiveness of the system. After designing a grid-connected PV system to fulfill the electrical energy needs of the hydroponic greenhouse (as depicted in Figure 14), it was determined that the maximum power required by the greenhouse is 2 kW. Consequently, a 2.1 kWp PV system was selected to satisfy this demand.



Fig. 14. PV installation.



Fig. 15. Battery storing system.

To validate the numerical model, an experimental analysis has been performed over a year with an experimental PV system equipped with the HG layout. The scatter plots of the measured and predicted values of PV power and ambient temperature are portrayed in Figure 16. The latter underscores the effectiveness of the Simulink model developed in this study. With an ambient temperature of around 16 °C in winter and spring, the measured and predicted values showcased a slight difference. Figure 16 illustrates the generated and forecasted power, visualized as yearly variations over a 12-month period. A polynomial fit is applied to both the experimental and simulated data sets. The two polynomial fits have almost the same appearance with a slight variation, confirming the effectiveness of the developed Simulink model. In summer and autumn, the curves of the experimental and numerical results are almost identical.

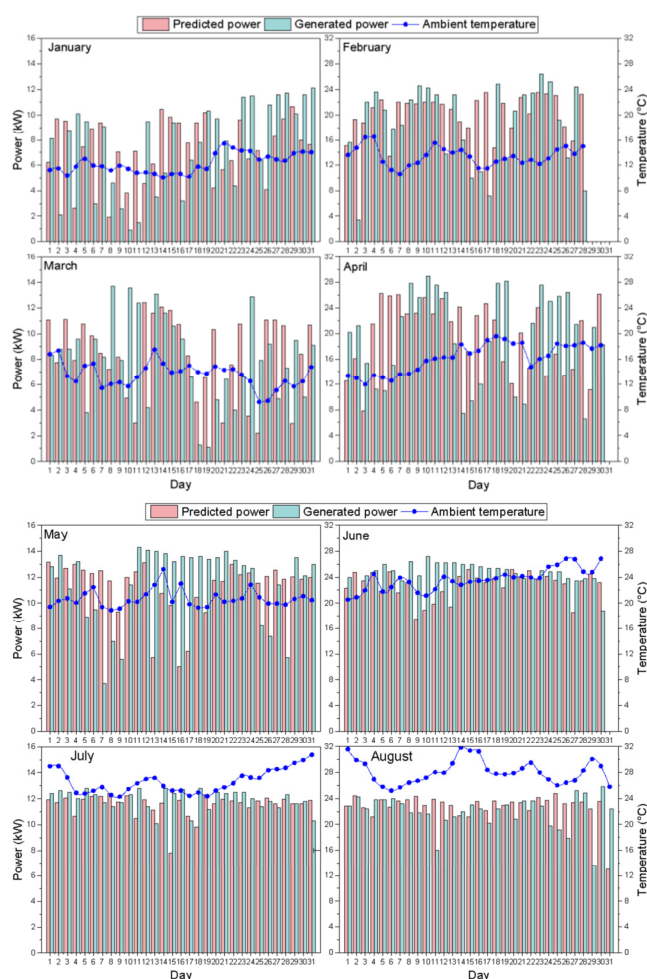


Fig. 16. Measured and predicted annual values of PV power and ambient temperature.

V. CONCLUSION

This study investigated the feasibility of implementing renewable energy systems, specifically solar and wind, to achieve the energy independence of a hydroponic greenhouse. The research was initiated with the design of a renewable energy system capable of fulfilling its energy requirements throughout the year. The efficiency of two renewable energy sources, photovoltaic panels and wind turbines has been demonstrated, and their integration within the agricultural chamber has been enhanced by using model simulations. Two scenarios were explored, each offering valuable insights into the optimization of energy management in renewable energy greenhouses. In the first scenario, the simulation results highlighted the dynamic behavior of the battery system, showcasing its resilience and reliability in addressing energy deficits during periods of low solar irradiation. The presented algorithm effectively optimized the combination of energy sources to minimize grid usage while maximizing photovoltaic cell utilization.

In the second scenario, it was observed that the wind turbine could cover 50% of the greenhouse's energy needs,

indicating the superiority of photovoltaics compared to wind turbines in meeting the electrical energy demands. Furthermore, the numerical analysis was validated through a one-year experimental study on a grid-connected photovoltaic installation with storage, affirming the effectiveness of the Simulink model developed in this study.

In comparison to existing works, the current research provides valuable insights into sustainable and self-sufficient energy solutions for agricultural applications. By integrating renewable energy sources with advanced energy management systems, not only does this study contribute to reducing greenhouse gas emissions, but it also highlights the potential for efficient environmental regulation and resource utilization in hydroponic greenhouse operations.

REFERENCES

- [1] W. Ahmed, J. A. Sheikh, A. Z. Kouzani, and M. A. P. Mahmud, "The Role of Single End-Users and Producers on GHG Mitigation in Pakistan—A Case Study," *Sustainability*, vol. 12, no. 20, Jan. 2020, Art. no. 8351, <https://doi.org/10.3390/su12208351>.
- [2] F. Salata *et al.*, "Heading towards the nZEB through CHP+HP systems. A comparison between retrofit solutions able to increase the energy performance for the heating and domestic hot water production in residential buildings," *Energy Conversion and Management*, vol. 138, pp. 61–76, Apr. 2017, <https://doi.org/10.1016/j.enconman.2017.01.062>.
- [3] S. Bouadila, S. Baddadi, S. Skouri, and R. Ayed, "Assessing heating and cooling needs of hydroponic sheltered system in mediterranean climate: A case study sustainable fodder production," *Energy*, vol. 261, Dec. 2022, Art. no. 125274, <https://doi.org/10.1016/j.energy.2022.125274>.
- [4] S. Baddadi, S. Bouadila, W. Ghorbel, and A. Guizani, "Autonomous greenhouse microclimate through hydroponic design and refurbished thermal energy by phase change material," *Journal of Cleaner Production*, vol. 211, pp. 360–379, Feb. 2019, <https://doi.org/10.1016/j.jclepro.2018.11.192>.
- [5] E. Cuce, D. Harjunowibowo, and P. M. Cuce, "Renewable and sustainable energy saving strategies for greenhouse systems: A comprehensive review," *Renewable and Sustainable Energy Reviews*, vol. 64, pp. 34–59, Oct. 2016, <https://doi.org/10.1016/j.rser.2016.05.077>.
- [6] C. Lee, P. Hoes, D. Cóstola, and J. L. M. Hensen, "Assessing the performance potential of climate adaptive greenhouse shells," *Energy*, vol. 175, pp. 534–545, May 2019, <https://doi.org/10.1016/j.energy.2019.03.074>.
- [7] A. S. Anifantis, A. Colantoni, and S. Pascuzzi, "Thermal energy assessment of a small scale photovoltaic, hydrogen and geothermal stand-alone system for greenhouse heating," *Renewable Energy*, vol. 103, pp. 115–127, Apr. 2017, <https://doi.org/10.1016/j.renene.2016.11.031>.
- [8] F. A. Alturki, A. A. Al-Shamma'a, H. M. H. Farh, and K. AlSharabi, "Optimal sizing of autonomous hybrid energy system using supply-demand-based optimization algorithm," *International Journal of Energy Research*, vol. 45, no. 1, pp. 605–625, 2021, <https://doi.org/10.1002/er.5766>.
- [9] C. Lu, S. Li, J. Gu, W. Lu, T. Olofsson, and J. Ma, "A hybrid ensemble learning framework for zero-energy potential prediction of photovoltaic direct-driven air conditioners," *Journal of Building Engineering*, vol. 64, Apr. 2023, Art. no. 105602, <https://doi.org/10.1016/j.job.2022.105602>.
- [10] K. G. Tataraki, K. C. Kavvadias, and Z. B. Maroulis, "Combined cooling heating and power systems in greenhouses. Grassroots and retrofit design," *Energy*, vol. 189, Dec. 2019, Art. no. 116283, <https://doi.org/10.1016/j.energy.2019.116283>.
- [11] A. Shaqour, H. Farzaneh, Y. Yoshida, and T. Hinokuma, "Power control and simulation of a building integrated stand-alone hybrid PV-wind-battery system in Kasuga City, Japan," *Energy Reports*, vol. 6, pp. 1528–1544, Nov. 2020, <https://doi.org/10.1016/j.egyr.2020.06.003>.
- [12] F. Attig-Bahar, U. Ritschel, P. Akari, I. Abdeljelil, and M. Amairi, "Wind energy deployment in Tunisia: Status, Drivers, Barriers and

- Research gaps—A Comprehensive review," *Energy Reports*, vol. 7, pp. 7374–7389, Nov. 2021, <https://doi.org/10.1016/j.egy.2021.10.087>.
- [13] S. Mahjoub, L. Chrifi-Alaoui, S. Drid, and N. Derbel, "Control and Implementation of an Energy Management Strategy for a PV–Wind–Battery Microgrid Based on an Intelligent Prediction Algorithm of Energy Production," *Energies*, vol. 16, no. 4, Jan. 2023, Art. no. 1883, <https://doi.org/10.3390/en16041883>.
- [14] H. ur Rehman, J. Hirvonen, and K. Siren, "Influence of technical failures on the performance of an optimized community-size solar heating system in Nordic conditions," *Journal of Cleaner Production*, vol. 175, pp. 624–640, Feb. 2018, <https://doi.org/10.1016/j.jclepro.2017.12.088>.
- [15] M. Ahliouati *et al.*, "Energetic and parametric studies of a basic hybrid collector (PV/T–Air) and a photovoltaic (PV) module for building applications: Performance analysis under El Jadida weather conditions," *Materials Science for Energy Technologies*, vol. 6, pp. 267–281, Jan. 2023, <https://doi.org/10.1016/j.mset.2023.02.001>.
- [16] F. Wadie, "Evaluative analysis for standardized protection criteria against single and multiple lightning strikes in hybrid PV-wind energy systems," *Electric Power Systems Research*, vol. 218, May 2023, Art. no. 109227, <https://doi.org/10.1016/j.epsr.2023.109227>.
- [17] D. Hidouri, R. Marouani, and A. Cherif, "Modeling and Simulation of a Renewable Energy PV/PEM with Green Hydrogen Storage," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12543–12548, Feb. 2024, <https://doi.org/10.48084/etasr.6492>.
- [18] N. Niveditha and M. M. Rajan Singaravel, "Optimal sizing of hybrid PV–Wind–Battery storage system for Net Zero Energy Buildings to reduce grid burden," *Applied Energy*, vol. 324, Oct. 2022, Art. no. 119713, <https://doi.org/10.1016/j.apenergy.2022.119713>.
- [19] M. A. Abdullah, A. H. M. Yatim, C. W. Tan, and R. Saidur, "A review of maximum power point tracking algorithms for wind energy systems," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 5, pp. 3220–3227, Jun. 2012, <https://doi.org/10.1016/j.rser.2012.02.016>.
- [20] A. Sohani *et al.*, "Techno-economic evaluation of a hybrid photovoltaic system with hot/cold water storage for poly-generation in a residential building," *Applied Energy*, vol. 331, Feb. 2023, Art. no. 120391, <https://doi.org/10.1016/j.apenergy.2022.120391>.
- [21] N. Arfaoui, S. Bouadila, and A. Guizani, "A highly efficient solution of off-sunshine solar air heating using two packed beds of latent storage energy," *Solar Energy*, vol. 155, pp. 1243–1253, Oct. 2017, <https://doi.org/10.1016/j.solener.2017.07.075>.
- [22] A. Alanazi, "Optimization of Concentrated Solar Power Systems with Thermal Storage for Enhanced Efficiency and Cost-Effectiveness in Thermal Power Plants," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12115–12129, Dec. 2023, <https://doi.org/10.48084/etasr.6381>.
- [23] R. Ben Ali, H. Schulte, and A. Mami, "Modeling and simulation of a small wind turbine system based on PMSG generator," in *Evolving and Adaptive Intelligent Systems*, Ljubljana, Slovenia, Jun. 2017, pp. 1–6, <https://doi.org/10.1109/EAIS.2017.7954833>.
- [24] E. N. F. Ltd, "ENF Ltd." <https://fr.enfsolar.com/pv/panel-datasheet/crystalline/10249>.
- [25] S.-W. Lee and K.-H. Chun, "Adaptive Sliding Mode Control for PMSG Wind Turbine Systems," *Energies*, vol. 12, no. 4, Jan. 2019, Art. no. 595, <https://doi.org/10.3390/en12040595>.
- [26] Y. Kassem, H. Camur, M. T. Adamu, T. Chikowero, and T. Apreala, "Prediction of Solar Irradiation in Africa using Linear-Nonlinear Hybrid Models," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11472–11483, Aug. 2023, <https://doi.org/10.48084/etasr.6131>.

A Hybrid Intelligent Controller for Extended-Range Electric Vehicles

Jayakumar Jayaraj

Department of Electrical and Electronics Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India
jayakumar@karunya.edu

Dakka Obulesu

Department of Electrical and Electronics Engineering, CVR College of Engineering, Hyderabad, India
dakkaobulesh@cvr.ac.in

Hemaprabha Govindaraj

Department of Electrical and Electronics Engineering, INFO institute of Engineering, Coimbatore, Tamilnadu, India
hemaprabhaeee@gmail.com

Francisxavier Thomas Josh

Department of Electrical and Electronics Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India
josh@karunya.edu

Nagalingam Rajeswaran

Department of Electrical and Electronics Engineering, Malla Reddy College of Engineering, Secunderabad, India
rajeswarann@gmail.com (corresponding author)

Chilakala Rami Reddy

Department of Electrical and Electronics Engineering, Joginpally B. R. Engineering College, Hyderabad, India | Applied Science Research Center, Applied Science Private University, Amman, Jordan
crreddy229@gmail.com

Abdullah S. Algarni

Department of Electrical Engineering, College of Engineering and Computing in Al-Lith, Umm Al-Qura University, Mecca, Saudi Arabia
asgarni@uqu.edu.sa

Abdullah Alwabli

Department of Electrical Engineering, College of Engineering and Computing in Alqunfudah, Umm Al-Qura University, Mecca, Saudi Arabia
aswabli@uqu.edu.sa

Saeed Faisal Malky

Department of Electrical and Electronics Engineering, University of Jeddah, Jeddah, Saudi Arabia
sfmalki@uj.edu.sa

Received: 30 January 2024 | Revised: 13 February 2024 | Accepted: 14 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6960>

ABSTRACT

A smart battery electric vehicle control framework is proposed in this paper. The specific controller empowers ceaseless observation and management of the battery's state with the scope of extending the vehicle's driving range under varying temperature and driving pattern conditions. The proposed method utilizes an incorporated scheme for dealing with a crossover energy stockpiling framework to expand a battery's lifespan while further ensuring its smooth activity.

Keywords-*electric vehicle; battery; energy storage system; hybrid electric vehicle controller; automation*

I. INTRODUCTION

The exhaust emissions of gasoline-powered automobiles are primary contributors to air pollution. It is vital to convert cars that run on fossil fuels into vehicles that run on green energy to accomplish the low-carbon objectives of green cities. Electric Vehicles (EVs) do not produce emissions or pollutants, so they play a significant part in this transition [1]. On the other hand, the marketing of electric cars is greatly hindered by technological restrictions, such as low battery power and limited driving range. Not only do Extended-Range Electric Vehicles (EREVs) possess the qualities of low emissions, but they also can expand the mileage that cars can tolerate for an extended period. In traditional Hybrid Electric Vehicles (HEVs), the battery functions as an energy buffer [2]. Over the course of the journey, there is a little fluctuation in the State Of Charge (SOC) of the battery. EREVs, on the other hand, are plug-in HEVs equipped with a large battery that can store electric power provided by a charging station. During its charging cycle, the EREV leaves a charging station with its battery energy at its maximum capacity until it arrives at another charging station [3]. If the EREV completes its journey and reaches the next charging station with a low amount of battery charge, it will be able to store more clean energy from the grid, thus decreasing the quantity of gasoline used. In the past, two different control procedures have been used to accomplish the control objectives indicated above [4]: The Charging Deplete/Charging Sustain (CD/CS) approach and the mixed control technique. The first part of the former approach is referred to as the CD stage, and it is achieved by cars operating in a mode that is inside the All-Electric Range (AER). In addition to the AER, the extender is activated as the primary energy source in order to meet the typical power requirements of the Traction Motor (TM) [5]. When there is a higher demand for power, such as when there is a rapid acceleration or when there is an upslope, the battery works in conjunction with the extender to ensure that the SOC of the battery is maintained in a narrow range over the lower bound. This stage is referred to as the CS stage [6]. An EV's only energy source is a battery. This can cause issues in case the latter is not joined with a regulator. During the EV design phase, the primary obstructions are: Inadequate charging framework, reliance on imported batteries, reliance on imported segments and parts, range anxiety, high EV cost, limited choices of superior EVs, inadequate power supply in some parts of India, lack of value support, and possible car industry slump.

II. ENERGY STORAGE REQUIREMENTS

The energy stockpiling mechanism of an EV should be lightweight, with high energy density, easy to control, must convey the initial current, and must have a long life cycle,

transient security, and low activity cost. The distance an EV can travel is restricted by numerous elements. Authors in [3] deem SOC and the wellbeing of a lithium-ion battery cell significant quality measures for EVs. With regard to driving patterns, one of the factors negatively influencing a battery's operation is the frequent beginning and halting due to traffic and street conditions. Therefore, a battery encounters pressure, bringing down its charging time and its lifespan. Outside temperature variances also affect the battery's performance [4]. To guarantee a legitimate energy stockpiling framework plan, misfortunes and working conditions should be considered. Additionally, an auxiliary fuel source should be utilized [5].

III. DRIVING PATTERNS

Lithium-ion battery efficiency is affected by driving patterns, which may alter the battery usage. Battery current fluctuates with driving standards. The main driving patterns are: city traffic (light and heavy), suburban, state highway, national highway, express highway, local/country loads, small towns, hilly/high range areas (Figure 1). The actual habits of a driver may diminish the battery's performance or life expectancy [6, 7].

Energy stockpiling is fundamental for EVs on the grounds of energy density and weight. Energy density is mostly subject to the vehicle's design, construction, and mileage. Authors in [8] offer an energy stockpiling methodology for EVs that is subject to driving examples. Authors in [9] investigated load difference and its importance for EVs.

IV. RANGE ANXIETY

Regarding the increment of the vehicle range, vehicle charging and an EV energy stockpiling framework are both required. To meet the vehicle's energy needs, an energy stockpiling framework that is fit to provide the required energy should be utilized. Diminishing release time is caused by expansions in battery strain. This suggests that in order to guarantee that optimum current is drawn from the battery, the current scheme should contain an approach to augment the current [10].

Reach anxiety is the concern drivers get when they are worried about the quantity of kilometers that their EV can cover before it is re-energized. The target of this paper is to reduce range anxiety utilizing improved battery controls. Authors in [12] examined the compromises of proficiency, energy utilization, and engine heat dispersal while creating and surveying a battery-EV power train. Controlling battery frameworks is essential. The effects of battery release control on vehicle range were researched in [13], with the aim of ensuring longer battery use times to upgrade the mileage per charge.

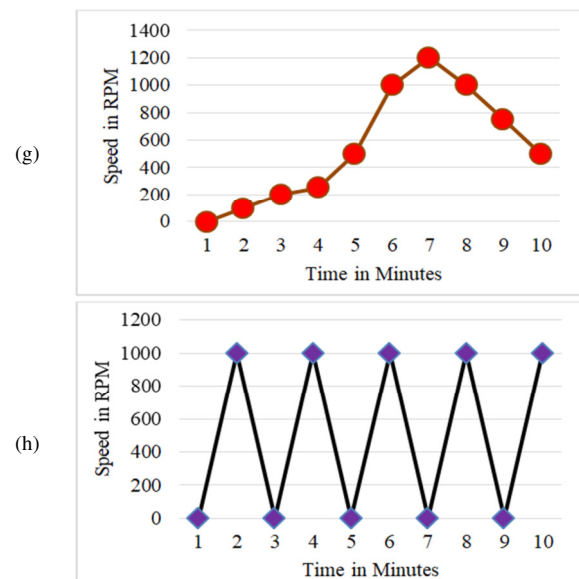
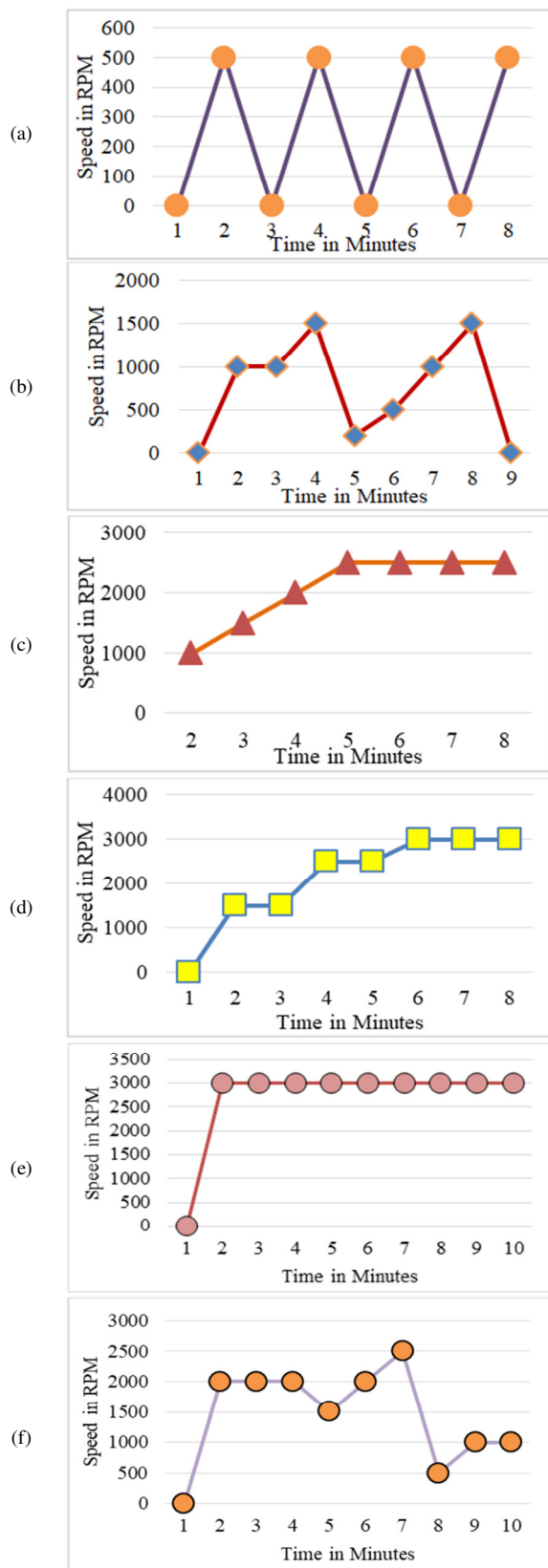


Fig. 1. Driving patterns: (a) city, (b) suburban, (c) state highway, (d) national highway, (e) express highway, (f) local/country roads, (g) small town, (h) hilly areas.

V. ELECTRIC VEHICLE RANGE EXTENSION

A complex energy regulator may assist with range anxiety in EVs. A range expansion procedure for EVs is proposed in [14]. To keep the cycle working, a shrewd regulator should screen information and input boundaries. Authors in [15] fostered a mixture energy stockpiling framework that utilizes fuel to extend the range of EVs. With this technique, the capacity devices are: (a) a super capacitor, (b) a main battery, (c) a backup battery, and (d) a charging source (photovoltaic cell). In the design, modeling, and hardware implementation of a next generation EREV, street conditions and vehicle load are considered [16], while a smart regulator uses boundaries for the battery and directs energy flow. The settings of the regulator administer the activity of the energy stockpiling mechanisms. The vehicle is halted when the speed is zero or when the supercapacitor is activated to move it. When the vehicle gets going, the capacitor that backs it off is deactivated and the main battery pack is turned on, whereas the secondary battery pack is charged through a sun-powered cell. If the main battery pack is depleted, the secondary battery will be turned on. By putting the battery on a sturdier establishment, the tension on the battery is limited. This has the effect of broadening the lifespan of the battery and improving its effectiveness. Authors in [17] examined the execution of a low-voltage-to-cell battery adjusting circuit. A productivity strategy for a DC-DC converter in a half breed energy stockpiling framework for EVs is presented in [18]. EV designers should be informed about a few issues, including the heaviness of the framework, the type of the EV, the need for energy stockpiling, driving patterns, and the scope of the vehicle. Each framework should uphold and be viable when combined with every other framework. The EV's amassing capacity, energy, framework plan, and boundary configuration should all be viable to assemble the vehicle.

VI. DESIGN OF THE RANGE EXTENDED ELECTRIC ENERGY STORAGE SYSTEM CONTROL ALGORITHM

Since it includes several different energy sources, this hybrid energy storage system has less dependency on a single energy source, offering the former better overall performance [19]. An unscented Kalman filter-based battery SOC estimation and the peak power prediction approach for the power distribution of hybrid electric automobiles for on-road applications is proposed. Changes in needs, switching, and charging will be handled by the controller, as shown in Figure 2.

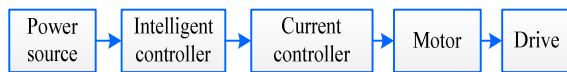


Fig. 2. Connections of the functional parts.

A. Functions of Each Part

- Power sources (lithium-ion batteries and super capacitor). They form the energy storage system which powers the micro controller and the EV.
- Intelligent controller: It controls and monitors all the parameters of the energy storage system and the EV in order to optimize the current from the battery/power source.
- Power electronic current controller/converter: It limits the maximum current given to the motor.
- The motor: It drives the vehicle.
- Battery management system: It records the parameters of the battery and the super capacitor.
- Speed sensor/driving pattern recognizer: It gives feedback parameters to the controller.

B. Block Diagram of the Proposed Controller

Figure 3 displays the block diagram of the novel intelligent integrated hybrid energy storage system controller that includes a primary battery, an auxiliary battery, a super capacitor, solar charging, driving patterns, traffic circumstances, and vehicle load. The use of current, input, and feedback parameters to optimize current are the primary roles of the intelligent controller. Data derived from the evaluation process (speed, driving patterns, gradient angle, load current) are considered when controlling the main battery bank, the auxiliary battery, the super capacitor, and the solar panel. The controller regulates the current drawn by the electronic converter and sent to the motor. Lithium-ion batteries and super capacitor power the EV. Display operations, including the controller, are powered by an extra battery. A solar panel is utilized to charge the auxiliary battery and the super capacitor. Even if the primary battery is completely depleted, it will be possible to utilize control and display features. The super capacitor will provide power when movement is initiated or ascending or at surges, whilst the main battery would power the vehicle at constant speeds. An optimum driving pattern is exhibited on the HMI when the power/torque need is recognized from the driving pattern, extracted from the user control block established by the user. This research suggests a controller that

will allow EVs to cover greater ranges. Authors in [20-23] describe how the acceleration and deceleration patterns of drivers affect the velocity control of EVs. Traffic and road conditions are given to the driver in order to be able to regulate the use of the accelerator and brake pedal [24-30].

C. Intelligent Algorithm based on the Fuzzy Logic Controller

The algorithm controls the priority sets (super capacitor, main battery) and acts according to a priority hierarchy based on the availability of sources for charging the battery and powering the motor. The constraints are: driving patterns, driver behavior, road conditions, and traffic conditions. The control system functions are depicted in Tables I and II. Equation (1) is used to represent the amount of power used and (2) demonstrates the distance covered based on the driving pattern.

If $f(x)$ is the power requirement, then:

$$f(x) = ax + b \quad (1)$$

where x is the power requirement from the super capacitor, a is the driving pattern factor, and b is the positive power.

The distance covered by the driving pattern is expressed by:

$$dj = \sqrt{x^2 - c^2} \dots \quad (2)$$

where d is the distance, x is the sampling point, c represents the cluster, and $j = 1, 2, 3$.

An integrated control algorithm for hybrid and electric cars called the Novel Hybrid Integrated Intelligent Control (HHIC) was utilized to lower the strain on the battery and to optimize the current flow to prolong range. The Energy Storage System (ESS) showcases better performance due to the battery discharge time extension regulated by the intelligent controller.

TABLE I. RULE-BASED FUZZY LOGIC MEMBERSHIP FUNCTIONS

Operating condition	Switch condition	Power allocation
Driving: Battery and super capacitor	$P_{min} < P_{demand}$ and $SOC_{UCmin} < SOC_{UC}$	$P_{uc} = aP_{demand} + bP_{bat}$ $t = -u$
Driving: Battery	$P_{min} < P_{demand}$ and $SOC_{UCmin} < SOC_{UC}$	$P_{bat} = P_{demand}$ $P_{uc} = 0$
Driving: Battery	$0 \leq P_{dem} \leq P_{min}$ $0 < P_{demand} < P_{min}$	$P_{bat} = P_{demand}$ $P_{uc} = 0$
Solar: Super capacitor	$P_{demand} < 0$ $SOC_{UCmax} < SOC_{UC}$	$P_{bat} = P_{demand}$ $P_{uc} = 0$
Braking: Neither	$P_{dem} < 0$ & $SOC_{UCmax} < SOC_{UC}$	$P_{bat} = 0$ $P_{uc} = 0$

TABLE II. RULE-BASED FUZZY LOGIC MEMBERSHIP FUNCTION

Input and output	Actual domain	Fuzzy domain	Membership function	Fuzzy subset levels
P_{demand}	0-90	[0 to 1]	Gauss type/Bilateral Gauss	3
V_{trend}	[-3 to 3]	[-3 to 3]	Rectangle	7
G_{road}	[-10 to 10]	[-1 to 1]	Rectangle	7
a_{uc_corr}	[-0.2 to 0.2]	[-0.2 to 0.2]	Gauss type/Bilateral Gauss	7

D. Performance Analysis of the Proposed Controller

To consider the impact of varying conditions on EVs, tests were conducted with a novel hybrid crossover energy stockpiling control framework. Energy utilization allows EVs

to cover longer distances [24-27]. The tests involved driving on various courses at different rates, with changing climate conditions, and speed patterns. The EV was running on super capacitor for movement initiation, and on the principle battery when moving with constant speed.

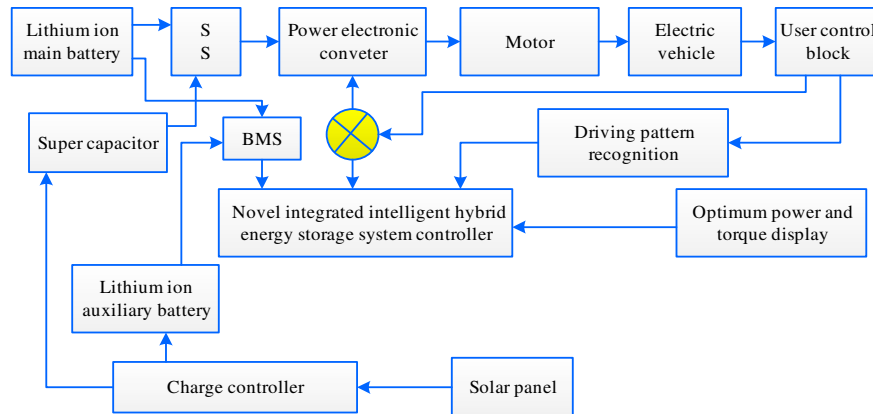


Fig. 3. Connections of the functional parts.

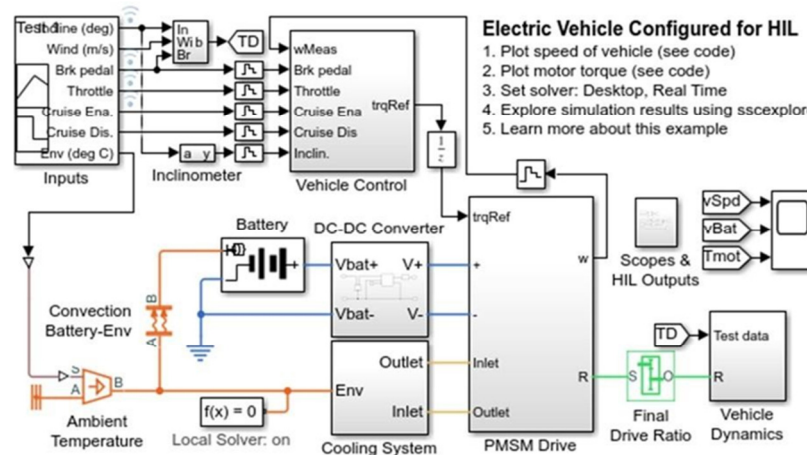


Fig. 4. Simulation hardware model of a generic EV.

VII. RESULTS AND DISCUSSION

The Matlab/Simulink tests on how progressing conditions change the mathematical limits and how those segments convert into setup features of the vehicle where conducted with the aid of an expanded vehicle model. The EV model's schematic outline is displayed in Figure 4. The vehicle's speed is constrained by the reenactment model that is related with the PC through an interface. Li-ion batteries and a super capacitor power the electric vehicle. Other devices, including the controller, are powered by a secondary battery. The points of interest of the EV model are indicated in Table III. When solar power is available, the secondary battery and the super capacitor are charged. Whether or not the main battery is completely depleted, utilizing control and showing features will be possible. Current spike occurrences are the occasions when the super capacitor controls the vehicle. The main battery controls the vehicle at constant speeds, while the solar charging device powers the super capacitor and the secondary battery.

The optimal driving model is shown on the HMI and the driving model is setup by the customer. Cloud data or the customer control block choose the traffic and road conditions. Using a super capacitor makes it easier to use a lithium battery, making it possible to initiate and accelerate simultaneously. Tables IV-VII portray the range extension at various conditions and Figures 5-7 visualize the results. The Matlab code copies the distinctive driving conditions that occur during the driving example of an EV, helping to select how the energy utilization changes under various conditions [22].

Regulator calculations that can upgrade and control the energy flow will permit EVs to cover longer distances. It was observed that the proposed control approach may accomplish greater vehicle range expansion than the elective control techniques. Figure 5 delineates the rate extension of the release time for various driving patterns. Table V shows the release time expansion under various temperature conditions, as depicted in Figure 6. When the SOC drops, the temperature of the lithium-ion battery also drops.

TABLE III. SPECIFICATIONS OF THE EV MODEL

Quantity / Capacity	Unit / Type
Weight (vehicle)	830 kg
Dimensions	3280×1514×1560 (Turning Radius=3.90)
Tyre Size	Front: R16
	Rear: R16
Speed (max)	81 km/h
Voltage	48 V
Main battery capacity (Lithium – ion)	210 Ah
Running time in km (battery)	120 (full charge)
Super capacitor capacity	500 F / 1000 A / 5.55 Ah
Auxiliary battery capacity	21 Ah
Solar panel capacity	480 W / 48 V / 10 A
Current controller	600 A (peak)
Communication module	Bluetooth

TABLE IV. DISCHARGE TIME EXTENSION IN PERCENTAGE FOR DIFFERENT DRIVING PATTERNS

SO C	Driving Patterns Constant (dv/dt)								
	T(°C)	City	Suburban	State highway	National highway	Express highway	High range/hilly area	Small town	Average discharge time increase (%)
	28	0.05	0.05	0.05	0.05	0.05	0.05	0.05	4.13
5	28	0.1	0.1	0.1	0.1	0.1	0.1	0.1	4.13
10	27.5	0.15	0.15	0.15	0.15	0.15	0.15	0.15	4.13
15	27	0.2	0.2	0.2	0.2	0.2	0.2	0.2	4.13
20	26.5	0.25	0.25	0.25	0.25	0.25	0.25	0.25	4.13
25	26	0.3	0.3	0.3	0.3	0.3	0.3	0.3	4.13
35	25.425	0.4	0.4	0.4	0.4	0.4	0.4	0.4	4.13
40	25.4	0.45	0.45	0.45	0.45	0.45	0.45	0.45	4.13
45	25.325	0.5	0.5	0.5	0.5	0.5	0.5	0.5	4.13
50	25.3	0.55	0.55	0.55	0.55	0.55	0.55	0.55	4.13
55	25.225	0.6	0.6	0.6	0.6	0.6	0.6	0.6	4.13
60	25.2	0.65	0.65	0.65	0.65	0.65	0.65	0.65	4.13
65	25.175	0.7	0.7	0.7	0.7	0.7	0.7	0.7	4.13
70	25.15	0.75	0.75	0.75	0.75	0.75	0.75	0.75	4.13
80	25.1	0.85	0.85	0.85	0.85	0.85	0.85	0.85	4.13
75	25.075	0.9	0.9	0.9	0.9	0.9	0.9	0.9	4.13
90	25.05	0.95	0.95	0.95	0.95	0.95	0.95	0.95	4.13
95	25.025	1	1	1	1	1	1	1	4.13
100	25	1.5	1.5	1.5	1.5	1	1.5	1.5	4.13

TABLE V. DISCHARGE TIME EXTENSION (%) FOR DIFFERENT BATTERY SOC VALUES

Rules	Main battery SOC (%)	Super capacitor SOC (%)	Temperature (°C)	Battery discharge time with/without controller (hr)	Range extension (%)
VERY LOW	0	0	28	1.48/1.54	4.13
VERY LOW	5	5	28		
VERY LOW	10	10	27.5		
VERY LOW	15	15	27		
VERY LOW	20	20	26.5		
LOW	25	25	26		
LOW	30	30	25.5		
LOW	35	35	25.425		
LOW	40	40	25.4		
LOW	45	45	25.325		
MEDIUM	50	50	25.3		
MEDIUM	55	55	25.225		
MEDIUM	60	60	25.2		
MEDIUM	65	65	25.175		
MEDIUM	70	70	25.15		
HIGH	75	75	25.125		
HIGH	80	80	25.1		
HIGH	75	85	25.075		
HIGH	90	90	25.05		
HIGH	95	95	25.025		
HIGH	100	100	25		

Figure 7 illustrates the acquired range extension for different driving patterns. The release period of the battery is affected by the measure of tension on the battery, which expands the measure of the smart regulator's impact on the battery and also expands the EV range.

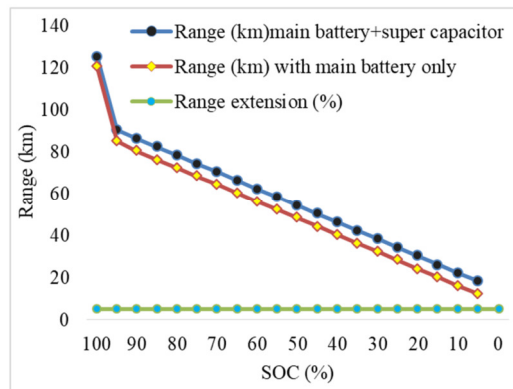


Fig. 5. SOC vs range and range extension.

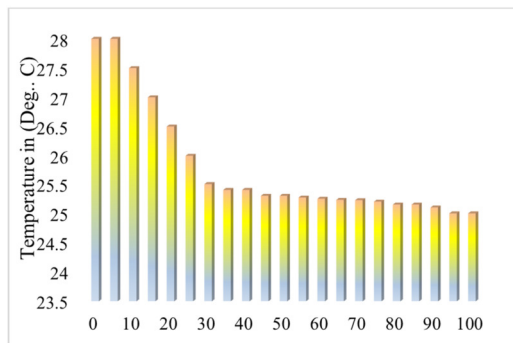


Fig. 6. Temperature variation graph.

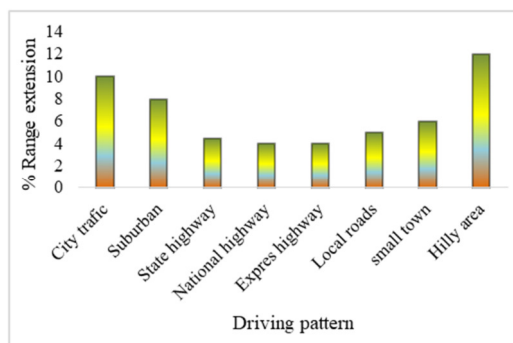


Fig. 7. Range Extension vs driving pattern.

TABLE VI. MODEL RANGE EXTENSION

Running time without controller (hr)	Running time with controller (hr)	Extension (%)
1	1.03	3

TABLE VII. RANGE EXTENSION COMPARISON

% Range extension in hardware model	% Range extension in simulation model	% Extension variation
3	4.13	1.13

VIII. CONCLUSIONS

The proposed methodology generated satisfactory results when compared to the ones acquired by an electric vehicle without following it. Tables VI and VII illustrate the simulation and hardware model comparison of range expansions in the proposed controller for electric vehicles. In the simulation model, when the discharge duration was prolonged by 4.13%, the range extension was raised by 4.13%. However, in the hardware model, the range extension was increased by 3%. Future work will include the integration of intelligent controllers on road vehicles.

REFERENCES

- [1] R. R. Kumar and K. Alok, "Adoption of electric vehicle: A literature review and prospects for sustainability," *Journal of Cleaner Production*, vol. 253, Apr. 2020, Art. no. 119911, <https://doi.org/10.1016/j.jclepro.2019.119911>.
- [2] M. E. Bendib and A. Mekias, "Solar Panel and Wireless Power Transmission System as a Smart Grid for Electric Vehicles," *Engineering, Technology & Applied Science Research*, vol. 10, no. 3, pp. 5683–5688, Jun. 2020, <https://doi.org/10.48084/etasr.3473>.
- [3] F. Orecchini, A. Santiangeli, and F. Zuccari, "Hybrid-electric system truth test: Energy analysis of Toyota Prius IV in real urban drive conditions," *Sustainable Energy Technologies and Assessments*, vol. 37, Feb. 2020, Art. no. 100573, <https://doi.org/10.1016/j.seta.2019.100573>.
- [4] S. Rhode, S. Van Vaerenbergh, and M. Pfriem, "Power prediction for electric vehicles using online machine learning," *Engineering Applications of Artificial Intelligence*, vol. 87, Jan. 2020, Art. no. 103278, <https://doi.org/10.1016/j.engappai.2019.103278>.
- [5] Y. Tian, H. Yang, W. Mo, S. Zhou, N. Zhang, and P. D. Walker, "Optimal coordinating gearshift control of a two-speed transmission for battery electric vehicles," *Mechanical Systems and Signal Processing*, vol. 136, Feb. 2020, Art. no. 106521, <https://doi.org/10.1016/j.ymssp.2019.106521>.
- [6] J. Zhu and I. Tasic, "Safety analysis of freeway on-ramp merging with the presence of autonomous vehicles," *Accident Analysis & Prevention*, vol. 152, Mar. 2021, Art. no. 105966, <https://doi.org/10.1016/j.aap.2020.105966>.
- [7] J. Zhao, T. Qu, and F. Xu, "A Deep Reinforcement Learning Approach for Autonomous Highway Driving," *IFAC-PapersOnLine*, vol. 53, no. 5, pp. 542–546, Jan. 2020, <https://doi.org/10.1016/j.ifacol.2021.04.142>.
- [8] C. Lu, J. Dong, A. Houchin, and C. Liu, "Incorporating the standstill distance and time headway distributions into freeway car-following models and an application to estimating freeway travel time reliability," *Journal of Intelligent Transportation Systems*, vol. 25, no. 1, pp. 21–40, Jan. 2021, <https://doi.org/10.1080/15472450.2019.1683450>.
- [9] Q. Wang, X. Yang, Z. Huang, and Y. Yuan, "Multi-Vehicle Trajectory Design During Cooperative Adaptive Cruise Control Platoon Formation," *Transportation Research Record*, vol. 2674, no. 4, pp. 30–41, Apr. 2020, <https://doi.org/10.1177/0361198120913290>.
- [10] D.-Q. Nguyen, L. D. Hai, D. B. Minh, and V. D. Quoc, "Analysis of Electromagnetic Parameters of Hybrid Externally Excited Synchronous Motors for Electric Vehicle Applications," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10670–10674, Jun. 2023, <https://doi.org/10.48084/etasr.5824>.
- [11] H. Cheng, Z. Shang, Y. Guo, C. Kang, W. Zeng, "System Dynamics Analysis of Vehicle-Grid Interaction Process", *Power Grid Technologies*, vol. 45, pp. 4125–4133, 2021.
- [12] G. Yuan, W. Su "Research on Virtual Power Plants Participating in AGC Frequency Regulation Service Considering the Uncertainty of Electric Vehicles", *Power Grid Technologies*, vol. 44, pp. 2538–2548, 2020.
- [13] W. Shi, L. Lu, H. Gao, H. Li, Y. Liu, J. Liu "Economic Dispatch of Active Distribution Grids Considering Demand Response and Participation of Electric Vehicles" *Automation of Electric Power Systems*, vol. 44, pp. 41–51, 2020.

- [14] Y. Yang *et al.*, "Research on the energy management strategy of extended range electric vehicles based on a hybrid energy storage system," *Energy Reports*, vol. 8, pp. 6602–6623, Nov. 2022, <https://doi.org/10.1016/j.egyr.2022.05.013>.
- [15] N. T. Diep and N. K. Trung, "Transmitting Side Power Control for Dynamic Wireless Charging System of Electric Vehicles," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 9042–9047, Aug. 2022, <https://doi.org/10.48084/etasr.4988>.
- [16] C. Heilmann and G. Friedl, "Factors influencing the economic success of grid-to-vehicle and vehicle-to-grid applications—A review and meta-analysis," *Renewable and Sustainable Energy Reviews*, vol. 145, Jul. 2021, Art. no. 111115, <https://doi.org/10.1016/j.rser.2021.111115>.
- [17] M. Inci, M. M. Savrun, and O. Celik, "Integrating electric vehicles as virtual power plants: A comprehensive review on vehicle-to-grid (V2G) concepts, interface topologies, marketing and future prospects," *Journal of Energy Storage*, vol. 55, Nov. 2022, Art. no. 105579, <https://doi.org/10.1016/j.est.2022.105579>.
- [18] B. Bibak and H. Tekiner-Mogulkoc, "A comprehensive analysis of Vehicle to Grid (V2G) systems and scholarly literature on the application of such systems," *Renewable Energy Focus*, vol. 36, pp. 1–20, Mar. 2021, <https://doi.org/10.1016/j.ref.2020.10.001>.
- [19] S. R. Salkuti, "Energy Storage and Electric Vehicles: Technology, Operation, Challenges, and Cost-Benefit Analysis," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 4, pp. 40–45, 2021, <https://doi.org/10.14569/IJACSA.2021.0120406>.
- [20] *Development Trend and Policy System of V2G Interaction*. Energy Research Institute of National Development and Reform Commission. Natural Resources Defense Council: Beijing, China, 2021.
- [21] Y. Zhang, W. Han, C. Song, and S. Yang, "Joint planning and operation optimization of photovoltaic-storage- charging integrated station containing electric vehicles," *Energy Storage Science and Technology*, vol. 11, no. 5, pp. 1502–1511, May 2022, <https://doi.org/10.19799/j.cnki.2095-4239.2021.0481>.
- [22] A. Khadhraoui, T. Selmi, and A. Cherif, "Energy Management of a Hybrid Electric Vehicle," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 8916–8921, Aug. 2022, <https://doi.org/10.48084/etasr.5058>.
- [23] Y. Wei *et al.*, "A Comprehensive Study of Degradation Characteristics and Mechanisms of Commercial Li(NiMnCo)O₂ EV Batteries under Vehicle-To-Grid (V2G) Services," *Batteries*, vol. 8, no. 10, Oct. 2022, Art. no. 188, <https://doi.org/10.3390/batteries8100188>.
- [24] F. Ahmad, A. Iqbal, I. Ashraf, M. Marzband, and I. Khan, "Optimal location of electric vehicle charging station and its impact on distribution network: A review," *Energy Reports*, vol. 8, pp. 2314–2333, Nov. 2022, <https://doi.org/10.1016/j.egyr.2022.01.180>.
- [25] L. M. Caro, G. Ramos, K. Rauma, D. F. C. Rodriguez, D. M. Martinez, and C. Rehtanz, "State of Charge Influence on the Harmonic Distortion From Electric Vehicle Charging," *IEEE Transactions on Industry Applications*, vol. 57, no. 3, pp. 2077–2088, Jun. 2021, <https://doi.org/10.1109/TIA.2021.3057350>.
- [26] P. Pradhan, I. Ahmad, D. Habibi, G. Kothapalli, and M. A. S. Masoum, "Reducing the Impacts of Electric Vehicle Charging on Power Distribution Transformers," *IEEE Access*, vol. 8, pp. 210183–210193, 2020, <https://doi.org/10.1109/ACCESS.2020.3040056>.
- [27] C. Garcia Veloso, K. Rauma, J. F. Orjuela, and C. Rehtanz, "Real-time agent-based control of plug-in electric vehicles for voltage and thermal management of LV networks: formulation and HIL validation," *IET Generation, Transmission & Distribution*, vol. 14, no. 11, pp. 2169–2180, 2020, <https://doi.org/10.1049/giet-gtd.2018.6547>.
- [28] S. Farhani, E. M. Barhoumi, and F. Bacha, "Design and hardware investigation of a new configuration of an isolated DC-DC converter for fuel cell vehicle," *Ain Shams Engineering Journal*, vol. 12, no. 1, pp. 591–598, Mar. 2021, <https://doi.org/10.1016/j.asej.2020.07.014>.
- [29] F. Slah, A. Mansour, A. Abdelkarim, and F. Bacha, "Analysis and Design of an LC Parallel-Resonant DC–DC Converter for a Fuel Cell Used in an Electrical Vehicle," *Journal of Circuits, Systems and Computers*, vol. 27, no. 08, Jul. 2018, Art. no. 1850119, <https://doi.org/10.1142/S0218126618501190>.
- [30] S. Farhani, A. N'Diaye, A. Djerdir, and F. Bacha, "Design and practical study of three phase interleaved boost converter for fuel cell electric vehicle," *Journal of Power Sources*, vol. 479, Dec. 2020, Art. no. 228815, <https://doi.org/10.1016/j.jpowsour.2020.228815>.

Optimal Placement of Superconducting Magnetic Energy Storages in a Distribution Network with Embedded Wind Power Generation

Steven Foday Sesay

Department of Electrical Engineering, Pan African University Institute for Basic Science, Technology and Innovation, Kenya
foday.steven@students.jkuat.ac.ke (corresponding author)

Cyrus Wabuge Wekesa

Department of Electrical Engineering, Pan African University Institute for Basic Science, Technology and Innovation, Kenya | School of Engineering, University of Eldoret, Kenya
cwekesa@uoeld.ac.ke

Livingstone M. H. Ngoo

Department of Electrical Engineering, Pan African University Institute for Basic Science, Technology and Innovation, Kenya | Department of Electrical Engineering, Multimedia University, Kenya
livingngoo@gmail.com

Received: 15 December 2023 | Revised: 15 January 2024 | Accepted: 28 January

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6754>

ABSTRACT

The prevalence of distributed generation in most power grids can negatively affect their performance in terms of power loss, voltage deviation, and voltage stability. Superconducting Magnetic Energy Storages (SMESs) can help in addressing this problem as long as they are optimally placed in the distribution network. This paper presents a hybrid Grasshopper Optimization Algorithm and a Simulated Annealing (GOA-SA) method to determine the optimal placement of SMESs in a distribution network with an embedded wind power generation system. The optimization was formulated as a multi-objective problem to minimize active power losses, reactive power losses, and voltage deviation and maximize the voltage stability index. An IEEE 57-node distribution network was employed and simulations were performed using MATLAB R2020b. Based on simulations using 200 kW SMESs in discharge mode, the active power loss decreased by 82.57%, the reactive power loss decreased by 80.71%, the average voltage deviation index decreased by 66.91%, and the voltage stability index improved by 34.97%. In the charging operation mode, the active power loss increased by 24.86%, the reactive power loss increased by 8.21%, the average voltage deviation increased by 12.86%, and the voltage stability index increased by 12.79%. These results show that SMESs can improve the technical performance of a distribution network.

Keywords-SMES; WTG; hybrid GOA-SA; optimal placement

I. INTRODUCTION

Stressed distribution networks lead to poor power quality, frequent power outages, high energy costs, and decreased reliability [1]. The amount of energy used today has grown significantly, whereas, at the same time, expenditures in the infrastructure supporting the power system and the use of fossil fuels have decreased. Conventional distribution networks have a strong hierarchical structure. However, the growing need for energy, the fast pace of technological development, fuel crises, blackouts, financial incentives, and public awareness of

environmental issues are pushing toward distributed energy resources with high reliability, stability, quality, and security. Today, Distributed Energy Resources (DERs) are a common component of many utilities worldwide and are increasingly integrated into power distribution networks. In addition to being environmentally friendly as green energy, their installation is crucial for the distribution network in terms of reduced power losses and improved voltage profile [2]. As renewable energy sources, such as wind power, are weather-dependent and erratic, they can affect the reliability of

distribution networks. An augmentation in load demand in a distribution network can result in a rise in voltage drop, which can impact power loss, voltage profile, and system stability. Distribution networks are rapidly integrating Energy Storage Systems (ESS) to address the aforementioned issues and offer other advantages. The former consists of flywheel storage [3], fuel cell storage [4], Compressed Air Energy Storage (CAES) [5], Compressed Carbon Dioxide Energy Storage (CCES) [6], Battery Energy Storage System (BESS) [7], and SMES [8]. Despite being the industry standard, BESSs have serious drawbacks, including a very short lifespan, voltage and current restrictions, potential environmental problems, and a slow response time. SMESs have several advantages over other ESS, involving long lifespan, no restrictions on the number of charging and discharging cycles, no moving components, high power density (0.1-10 MW), and high storage efficiency (95-98%) [9]. The primary disadvantage of an SMES is its high costs. The use of High-Temperature Superconductors (HTS) can lessen this drawback. HTSs are cooled with liquid nitrogen at 77°K, as opposed to 4.2°K for Low-Temperature Superconductors (LTS). Compared to LTSs, HTSs provide greater system reliability and fewer refrigeration expenses [10]. A recent review examined the design and development of high-temperature SMESs for power applications. Numerous studies have been performed on SMES cooling methods and thermal management techniques to reduce the degradation of superconductor performance at high temperatures, based on thermal energy storage materials. A Direct Current (DC) flowing through a superconducting coil can store electrical energy in the magnetic field it creates. The SMES coil can absorb or release real and reactive power based on the power demand of the distribution network. This power can be released in seconds to several hours, given that it is stored as a circulating current [11]. SMES can be utilized as energy storage in the distribution network to achieve several objectives, entailing improving power quality and transient voltage dips, controlling reactive power flow and voltage, stabilizing wind generators, spinning reserve, as well as minimizing power and voltage fluctuations of wind generators.

Several studies have investigated the effectiveness of SMES in ameliorating transient stability [12-16] and addressing variations in the output power and voltage of Wind Turbine Generators (WTGs) [17-19]. However, there is a lack of research on the optimal placement of SMESs in a power grid. In [20], a genetic algorithm was applied to optimally position an SMES in an IEEE 14-bus transmission system to enhance voltage stability, solely focusing on the voltage stability index. In [21], a method was presented to determine the optimal location of a superconducting device and so minimize system losses. This approach employed loss sensitivity analysis in a typical power system, considering a daily load profile, and exclusively focused on power loss when determining the optimal location of the SMES. In [22], a novel approach was presented to optimally place WTGs and SMESs in a distribution network, using the Equilibrium Optimizer (EO) and the loss sensitivity factor in the optimization process. This innovative method was applied to an IEEE 33-node distribution network. In [23], a technique was proposed for optimal placement of SMES and superconducting fault current

limiters in interconnected microgrids. The objective function included the voltage deviation of the doubly fed induction generator, the power deviation at the point of common coupling, the fault current in transmission lines, the features of the superconducting fault current limiter, and SMES, which were optimized utilizing the PSO algorithm.

This study proposes an optimal placement method for SMESs to minimize active and reactive power losses and voltage deviation and improve the voltage stability index. The hybrid Grasshopper Optimization Algorithm and Simulated Annealing (GOA-SA) technique were used in an IEEE 57-node distribution network with an embedded wind power generation system. The main contributions of this study are:

- Investigating the optimal placement of the SMES in the IEEE 57-node distribution network with an embedded wind power generation system considering four objective functions: active and reactive power losses, average voltage deviation, and voltage stability index.
- Improving the performance of the distribution network in terms of decreasing active and reactive power losses, voltage deviation, and voltage stability.
- Proposing a novel hybrid optimization algorithm and comparing it with its derivatives to solve optimization problems.
- Considering a distribution network with an embedded wind power generation system, representing a real-life scenario.

II. MATERIALS AND METHODS

A. Network Component Modeling

The model consists of the IEEE 57-node distribution network, SMES, and a wind power generation system.

1) IEEE 57-Node Distribution Network and Study Area

Figure 1 shows the single-line diagram of the IEEE 57-node distribution [24], with system voltages of 12.66 kV and total active and reactive power loads of 2694.600 kVA_r and 3802.190 kW, respectively. This system includes one slack bus, 57 buses, and 56 branches and load buses, correspondingly. The network is considered to be a residential/commercial network with 85% of the load being commercial loads.

2) SMES

From the power system point of view, an SMES can be seen as a generator, load, or storage system. SMES can be studied as a generator, constant charging equipment, a load, constant discharging equipment, and a storage device that can discharge and charge depending on the system's characteristics. This study examines the SMES as a generator and a load. Table I illustrates the generator (discharging) and load (charging) characteristics of the SMES, and its kW ratings are used to determine their optimal position [25-26].

3) Wind Power Generation System

The wind power generation system is modeled as a Type 3 Distributed Generator (DG) to incorporate its active and

reactive power consumption and injection capabilities [27]. The wind power generation system, known as WTG, has a penetration level of 30% [28]. Although there are several ways to calculate wind power penetration levels, the percentage of wind turbine penetration could be obtained as the ratio of [29]:

- The total production of the wind power generation systems to the total generation.
- The peak wind power generation capacity to the loads' peak apparent power.
- The wind-rated power to the loads' active power demand.

The penetration percentage of wind power was calculated utilizing the last method to determine the total wind-rated power needed for 30% penetration or 1125 kW. A Newton-Raphson load flow simulation was performed in MATLAB/PSAT to identify the chosen location of the WTG. Node 46 was selected as the location of the WTG in the distribution network.

TABLE I. DISCHARGING AND CHARGING CASES

Cases	Mode of Operation	Quantity
Case I	Discharge 100 kW power	3
Case II	Discharge 200 kW power	3
Case III	Charge 100 kW power	3
Case IV	Charge 200 kW power	3

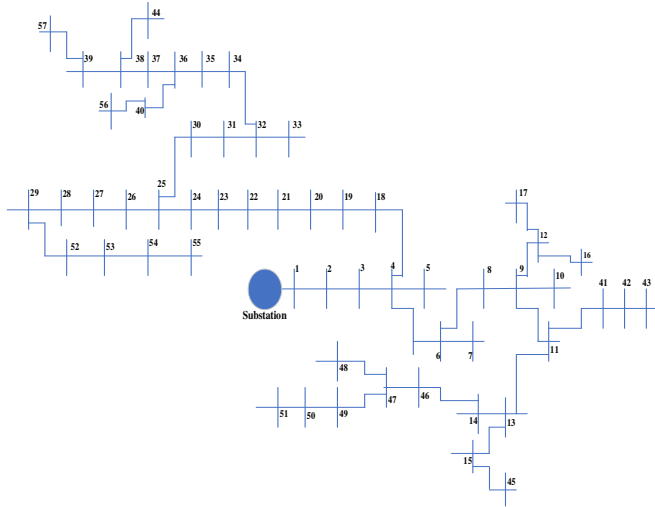


Fig. 1. Single line diagram for the IEEE 57-node distribution network.

B. Problem Formulation

The objective of this optimization problem is to minimize the active and reactive power losses of the network, the average voltage deviation index, and the voltage stability index when the SMESs are placed in the distribution network under discharging and charging conditions. Radial distribution networks have a large R/X ratio, and because of this, basic load flow tools such as Newton-Raphson or fast decoupled approaches do not provide accurate results. An efficient load flow method is presented based on the forward-backward method to solve the power flows of a radial distribution network [30-31].

1) Objective Function

a) Active and Reactive Power Loss Minimization

The active and reactive power losses in $P_{loss}(a, a+1)$ and $Q_{loss}(a, a+1)$ in a branch $a, a+1$ of the network is given by:

$$P_{loss(a,a+1)} = \left(\frac{P_{a+1}^2 + Q_{a+1}^2}{|V_{a+1}|^2} \right) * R_{br} \quad (1)$$

$$Q_{loss(a,a+1)} = \left(\frac{P_{a+1}^2 + Q_{a+1}^2}{|V_{a+1}|^2} \right) * X_{br} \quad (2)$$

where $P(a+1)$ and $Q(a+1)$ are the receiving end active and reactive powers, respectively, $V(i+1)$ is the receiving end voltage, R_{br} is the branch resistance, and X_{br} is the branch reactance. Therefore, the total power loss minimization function is given by:

$$f_1(j) = \min \sum_{b=1}^{br} [P_{loss}(b) + Q_{loss}(b)] \quad (3)$$

b) Minimization of Average Voltage Deviation Index (AVDI)

AVDI is the voltage deviation for 1.0 p.u., which is the reference voltage, and is defined in terms of the voltage magnitudes at all the nodes, given by:

$$f_2(j) = \frac{1}{N_i} \sum_{a=1}^{N_a} |1 - v_a|^2 \quad (4)$$

where N_a is the number of nodes in the network and v_a is the voltage at node a .

c) Maximization of the Voltage Stability Index (VSI)

At a receiving node b , VSI is given by:

$$f_3(j) = \frac{|V_b|^4 - 4(P_b x_{ab} + Q_b r_{ab})^2 - 4(P_b r_{ab} + Q_b x_{ab})|V_b|^2}{|V_b|^4} \quad (5)$$

where v_b is the voltage at node b , P_b is the active power demand at node b , Q_b is the reactive power demand at node b , r_{ab} is the resistance branch $a-b$, and x_{ab} is the reactance of branch $a-b$. Therefore, converting (1) into a minimization function and combining it with the former two equations gives the following multi-objective function:

$$F(j) = \min \left\{ w_1 f_1 + w_2 f_2 + w_3 \frac{1}{f_3(j)} \right\} \quad (6)$$

where w_1 , w_2 , and w_3 are weights assigned to the individual objective functions.

2) Constraints

The multi-objective function for this optimization problem is subject to the following constraints:

a) Power Balance (Equality Constraints)

The algebraic sum of all input and output active and reactive power flows through the system should be equal:

$$P_{sub} + \sum_{z=1}^{N_{WTG}} P_{WTG} + \sum_{z=1}^{N_s} P_{SMES} - \sum_{z=1}^{N_{br}} P_{loss} - \sum_{z=1}^{N_{al}} P_L = 0 \quad (7)$$

$$Q_{sub} + \sum_{z=1}^{N_{WTG}} Q_{WTG} + \sum_{z=1}^{N_s} Q_{SMES} - \sum_{z=1}^{N_{al}} Q_{loss} - \sum_{z=1}^{N_{al}} Q_L = 0 \quad (8)$$

where P_{sub} and Q_{sub} represents substation active and reactive power, respectively, P_L and Q_L are active and reactive load demands, P_{loss} and Q_{loss} are active and reactive power losses in the z^{th} branch, P_{SMES} and Q_{SMES} are the active and reactive powers of the SMES, P_{WTG} and Q_{WTG} are the active and reactive powers of WTG [27].

b) Inequality Constraints

The voltage magnitude at each bus must be kept within an acceptable range at all times:

$$V_i^{min} \leq V_i \leq V_i^{max} \quad (9)$$

where V_{min} and V_{max} are the minimum and the maximum voltage limits. The current magnitude of each line I_r must remain within acceptable operating limits to avoid any excessive thermal stress of the line:

$$I_r \leq I_r^{max} \quad (10)$$

The maximum-minimum power limit constraint for SMES is:

$$P_s^{min} \leq P_s \leq P_s^{max} \quad (11)$$

and $P_s^{min} = P_s = P_s^{max}$ as the power rating of each SMES is fixed.

3) Grasshopper Optimization Algorithm (GOA)

GOA mimics the swarming behavior of grasshoppers. Each grasshopper in the swarm has its position that corresponds to a possible solution to the optimization problem [32]. X_i represents the position of the i^{th} grasshopper as:

$$X_i = S_i + G_i + A_i \quad (12)$$

where S_i is the social interaction, G_i is the force of gravity on the i^{th} grasshopper, and A_i is the wind advection. Social interaction is the dominant part all coming from grasshoppers themselves defined in:

$$S_i = \sum_{j=1}^N s(d_{ij}) \widehat{d}_{ij} \quad (13)$$

where N denotes the number of grasshoppers, $d_{ij} = |X_j - X_i|$ defines the Euclidean distance between the i^{th} and the j^{th} grasshopper, $\widehat{d}_{ij} = \frac{X_j - X_i}{d_{ij}}$ is a unit vector from the i^{th} to the j^{th} grasshopper, and s is a function to define the strength of social forces represented by:

$$s(r) = f \exp \frac{-r}{l} - \exp^{-r} \quad (14)$$

where f indicates the intensity of attraction, r is the force of repulsion, and l is the attractive length scale that has a value in $[0, 4]$ and controls the attraction or repulsion between grasshoppers. The area where there is no attraction or repulsion is called a comfort area. The comfort area exists at an exact distance of 2.079. The distance should be normalized to the interval $[1, 4]$, as the s function cannot handle strong forces with large distances. The G component has two parts, as g is the gravitational constant, and \widehat{e}_g shows a unity vector towards the center of the earth. The mathematical definition is given by:

$$G_i = -g \widehat{e}_g \quad (15)$$

The wind advection A_i is given by:

$$A_i = u \widehat{e}_w \quad (16)$$

where u represents the drift constant and \widehat{e}_w is a unit vector in the wind direction. Using components, (1) can be written as:

$$X_i = \sum_{j=1}^N s(|x_j - x_i|) \frac{x_j - x_i}{d_{ij}} - g \widehat{e}_g + u \widehat{e}_w \quad (17)$$

An improved version of this equation can be:

$$X_i^d = c \left(\sum_{j=1}^N c \frac{u_{bd} - l_{bd}}{2} s(|x_j^d - x_i^d|) \frac{x_j^d - x_i^d}{d_{ij}} \right) + \widehat{T}_d \quad (18)$$

where u_{bd} is the upper bound, l_{bd} is the lower bound in the d^{th} dimension and \widehat{T}_d is the value of the d^{th} dimension in the target (best solution found so far). The G component is ignored assuming no gravitational force and wind direction is always towards a target. The decreasing coefficient c is used twice in (18) for controlling forces between grasshoppers and is updated by (19). The outer c maintains the balance between exploration and exploitation, while the inner c reduces repulsion/attraction forces between grasshoppers proportional to the number of iterations.

$$c = c_{max} - l \frac{c_{max} - c_{min}}{L} \quad (19)$$

where $c_{max} = 1$ is the maximum value, $c_{min} = 0.00001$ is the minimum value, l indicates the current iteration, and L is the maximum number of iterations. GOA is extended by MOGOA to address multi-objective optimization problems. The target selection is based on crowding distance, similar to one in MOPSO [33] using:

$$P_i = \frac{1}{M_i} \quad (20)$$

where P_i is the probability of choosing the target from the archive and M_i is the number of solutions in the neighborhood of the i^{th} solution. Later this probability helps find the target using roulette wheel selection.

4) Simulated Annealing (SA)

SA is a stochastic search technique that originated from the principles of Monte Carlo simulation. This algorithm demonstrates the capability to effectively address complex combinatorial optimization problems. The annealing process involves simulating the thermal motion of atoms in the presence of a heat bath, where the temperature gradually decreases from a higher value to a lower one [34]. SA can avoid local optima by adjusting the temperature and modifying the solution according to a probability function:

$$P(\Delta E) = e^{\frac{-\Delta E}{TK_B}} \quad (21)$$

where K_B is Boltzmann's constant, T is the current temperature and ΔE is the change in the energies of atoms.

5) Hybrid Multiobjective GOA and SA

A new hybrid model of MOGOA was introduced, incorporating SA [33]. This approach involves mapping the position of the new grasshopper to the current optimal position within a symmetrical interval, which is determined by the product of the current temperature and a random number

mapped to the dimensional space. By applying SA, the algorithm can randomly alter the current value of the control parameter c . This adaptation helps improve the search process, leading to the discovery of high-quality solutions, as shown in:

$$C_{new} = C_{old} * (1 + ns) * e^{-ns \cdot \frac{1}{N}} \quad (22)$$

where C_{new} is the new perturb c , C_{old} is its value in the previous iteration, constant in the first iteration but later updated in every iteration, ns is the number of steps in SA, and N is the number of grasshoppers in the swarm. In the annealing process, the temperature is adjusted by:

$$T_{new} = T_{old} * \alpha \quad (23)$$

where α is the cooling coefficient which decreases temperature in each iteration. SA was used to change the value of inertia weight. When the fitness of the population increases, the new value of c is accepted, otherwise, the probability is calculated by applying the Gaussian probability function:

$$G(t) = \min \left(1.0, e^{-\left(\frac{fitness_{new} - fitness_{old}}{K_B T} \right)} \right) \quad (24)$$

where $fitness_{new}$ is the fitness after obtaining the new value of c using (12), $fitness_{old}$ is the fitness in the previous iteration, T is the annealing temperature, and K_B is Boltzmann's constant. The following equation changes c using $G(t)$ and the next iteration starts:

$$C_{new} = C_{old} * G(t) \quad (25)$$

The updated values of c obtained through the SA process are used in MOGOA to adjust the positions of grasshoppers. Through the optimization process, the SA search component aids the hybrid MOGOA in escaping local optima and reaching global solutions. Figure 2 displays a flowchart of the steps of the hybrid MOGOA and SA. These steps are summarized as follows:

- Step 1: Input the network line and bus data.
- Step 2: Run load flow calculation and record the results (power losses, node voltage, AVDI, VSI).
- Step 3: Place the WTG systems at the specified load, run load flow calculation, and record the results (power losses, node voltage, AVDI, VSI)
- Step 4: Initialize the MOGOA-SA parameters: Number of grasshoppers, maximum number of iteration, minimum value of c , maximum value of c , lower bound, upper bound, dimension, change in the energies of the atoms, number of steps in SA, cooling coefficient, and annealing temperature, as observed in Table II.
- Step 5: Calculate the fitness function of each search individual. The fitness function is evaluated by (6) for each individual.
- Step 6: Evaluate the population fitness using (21). The selection operator is then employed to choose parents for reproduction. With the crossover and mutation operators, new offspring are generated. The Pareto dominance

operator is utilized to identify non-dominated solutions, while the archive operator stores these Pareto-optimal solutions. The crowded distance operator is employed to maintain diversity within the population. The roulette wheel selection operator is engaged to select a target from the archive.

- Step 7: update coefficient c using (22) to achieve a balance between exploration and exploitation. Also, update the position of a current individual (i.e. the SMESs) using (25).
- Step 8: Output load flow results (power losses, node voltage, AVDI, VSI) and the optimal locations for the SMESs by checking the constraints and storing the solution according to the best fit. Then, update the best search individual and repeat steps 5, 6, and 7 until reaching the maximum iteration.

TABLE II. HYBRID MOGOA-SA PARAMETER VALUES

Parameters	Symbol	Values
Number of grasshoppers	N	10
Maximum number of Iteration	L	20
Minimum value of c	C_{min}	0.00004
Maximum value of c	C_{max}	0.1
Lower bound	l_{bd}	-7
Upper bound	u_{bd}	7
Dimension	d	2
Change in the energies of the atoms	ΔE	1.67
Number of steps in SA	ns	5
Cooling coefficient	α	0.95
Annealing temperature	T	200

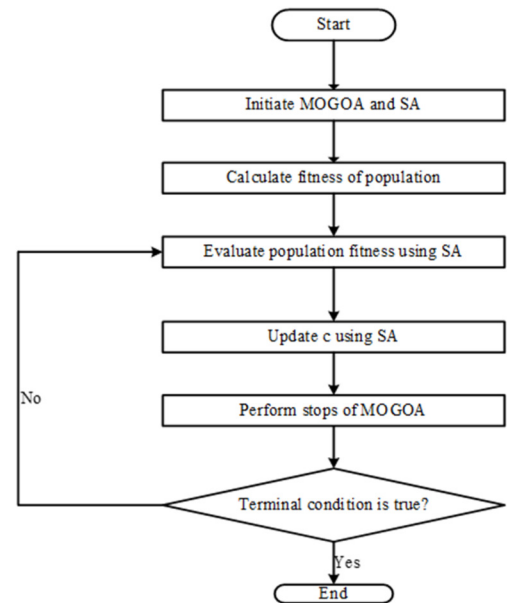


Fig. 2. Flowchart for hybrid MOGOA and SA.

III. RESULTS AND DISCUSSION

A. Optimal SMES Placement

Table III exhibits the optimal placement for the SMESs determined by the hybrid GOA-SA algorithm.

TABLE III. SUMMARY RESULTS FOR OPTIMAL PLACEMENT OF SMES

-	Base case	Case I	Case II	Case III	Case IV
Placement	-	47,22,8	14,10,25	47,22,8	14,10,25

TABLE IV. VOLTAGE MAGNITUDE FOR THE CASES

Node No.	Base Case	WTG	Case I	Case II	Case III	Case IV
1	0.999	0.999	0.999	1.020	0.961	0.929
2	0.999	0.999	0.999	1.017	0.961	0.928
3	0.999	0.999	0.999	1.013	0.961	0.928
4	0.999	0.999	0.999	1.002	0.961	0.928
5	0.999	0.999	0.999	1.001	0.961	0.928
6	0.999	0.997	0.999	1.000	0.961	0.928
7	0.998	0.989	0.989	1.000	0.960	0.928
8	0.998	0.976	0.989	1.000	0.960	0.928
9	0.998	0.975	0.989	1.000	0.960	0.913
10	0.999	0.973	0.990	1.000	0.961	0.913
11	0.995	0.972	0.986	1.000	0.957	0.913
12	0.993	0.970	0.984	1.000	0.955	0.913
13	0.984	0.964	0.975	1.000	0.946	0.910
14	0.986	0.962	0.977	1.000	0.948	0.901
15	0.987	0.961	0.978	1.000	0.949	0.901
16	0.986	0.960	0.977	1.000	0.948	0.901
17	0.985	0.957	0.976	1.000	0.947	0.900
18	0.984	0.957	0.975	1.000	0.946	0.900
19	0.983	0.957	0.974	1.000	0.945	0.900
20	0.982	0.956	0.973	1.000	0.944	0.900
21	0.981	0.956	0.972	1.000	0.943	0.900
22	0.985	0.956	0.976	1.000	0.947	0.900
23	0.982	0.956	0.973	1.000	0.944	0.909
24	0.980	0.956	0.971	0.998	0.942	0.909
25	0.981	0.956	0.972	0.997	0.943	0.909
26	0.979	0.955	0.970	0.989	0.941	0.920
27	0.978	0.955	0.969	1.000	0.940	0.920
28	0.976	0.955	0.967	1.000	0.938	0.923
29	1.000	1.000	0.991	1.000	0.962	0.923
30	0.999	0.999	0.999	1.000	0.961	0.923
31	0.999	0.999	0.999	1.000	0.961	0.923
32	0.999	0.999	0.999	1.000	0.961	0.923
33	0.999	0.999	0.999	1.000	0.961	0.923
34	0.999	0.999	0.999	1.000	0.961	0.923
35	0.999	0.999	0.999	1.000	0.961	0.923
36	0.999	0.999	0.999	1.000	0.961	0.923
37	0.999	0.999	0.999	1.000	0.961	0.923
38	0.998	0.999	0.989	1.000	0.960	0.923
39	0.998	0.999	0.989	1.000	0.960	0.923
40	0.998	0.999	0.989	1.000	0.960	0.923
41	0.996	0.998	0.987	1.000	0.958	0.923
42	0.995	0.996	0.986	1.000	0.957	0.923
43	0.995	0.997	0.986	1.000	0.957	0.923
44	0.993	0.997	0.984	1.000	0.955	0.923
45	0.988	0.996	0.979	1.000	0.950	0.923
46	0.987	0.996	0.978	1.000	0.949	0.923
47	0.994	0.994	0.985	1.000	0.956	0.923
48	0.994	0.994	0.985	1.000	0.956	0.923
49	0.993	0.993	0.984	1.000	0.955	0.921
50	0.993	0.993	0.984	1.000	0.955	0.921
51	0.992	0.987	0.983	1.000	0.954	0.921
52	0.991	0.975	0.982	1.000	0.953	0.920
53	0.990	0.971	0.981	1.000	0.952	0.918
54	0.987	0.967	0.978	1.000	0.949	0.916
55	0.984	0.962	0.975	1.000	0.946	0.914
56	0.985	0.940	0.976	1.000	0.947	0.909
57	0.984	0.930	0.975	1.000	0.946	0.901

It was assumed that the optimal placement obtained for the discharge operations of SMESs is the same as for the charging operation. Cases I and II have moderately dispersed SMESs. The reason behind this could be that, in the discharging operation, an effective minimization of active power loss, reactive power loss, AVDI, and maximum VSI might occur. This can cause a minimal deterioration in technical performance, which is expected in the charging mode of operation of the SMESs.

1) Network Voltage Profile

The network voltage magnitude for all four simulation cases is shown in Table IV. Introducing the WTG at a 30% penetration level into the distribution network leads to deterioration of the voltage profile, with a minimum voltage of 0.930 p.u. at node 57. When the SMESs are placed in the network in the discharge mode of operations, improvement is noticed in cases I and II. This improvement in the network voltage magnitude is a result of the SMESs serving as generators, thereby counteracting the drop in the voltage magnitudes. Maximum voltage magnitude of 1.020 p.u. at node 1 and minimum voltage magnitude of 0.989 p.u. at node 26 are obtained. In the charging mode of operation for case II, it was assumed that the SMESs are in the optimal location, as in the discharge case. Cases III and IV have a minimum voltage magnitude of 0.901 p.u. at node 57. It was observed that the voltage magnitude of the nodes in the network drops as the sizes of the SMESs increase from 100kW to 200kW. Using the hybrid GOA-SA algorithm, we can say that when SMESs are placed optimally, there is significant improvement in the voltage magnitude during the discharging modes of operation of the SMESs.

2) Average Voltage Deviation Index (AVDI)

The VDI of a node can be defined as the difference between the actual voltage value seen at the node and the reference voltage, which is often represented as 1 p.u. The average VDI value across all network nodes is used to compute AVDI. A decrease in value corresponds to an increase in the voltage stability of the network. Table V shows that the base case's AVDI increased by 30.08% as an outcome of WTG being integrated into the network, since it rose from 0.0186 in the base case to 0.0266 when the WTG was embedded at node 46. When SMESs are optimally placed using the proposed GOA-SA method and operate in the discharge mode, the AVDI for Case I was 0.0132, reduced by 50.86%. In Case II, the AVDI was 0.0088, reduced by 66.91%. In the charging mode, the AVDI for Case III was 0.0278, and for Case IV was 0.0289, increasing by 9.02 and 12.86%, respectively.

TABLE V. AVERAGE VOLTAGE DEVIATION INDEX

Scenario	Case I	Case II	Case III	Case IV
Base Case	0.0186	0.0186	0.0186	0.0186
WTG	0.0266	0.0266	0.0266	0.0266
WTG+SMES	0.0132	0.011	0.0278	0.0289

3) Voltage Stability Index (VSI)

VSI measures network stability. The network's voltage and current magnitudes drive the VSI to estimate the distance between the operating point of the current and the collapse

point of the voltage [35]. As it decreases, the network's vulnerability to voltage instability increases. A greater VSI value raises network stability, unlike VDI. Table VI portrays the minimum VSIs obtained from the simulation.

TABLE VI. MINIMUM VOLTAGE STABILITY INDEX

Scenario	Case I	Case II	Case III	Case IV
Base Case	0.739	0.739	0.739	0.739
WTG	0.735	0.735	0.735	0.735
WTG+SMES	0.931	0.988	0.728	0.671

VSI decreased by 14.14%, from 0.839 to 0.735, as a result of the WTG's integration into the network. During the discharge mode of the SMESs, for Case I, the minimum VSI was 0.931, increased by 26.67%. In Case II, the minimum VSI was 0.988, increased by 34.97%. In the charging mode, for Case III, the minimum VSI was 0.702, reduced by 4.49%. In Case IV, the minimum VSI was 0.641, reduced by 12.79%. The reduction in the minimum VSI of the network occurs as the size of SMESs increases under the charging mode of operation. The placement of SMESs using the GOA-SA method improved the net VSI of the network.

4) Active and Reactive Power Losses

In each simulation case, the integration of the WTG results led in a significant reduction in the overall active power loss, as it decreased from 158.06 to 56.11 kW, reducing the active power loss by 64.43% for the four cases, as illustrated in Table VII. In the discharging mode of operation of SMESs, for Case I, the active power loss decreased by 76.03%, from 56.11 to 13.45 kW. In Case II, the active power loss decreased by 82.57%, from 56.11 to 9.78 kW. In the charging mode of operation for Case III, the active power loss increased by 14.86%, from 56.11 to 64.45 kW. In Case IV, the active power loss increased by 24.86%, from 56.11 to 70.06 kW.

Table VIII indicates that in the discharging mode of SMESs, the reactive power loss increased by 24.88%, from 99.86 to 75.01 kVar, when the WTG was incorporated into the network. This is because WTG can generate reactive power as well as active power. When SMESs are optimally placed in the

network for Case I, the reactive power loss decreased by 67.26%, from 75.01 to 24.56 kVar. In Case II, the reactive power loss decreased by 80.71%, to 14.56 kVar. In the charging mode of operation of SMESs, for Case III, the reactive power loss increased by 6.18%, from 75.01 to 58.78 kVar. In Case IV, reactive power loss increased by 8.21%, to 81.16 kVar. The active and reactive power losses under both discharging and charging modes of operation were less compared to the base case.

TABLE VII. TOTAL ACTIVE POWER LOSS

Cases	Base Case (kW)	WTG (kW)	WTG+SMES (kW)
Case I	158.654	56.11	13.45
Case II	158.654	56.11	9.78
Case III	158.654	56.11	64.45
Case IV	158.654	56.11	70.06

TABLE VIII. TOTAL REACTIVE POWER LOSS

Cases	Base Case (kVar)	WTG (kVar)	WTG +SMES (kVar)
Case I	99.86	75.01	24.56
Case II	99.86	75.01	14.56
Case III	99.86	75.01	58.78
Case IV	99.86	75.01	81.16

B. Validation of the Proposed Hybrid GOA-SA

Tables IX and X show the comparison of the proposed hybrid GOA-SA algorithm for the optimal placement of SMESs in a distribution network with embedded wind power generation systems with GOA and SA independently for the same objective. The proposed method for placing SMESs leads to reduced active power losses, reactive power losses, AVDI, and an improved VSI. In contrast to using the algorithms independently, their hybrid usage results in lower active and reactive power losses, reduced voltage deviation, and improved VSI. These findings demonstrate the effectiveness of the proposed hybrid GOA-SA in identifying the optimal nodes for SMESs. Furthermore, this shows the efficiency of the proposed hybrid method in leveraging the strengths of one algorithm to address the limitations of the other.

TABLE IX. COMPARISON OF THE RESULTS OBTAINED FOR DISCHARGE OPERATIONS USING CASES I AND II

Algorithms	Case I						Case II					
	Opt. Location	Min Volt (p.u)	kW Loss	kVar Loss	Min VSI	AVDI	Opt. Location	Min Volt (p.u)	kW Loss	kVar Loss	Min VSI	AVDI
GOA- SA	47,22,8	0.9890	13.45	23.56	0.931	0.0132	14,10,25	0.9970	9.78	14.54	0.988	0.0088
GOA	47,23,9	0.9768	13.71	25.03	0.927	0.0143	7,20,31	0.9871	10.34	15.12	0.979	0.0098
SA	23,45,12	0.9618	14.32	24.56	0.928	0.0139	19,8,34	0.9862	11.73	14.83	0.968	0.0120

TABLE X. COMPARISON OF THE RESULTS OBTAINED FOR DISCHARGE OPERATIONS USING (CASES III AND IV)

Algorithms	Case III						Case IV					
	Opt. Location	Min Volt (p.u)	kW Loss	kVar Loss	Min VSI	AVDI	Opt. Location	Min Volt (p.u)	kW Loss	kVar Loss	Min VSI	AVDI
GOA-SA	47,22,8	0.9294	64.45	79.34	0.702	0.0279	14,10,25	0.9000	70.05	81.16	0.0641	0.0289
GOA	47,23,9	0.9133	66.23	80.58	0.701	0.0286	7,20,31	0.8951	72.19	83.26	0.0638	0.0293
SA	23,45,12	0.9324	67.59	82.09	0.689	0.0286	19,8,34	0.898	73.00	83.29	0.0636	0.0297

IV. CONCLUSION AND RECOMMENDATIONS

This study focused on the optimal placement of SMESs in a distribution network with embedded WTGs using a novel

hybrid GOA-SA optimization technique, considering four objective functions: active and reactive power losses, AVDI, and VSI. The objective was to optimally place SMESs to significantly improve the technical performance when

discharging, while considering its effect under the charging mode of operations using 100 and 200 kW SMESs. Integrating WTGs into the network significantly reduces the active and reactive power losses, but deteriorates the voltage profile in the network. The 200 kW SMESs in the discharge mode of operations performed the best in reducing active power losses, reactive power losses, and average voltage deviations and improved VSI. However, this resulted in a higher deterioration in the aforementioned technical parameters under the charging mode of operations. The novel hybrid GOA-SA method clearly shows its effectiveness in optimally placing SMESs in the network compared to GOA and SA independently. This concept can be utilized to determine the optimal placement of the SMESs for maximum savings in system operation costs.

ACKNOWLEDGEMENT

Steven Foday Sesay wishes to acknowledge the African Union Commission for his scholarship.

REFERENCES

- [1] M. J. Burke and J. C. Stephens, "Political power and renewable energy futures: A critical review," *Energy Research & Social Science*, vol. 35, pp. 78–93, Jan. 2018, <https://doi.org/10.1016/j.erss.2017.10.018>.
- [2] S. M. Said, A. Selim, and B. Hartmann, "Enhancement of voltage profile for unbalanced distribution system with wind energy and superconducting magnetic energy storage," in *2018 International Conference on Innovative Trends in Computer Engineering (ITCE)*, Aswan, Egypt, Feb. 2018, pp. 289–295, <https://doi.org/10.1109/ITCE.2018.8316640>.
- [3] A. H. Yakout, H. M. Hasanien, and H. Kotb, "Proton Exchange Membrane Fuel Cell Steady State Modeling Using Marine Predator Algorithm Optimizer," *Ain Shams Engineering Journal*, vol. 12, no. 4, pp. 3765–3774, Dec. 2021, <https://doi.org/10.1016/j.asej.2021.04.014>.
- [4] W. Buckles and W. V. Hassenzahl, "Superconducting magnetic energy storage," *IEEE Power Engineering Review*, vol. 20, no. 5, pp. 16–20, May 2000, <https://doi.org/10.1109/39.841345>.
- [5] X. Luo, J. Wang, M. Dooner, J. Clarke, and C. Krupke, "Overview of Current Development in Compressed Air Energy Storage Technology," *Energy Procedia*, vol. 62, pp. 603–611, Jan. 2014, <https://doi.org/10.1016/j.egypro.2014.12.423>.
- [6] F. Chaychizadeh, H. Dehghandorost, A. Aliabadi, and A. Taklifi, "Stochastic dynamic simulation of a novel hybrid thermal-compressed carbon dioxide energy storage system (T-CCES) integrated with a wind farm," *Energy Conversion and Management*, vol. 166, pp. 500–511, Jun. 2018, <https://doi.org/10.1016/j.enconman.2018.04.050>.
- [7] R. Sadiq, Z. Wang, C. y. Chung, C. Zhou, and C. Wang, "A review of STATCOM control for stability enhancement of power systems with wind/PV penetration: Existing research and future scope," *International Transactions on Electrical Energy Systems*, vol. 31, no. 11, 2021, Art. no. e13079, <https://doi.org/10.1002/2050-7038.13079>.
- [8] M. H. Qais, H. M. Hasanien, and S. Alghuwainem, "Output power smoothing of wind power plants using self-tuned controlled SMES units," *Electric Power Systems Research*, vol. 178, Jan. 2020, Art. no. 106056, <https://doi.org/10.1016/j.epsr.2019.106056>.
- [9] M. H. Ali, B. Wu, and R. A. Dougal, "An Overview of SMES Applications in Power and Energy Systems," *IEEE Transactions on Sustainable Energy*, vol. 1, no. 1, pp. 38–47, Mar. 2010, <https://doi.org/10.1109/TSTE.2010.2044901>.
- [10] E. A. Gouda, A. Abd-Alaziz, and M. El-Saadawi, "Design modeling, and control of multi-stage SMES integrated with PV system," *Journal of Energy Storage*, vol. 29, Jun. 2020, Art. no. 101399, <https://doi.org/10.1016/j.est.2020.101399>.
- [11] M. C. Argyrou, P. Christodoulides, and S. A. Kalogirou, "Energy storage for electricity generation and related processes: Technologies appraisal and grid scale applications," *Renewable and Sustainable Energy Reviews*, vol. 94, pp. 804–821, Oct. 2018, <https://doi.org/10.1016/j.rser.2018.06.044>.
- [12] H. M. Hasanien and S. M. Mueen, "Particle Swarm Optimization-based Superconducting Magnetic Energy Storage for Low-voltage Ride-through Capability Enhancement in Wind Energy Conversion System," *Electric Power Components and Systems*, vol. 43, no. 11, pp. 1278–1288, Jul. 2015, <https://doi.org/10.1080/15325008.2015.1027017>.
- [13] H. M. Hasanien, S. Q. Ali, and S. M. Mueen, "Wind generator stability enhancement by using an adaptive artificial neural network-controlled superconducting magnetic energy storage," in *2012 15th International Conference on Electrical Machines and Systems (ICEMS)*, Sapporo, Japan, Oct. 2012.
- [14] S. M. Said and B. Hartmann, "Alleviation of Extremely Power and Voltage Variations Caused by Wind Power and Load Demand Using SMES," *Periodica Polytechnica Electrical Engineering and Computer Science*, vol. 63, no. 3, pp. 134–143, Mar. 2019, <https://doi.org/10.3311/PPee.13718>.
- [15] M. H. Ali, T. Murata, and J. Tamura, "Transient Stability Enhancement by Fuzzy Logic-Controlled SMES Considering Coordination With Optimal Reclosing of Circuit Breakers," *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 631–640, May 2008, <https://doi.org/10.1109/TPWRS.2008.920045>.
- [16] R. Khanna, G. Singh, and T. K. Nagsarkar, "Artificial neural network based SMES unit for transient stability improvement," in *2010 Joint International Conference on Power Electronics, Drives and Energy Systems & 2010 Power India*, New Delhi, India, Sep. 2010, pp. 1–7, <https://doi.org/10.1109/PEDES.2010.5712476>.
- [17] W. Kreeumporn and I. Ngamroo, "Optimal Superconducting Coil Integrated Into PV Generators for Smoothing Power and Regulating Voltage in Distribution System With PHEVs," *IEEE Transactions on Applied Superconductivity*, vol. 26, no. 7, Jul. 2016, <https://doi.org/10.1109/TASC.2016.2591981>.
- [18] T. Chaiyatham and I. Ngamroo, "Optimal Fuzzy Gain Scheduling of PID Controller of Superconducting Magnetic Energy Storage for Power System Stabilization," *International Journal of Innovative Computing, Information and Control*, vol. 9, no. 2, pp. 651–666, Feb. 2013.
- [19] A. Boudia, S. Messalti, A. Harrag, and M. Boukhniher, "New hybrid photovoltaic system connected to superconducting magnetic energy storage controlled by PID-fuzzy controller," *Energy Conversion and Management*, vol. 244, Sep. 2021, Art. no. 114435, <https://doi.org/10.1016/j.enconman.2021.114435>.
- [20] X. Huang, G. Zhang, and L. Xiao, "Optimal Location of SMES for Improving Power System Voltage Stability," *IEEE Transactions on Applied Superconductivity*, vol. 20, no. 3, pp. 1316–1319, Apr. 2010, <https://doi.org/10.1109/TASC.2010.2044988>.
- [21] S. J. Lee, "Location of a Superconducting Device in a Power Grid for System Loss Minimization Using Loss Sensitivity," *IEEE Transactions on Applied Superconductivity*, vol. 17, no. 2, pp. 2351–2354, Jun. 2007, <https://doi.org/10.1109/TASC.2007.898435>.
- [22] M. Hashem, M. Abdel-Salam, M. Th. El-Mohandes, M. Nayel, and M. Ebeed, "Optimal Placement and Sizing of Wind Turbine Generators and Superconducting Magnetic Energy Storages in a Distribution System," *Journal of Energy Storage*, vol. 38, Jun. 2021, Art. no. 102497, <https://doi.org/10.1016/j.est.2021.102497>.
- [23] A. Komijani, M. Kheradmandi, and M. Sedighizadeh, "Optimal Allocation and Control of Superconducting Fault Current Limiter and Superconducting Magnetic Energy Storage in Mesh Microgrid Networks to Improve Fault Ride Through," *Journal of Operation and Automation in Power Engineering*, vol. 11, no. 1, pp. 22–32, Apr. 2023, <https://doi.org/10.22098/joape.2023.9577.1668>.
- [24] M. M. Ansari, C. Guo, M. S. Shaikh, N. Chopra, I. Haq, and L. Shen, "Planning for Distribution System with Grey Wolf Optimization Method," *Journal of Electrical Engineering & Technology*, vol. 15, no. 4, pp. 1485–1499, Jul. 2020, <https://doi.org/10.1007/s42835-020-00419-4>.
- [25] I. Hadjipaschalis, A. Poullikkas, and V. Efthimiou, "Overview of current and future energy storage technologies for electric power applications," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 6, pp. 1513–1522, Aug. 2009, <https://doi.org/10.1016/j.rser.2008.09.028>.

- [26] Y. Huang, Y. Ru, Y. Shen, and Z. Zeng, "Characteristics and Applications of Superconducting Magnetic Energy Storage," *Journal of Physics: Conference Series*, vol. 2108, no. 1, Aug. 2021, Art. no. 012038, <https://doi.org/10.1088/1742-6596/2108/1/012038>.
- [27] A. D. Petropoulos, N. C. Koutsoukis, E. S. Karapidakis, and P. S. Georgilakis, "Optimal Mix of Wind Generation and Energy Storage Systems in Power Distribution Networks," presented at the MedPower 2014, Athens, Greece, Nov. 2014, <https://doi.org/10.1049/cp.2014.1640>.
- [28] Y. Kassem, H. Gokcekus, H. Camur, and A. H. A. Abdelnaby, "Wind Power Generation Scenarios in Lebanon," *Engineering, Technology & Applied Science Research*, vol. 12, no. 6, pp. 9551–9559, Dec. 2022, <https://doi.org/10.48084/etasr.5258>.
- [29] S. Sulaeman, F. T. Alharbi, M. Benidris, and J. Mitra, "A new method to evaluate the optimal penetration level of wind power," in *2017 North American Power Symposium (NAPS)*, Morgantown, WV, USA, Sep. 2017, pp. 1–6, <https://doi.org/10.1109/NAPS.2017.8107332>.
- [30] V. K. B. Ponnamp and K. Swarnasri, "Multi-Objective Optimal Allocation of Electric Vehicle Charging Stations and Distributed Generators in Radial Distribution Systems using Metaheuristic Optimization Algorithms," *Engineering, Technology & Applied Science Research*, vol. 10, no. 3, pp. 5837–5844, Jun. 2020, <https://doi.org/10.48084/etasr.3517>.
- [31] T. N. Ton, T. T. Nguyen, A. V. Truong, and T. P. Vu, "Optimal Location and Size of Distributed Generators in an Electric Distribution System based on a Novel Metaheuristic Algorithm," *Engineering, Technology & Applied Science Research*, vol. 10, no. 1, pp. 5325–5329, Feb. 2020, <https://doi.org/10.48084/etasr.3372>.
- [32] L. Abualigah and A. Diabat, "A comprehensive survey of the Grasshopper optimization algorithm: results, variants, and applications," *Neural Computing and Applications*, vol. 32, no. 19, pp. 15533–15556, Oct. 2020, <https://doi.org/10.1007/s00521-020-04789-8>.
- [33] F. Sajjad *et al.*, "An efficient hybrid approach for optimization using simulated annealing and grasshopper algorithm for IoT applications," *Discover Internet of Things*, vol. 3, no. 1, Jul. 2023, Art. no. 7, <https://doi.org/10.1007/s43926-023-00036-3>.
- [34] M. O. Genc and N. Kaya, "Vibration Damping Optimization using Simulated Annealing Algorithm for Vehicle Powertrain System," *Engineering, Technology & Applied Science Research*, vol. 10, no. 1, pp. 5164–5167, Feb. 2020, <https://doi.org/10.48084/etasr.3242>.
- [35] M. S. S. Danish, T. Senju, S. M. S. Danish, N. R. Sabory, N. K., and P. Mandal, "A Recap of Voltage Stability Indices in the Past Three Decades," *Energies*, vol. 12, no. 8, Jan. 2019, Art. no. 1544, <https://doi.org/10.3390/en12081544>.

Efficiency Assessment of an Inverter based on Solar PV Energy in Baghdad City

Bilal A. Nasir

Northern Technical University, Iraq
bilalalnasir@ntu.edu.iq

Received: 22 January 2024 | Revised: 11 February 2024 | Accepted: 14 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6948>

ABSTRACT

The yearly energy yield of a Solar Photovoltaic (SPV) system is a rendition pointer utilized by the erector to determine the output energy generated by it. From the energy speculation, the payback period and the return on investment can be contemplated. The system energy yield formula consists of many parameters, the most important of which is the SPV inverter efficiency. The European and peak (maximum) efficiency factors from the inverter data sheet are typically utilized, but this utilization is unsound because the SPV does not always work at the peak of its effectiveness due to varying irradiance. The inverter's weighted efficiency is considered more sound as it deems the inverter output power peculiarities. The European weighted efficiency is the most widely accepted inverter efficiency determination. Since it is derived and documented on a rimmed European irradiance profile, it may not be appropriate for inverters constructed in different climatic conditions, especially in the equatorial and subtropical environmental regions. This work aims to formulate a fangled weighted efficiency equation for the inverter's work in the Iraqi environment (especially in Baghdad city as a case study) documented on the IEC 61683: 1999 Standard and Irradiance-Duration curve. The sophisticated formula is endorsed on experimental data from the field using an SMA-SB-4000-TL inverter. It was found that the speculated energy yield using the derived efficiency formula for the Baghdad environment closely matches the energy yield of an original 4.0 KW SPV inverter system with only 1% difference between the determined and acquired values. This means that the employment of the Baghdad weighted efficiency in place of the European or peak weighted efficiency will result in a sounder speculation of the system energy yield, return on investment, and payback duration of the SPV system project.

Keywords-solar PV system; PV inverter; weighted efficiency; energy yield; irradiance duration curve

I. INTRODUCTION

Before the implementation of an SPV system, the designer must estimate the suitability of the site position and deliver the energy yield (E_{sys}) report to the client. The report allows for payback duration and Return on Investment (RoI) speculation. The most important parameter in E_{sys} calculation is the SPV inverter efficiency. Many designers of SPV systems utilize the peak efficiency of the inverter from the data sheet to compute the energy yield from the system. The maximum efficiency (η_{max}) of the inverter is accomplishable under the Standard Test Conditions (STC), which are determined at 1000 W/m² irradiance and 25 °C, although the actual SPV system may rarely function at STC [1-2]. The inverter weighted efficiency estimates, such as the European (η_{EURO}) and California Energy Commission (η_{CEC}), can be utilized instead of the system's energy yield determination. The weighted efficiency determination is documented on the fact that the inverter does not always work at STC and is exposed to variations in irradiation and temperature. The efficiency of the inverter is affected by its own losses (due to the control circuits) and by losses deriving from the load variation. The weighted efficiency is considered more reliable for the inverter's efficiency estimations [3]. The weighted efficiency refers to a

certain climatic area and its utilization in varying weather conditions, which can produce unsound energy yield (E_{sys}) estimation. For example, an inverter function in Europe will not provide the same power output if it is constructed in Equatorial or Subtropical areas because of the two areas' difference in the irradiance profile and temperature. Up to date, most SPV system designers still deem η_{EURO} and η_{max} as the benchmark for efficiency determination. With the recent rapid growth of SPV systems in subtropical regions, such as Iraq, there is a need for a more solid energy yield speculation from the SPV systems.

The current work has two aims: to prove the assumption that if the SPV system is constructed in a subtropical area like Iraq, the employed η_{max} and η_{EURO} from the data sheet will not be sound and to establish a formula for inverter weighted efficiency for the Baghdad area based on the IEC 61683: 1999 Standard. To fulfill these targets, acquiring a one-year in-plane irradiance dataset from a weather station is crucial. The irradiance data are used as the input to the solar PV array system, which then feeds the inverter under test. The output of the inverter is linked to the AC grid, whereas the input DC power to the inverter and the output power from the inverter are categorized to rimmed actuation classes. From these data,

the weighted efficiency of the inverter type SMA-4000-SB-TL is determined and compared with the values given in the datasheet of the inverter. To assert the correct Baghdad weighted efficiency (η_{BAG}), the derived formula is applied to calculate the E_{sys} and the result is compared with that calculated by η_{EURO} and η_{max} . These calculations are then compared with the feasible measurement taken from a 4.0 Kw inverter solar PV system.

II. SOLAR PV INVERTER WEIGHTED EFFICIENCY

The solar PV inverter is the most important part of the SPV system due to the long living cycle of the system, i.e. over 25 years. Also, the inverter efficiency has affected the gross system performance. The energy yield of the system in kWh can be calculated as [1, 2]:

$$E_{sys} = P_{panels} \cdot PSH \cdot f_{mm} \cdot f_{temp} \cdot f_{dirt} \cdot \eta_{cable} \cdot \eta_{inv} \quad (1)$$

P_{panels} indicates the rated power of the system modules, PSH represents the maximum hours of irradiance during the designated month or year, f_{mm} is the mismatch factor of the manufacturer, f_{temp} is the wastage due to temperature, f_{dirt} is the wastage due to the accretion of dirt on the panel surface, and η_{cable} is the efficiency of the DC cable link between the panels and the inverter input side. The last term (η_{inv}) is the inverter efficiency. An inverter that is 1% less efficient is deduced to be 10% cheaper [4]. The inverter efficiency is the ratio of the output A.C. power to the input DC power. A Maximum Power Point Tracker (MPPT) is used to track the maximum point of the module's P-V characteristic curve of the system modules as input power to the inverter at any given temperature and irradiance level. In a simplified determination of E_{sys} , the designers assume η_{inv} is equal to η_{max} or η_{EURO} .

Maximum efficiency and Euro-weighted efficiency are usually given in the inverter data sheet, which is determined at STC. The usage of these data sheet efficiencies is not realistic because the inverter rarely functions at STC. The weighted efficiency takes into consideration the inverter's exposure to climate changes and the irradiance profile, since the irradiance is proportional to the inverter input power and the inverter efficiency differs according to the power classes of operation. So, the weighted efficiency is a sound characterization of the inverter efficiency due to its conformity to the working conditions.

III. FORMULATION OF WEIGHTED EFFICIENCY

Calculation of the inverter weighted efficiency is conducted according to in the IEC 61683:1999 Standard [3]. This standard recommends that the weighted efficiency formula should be derived with the use of the PV emulator, which emulates the output of the actual PV and inverter system. The emulator is fed by one-year irradiance data and the input and output power to the inverter are recorded. The main assumption is that the panel's temperature is kept constant at 25 °C. The irradiance is made variable to compare the inverter efficiency when tested in different environmental conditions [5].

The weighted average energy efficiency (η_{WT}) of the inverter is based on the irradiance-duration curve for an ON-grid connected inverter, which has no storage, and the reverse

power flow is accepted and allowed. On the contrary, the weighted efficiency determination documented on the load power-duration curve in case of stand-alone or off-grid operation of the inverter, which has a storage subsystem and some of the generated power is dissipated in the inverter and batteries.

The formula for the weighted energy efficiency (η_{WT}) of the on-grid operation of the inverter is determined as the summation of the inverter efficiency product of each irradiance operation class and the debating weightage factor (K_i). The weighted factor is related to the sum of time (T_i) that the inverter works at an irradiance class by the whole duration time (T_{WT}) of the inverter actuation. By using T_i as the duration of the DC input to the inverter power class (P_{dci}) or irradiance class (I_i), P_{outi} as the output power class, η_i as the inverter efficiency at class i , then the inverter weighted efficiency is defined as [3]:

$$\eta_{WT} = \frac{\sum P_{outi} \cdot T_i}{\sum P_{dci} \cdot T_i} \quad (2)$$

$$= K_1 \cdot \eta_1 + K_2 \cdot \eta_2 + K_3 \cdot \eta_3 + \dots + K_n \cdot \eta_n \quad (3)$$

$$K_i = \frac{P_{dci} \cdot T_i}{\sum P_{dci} \cdot T_i} \quad (4)$$

where K_i is the weigh coefficient for each DC input or irradiance level of operation i . If the irradiance-duration curve is given as shown in Figure C.1 in [3], the weighted efficiency of the on-grid inverter can be written as [3]:

$$\eta_{WT} = \sum_{i=1}^n \frac{i \cdot T_i}{T_{WT}} \cdot \eta_{i\%} \quad (5)$$

where n is the number of irradiance levels, $\eta_{i\%}$ is the efficiency of the inverter when its DC input power or irradiance is the $i\%$ of its rated value.

$$T_{WT} = 1 \cdot T_1 + 2 \cdot T_2 + 3 \cdot T_3 + \dots + n \cdot T_n \quad (6)$$

The weigh factors K_i in (4) can be written as:

$$K_i = \frac{i \cdot T_i}{T_{WT}} \quad (7)$$

IV. A REVIEW ON INVERTER WEIGHTED EFFICIENCY

Up to date, the most used weighted efficiency for energy yield determination is the η_{EURO} . It is formulated regarding inverters working in European areas that accept moderate irradiance penetration and it is given by [6]:

$$\eta_{EURO} = 0.03 * \eta_{5\%} + 0.06 * \eta_{10\%} + 0.13 * \eta_{20\%} + 0.1 * \eta_{30\%} + 0.48 * \eta_{50\%} + 0.2 * \eta_{100\%} \quad (8)$$

The η_{EURO} has six levels of operation: 5%, 10%, 20%, 30%, 50%, and 100%. The irradiance divided by the rated minus the irradiance forms the class of booting, and its values are shown in Table I. The highest weigh factor 0.48 occurs at the 50% operation class, which means that for half of the operation time, the inverter works at 50% of its rated power. The California Energy Commission inverter weighted efficiency (η_{CEC}) [7] has become a demotic pointer to specify the inverter rendition in regions with high irradiance levels [6].

Analogous to η_{EURO} , η_{CEC} has six classes of operation and each class imposes a different value of weigh factor:

$$\eta_{CEC} = 0.04.\eta_{10\%} + 0.05.\eta_{20\%} + 0.12.\eta_{30\%} + 0.21.\eta_{50\%} + 0.53.\eta_{75\%} + 0.05.\eta_{100\%} \quad (9)$$

TABLE I. RANGE OF BOOTING AND WEIGH FACTOR FOR EACH CLASS OF OPERATION

Booting class (%)	Booting range	Weigh factor
5	0-7.5	0.03
10	7.5-15	0.06
20	15-25	0.13
30	25-40	0.10
50	40-75	0.48
100	> 75	0.2

By comparing (8) and (9), the η_{CEC} is further depended on the efficiency of the irradiance levels that the inverter operates. Apart from η_{EURO} and η_{CEC} , there were other formulations of inverter weighted efficiency according to respective local environmental climate changes, such as Izmir efficiency (η_{IZM}) in Turkey [8-9], Chennai weighted efficiency (η_{CHE}) in India [10], Kanpur weighted efficiency (η_{KAN}) in India [11-13], Brazilian weighted efficiency [14-15], and Equatorial weighted efficiency (η_{EQUA}) [16].

Unluckily, the procedures of these weighted efficiency calculations do not comply with the IEC61683:1999 standard, in which the efficiency must be determined by the SPV generated power fed by the regional irradiance profile at constant temperature to the emulator. In reality, both irradiance and temperature vary and the irradiance was horizontally estimated, which did not obey with IEC61683 [3]. Also, due to climate variations, those studies are used as case studies for each related city.

V. METHODOLOGY OF INVERTER WEIGHTED EFFICIENCY FOR BAGHDAD CITY (η_{BAG})

Baghdad is a city located in the southwest of the Asia continent at 33°13' N latitude, 44°13' E longitude, and 34 m altitude. The climate of Baghdad is marked as BWH: subtropical desert with dry and hot arid climate [9]. We can suppose that the accurate performance of SPV inverter systems operating in subtropical climates differentiates from the performance of those tested in other climates. This work assumes that η_{max} and η_{EURO} given in the inverter data sheet in Table II are unusable for the energy yield estimation of inverters installed out of the European or Equatorial climates. Therefore, a fangled weighted efficiency formula for SPV inverter in subtropical (Baghdad) climate is derived. To obtain this objective:

- One-year irradiance data from a weather station located in Baghdad city [17-18] were considered.
- The SPV inverter type SMA-4000-SB-TL was run with the SPV array emulator using the combined irradiance-time profile data.
- The input and output power of the inverter were calculated.
- The Baghdad weighted efficiency (η_{BAG}) was calculated.

- The η_{BAG} was validated by using the measured data of one year for the system energy yield (E_{sys}) and the total measured energy with that calculated in (1) was compared using η_{BAG} , η_{max} , and η_{EURO} , to show the difference between these calculations and the actual measurements.

TABLE II. SPECIFICATIONS OF SMA-SB-4000-TL-21 SPV INVERTER

DC Input	
	SB 4000TL-21
Maximum DC power $\cos \varphi = 1$	4200 W
Maximum input voltage	750 V
MPP voltage range	175 V to 500 V
Rated input voltage	400 V
Minimum input voltage	125 V
Initial input voltage	150 V
Maximum input current, input A	15 A
Maximum input current, input B	15 A
Maximum short-circuit current per input*	20 A
Maximum reverse current from the inverter in the system for max. 1 ms	0 A
Number of independent MPP inputs	2
Strings per MPP inputs	2
Overvoltage category following IEC 60664-1	II
*Input following IEC 62109-2: I_{SCP}	
AC Output	
	SB 4000TL-21
Rated power at 230 V, 50 Hz	4000 W
Maximum apparent AC power	4000 VA
Rated grid voltage	230 V
Nominal AC voltage	220 V / 230 V / 240 V
AC voltage range*	180 V to 280 V
	SB 4000TL-21
Nominal AC at 220 V	18.2 A
Nominal AC at 230 V	17.4 A
Nominal AC at 240 V	16.7 A
Maximum output current	22 A
Total harmonic distortion of the output current with total harmonic distortion of the AC voltage < 2% and AC power > 50% of the rated power	≤ 4%
Inrush current	< 20% of the nominal AC for a maximum of 10 ms
Maximum output current under fault conditions	34 A
Rated power frequency	50 Hz
AC power frequency*	50 Hz / 60 Hz
Operating range at AC power frequency 50 Hz	45 Hz to 55 Hz
Operating range at AC power frequency 60 Hz	55 Hz to 65 Hz
Power factor at rated power	1
Displacement power factor $\cos \varphi$, adjustable	0.8 under excited to 1 to 0.8 overexcited
Feed-in phases	1
Connection phases	1
Overvoltage category as per IEC 60664-1	III
* Depending on the configured country data set	
Efficiency	
	SB 4000TL-21
Maximum efficiency, η_{max}	97.0%
European weighted efficiency, η_{EU}	96.4%

To formulate the weighted efficiency of the grid-connected SPV inverter in the subtropical climate of Baghdad, a fangled set of weigh factors must be determined according to the IEC83:1999. However, the irradiance classes (levels) are aggregated into seven classes of operation according to the irradiance-duration curve produced by the SMA-4000-SB-TL

inverter operated in Baghdad, as displayed in Figure 1, which was erected from one-year irradiance data of Baghdad weather station [17-18]. The period for each class is respectively T_1, T_2, \dots, T_7 . The weigh factor for each level can be seen in Figure 1 and (6) and (7) as illustrated in Table III. The weighted efficiency of the inverter is calculated by (3). Figure 2 shows the relation of inverter DC input power levels with inverter efficiency. The weighted efficiency values are given in Table III. From Table III, the inverter η_{BAG} is 91.8% while from Table II of the inverter data sheet, the η_{max} is 97% and η_{EURO} is 96.4%. An inverter that is 1% less efficient is deduced to be 10% cheaper than other inverters [19-24].

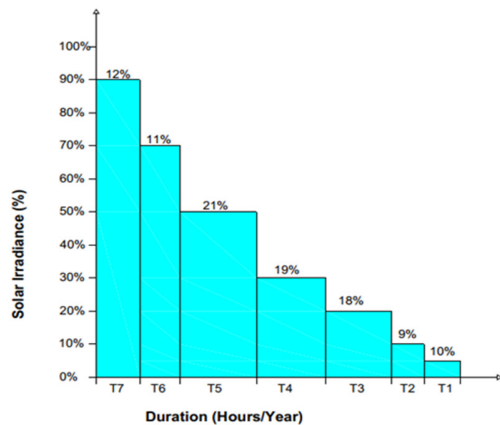


Fig. 1. Solar irradiance-duration curve in Baghdad city.

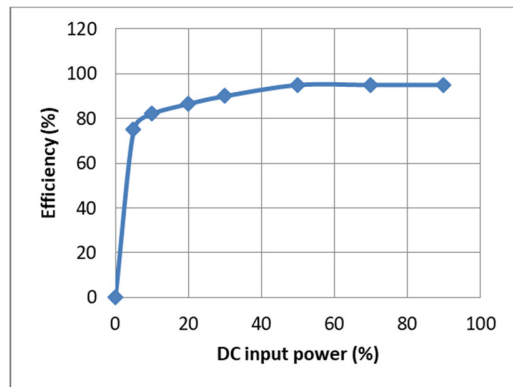


Fig. 2. DC input power levels vs inverter efficiency.

TABLE III. ON-GRID SOLAR PV INVERTER WEIGHTED EFFICIENCY

Solar irradiance (%)	Inverter efficiency (%)	On-grid weigh factor	On-grid inverter weighted efficiency (%)
5	75	$K_1 = 0.0237$	1.77
10	82	$K_2 = 0.0459$	3.76
20	86	$K_3 = 0.1290$	11.1
30	90	$K_4 = 0.1836$	16.52
50	95	$K_5 = 0.2600$	24.70
70	95	$K_6 = 0.1570$	14.91
90	95	$K_7 = 0.2008$	19.07
Total Weighted Efficiency			91.8 %

VI. RESULT VALIDATION

Since the temperature is rarely constant during the system operation and the weighted efficiency is derived at constant temperature, the result of η_{BAG} must be validated. The annually measured energy yield of the system (E_{sys}) throughout the year 2022, as shown in Figure 3, can be compared with the calculated energy yield using η_{BAG} , η_{max} , and η_{EURO} in (1).

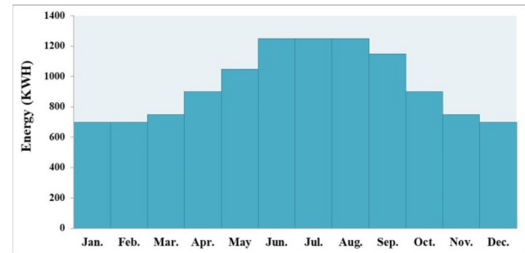


Fig. 3. Yearly measured energy (KWH) produced by SMA-4000-TL.

To determine the E_{sys} , the number of sun hours in one year is (3380) [25], P_{array} is equal to 4.0 Kw, f_{temp} is equal to 98%, f_{mm} is equal to 96%, f_{dirt} is equal to 97% and η_{cable} is equal to 98%. These values make error to be less than 0.8% and are verified in [1]. The η_{inv} is varied using either η_{BAG} , η_{max} , or η_{EURO} . Table IV shows the results of the three computed energy yields, and from the comparison with the measured energy, it is clear that η_{BAG} is the best option to represent the inverter efficiency with only a 1% difference between calculated and measured energy. The real SPV system is a 4 kW roof-top grid-connected system with the SMA-4000-SB-TL inverter. The results show that every region in the world must have its inverter weighted efficiency and SPV system energy calculations to get the correct payback duration and RoI.

TABLE IV. RESULTS OF E_{sys} CALCULATION FOR VARIOUS SPV INVERTER EFFICIENCIES

Efficiency	Efficiency percentage (%)	Calculated E_{sys} (kWh)	Measured E_{sys} (kWh)	E_{sys} divergence (kWh)	E_{sys} divergence (%)
η_{BAG}	91.8	11483	11350	133	1%
η_{max}	97.0	12147	11350	797	7%
η_{EURO}	96.4	12071	11350	721	6%

VII. CONCLUSION

Due to the fact that any region has its own irradiance profile, the SPV inverter weighted efficiency is affected by the irradiance-duration curve, hence, Baghdad weighted efficiency (η_{BAG}) is derived in this paper to best represent the real inverter efficiency in the specific subtropical region, based on the IEC61683:1999 Standard. The derived formula was further validated when it calculated the energy yield of the SPV system ($E_{sys-cal}$), which was compared with the measured energy yield of the system for one year data of a real grid-connected type SMA-4000-SB-TL inverter. The comparison shows that the $E_{sys-cal}$ with η_{BAG} instead of η_{max} or η_{EURO} (given from the data sheet of the inverter) matches the real measured energy

yield ($E_{sys-meas}$) more closely with only a difference of 1% between the calculated and the measured values. This means that the use of η_{BAG} instead of the efficiency given in the inverter data sheet (η_{max} or η_{EURO}) in the determination of the energy yield of the system will describe the predicted output power from a constructed SPV system and will more accurately forecast the payback duration and return of investment.

REFERENCES

- [1] K. Sopian, S. Shaari, N. Amin, R. Zulkifli, and M. N. A. Rahman, "Performance of a grid-connected photovoltaic system in Malaysia," *International Journal of Engineering and Technology*, vol. 4, no. 1, pp. 57–65, 2007.
- [2] *Grid-Connected Pv Systems (no Battery Storage) System Design Guidelines*. Melbourne, Australia: Pacific Power Association, 2012.
- [3] "IEC 61683:1999 Energy efficiency, rural electrification, solar power, LVDC," IEC. <https://webstore.iec.ch/publication/5720>.
- [4] B. Bletterie, R. Brundlinger, and G. Lauss, "On the characterisation of PV inverters' efficiency—introduction to the concept of achievable efficiency," *Progress in Photovoltaics: Research and Applications*, vol. 19, no. 4, pp. 423–435, 2011, <https://doi.org/10.1002/pip.1054>.
- [5] Z. Salam and A. Ab. Rahman, "Efficiency for photovoltaic inverter: A technological review," in *IEEE Conference on Energy Conversion*, Johor Bahru, Malaysia, Oct. 2014, pp. 175–180, <https://doi.org/10.1109/CENCON.2014.6967497>.
- [6] B. Burger *et al.*, "Are we benchmarking inverters on the basis of outdated definitions of the European and CEC efficiency?," in *24th European Photovoltaic Solar Energy Conference*, Hamburg, Germany, Sep. 2009, pp. 3638–3643, <https://doi.org/10.4229/24thEUVPVSEC2009-4BV.1.10>.
- [7] J. Newmiller, W. Erdman, J. S. Stein, and S. Gonzalez, "Sandia Inverter Performance Test Protocol efficiency weighting alternatives," in *40th Photovoltaic Specialist Conference*, Denver, CO, USA, Jun. 2014, pp. 897–900, <https://doi.org/10.1109/PVSC.2014.6925058>.
- [8] I. Ongun and E. Ozdemir, "Weighted efficiency measurement of PV inverters: introducing η_{ZMIR} ," *Journal of Optoelectronics and Advanced Materials*, vol. 15, no. 5–6, pp. 550–554, Jun. 2013.
- [9] M. C. Peel, B. L. Finlayson, and T. A. McMahon, "Updated world map of the Koppen-Geiger climate classification," *Hydrology and Earth System Sciences*, vol. 11, no. 5, pp. 1633–1644, Oct. 2007, <https://doi.org/10.5194/hess-11-1633-2007>.
- [10] A. Kalathil and H. Krishnamurthy, "Quantification of solar inverter efficiency for Indian tropical climatic conditions," in *Region 10 Humanitarian Technology Conference*, Chennai, India, Aug. 2014, pp. 14–18, <https://doi.org/10.1109/R10-HTC.2014.7026316>.
- [11] G. Pillai, N. Pearsall, G. Putrus, R. S. Anand, and R. Pon Perumal, "Performance assessment of grid-connected photovoltaic inverters based on field monitoring in India," in *5th International Symposium on Power Electronics for Distributed Generation Systems*, Galway, Ireland, Jun. 2014, pp. 1–8, <https://doi.org/10.1109/PEDG.2014.6878689>.
- [12] Md. Ehtesham, M. Ahmad, S. Kirmani, and M. Jamil, "Enhancing Solar Inverter Performance for both Stand-Alone and Grid Connected Systems," *Applied Solar Energy*, vol. 59, no. 3, pp. 201–211, Jun. 2023, <https://doi.org/10.3103/S0003701X22601259>.
- [13] A. Harish, A. Ramanan, P. K. Das, G. Giridhar, K. Balaraman, and T. Prabu, "Formulation of Efficiency of Inverters for Solar Photovoltaic Power Plants - Indian Case Study," in *2nd International Conference on Power Energy, Environment and Intelligent Control*, Greater Noida, India, Oct. 2019, pp. 196–201, <https://doi.org/10.1109/PEEIC47157.2019.8976638>.
- [14] F. H. Dupont, C. Rech, and J. R. Pinheiro, "A methodology to obtain the equations for the calculation of the weighted average efficiency applied to photovoltaic systems," in *10th IEEE/IAS International Conference on Industry Applications*, Fortaleza, Brazil, Nov. 2012, pp. 1–8, <https://doi.org/10.1109/INDUSCON.2012.6453445>.
- [15] A. Kellermann, R. Alfonso Reiter, and A. Peres, "Calculation of the weighted average efficiency of photovoltaic systems in the Brazilian State of Santa Catarina," in *International Conference on Renewable Energies and Power Quality*, La Coruna, Spain, Mar. 2015, pp. 514–518, <https://doi.org/10.24084/repqj13.382>.
- [16] A. Ab. Rahman, Z. Salam, S. Shaari, and M. Z. Ramli, "Methodology to Determine Photovoltaic Inverter Conversion Efficiency for the Equatorial Region," *Applied Sciences*, vol. 10, no. 1, Jan. 2020, Art. no. 201, <https://doi.org/10.3390/app10010201>.
- [17] K. J. Al-Jumaily, M. F. Al-Zuhairi, and Z. S. Mahdi, "Estimation of clear sky hourly global solar radiation in Iraq," *International Journal of Energy and Environment*, vol. 3, no. 5, pp. 659–666, Jul. 2012.
- [18] B. K. Ahmed, G. G. Younis, and Z. Abdalwahid, "Estimation and Analysis of Solar Radiation on Horizontal and Inclined Surface for Baghdad City," *Iraqi Journal of Science*, vol. 62, no. 11, pp. 4249–4259, Dec. 2021, [https://doi.org/10.24996/ijis.2021.62.11\(SI\).5](https://doi.org/10.24996/ijis.2021.62.11(SI).5).
- [19] A. Nouaiti, M. Reddak, C. Boutahiri, A. Mesbahi, A. M. Hsaini, and A. Bouazi, "A Single Stage Photovoltaic Solar Pumping System based on the Three Phase Multilevel Inverter," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12145–12150, Dec. 2023, <https://doi.org/10.48084/etasr.6403>.
- [20] B. A. Nasir, "Determination of the Harmonic Losses in an Induction Motor Fed by an Inverter," *Engineering, Technology & Applied Science Research*, vol. 12, no. 6, pp. 9536–9545, Dec. 2022, <https://doi.org/10.48084/etasr.5012>.
- [21] F. Z. Kebbab, L. Sabah, and H. Nouri, "A Comparative Analysis of MPPT Techniques for Grid Connected PVs," *Engineering, Technology & Applied Science Research*, vol. 12, no. 2, pp. 8228–8235, Apr. 2022, <https://doi.org/10.48084/etasr.4704>.
- [22] Y. Kassem, H. Gokcekus, and F. A. R. Agila, "Techno-Economic Feasibility Assessment for the promotion of Grid-Connected Rooftop PV Systems in Botswana: A Case Study," *Engineering, Technology & Applied Science Research*, vol. 13, no. 2, pp. 10328–10337, Apr. 2023, <https://doi.org/10.48084/etasr.5668>.
- [23] Z. Ali, Z. M. Abdullah, B. A. Naser, R. W. Daoud, and A. H. Ahmed, "Design of a Single-Phase Inverter for Solar Energy Conversion System," *NTU Journal of Renewable Energy*, vol. 1, no. 1, pp. 38–42, Sep. 2021, <https://doi.org/10.56286/ntujre.v1i1.13>.
- [24] M. M. Rashid, M. A. Sharif, and F. Korkmaz, "Effects of Weather and Environmental Conditions on the Power Productivity of Photovoltaic Module in Kirkuk City," *NTU Journal of Renewable Energy*, vol. 4, no. 1, pp. 1–6, 2023, <https://doi.org/10.56286/ntujre.v4i1>.
- [25] "Bagdad Climate By Month | A Year-Round Guide," *World Weather & Climate Information*. <https://weather-and-climate.com:80/average-monthly-Rainfall-Temperature-Sunshine,Bagdad,Iraq>.
- [26] O. M. Al-Barbarawi, "Improving Performance of the Braking Process, and Analysis Torque-Speed Characteristics of the Induction Motor," *Engineering, Technology & Applied Science Research*, vol. 8, no. 6, pp. 3585–3591, Dec. 2018, <https://doi.org/10.48084/etasr.2325>.
- [27] N. H. Mugheri and M. U. Keerio, "An Optimal Fuzzy Logic-based PI Controller for the Speed Control of an Induction Motor using the V/F Method," *Engineering, Technology & Applied Science Research*, vol. 11, no. 4, pp. 7399–7404, Aug. 2021, <https://doi.org/10.48084/etasr.4255>.
- [28] A. Elgharbi, D. Mezghani, and A. Mami, "Intelligent Control of a Photovoltaic Pumping System," *Engineering, Technology & Applied Science Research*, vol. 9, no. 5, pp. 4689–4694, Oct. 2019, <https://doi.org/10.48084/etasr.2982>.
- [29] A. Mohamed Yassine, F. Tadeo, D. Mezghani, and A. Mami, "FPGA Implementation of a Robust MPPT of a Photovoltaic System Using a Fuzzy Logic Controller Based on Incremental and Conductance Algorithm," *Engineering, Technology and Applied Science Research*, vol. 9, no. 4, pp. 4322–4328, Aug. 2019, <https://doi.org/10.48084/etasr.2771>.
- [30] Y. Aljarhizi, A. Nouaiti, E. A. Ibrahim, C. Boutahiri, A. Hassoune, and A. Mesbahi, "Optimized Wind Turbine Emulator based on an AC to DC Motor Generator Set," *Engineering, Technology & Applied Science Research*, vol. 13, no. 2, pp. 10559–10564, Apr. 2023, <https://doi.org/10.48084/etasr.5775>.

Temperature Dependency of Photoelectronic Properties of Group III-V Arsenide Solar Cell

Md. Abdullah Al Humayun

Department of EEE, Eastern University, Bangladesh
humayun0403063@gmail.com

Masum Hossen

Department of EEE, Green University of Bangladesh, Bangladesh
masum.eee007@gmail.com

Md. Zamil Haider

Department of EEE, Eastern University, Bangladesh
mzhs.eee@gmail.com

Bedir Yousif

Department of Electrical Engineering, College of Engineering and Information Technology, Onaizah Colleges, Qassim, Saudi Arabia | Electrical Engineering Department, Faculty of Engineering, Kafrelsheikh University, Egypt
bedir.yousif@eng.kfs.edu.eg

Muhammad Tajammal Chughtai

Department of Electrical Engineering, College of Engineering, University of Hail, Saudi Arabia
mt.chughtai@uoh.edu.sa

Muhammad Islam

Department of Electrical Engineering, College of Engineering, Qassim University, Buraydah 51452, Saudi Arabia
muha.khan@qu.edu.sa (corresponding author)

Sheroz Khan

Department of Electrical Engineering, College of Engineering and Information Technology, Onaizah Colleges, Qassim, Saudi Arabia
cnar32.sheroz@gmail.com

Received: 19 August 2023 | Revised: 30 September 2023 and 5 October 2023 | Accepted: 8 October 2023

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6293>

ABSTRACT

This study explores the effect of temperature on different characteristics of Solar Cells (SC) composed of a structured III-V arsenide group. The temperature dependence of the SC characteristics was investigated numerically and by simulation. In both approaches, each characteristic was compared with a conventional Si SC. InAs showed superior stability and lower temperature sensitivity, as it has a negligible decrease of 0.098 eV in the energy bandgap, while the energy bandgaps of Si, AlAs, and GaAs are 0.129, 0.186, and 0.200 eV, respectively. Moreover, with a decay rate of 81.911 mV/°K, InAs exhibited the lowest temperature sensitivity in open-circuit voltage. InAs additionally demonstrated the least increase in degradation rate, while the SC power output is still a cause of concern. AlAs, Si, and GaAs had a total accumulative gradient change of 0.162, 0.136, and 0.034% in the degradation rate, respectively, while InAs showcased the highest stability by displaying a change of only 0.008%. A comparative analysis illustrated

that among these III-V arsenide compounds, InAs had a rock-bottom sensitivity to temperature changes and better temperature stability in both numerical and simulation approaches.

Keywords-density of state; feedback level; frequency fluctuation; momentum relaxation time; laser

I. INTRODUCTION

Conventional energy sources are limited and are gradually diminishing day by day. In addition, the consumption of fossil fuels is producing CO₂ and other toxic gases that pollute the environment at an alarming rate. To overcome dependence on traditional energy sources, such as coal, gas, biomass, and others, it is necessary to satisfy our daily energy demands through alternative sources. Today, energy needs are constantly increasing in parallel with the raise in industrial development, the population augmentation, and lifestyle. Solar energy harvesting is an alternative source. The light energy coming from the Sun on Earth (120,000 TW of sunlight) is almost 20000 times more than the energy required to maintain daily activities on Earth [1-4]. SC is the "Eureka" for using this abundant amount of solar energy since it converts light into electrical energy. Moreover, it is a means of maneuvering energy from a natural fusion reactor in the sky, the Sun, and constitutes a possible solution to prevent climate change and environmental pollution. However, there are several challenges, such as high costs, low efficiency, and complex manufacturing processes, that slow down the growth of this technology [5]. Therefore, new materials are thoroughly studied to bypass these deficiencies. The use of new materials may advance the development and manufacturing of SCs and create an easy pathway to access clean, abundant, and affordable energy for all [6].

Silicon (Si) is the most widely employed material for SC fabrication. It has become convenient for the generation of electron-hole pairs of the depletion region in the crystal to move up and down to opposite polarities, creating a stronger electric field when exposed to light rather than vibrating in place and generating heat [7-8]. However, the process of developing or cultivating large crystals is very expensive due to the difficulty in growing them. In addition, it comes with material waste as a result of the need to slice it into thin films from its single-crystalline form of ingot, which are reshaped and converted into very thin slices called silicon wafers. The latter are called the hearts of an SC. SCs consisting of silicon are the first generation of solar cells, which are manufactured and used on a large scale and divided into monocrystalline and polycrystalline silicon SCs [9-10]. The absorption coefficient of Si decreases with the increasing wavelength of incident light photons. Therefore, researchers have introduced binary semiconductor compounds of GaAs, InAs, and AlAs as active layers for use in SCs to enhance cell efficiency, thus making solar energy more affordable. The theoretical limit for solar efficiency has been reported to will have achieved 40% by 2040 [11-12].

However, researchers are putting their efforts into increasing this efficiency figure so that this limitation will soon be overcome by tandem cells and thin-film SCs [13-15]. The highly intrinsic features of group III-V elements, such as direct energy bandgaps, high optical absorption coefficients, and good values for minority carrier lifetime and mobility, make

III-V arsenide compounds of GaAs, InAs, AlAs excellent candidate materials for the fabrication of high-efficiency SCs [14-15]. Many researchers have acknowledged that low-cost materials have the potential to be applied in SCs because they are suitable for absorbing energy from the solar spectrum [16-17]. The bandgap energy of the material is the most important characteristic that controls the properties of the SC, namely open circuit voltage, short-circuit current, and output power. Therefore, this study investigated the temperature dependence of these parameters for SCs using Group-III Arsenide and examined the performance of single-junction SCs. This study also considered an improved technique for generating solar energy from alarmingly high temperatures. The analysis which was conducted can be used to further study the temperature-dependent performance of thin-film multijunction SCs.

II. MATERIAL AND METHOD

A. Device Structure

Figure 1 shows a Quantum Dot SC (QDSC) grown with InAs QD layers [17-18] in 20 iterations. Each layer is isolated from the neighboring by a 15 nm GaAs spacer [19]. The QD layers are made from 2.5ML of InAs, covered with a 15 nm GaAs coating to sandwich the QD layers. In addition, the p-i-n structure of the QDSC was designed with a dielectric having a thickness of 200 nm of n-GaAs buffer layer, n-Al_{0.33}Ga_{0.67}As of thickness 30 nm, 300 nm of n-GaAs base, emitter with 250 nm thick p-GaAs, 30 nm p-Al_{0.75}Ga_{0.25}As window, and 250 nm thick p+-GaAs contact layer.

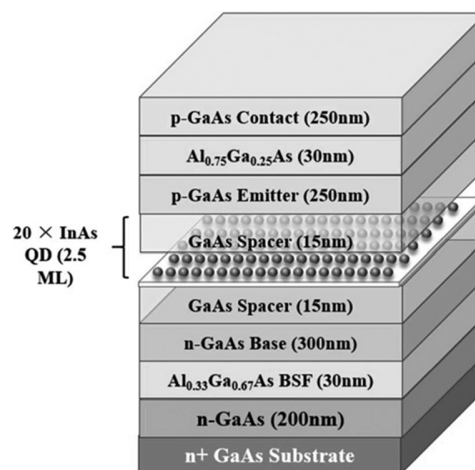


Fig. 1. Schematic diagram of the QDSC structure.

B. Numerical Analysis

The major characteristics of QDSC were analyzed using Si, InAs, GaAs, and AlAs QD as the active layer material of the QDSC structure. The characteristics of the QDSC analyzed were open-circuit voltage, short-circuit current, output power, and temperature stability. The effect of temperature on the

energy bandgap of semiconductors has been explored in Varshni's model, given by [18]:

$$E_g(T) = E_g(0) - \frac{\gamma T^2}{(T + \beta)} \quad (1)$$

where $E_g(T)$ is the semiconductor bandgap energy at a given temperature T , $E_g(0)$ is the bandgap value at 0°K, and γ and β are Varshni constants. The bandgap is one of the most important semiconductor parameters in solid-state physics. Variations in bandgap energy with temperature changes affect the overall performance of the solar cell, as it determines the absorption band of the cells.

By performing the derivation of the band gap $E_g(T)$ for temperature, the rate of change or the bandgap temperature reliance can be discussed with the help of the following calculations:

$$\frac{dE_g}{dT} = -\gamma \frac{T(T+2\beta)}{(T+\beta)^2} \quad (2)$$

The negative sign in (2) indicates that the bandgap energy of the semiconductor decreases with increasing temperature values and vice-versa. To determine the magnitude of the rate of fluctuation of E_g volatility by taking the scales on both sides of (2) gives:

$$\left| \frac{dE_g}{dT} \right| = \gamma \frac{T(T+2\beta)}{(T+\beta)^2} \quad (3)$$

The open circuit voltage, which is an important characteristic parameter of an SC device, can be determined according to:

$$V_{oc} = \frac{E_g(T)}{q} \quad (4)$$

where q is the electric charge and $E_g(T)$ is the bandgap energy at temperature T of the active layer material of the SC structure. To analyze the temperature effect on the rate of change of open circuit voltage (V_{oc}), it is necessary to differentiate the V_{oc} equation to temperature:

$$\frac{dV_{oc}}{dT} = \frac{1}{q} \frac{dE_g}{dT} \quad (5)$$

The open circuit voltage of the solar cell depends on the band gap energy of the active layer material in the solar cell and can be evaluated using:

$$I_{sc} = I_s e^{\left(\frac{qV_{oc}}{KT} - 1\right)} \quad (6)$$

where I_s is the saturation current of the binary material used in the active layer, q is the electric charge, K is Boltzmann's constant, and T is the absolute temperature in K. Both sides of (6) should be derived to explore the rate of short-circuit current degradation as a function of temperature, leading to:

$$\frac{dI_{sc}(T)}{dT} = \frac{qI_s}{KT^2} \left(\frac{TdV_{oc}(T)}{dT} - V_{oc} \right) e^{\left(\frac{qV_{oc}}{KT} - 1\right)} \quad (7)$$

One of the parameters that affect the performance of the SC is the output power P_{out} , which can be obtained by:

$$P_{out}(T) = V_{oc}(T) * I_{sc}(T) \quad (8)$$

where V_{oc} and I_{sc} are the open circuit voltage and short circuit current, respectively. To examine the change in the output

power of the solar cell caused by temperature fluctuations, (8) is differentiated by the temperature to obtain:

$$\frac{dP_{out}(T)}{dT} = V_{oc}(T) \frac{dI_{sc}(T)}{dT} + I_{sc}(T) \frac{dV_{oc}(T)}{dT} \quad (9)$$

C. Simulation Model

This section presents the simplified schematic diagram of the simulation model of the proposed QDSC using MATLAB Simulink. This study associated the achieved results of SC characteristics through the numerical approach with the characteristics obtained from simulation results to show the accuracy of the QDSC model. The SC characteristics, such as V_{oc} , I_{sc} , and P_{out} , have been measured at ambient conditions for four different QD materials as the active layer materials of the device structure. Figure 2 presents the QDSC models developed using Simulink.

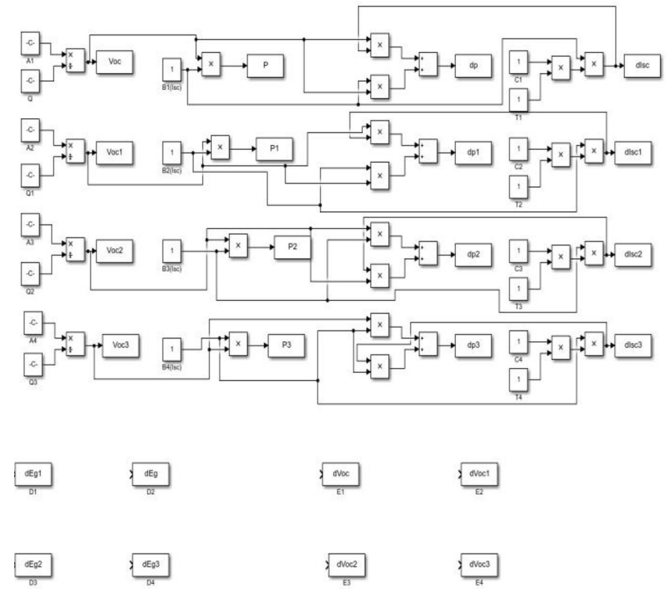


Fig. 2. Simplified schematic diagram of the simulation model developed for the proposed QDSC structure using MATLAB Simulink.

III. RESULTS AND DISCUSSION

The numerical analysis discloses the effects of temperature on the bandgap energy of the active layer material of SC, open circuit voltage, short circuit current, and the output power of SC. Table I depicts the values of some of the primary parameters necessary to analyze the performance characteristics of SC [21-23].

TABLE I. BANDGAP ENERGY PARAMETERS FOR SI, GAAS, INAS, ALAS

Material	$E_g(0)$ [eV]	γ [eV°K]	β [°K]
Si	1.17	4.73×10^{-4}	636
GaAs	1.521	5.58×10^{-4}	220
InAs	0.414	2.5×10^{-4}	75
AlAs	2.239	6.0×10^{-4}	408

Figure 3 illustrates the degradation of bandgap energy E_g for GaAs, AlAs, and InAs for the temperature over the range of

300-700°K along with the most widely used material Si to design SCs. The photovoltaic properties of the device are highly dependent on the E_g of the material used in the active layer of the device. In this figure, the dotted, solid, dash-dot, and dashed lines represent the respective E_g of Si, InAs, GaAs, and AlAs. The E_g of the materials utilized in this numerical analysis is gradually reducing nonlinearly with a uniform increase in temperature over the range of 300-700°K. The decrease in E_g occurs as the spacing between atoms increases, and this happens when the amplitude of atomic vibrations increases due to the rise in thermal energy. The resulting plot shows that Si experienced a decrease of 0.129 eV in the bandgap energy, from 1.125 eV to 0.997 eV over the temperature range. AlAs decreased by 0.186 eV. Remarkably, GaAs had the highest downward trend in the energy bandgap of 0.200 eV, from 1.424 eV to 1.224 eV. However, InAs has experienced the lowest bandgap drop of 0.098 eV. The results obtained through numerical analysis and simulation show a good agreement.

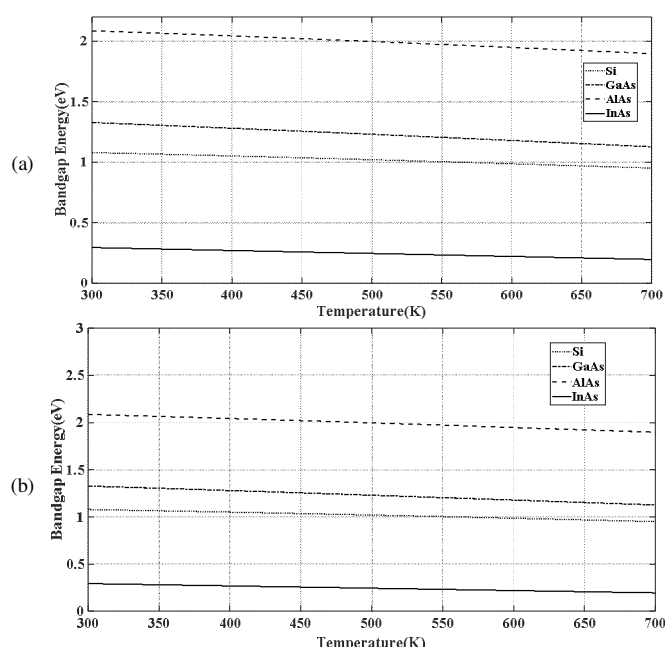


Fig. 3. Degradation of bandgap energy of III-V arsenide material and Si with the increase in temperature: (a) Numerical analysis and (b) simulation results.

Figure 4 portrays the rate of change in E_g of Si, GaAs, AlAs, and InAs to temperatures in the range of 300-700°K. The rate of change of E_g of the active layer material of the SC decreases with an increase in temperature. InAs experiences a flat rate of change of 0.82%, while it has a rate of change from 24, 24.56, and 24.77 at 300, 500, and 700°K respectively. After undergoing extreme fluctuations of 18.75% and 7.37% against the temperature changes over the 300-500°K range, both AlAs and GaAs tend to have a relatively planar value throughout 500-700°K. However, compared to III-V arsenide materials, Si demonstrated the lowest stability against temperature. An overall volatility of 28% was observed, while the fluctuation in the first half was more than in the past. Si has a rate of change

of 25.59 meV/°K at 300°K and 32.63 meV/°K at 500°K. Using InAs in the SC active layer, the fluctuations in the rate of change of E_g were greatly reduced. The numerical and simulation results show a good agreement.

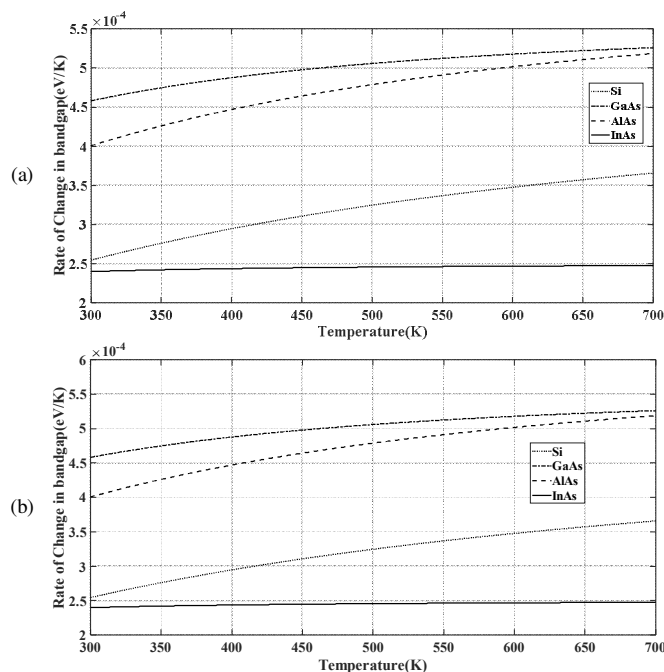


Fig. 4. Rate of change in bandgap energy of III-V arsenide material and Si with the increase in temperature: (a) Numerical analysis and (b) simulation output.

Figure 5 exhibits the temperature reliances (V_{oc}) of the SC using Si, GaAs, AlAs, and InAs in the active layer for temperatures within the 300-700°K range. The corresponding changes in the V_{oc} of SC for Si, InAs, GaAs, and AlAs as active layers are represented by the dotted, solid, dash-dot, and dashed lines, respectively. The V_{oc} of SC undergoes a decrease as the temperature reaches higher values. The V_{oc} of the AlAs SC suddenly decreased from 2.09 to 1.91 V as the temperature increased from 300 to 700°K. However, it showed a gradual decrease after the temperature of 500°K, where V_{oc} is 2.01 V. The GaAs SC had a total decrease of 0.19 V in V_{oc} , while for Si this value was 0.14 V. The InAs SC had a minimum drop in V_{oc} and experienced a total drop of 0.10 V with an increase in temperature from 300 to 700°K. The results obtained through numerical analysis and simulation reveal a good agreement.

Figure 6 depicts the rate of change of V_{oc} using Si, GaAs, AlAs, and InAs in the active layer of the SC structure for temperatures in the 300-700°K range. The rate of change in V_{oc} decreased significantly by employing InAs, while it had an overall increase of 0.446 $\mu\text{V}/^\circ\text{K}$ in the degradation rate. However, both AlAs and GaAs had a considerable increase in the degradation rate of 0.595 $\mu\text{V}/^\circ\text{K}$ and 0.776 $\mu\text{V}/^\circ\text{K}$, accordingly, for the total temperature variation. In contrast to the application of III-V Arsenide as the active layer material, conventional Si possesses poor stability against temperature variations. The rate of change in V_{oc} of SC is greatly increased for Si after 500°K. The slope of the curve experienced an

exponential of $480.81 \mu\text{V}/^\circ\text{K}$ in the rate of change of V_{oc} , where it decreased by a rate of $1036.36 \mu\text{V}/^\circ\text{K}$ in V_{oc} at 700°K and $555.55 \mu\text{V}/^\circ\text{K}$ mV increase at 500°K . Again, the outcomes obtained through numerical analysis and simulation disclose good agreement.

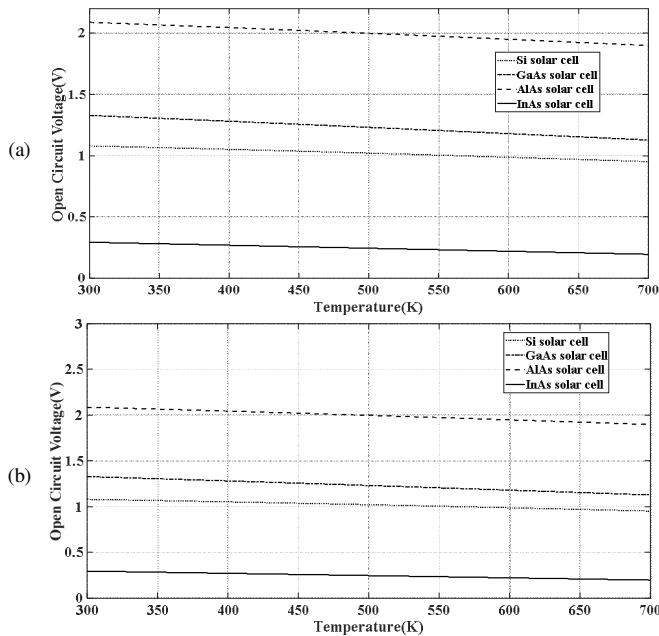


Fig. 5. Dependency of open circuit voltage (V_{oc}) of SC on temperature fluctuation: (a) Numerical analysis and (b) simulation output.

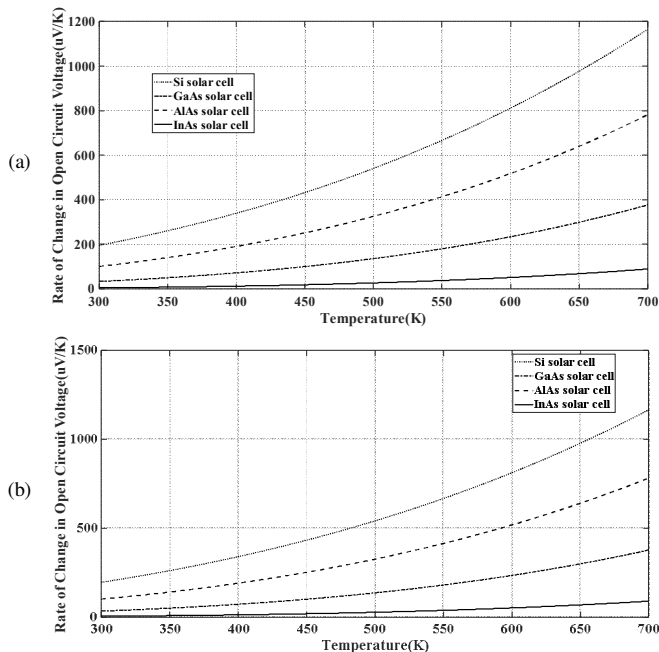


Fig. 6. Rate of change of open circuit voltage against temperature fluctuations: (a) Numerical analysis and (b) simulation output.

Figure 7 illustrates the dependency of I_{sc} of SCs with Si, GaAs, AlAs, and InAs in the active layer of their structure for

temperatures within 300°K - 700°K . The dotted, solid, dash-dot, and dashed lines represent the SCs having Si, InAs, GaAs, and AlAs in their active layer. The AlAs SC decreased by 0.40% after 500°K . Si and GaAs SCs presented quite similar sensitivity to temperature and had a gradual decrease in short circuit currents of 0.21% and 0.27%, correspondingly. The InAs SC demonstrated better stability, as it experienced only a decrease of 0.07% in the same temperature range. Again, the results obtained through numerical analysis and simulation show good agreement.

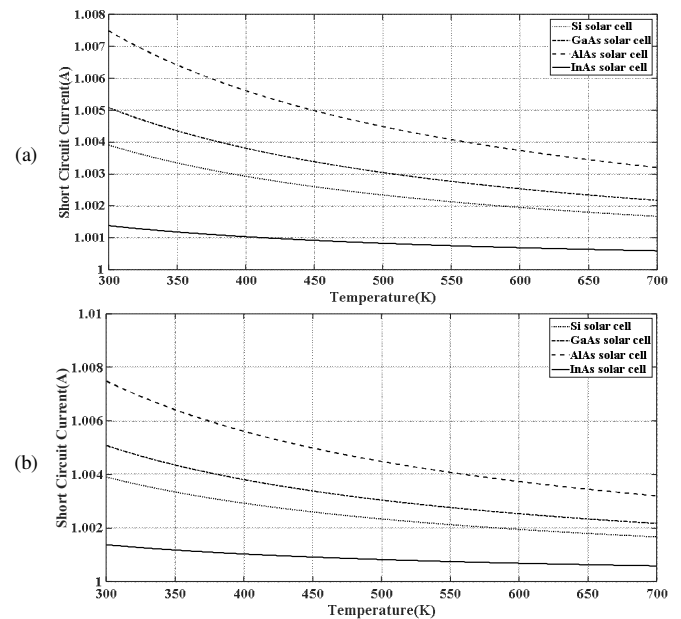


Fig. 7. Dependency of short circuit current of SC on the temperature fluctuations: (a) Numerical analysis and (b) simulation output.

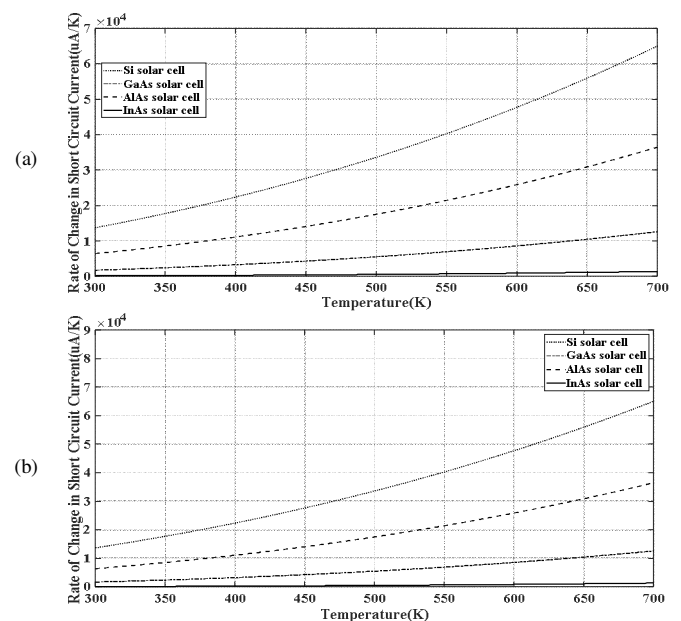


Fig. 8. Rate of change in short circuit current of SC due to temperature fluctuations: (a) Numerical analysis and (b) simulation output.

Figure 8 verifies the nominal decrease in the rate of change of I_{sc} of SCs as the temperature increases. Conventional Si SC showed the worst stability at high temperatures and had the most upward fluctuations of $6.01 \mu\text{Amp}/^\circ\text{K}$ in the rate of change of I_{sc} as temperature increased from 300 to 700°K . The instability gradually increased for AlAs and GaAs SCs, although GaAs SC displayed a lower drift rate compared to the AlAs SC. However, InAs brings out as a promising compound for use in SCs by exhibiting an overall lower increase of $0.11 \mu\text{Amp}/^\circ\text{K}$ in the I_{sc} value compared to other widely used AlAs, GaAs, and Si. Again, the numerical and simulation results show good agreement.

Output power P_{out} is generally considered to be one of the most reliable variables for measuring how well a device is performing for practical applications. Figure 9 presents the effect of temperature changes on the P_{out} of SC. In this plot, dotted, solid, dash-dot, and dashed lines illustrate the output power for SCs using Si, InAs, GaAs, and AlAs in the active layer of their structure. The general trend of P_{out} is downward with increasing temperature. The Si SC had a decrease of approximately 12%, as it has an output power of 1.09 and 0.96 nW at 300 and 700°K , respectively. InAs displayed a total output power drop of 0.10 nW, while GaAs SC had the highest loss of 0.19495 nW on P_{out} for the respective temperature change. Again, the results obtained through numerical analysis and simulation show good agreement.

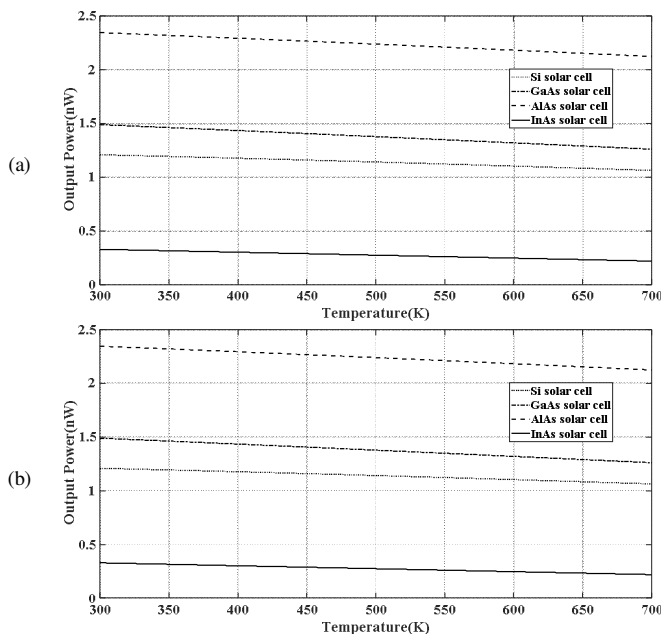


Fig. 9. Change in output power of SC against temperature fluctuations: (a) Numerical analysis and (b) simulation output.

Figure 10 shows the amount of variation in P_{out} of SCs for temperature fluctuation per unit. Si and AlAs SCs showed a significant increase in the rate of change in P_{out} , but the AlAs SC exhibited significant sensitivity after 500°K . The AlAs SC had a lower P_{out} rate of change value than Si at 300°K ($0.16 \text{ nW}/^\circ\text{K}$) and ended with a value of $0.79 \text{ nW}/^\circ\text{K}$ at 700°K . The

rate gradually increased for GaAs and ended with $0.17 \text{ nW}/^\circ\text{K}$ at 700°K . However, the analysis confirmed that the rate of change in the resulting P_{out} degradation was significantly reduced for the SC having InAs in its active layer structure, remaining nearly stable throughout the temperature fluctuation. Figure 10 clearly indicates that the GaAs SC had an acceptable reduction rate in P_{out} . However, conventional Si and AlAs had an intense upward rate of power degradation. Again, the numerical and simulation results show good agreement.

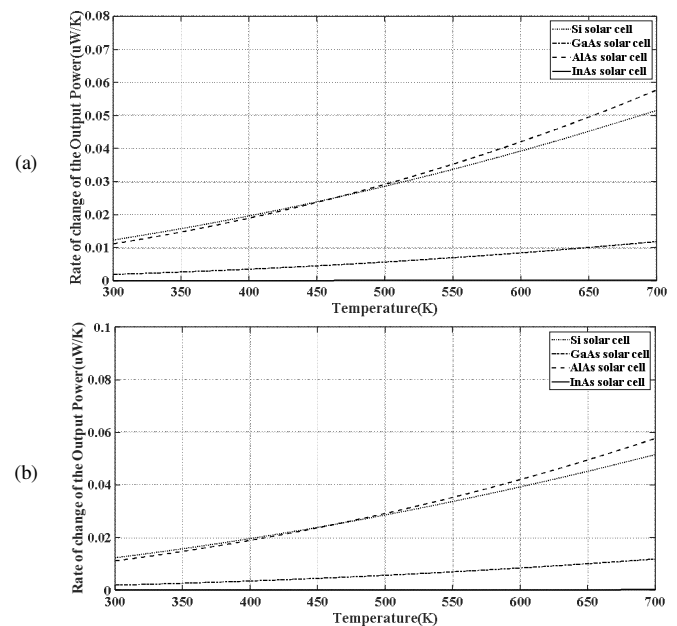


Figure 10: Rate of change of the output power of SC due to temperature fluctuations: (a) Numerical analysis and (b) simulation output.

IV. CONCLUSION

This study investigated the temperature dependence of the open circuit voltage (V_{oc}) and short circuit current (I_{sc}) of SCs within a temperature range of 300 to 700°K to determine their output power (P_{out}). The numerical results acquired were compared with an SC using Si as active layer material. To examine these characteristics, the temperature dependence of the bandgap energy of III-V Arsenide material and Si was analyzed with a temperature range of 300 to 700°K , since it governs the SC characteristics. Much of the solar energy is not taken into account, as the SC cannot cover the entire solar band (specifically, ultraviolet, infrared, and low or diffused light) into electrical energy. Moreover, the stability of the photovoltaic properties is also very crucial for SCs. Materials with high stability and low bandgap can be a concern for tandem cells, which split light into different wavelength bands, produce a different color of light, and direct the beams to different cells tuned to those bands. GaAs demonstrated good integrity at low temperatures and had a satisfactory bandgap energy for SC designs. Therefore, GaAs can be considered as a potentially promising candidate material to be applied in SC design. SCs have the highest efficiency within an energy bandgap around 1.35-1.5 eV [24].

ACKNOWLEDGMENT

The researchers would like to thank the Deanship of Scientific Research, Qassim University for funding the publication of this project.

REFERENCES

- [1] "The Balance of Power in the Earth-Sun System," National Aeronautics and Space Administration, 2005. [Online]. Available: https://www.nasa.gov/wp-content/uploads/2015/03/135642main_balance_trifold21.pdf.
- [2] Y. Liang *et al.*, "Co₃O₄ nanocrystals on graphene as a synergistic catalyst for oxygen reduction reaction," *Nature Materials*, vol. 10, no. 10, pp. 780–786, Oct. 2011, <https://doi.org/10.1038/nmat3087>.
- [3] "Solar Energy Source: Pros and Cons," *Canadian Institute For Knowledge Development*, Nov. 25, 2019. <https://cikd.ca/2019/11/25/solar-energy-source-pros-and-cons/>.
- [4] N. S. Lewis and D. G. Nocera, "Powering the planet: Chemical challenges in solar energy utilization," *Proceedings of the National Academy of Sciences*, vol. 103, no. 43, pp. 15729–15735, Oct. 2006, <https://doi.org/10.1073/pnas.0603395103>.
- [5] P. Balling *et al.*, "Improving the efficiency of solar cells by upconverting sunlight using field enhancement from optimized nano structures," *Optical Materials*, vol. 83, pp. 279–289, Sep. 2018, <https://doi.org/10.1016/j.optmat.2018.06.038>.
- [6] K. Kumari, T. Chakrabarti, A. Jana, D. Bhattachartjee, B. Gupta, and S. K. Sarkar, "Comparative Study on Perovskite Solar Cells based on Titanium, Nickel and Cadmium doped BiFeO₃ active material," *Optical Materials*, vol. 84, pp. 681–688, Oct. 2018, <https://doi.org/10.1016/j.optmat.2018.07.071>.
- [7] Q. H. Fan *et al.*, "High efficiency silicon–germanium thin film solar cells using graded absorber layer," *Solar Energy Materials and Solar Cells*, vol. 94, no. 7, pp. 1300–1302, Jul. 2010, <https://doi.org/10.1016/j.solmat.2010.03.006>.
- [8] S. Kaci *et al.*, "Impact of porous SiC-doped PVA based LDS layer on electrical parameters of Si solar cells," *Optical Materials*, vol. 80, pp. 225–232, Jun. 2018, <https://doi.org/10.1016/j.optmat.2018.05.006>.
- [9] C. Ji *et al.*, "Recent Applications of Antireflection Coatings in Solar Cells," *Photonics*, vol. 9, no. 12, Dec. 2022, Art. no. 906, <https://doi.org/10.3390/photonics9120906>.
- [10] M. S. Almomani *et al.*, "Performance Improvement of Graded Bandgap Solar Cell via Optimization of Energy Levels Alignment in Si Quantum Dot, TiO₂ Nanoparticles, and Porous Si," *Photonics*, vol. 9, no. 11, 2022, <https://doi.org/10.3390/photonics9110843>.
- [11] A. M. Mouafki, F. Bouaïcha, A. Hedibi, and A. Gueddim, "Porous Silicon Antireflective Coatings for Silicon Solar Cells," *Engineering, Technology & Applied Science Research*, vol. 12, no. 2, pp. 8354–8358, Apr. 2022, <https://doi.org/10.48084/etasr.4803>.
- [12] S. V. Boriskina and G. Chen, "Exceeding the solar cell Shockley–Queisser limit via thermal up-conversion of low-energy photons," *Optics Communications*, vol. 314, pp. 71–78, Mar. 2014, <https://doi.org/10.1016/j.optcom.2013.10.042>.
- [13] S. M. Ho, "Fabrication of Cu₄SnS₄ Thin Films: A Review," *Engineering, Technology & Applied Science Research*, vol. 10, no. 5, pp. 6161–6164, Oct. 2020, <https://doi.org/10.48084/etasr.3663>.
- [14] M. Sojoudi, R. Madatov, T. Sojoudi, and P. Farhadi, "Achieving Steady and Stable Energy from AlGaAsGaAs Solar Cells," *Engineering, Technology & Applied Science Research*, vol. 1, no. 6, pp. 151–154, Dec. 2011, <https://doi.org/10.48084/etasr.93>.
- [15] S. Guterma *et al.*, "Optimized flexible cover films for improved conversion efficiency in thin film flexible solar cells," *Optical Materials*, vol. 79, pp. 243–246, May 2018, <https://doi.org/10.1016/j.optmat.2018.03.034>.
- [16] S. K. Tripathy and A. Pattanaik, "Optical and electronic properties of some semiconductors from energy gaps," *Optical Materials*, vol. 53, pp. 123–133, Mar. 2016, <https://doi.org/10.1016/j.optmat.2016.01.012>.
- [17] E. V. Kunitsyna *et al.*, "Narrow gap III–V materials for infrared photodiodes and thermophotovoltaic cells," *Optical Materials*, vol. 32, no. 12, pp. 1573–1577, Oct. 2010, <https://doi.org/10.1016/j.optmat.2010.06.010>.
- [18] S. Takatori *et al.*, "Investigation of the terahertz emission characteristics of MBE-grown GaAs-based nanostructures," *Optical Materials*, vol. 32, no. 7, pp. 776–779, May 2010, <https://doi.org/10.1016/j.optmat.2010.02.014>.
- [19] P. M. Lam *et al.*, "Effect of rapid thermal annealing on InAs/GaAs quantum dot solar cells," *IET Optoelectronics*, vol. 9, no. 2, pp. 65–68, 2015, <https://doi.org/10.1049/fiet-opt.2014.0079>.
- [20] A. Mahfoud, F. Mohamed, S. Mekhilef, and F. Djahli, "Effect of Temperature on the GaInP/GaAs Tandem Solar Cell Performances," *International Journal Of Renewable Energy Research*, vol. 5, no. 2, pp. 629–634, Jun. 2015.
- [21] J. Singh, *Physics of Semiconductors and Their Heterostructures*. New York, NY, USA: McGraw-Hill College, 1992.
- [22] M. V. Fischetti and S. E. Laux, "Monte Carlo simulation of transport in technologically significant semiconductors of the diamond and zinc-blende structures. II. Submicrometer MOSFET's," *IEEE Transactions on Electron Devices*, vol. 38, no. 3, pp. 650–660, Mar. 1991, <https://doi.org/10.1109/16.75177>.
- [23] V. A. Wilkinson and A. R. Adams, "The effect of temperature and pressure on InGaAs band structure," *EMIS Data Reviews Series*, vol. 8, 1993.
- [24] G. Saint-Girons, A. Mereuta, G. Patriarche, J. M. Gérard, and I. Sagnes, "Influence of the thermal treatment on the optical and structural properties of 1.3 μm emitting LP-MOVPE grown InAs/GaAs quantum dots," *Optical Materials*, vol. 17, no. 1, pp. 263–266, Jun. 2001, [https://doi.org/10.1016/S0925-3467\(01\)00089-1](https://doi.org/10.1016/S0925-3467(01)00089-1).

Using Artificial Neural Networks with GridSearchCV for Predicting Indoor Temperature in a Smart Home

Talal Alshammari

Department of Information and Computer Science, College of Computer Science and Engineering,
University of Ha'il, Saudi Arabia
talal.alshammari@uoh.edu.sa

Received: 4 February 2024 | Revised: 16 February 2024 | Accepted: 19 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7008>

ABSTRACT

The acceleration of house technology via the use of mobile phones has made it easier to control houses, where occupants (especially older people) spend most of their time. The climate of Saudi Arabia, especially in the northern area, is too hot during summer and cold during winter. Control of the indoor environment in a smart home is a preferable choice that can reduce power consumption to operate heating, ventilation, and air-conditioning. Machine learning algorithms have been used to predict physical variables of indoor environment, such as temperature and humidity. The model can be trained, learn, and make predictions using historical data. Machine learning techniques can automate temperature monitoring and control. This paper proposes an algorithm that combines Artificial Neural Networks (ANNs) and GridSearchCV to predict physical variables in indoor environments in Saudi Arabia. GridSearchCV was utilized to tune the parameters of the machine learning algorithm. The assessment of the proposed algorithm involved its performance comparison to state-of-the-art machine learning algorithms. A real-world dataset was generated to estimate the performance of the considered algorithms. The room data were collected every 5 min for 31 days during July 2022. The dataset contains 6 columns and 8,910 records from 6 sensors (timestamps, light, temperature, humidity, pressure, and altitude). Random Forest (RF), Decision Tree (DT), and ANN methods were compared with the proposed algorithm. The RF had the highest R^2 value of 0.84 and the lowest Mean Square Error (MSE) of 0.43. The DT achieved an R^2 score of 0.78, while the ANN achieved R^2 score of 0.61, MSE of 1.04, and Mean Absolute Error (MAE) of 0.75. The proposed algorithm achieved an R^2 of 0.69, MSE of 0.87, and MAE of 0.67.

Keywords-artificial neural networks; deep learning; machine learning algorithms; smart home

I. INTRODUCTION

Older people are more susceptible to heat-related illnesses, such as heat exhaustion, heat stroke, and heat cramps. Most older adults take medication that make them more sensitive to heat [1]. Clinical reports have noted that people over 60 have a higher potential to experience heat-related illnesses [2]. Moreover, by 2030, more than 60% of the world's population will live in urban environments, and will be exposed to more increased temperatures than rural areas [3]. According to the statistics, the energy consumption of buildings is around 40% of the total global energy consumption, and buildings emit more than 36% of the total global emissions contributing to carbon dioxide emissions [4, 5]. Consequently, optimal energy use can facilitate the reduction of energy consumption and CO₂ emissions [5]. In building technology, the essential goals are reducing energy consumption in buildings and developing a quality indoor environment [10, 11]. Physical variable quality in indoor environments is essential, especially for older people and patients. Hospitals, homes, and healthcare facilities require sufficient monitoring and control of the indoor environment [6,

7]. Occupancy prediction in indoor environments is essential to reduce energy consumption and maintain occupant comfort. Moreover, occupant behavior is important in physical variables, such as temperature, lighting, heating, and air-conditioning [8, 9].

Machine learning algorithms are essential techniques for efficaciously dealing with numerous data. Different applications can predict load forecasting, solar energy, wind energy, and other variables [12, 13]. Even though elevated temperatures are potentially harmful, to the best of our knowledge, there is a lack of an accurate temperature prediction model for older people. For temperature prediction in healthcare, machine learning algorithms are crucial in controlling and monitoring temperatures in smart home environment and can improve accuracy and efficiency in temperature prediction. There are many advantages in the use of machine learning algorithms, such as:

- Recognizing the home temperature for older adults: accurately predicting temperature variations in homes can help the timely intervention of healthcare providers.

- Reducing physical visits: Fewer visits from healthcare providers to check the temperature in the home are needed. In smart homes, smart sensors produce massive sensory data that assists stakeholders to make decisions.

Numerical weather prediction is important to provide vital knowledge for inhabitants and healthcare providers [13]. Therefore, machine learning algorithms have been used to rectify weather forecasting shortcomings with imperfect physical parameterizations [13]. However, measuring the indoor environment for smart homes is challenging. Traditional measurement solutions are insufficient because internal measurements vary and require a massive volume of data for long periods [6, 14]. Internet of Things (IoT) applications provide real-time measurements, and each sensor generates numerous data to apply prediction models to reduce energy consumption and improve environmental quality [6, 14]. Machine learning techniques can predict the indoor environment to control and monitor the temperature and humidity variables and the lighting and Heating, Ventilation, and Air-Conditioning (HVAC) systems. In addition, machine learning techniques are applied to recognize the inhabitants' daily living habits in an indoor environment, monitor health and wellness (especially in older people), and give recommendations for healthcare providers to make decisions. To the best of our knowledge, there is a shortage of investigations and assessments of machine learning algorithms in the scope of indoor environment. Although existing studies on indoor temperature prediction have made good progress, there are many challenges and constraints, such as ensuring data and obtaining the highest performance accuracy [15].

A comprehensive review of the use of ANNs for air temperature prediction is presented in [16]. Different types of ANNs are introduced, such as Recurrent Neural Networks (RNNs) and Long-Short Term Memory (LSTM) Networks. ANNs have the ability to learn complex non-linear relationships from the input data set. The conventional prediction models and their limitations when using ANNs for meteorological data are pointed out. Authors in [17] used ANNs to predict the temperatures inside the heated foil tunnel. The air temperature inside the tunnel plays an important role in managing the heating system, improving crop yield, and reducing energy consumption. The efficiency of ANN models is discussed in forecasting temperature changes inside the heated foil tunnel and the optimized conditions in different industries depend on the monitored environment. Root Mean Square Error (RMSE) was used to evaluate the performance of the ANN model. Authors in [18] proposed a probabilistic ANN that was a combination of yield prediction and optimization, using the hybrid Moth Flame Optimization algorithm with a machine learning algorithm for crop recommendations and yield predictions. The utilized dataset collected information on weather, rain and fertilizer in India. They achieved a prediction accuracy of 99.67% and an R2 value of 98.82%. In [19], DT was applied to develop an automatic irrigation system. The authors installed many sensors under the roots to measure the moisture content in the soil. The sensors recorded the temperature and humidity that were done every 10 min to control the automatic irrigation based on the humidity level to save irrigation water on the agricultural field. The result

achieved prediction accuracy that approached 97.86%. Authors in [20] used ANN algorithms to predict maize yield. Six datasets were used to validate the proposed algorithm. The datasets comprise of features of climate, soil water balance and agricultural features of maize for two crop years. The ANN models were able to capture the non-linear effects of the datasets. However, the amount of data for the agricultural environment was limited. RMSE, MSE, and R^2 were used to calculate performance of ANN models for predicting maize productivity. Authors in [21] used deep learning to assess the energy consumption of buildings. They pointed out that prediction is crucial in estimating the energy consumption of buildings to save energy. A benchmark dataset was used to estimate their models. The dataset was generated from the electricity consumption of a single occupant measured every minute. FCRBM achieved the best prediction compared to the other machine learning algorithms. Their article gives an overview of different machine learning algorithms that are used to predict air temperature. They show that the deep learning approaches have a lower MSE compared to conventional ANNs. Authors in [22] emphasized the importance of feature selection, data pre-processing and algorithm evaluation in temperature prediction. They used ANN algorithms to predict the air temperature in August in the southern part of the Iberian Peninsula. Authors in [23] used the ERA5 Reanalysis dataset, which integrates monitoring data with numerical weather forecasting techniques to capture temporal and spatial patterns of temperature. The results show that the XAI models are able to efficiently predict air temperature one month ahead with reasonable performance and interpretability.

The main contributions of this paper are: a review of the previous studies on indoor temperature prediction and a comprehensive evaluation of the state-of-the-art machine learning algorithms accompanied with a comparison with the proposed algorithm on the prediction of indoor temperature in a smart home domain based on certain evaluation metrics.

II. METHODOLOGY

The input data are preprocessed (cleaned) and missing data values are filled in. Real datasets must be used to evaluate the performance of the machine learning algorithms. Only a few real datasets for indoor environments are publicly available. Datasets representing indoor environments with occupancy performance metrics must be generated to assess the models. Predicting indoor environments involves feature importance scores and machine learning algorithms.

III. REAL INDOOR ENVIRONMENTAL DATASET

The dataset includes measurements of an indoor environment from 6 sensors (timestamp, light, temperature, humidity, pressure, and altitude) from July 2022, consisting of 8,910 real-time records [24]. Samples were captured every 5 min. Table I shows a sample of the dataset.

IV. EXPERIMENT DESIGN

This section proposes an estimation algorithm for the prediction of temperature. The methodology involves two approaches. In the first approach, the models learn from data without tuning and optimizing the parameters, as in the second

approach. The dataset was divided into 80% training and 20% testing subsets. All the input features were fed to the machine learning algorithms with the exception of the timestamp column, because some of the models were not able to handle this data type. A diagram of the proposed model is shown in Figure 1. The proposed model was trained and tested using the optimization method. GridSearchCV was used for finding the best hyperparameters. The GridSearchCV and ANN approach was implemented to improve the performance of the model and apply feature selection to the training and testing sets. The proposed model learns from the data using hyperparameters in the training phase and performs predictions using the trained GridSearch with ANNs. To assess the prediction of the performance of the model, GridSearchCV was employed from Sklearn to carry out a grid search to find the best combination of hyperparameters and estimate classifier's performance using cross-validation.

TABLE I. A DATASET SAMPLE

Time	Light	Temperature	Humidity	Pressure	Attitude
12:02:08	357	37.5	11	89535	1031.2
12:07:09	352	37.6	11	89536	1031.1
12:12:10	349	37.6	11	89532	1031.47
12:33:54	339	37.7	11	89510	1033.5
12:38:55	336	37.7	11	89512	1033.31
12:43:56	330	37.7	11	89507	1033.77
12:48:57	325	37.7	11	89501	1034.33
12:53:58	322	37.9	11	89502	1034.23
12:58:59	315	37.9	11	89499	1034.51
13:04:01	312	37.9	11	89492	1035.15
13:09:02	304	38	12	89495	1034.88
13:14:03	300	38	12	89487	1035.61

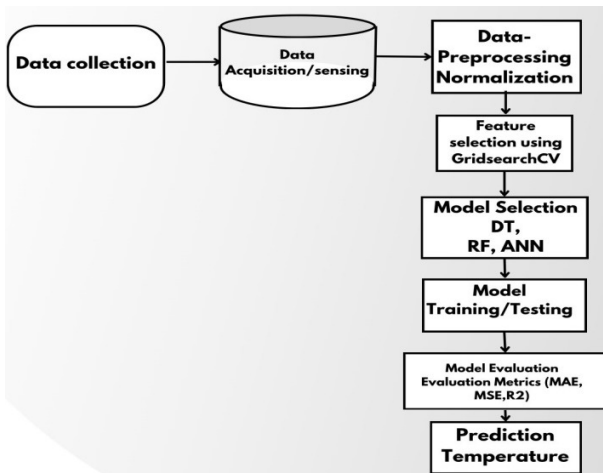


Fig. 1. The proposed optimized model for temperature prediction.

This paper used the Keras library to create a sequential model that illustrates fully-connected layers in the ANN. The model was compiled with the MSE loss function and the Adam optimizer. The model learned from the training data and used GridSearchCV to optimize the parameters. The model was validated with 5-fold cross-validation on the predictions of the testing data. MSE, MAE, and R^2 were used as evaluation metrics. Finally, the loss curve of the model during the training phase was obtained and a comparison between the predicted and the actual values was made.

V. PERFORMANCE METRICS

The MAE between the predictions and the target values is used in regression analysis to measure the model prediction accuracy. The closer the MAE is to zero, the more accurate the predictions and the nearer they are to the target values. MSE is the average of the squared error difference between the predicted and target values. The closer the MSE is to zero, the more accurate the predictions are.

R^2 (or the coefficient of determination) measures the estimated value accuracy calculated in a regression model, determining the ratio of variance in the dependent variable that the independent variable can predict. The R^2 value is always between 0 and 1. An R^2 of 1 indicates that the regression predictions are appropriate for the data, whereas R^2 values closer to zero indicate a greater spread [25].

A. Decision Tree (DT) Algorithm

The DT regression method is a supervised learning algorithm for classification and regression problems. It is a nonparametric learning algorithm that predicts continuous values instead of categorical target variables. The tree is structured by recursively dividing each node's data into smaller parts. It aims to reduce the MSE for each child node. The DT model can deal with nonlinear relationships between features and labels and uses categorical and numerical value datasets. However, DTs can be prone to overfitting and cannot handle missing values. Thus, the dataset must be prepared. The tree has nodes linked by edges, and one root node. Each node is connected to a parent node. Figure 2 shows the tree data structure in Python. This method assigns nodes to a root and adds smaller subsets as child nodes [26, 27].

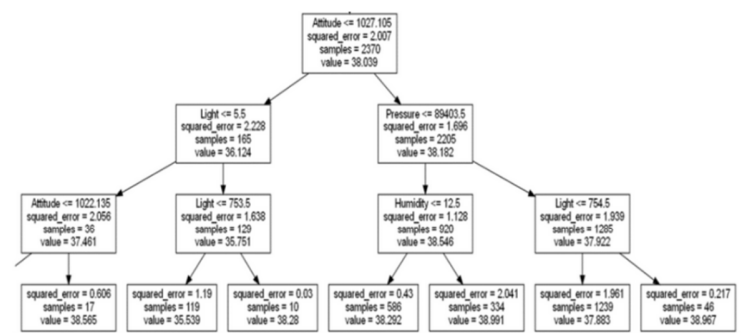


Fig. 2. Export graphviz flow chart.

B. Random Forest (RF)

The RF regression method is a supervised learning algorithm used in regression problems. It is an ensemble of DTs to predict a continuous value. Multiple trees form an RF using a random subset of data and features, and the trees are generated from the samples of features. At each node, a sample of features is selected for splitting, assisting in decreasing overfitting [27, 28].

C. Artificial Neural Network

An ANN is a feedforward system for classification and regression problems inspired by the human brain. Every node is a neuron. ANNs consist of interconnected neurons arranged

in layers, which include several hidden layers between the input and the output. Each layer is completely connected to the following layer by weights, and the output is passed to an activation function. Regression ANNs predict a dependent variable value based on the input characteristics (independent variables) [29].

D. GridSearchCV

GridSearchCV (Grid Search Cross-Validation) is a technique used to tune hyperparameters and model selection [30]. It automates the process to find an optimal set of parameters and avoid overfitting. The proposed model learns from the data in the training phase and the prediction is performed using GridSearchCV with the proposed model. The hyperparameters used in ANNs are batch_size, epochs, hidden_layers, learning_rate and units.

VI. RESULTS AND DISSCUSION

This study demonstrates the effectiveness of the proposed model in predicting the temperature indoor environment and highlights its potential for optimizing states in diverse indoor environments in a smart home. The combination of ANN and GridSearchCVh in the proposed model showed good error evaluation, which produced the lowest MSE in the considered dataset. The proposed model obtained optimal hyperparameters determined through GridSearchCV and k-fold cross-validation in the dataset. At first, experiments were performed without the use of Grid Search. The accuracy of the machine learning algorithms was evaluated using MAE, MSE, and R^2 on the dataset and was determined during the training phase (Table II). Table III shows the results achieved by 5-fold cross-validation. Table IV shows the results revealing the competitive performance of the proposed model and the state-of-the-art machine learning algorithms during the testing phase. The proposed model obtained lower error than the ANN model with normal parameters. The Grid Search was performed for optimizing the proposed algorithm and various parameter settings were used, as shown in Table V. The proposed model's performance had an MSE of roughly 0.87 for testing datasets, MAE of is around 0.67, and R^2 of approximately 0.69, demonstrating its ability to predict temperature accurately. Figures 3-6 show the cross-validated predictions during training and the predicted values versus the actual temperature values.

TABLE II. PERFORMANCE RESULTS OF THE EVALUATED MODELS DURING THE TRAINING PHASE

Model	MSE	MAE	R^2
RF	0.1538	0.2058	0.9425
DT	0.1157	0.1206	0.9568
ANN	1.0026	0.7425	0.6256
Proposed	0.8937	0.6979	0.6663

TABLE III. PERFORMANCE RESULTS OF THE EVALUATED MODELS BY 5-FOLD CROSS-VALIDATION

Model	MSE	MAE	R^2
RF	0.4696	0.3865	0.8254
DT	0.6165	0.3982	0.7709
ANN	1.0739	0.7494	0.5885
Proposed	1.0391	0.7451	0.6141

TABLE IV. RESULTS OF THE PERFORMANCE OF ALL EVALUATED MODELS BY TESTING PHASE.

Model	MSE	MAE	R^2
RF	0.4398	0.3819	0.8404
DT	0.5970	0.3961	0.7834
ANN	1.0496	0.7523	0.6193
Proposed	0.87075	0.6789	0.6913

TABLE V. GRIDSEARCHCV SETTINGS USED

Parameter	Value
batch_size	16
epochs	100
hidden_layers	3
learning_rate	0.01
units	3

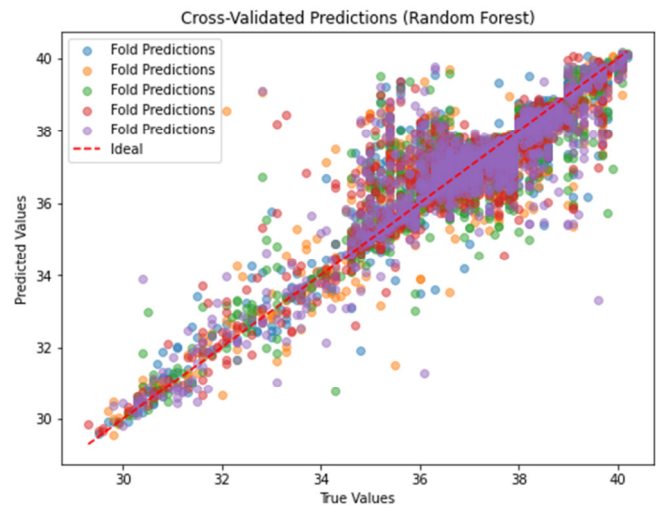


Fig. 3. Cross-validated prediction (RF).

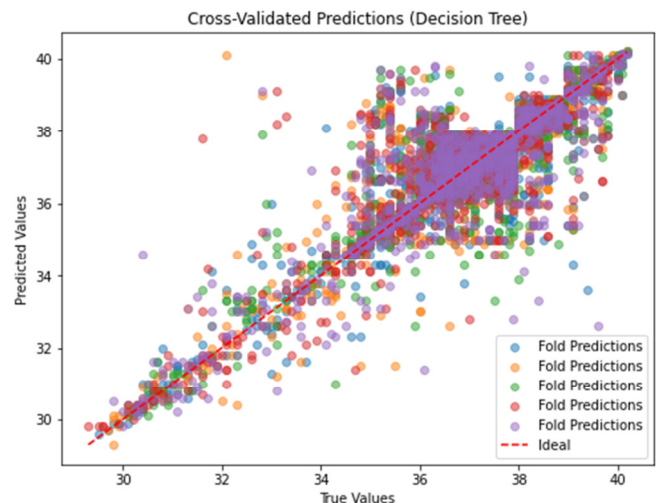


Fig. 4. Cross-validated rediction (DT).

Figures 7-10 show the actual against the predicted temperature during the testing phase. The plot of the training and validation loss obtained during training is shown in Figure 11.

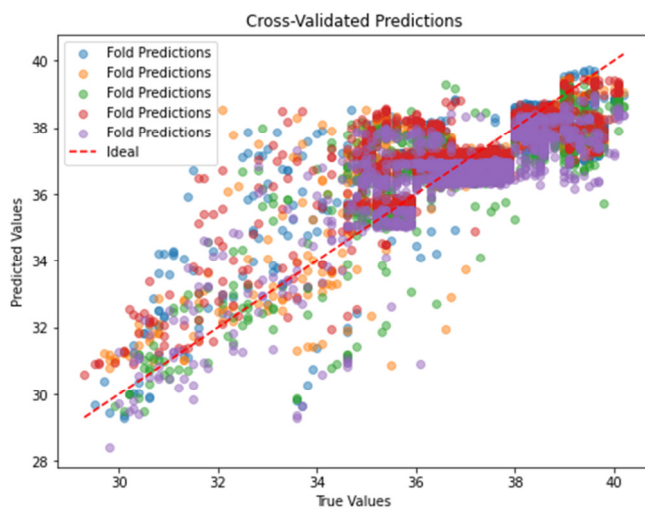


Fig. 5. Cross-validated prediction (ANN).

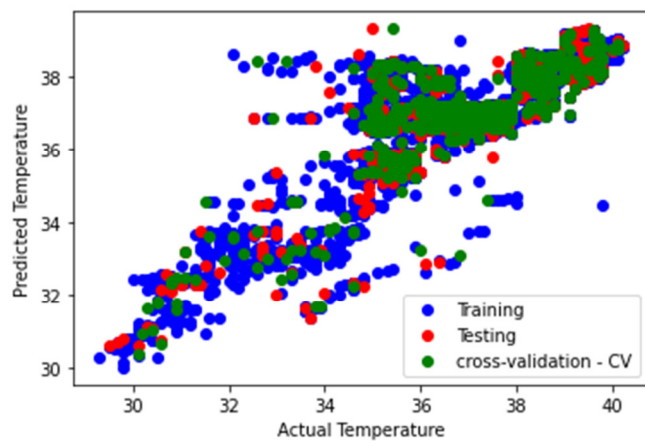


Fig. 6. Cross-validated prediction (proposed model).

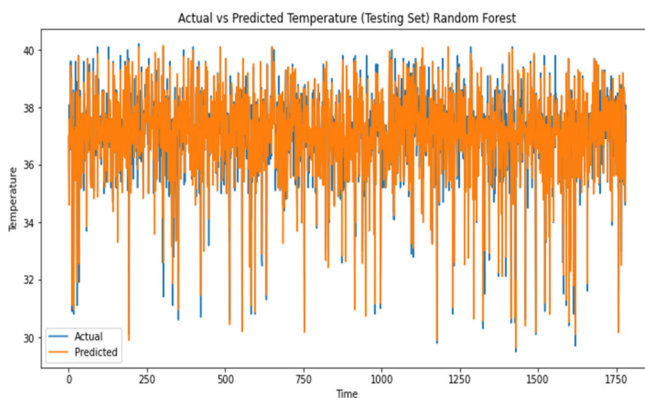


Fig. 7. Actual vs predicted temperature (RF).

VII. CONCLUSION

This paper presents a useful application for predicting the indoor temperature in a smart home. This study proposes an optimized machine learning algorithm to predict the temperature in the home environment. Various machine

learning algorithms were evaluated on a real-world dataset generated from a smart room. Learning from historical indoor physical measurements and predicting future temperature is crucial for smart homes.

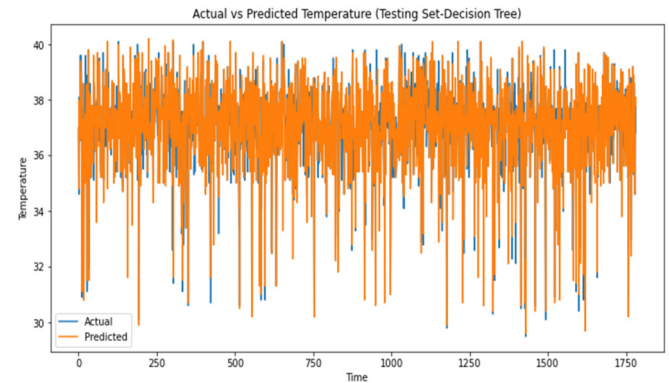


Fig. 8. Actual vs predicted temperature (DT).

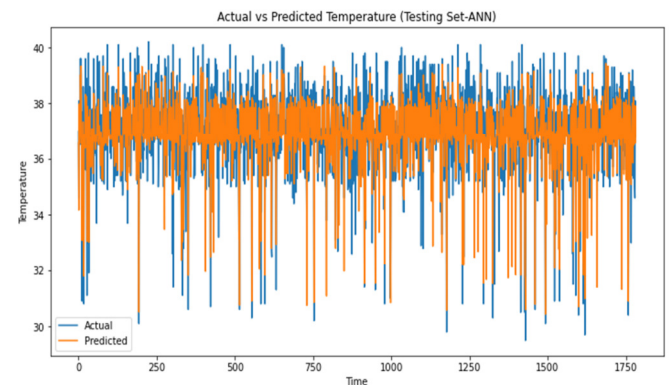


Fig. 9. Actual vs predicted temperature (ANN).

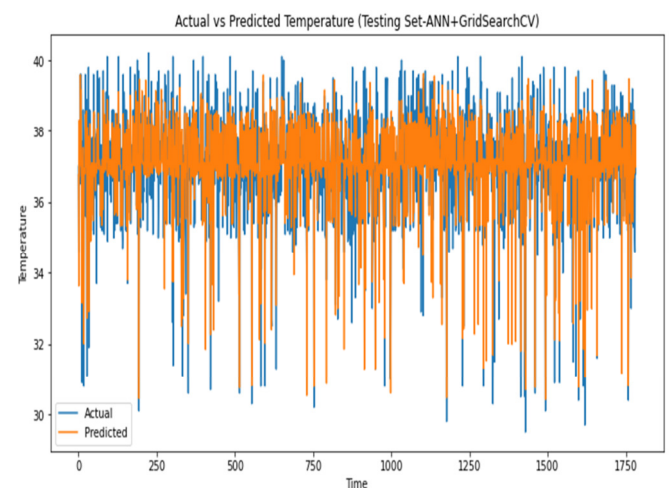


Fig. 10. Actual vs predicted temperature (proposed model).

The considered algorithms were evaluated with R^2 , MSE, and MAE metrics, including the proposed algorithm against RF, DT, and ANN. RF achieved the highest R^2 score of 0.84

whereas the proposed algorithm obtained an R^2 of 0.69, MSE of 0.87, and MAE of 0.67. The ANN algorithm achieved an R^2 score of 0.61, MSE of 1.04, and MAE of 0.75. However, the results showed that the proposed algorithm obtained better result than the unoptimized ANN. Future work will focus on using reinforcement learning for the prediction of the indoor temperature.

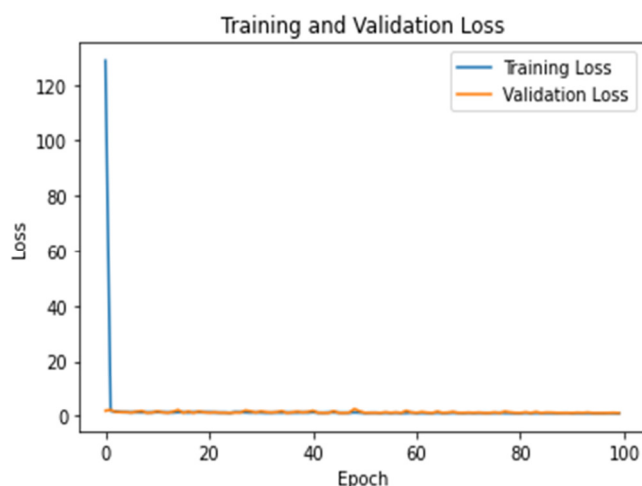


Fig. 11. Training and validation loss (proposed model).

REFERENCES

- [1] R. K. Nath and H. Thapliyal, "Wearable Health Monitoring System for Older Adults in a Smart Home Environment," in *2021 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, Tampa, FL, USA, Jul. 2021, pp. 390–395, <https://doi.org/10.1109/ISVLSI51109.2021.00077>.
- [2] B. N. Balmain, S. Sabapathy, M. Louis, and N. R. Morris, "Aging and Thermoregulatory Control: The Clinical Implications of Exercising under Heat Stress in Older Individuals," *BioMed Research International*, vol. 2018, 2018, Art. no. 8306154, <https://doi.org/10.1155/2018/8306154>.
- [3] S. Vulova, F. Meier, D. Fenner, H. Nouri, and B. Kleinschmit, "Summer Nights in Berlin, Germany: Modeling Air Temperature Spatially With Remote Sensing, Crowdsourced Weather Data, and Machine Learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5074–5087, 2020, <https://doi.org/10.1109/JSTARS.2020.3019696>.
- [4] Z. Li, P. Wang, J. Zhang, and S. Mu, "A strategy of improving indoor air temperature prediction in HVAC system based on multivariate transfer entropy," *Building and Environment*, vol. 219, Jul. 2022, Art. no. 109164, <https://doi.org/10.1016/j.buildenv.2022.109164>.
- [5] W. Zhang, Y. Wu, and J. K. Calautit, "A review on occupancy prediction through machine learning for enhancing energy efficiency, air quality and thermal comfort in the built environment," *Renewable and Sustainable Energy Reviews*, vol. 167, Oct. 2022, Art. no. 112704, <https://doi.org/10.1016/j.rser.2022.112704>.
- [6] M. Marzouk and M. Atef, "Assessment of Indoor Air Quality in Academic Buildings Using IoT and Deep Learning," *Sustainability*, vol. 14, no. 12, Jan. 2022, Art. no. 7015, <https://doi.org/10.3390/su14127015>.
- [7] A. Fonseca, I. Abreu, M. J. Guerreiro, C. Abreu, R. Silva, and N. Barros, "Indoor Air Quality and Sustainability Management—Case Study in Three Portuguese Healthcare Units," *Sustainability*, vol. 11, no. 1, Jan. 2019, Art. no. 101, <https://doi.org/10.3390/su11010101>.
- [8] S. H. Ryu and H. J. Moon, "Development of an occupancy prediction model using indoor environmental data based on machine learning techniques," *Building and Environment*, vol. 107, pp. 1–9, Oct. 2016, <https://doi.org/10.1016/j.buildenv.2016.06.039>.
- [9] C. Wang, D. Yan, H. Sun, and Y. Jiang, "A generalized probabilistic formula relating occupant behavior to environmental conditions," *Building and Environment*, vol. 95, pp. 53–62, Jan. 2016, <https://doi.org/10.1016/j.buildenv.2015.09.004>.
- [10] C. Li, H. Zhu, X. Lian, Y. Liu, X. Li, and Y. Feng, "Study of 'time-lag' of occupant behavior occurrences for establishing an occupant-centric building control system," *Building and Environment*, vol. 216, May 2022, Art. no. 109005, <https://doi.org/10.1016/j.buildenv.2022.109005>.
- [11] S.-Y. Lin, S.-C. Chiu, and W.-Y. Chen, "Simple automatic supervisory control system for office building based on energy-saving decoupling indoor comfort control," *Energy and Buildings*, vol. 86, pp. 7–15, Jan. 2015, <https://doi.org/10.1016/j.enbuild.2014.09.066>.
- [12] S. Lee and D.-H. Choi, "Reinforcement Learning-Based Energy Management of Smart Home with Rooftop Solar Photovoltaic System, Energy Storage System, and Home Appliances," *Sensors*, vol. 19, no. 18, Jan. 2019, Art. no. 3937, <https://doi.org/10.3390/s19183937>.
- [13] J. A. Weyn, D. R. Durran, and R. Caruana, "Can Machines Learn to Predict Weather? Using Deep Learning to Predict Gridded 500-hPa Geopotential Height From Historical Weather Data," *Journal of Advances in Modeling Earth Systems*, vol. 11, no. 8, pp. 2680–2693, 2019, <https://doi.org/10.1029/2019MS001705>.
- [14] S. Serroni, M. Arnesano, L. Violini, and G. M. Revel, "An IoT measurement solution for continuous indoor environmental quality monitoring for buildings renovation," *Acta IMEKO*, vol. 10, no. 4, pp. 230–238, Dec. 2021, https://doi.org/10.21014/acta_imeko.v10i4.1182.
- [15] T. Xing, K. Sun, and Q. Zhao, "MITP-Net: A deep-learning framework for short-term indoor temperature predictions in multi-zone buildings," *Building and Environment*, vol. 239, Jul. 2023, Art. no. 110388, <https://doi.org/10.1016/j.buildenv.2023.110388>.
- [16] T. T. K. Tran, S. M. Bateni, S. J. Ki, and H. Vosoughifar, "A Review of Neural Networks for Air Temperature Forecasting," *Water*, vol. 13, no. 9, Jan. 2021, Art. no. 1294, <https://doi.org/10.3390/w13091294>.
- [17] S. Francik and S. Kurpaska, "The Use of Artificial Neural Networks for Forecasting of Air Temperature inside a Heated Foil Tunnel," *Sensors*, vol. 20, no. 3, Jan. 2020, Art. no. 652, <https://doi.org/10.3390/s20030652>.
- [18] S. R. Gopi and M. Karthikeyan, "Effectiveness of Crop Recommendation and Yield Prediction using Hybrid Moth Flame Optimization with Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11360–11365, Aug. 2023, <https://doi.org/10.48084/etasr.6092>.
- [19] A. H. Blasi, M. A. Abbadi, and R. Al-Huweimel, "Machine Learning Approach for an Automatic Irrigation System in Southern Jordan Valley," *Engineering, Technology & Applied Science Research*, vol. 11, no. 1, pp. 6609–6613, Feb. 2021, <https://doi.org/10.48084/etasr.3944>.
- [20] P. V. D. de Souza, L. P. de Rezende, A. P. Duarte, and G. V. Miranda, "Maize Yield Prediction using Artificial Neural Networks based on a Trial Network Dataset," *Engineering, Technology & Applied Science Research*, vol. 13, no. 2, pp. 10338–10346, Apr. 2023, <https://doi.org/10.48084/etasr.5664>.
- [21] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, "Deep learning for estimating building energy consumption," *Sustainable Energy, Grids and Networks*, vol. 6, pp. 91–99, Jun. 2016, <https://doi.org/10.1016/j.segan.2016.02.005>.
- [22] J. Cifuentes, G. Marulanda, A. Bello, and J. Reneses, "Air Temperature Forecasting Using Machine Learning Techniques: A Review," *Energies*, vol. 13, no. 16, Jan. 2020, Art. no. 4215, <https://doi.org/10.3390/en13164215>.
- [23] A. M. Gómez-Orellana, D. Guijo-Rubio, J. Pérez-Aracil, P. A. Gutiérrez, S. Salcedo-Sanz, and C. Hervás-Martínez, "One month in advance prediction of air temperature from Reanalysis data with eXplainable Artificial Intelligence techniques," *Atmospheric Research*, vol. 284, Mar. 2023, Art. no. 106608, <https://doi.org/10.1016/j.atmosres.2023.106608>.
- [24] T. Alshammari, R. A. Ramadan, and A. Ahmad, "Temporal Variations Dataset for Indoor Environmental Parameters in Northern Saudi Arabia," *Applied Sciences*, vol. 13, no. 12, Jan. 2023, Art. no. 7326, <https://doi.org/10.3390/app13127326>.

- [25] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ. Computer Science*, vol. 7, 2021, Art. no. e623, <https://doi.org/10.7717/peerj-cs.623>.
- [26] M. Abdurrohman, A. G. Putrada, and M. M. Deris, "A Robust Internet of Things-Based Aquarium Control System Using Decision Tree Regression Algorithm," *IEEE Access*, vol. 10, pp. 56937–56951, 2022, <https://doi.org/10.1109/ACCESS.2022.3177225>.
- [27] "scikit-learn: machine learning in Python — scikit-learn 1.4.1 documentation." [Online]. Available: <https://scikit-learn.org/stable/>.
- [28] S. Murugan, B. M. Kumar, and S. Amudha, "Classification and Prediction of Breast Cancer using Linear Regression, Decision Tree and Random Forest," in *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*, Mysore, India, Sep. 2017, pp. 763–766, <https://doi.org/10.1109/CTCEEC.2017.8455058>.
- [29] H. Salem, A. E. Kabeel, E. M. S. El-Said, and O. M. Elzeki, "Predictive modelling for solar power-driven hybrid desalination system using artificial neural network regression with Adam optimization," *Desalination*, vol. 522, Jan. 2022, Art. no. 115411, <https://doi.org/10.1016/j.desal.2021.115411>.
- [30] T. Yan, S.-L. Shen, A. Zhou, and X. Chen, "Prediction of geological characteristics from shield operational parameters by integrating grid search and K-fold cross validation into stacking classification algorithm," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 14, no. 4, pp. 1292–1303, Aug. 2022, <https://doi.org/10.1016/j.jrmge.2022.03.002>.

Numerical and Experimental Investigation of Performance and Flooding Phenomena of a PEM Fuel Cell with and without Micro-Porous Layers

Nguyen Ha Hiep

Institute of Vehicle and Energy Engineering, Le Quy Don Technical University, Vietnam
hahiepshippower@lqdtu.edu.vn

Vu Duong

School of Engineering and Technology, Duy Tan University, Vietnam
duongvuaustralia@gmail.com (corresponding author)

Received: 2 February 2024 | Revised: 13 February 2024 | Accepted: 21 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6996>

ABSTRACT

This work presents the results of manufacturing a single Proton Exchange Membrane Fuel Cell (PEMFC) with Micro-Porous Layers (MPLs) and an active area of 25 cm², and the experimental study required to build its polarization curve. Based on the physical model data, a numerical model of this PEMFC is created in the ANSYS PEM Fuel Cell module. Numerical simulations were performed with boundary conditions consistent with the experimental conditions on the test station. The calculation and experimental result comparison of the polarization curves for voltages ranging from 0.29 V to 0.94 V proved that the utilized numerical model is highly reliable. The simulation of PEMFC without MPLs was conducted according to such stable input parameters and boundary conditions. The results show that the PEMFC performance decreases significantly due to the flooding phenomenon inside PEMFC without MPLs compared to PEMFC with MPLs. Such phenomena are challenging to observe experimentally. Numerical modeling can be further used to optimize the fuel cell components.

Keywords-*fuel cell; micro-porous layer; current; flooding phenomenon; water content*

I. INTRODUCTION

On December 13, 2023, the COP28 adopted a climate agreement calling for the world to move away from fossil fuels [1]. Alternative energy sources, such as hydrogen, and replacing burning fossil fuels energy conversion methods with electrochemical processes implemented in Fuel Cells (FCs), are encouraged to be utilized. Currently, there are several FC types: Proton Exchange Membrane (PEM), phosphoric acid, solid oxide, alkaline, and molten carbonate [2]. PEMFCs are more popular because they provide higher efficiency and energy density with low weight, volume, and noise and have a more straightforward structure [3]. PEMFCs can be used as a power source for FC electric vehicles and can be combined with batteries in hybrid electric vehicles [3]. In several studies, researchers have used FCs in power systems for vehicles [4-11], and physicochemical phenomena in FC and the influence of the components of FCs on their properties have been investigated [12-16]. Research on perfecting FCs and FC propulsion systems is increasingly vibrant and developing.

Through a PEMFC, the electric energy is derived from hydrogen and oxygen by electrochemical processes. The reagents are transferred to the reaction site in a PEMFC via supply channels. The hydrogen/oxygen reactions producing the oxidation of hydrogen and the reduction of oxygen are promoted in the Catalytic Layers (CLs). For a profitable PEMFC process, delivering a consistent allocation of the reagents in the electrode region and shunting to the collector of all catalyst particles is required to reduce ohmic losses. This procedure is continued by the Gas Diffusion Layer (GDL), which is placed suitably from the bipolar plate to the CL. An extra Micro-Porous Layer (MPL) can be adhered to the inner surface of a GDL, contacting with the CL to enhance the performance of GDL. The contacting resistance between the GDL and the CL can be decreased due to the MPL leads [14]. For PEMFC development, properties such as CLs, PEM, GDLs, MPLs, the number and configuration of channels on flow plates, operating conditions, etc. are the target of researchers to highlight their role in influencing PEMFC performance and internal phenomena.

This work aims to investigate the influence of the MPLs on PEMFC performance and the flooding phenomenon in PEMFCs. To do this, a PEMFC prototype was built using commercial materials and tests were conducted to measure its performance. The creation and calibration of this prototype model to obtain a polarization curve close to the experimental polarization curve and evaluate the model's reliability were conducted with the ANSYS PEM Fuel Cell Module (17.2 version, including upgrades). Then, the ANSYS model was numerically simulated and the flooding phenomenon inside the PEMFC was studied for both cases, with and without MPLs.

II. FABRICATING AND TESTING A FUEL CELL

A single PEMFC with MPL and an activated area of 25 cm² was developed using PEM (Nafion 212), GDLs + MPL (Sigracet 35BC), PTFE loading of 5 wt% and 25 wt% of GDL and MPL, respectively, 0.8 porosity [17, 18], CLs with 40 wt% Pt/C, and 40 wt% Nafion, and CL binder (Solution D520 5 wt%). The Membrane Electrode Assembly (MEA) was fabricated by hot pressing two electrodes onto both membrane sides at a temperature of 403 K and a pressure of 1.96 MPa for 180 s. The thickness of the MEA was 662 μm after hot pressing. The MEA was installed between two graphite bipolar plates (Graphite HK3), and 0.1 mm thick PTFE gaskets were placed between the contact surfaces. The graphite bipolar plates have serpentine grooves 1 mm wide and 1 mm deep. The MEA and graphite bipolar plate assembly was installed between the collector plates and then secured to the two housing plates with eight bolts (Figure 1). The PEMFC prototype was installed to the test bench, as shown in Figure 2. The mass flow controllers regulate the oxygen and hydrogen gas from the cylinders and pass through the hot water in humidifiers. Oxygen and hydrogen feed PEMFC has a constant relative humidity of 100% guaranteed by temperature regulation of the humidifier. A polarization curve of the PEMFC was evaluated in the Potentiostat.

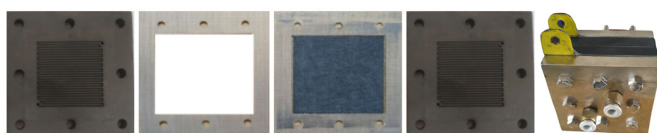


Fig. 1. The components and single PEMFC prototype.

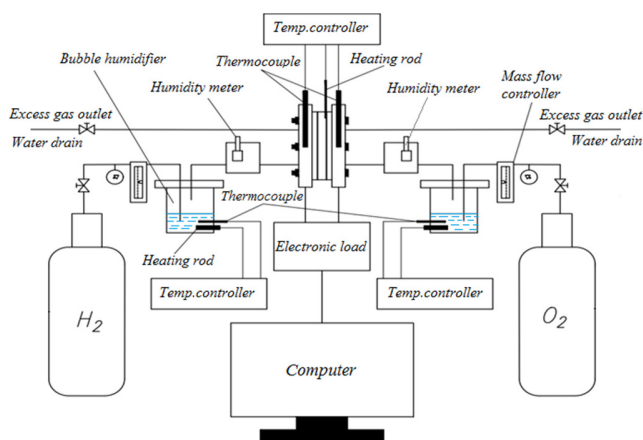


Fig. 2. The testing diagram of the PEMFC.

III. NUMERICAL STUDY

Assumptions and mathematical models for ANSYS PEMFC modeling are presented in [19]. Based on the physical model (Figure 1), a numerical model of PEMFC was created in the ANSYS. The geometric and mesh models (Figure 3) was built from about $7.4 \cdot 10^6$ elements and underwent many adjustments to obtain the desired convergence based on the meshing techniques mentioned in [20-24]. The simulation settings for the unsteady coefficients are pressure of 0.7, momentum of 0.3, and species of 0.95. The other parameters are equal to 1. The SIMPLE algorithm relates pressure and velocity to reference pressure discretization. A second-order bottom-up scheme was applied to discretize the other physicochemical and electrochemical quantities inside PEMFC. Simulations were performed until the current densities on the electrodes differed by less than 10^{-5} mA/cm², and the other criteria converged with 0.001% error. Cathode current densities were calculated by adjusting voltages from 0.29 to 0.94 V.

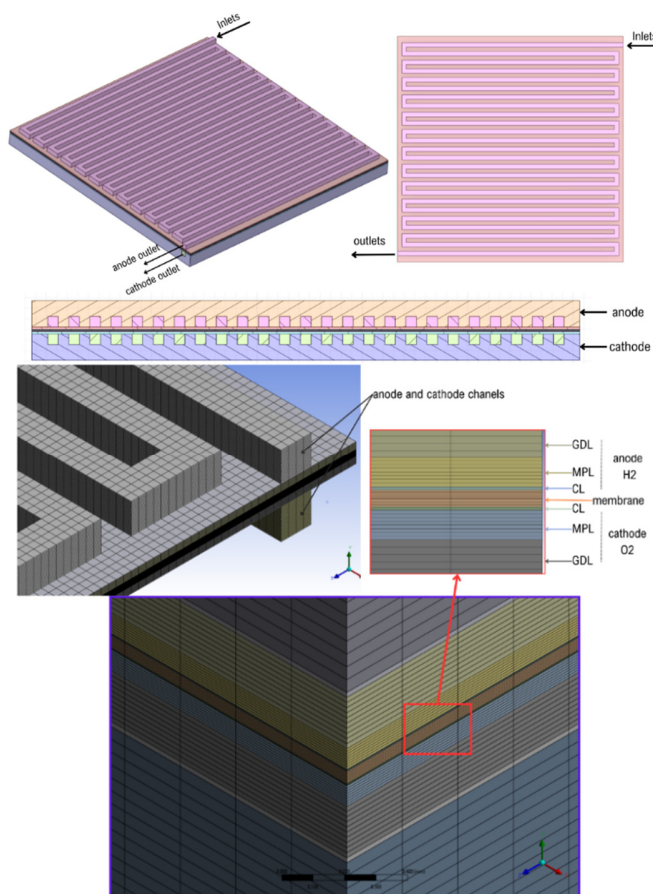


Fig. 3. The PEMFC geometric and meshing model

The input parameters for the simulation model are determined by (1)-(7).

The porosity of GDLs is determined by [25]:

$$\varepsilon_{GDL} = 1 - (1 - \varepsilon_0) \frac{\delta_0}{\delta} \quad (1)$$

The gas permeability of GDLs [m^2] is evaluated by [26]:

$$K_{GDL} = \frac{\varepsilon_{GDL}}{8(\ln \varepsilon_{GDL})^2} \frac{(\varepsilon_{GDL} - \varepsilon_p)^{(\alpha+2)} d_f^2}{(1 - \varepsilon_p)^\alpha [(\alpha + 1)\varepsilon_{GDL} - \varepsilon_p]^2} \quad (2)$$

The MPL porosity and absolute permeability are also defined by (1) - (2), respectively. The CL volume fraction is calculated by [27]:

$$\varepsilon_{CL} = 1 - L_{Pt/C} - L_i \quad (3)$$

where $L_{Pt/C}$ is determined by [15] (4) and f is the mass fraction of Pt, calculated by (5) [28]:

$$L_{Pt/C} = \frac{m_{pt}}{t_{CL}} \left(\frac{1}{\rho_{pt}} + \frac{1-f}{f} \frac{1}{\rho_c} \right) \times 10^1 \quad (4)$$

$$f = m_{pt} / (m_{pt} + m_c) \quad (5)$$

The ionomer content of Nafion in CLs is [29]:

$$W_{L_i} = V_N \rho_N \cdot 0.05 \quad (6)$$

$$L_i (\text{wt}\%) = 100 \cdot W_{L_i} / (W_{L_i} + W_{Pt/C})$$

The CL absolute permeability is [25]:

$$K_{CL} = \frac{d_p^2 \varepsilon_{CL}^3}{150(1 - \varepsilon_{CL})^2} \quad (7)$$

Table I presents the PEMFC component properties and the boundary conditions. The boundary conditions for the simulation model are consistent with the experimental conditions.

TABLE I. INPUT PARAMETERS AND BOUNDARY CONDITIONS

Parameter [Unit]	Symbol	Value	Reference
GDL (with MPL) thickness [m]	δ_0	$325 \cdot 10^{-6}$	Fabricated
GDL porosity	ε_0	0.8	Fabricated
GDL thickness after hot pressing [m]	δ	$275 \cdot 10^{-6}$	Fabricated
GDL fiber diameter [m]	d_f	$6.4 \cdot 10^{-6}$	[17]
Constants	ε_p, α	0.037; 0.661	[26]
Platinum loading of CL [mg/cm^2]	m_{pt}	0.4	Fabricated
Carbon loading of CL [mg/cm^2]	m_c	0.6	Fabricated
CL initial thickness [m]	t_{CL}	10^{-5}	Fabricated
CL area [cm^2]	A_{CL}	25	Fabricated
Pt/C weight of CL [g]	$W_{Pt/C}$	25	Fabricated
Pt and C densities [kg/m^3]	ρ_{pt}, ρ_c	21450; 1800	[30]
Particle diameter of CL [m]	d_p	$2 \cdot 10^{-7}$	[31]
Nafion solution volume [ml]	V_N	363	Fabricated
Nafion solution density [g/cm^3]	ρ_N	0.92	[29]
Relative humidity in electrodes [%]	-	100	Tested
Pressure in electrodes [Pa]	-	$0.2 \cdot 10^5$	Tested
Anode H_2 stoichiometry	-	1.5	Tested
Anode O_2 stoichiometry	-	2.0	Tested
Anode mass flow rate [g/s]	-	$7.8 \cdot 10^{-4}$	Tested
Cathode mass flow rate [g/s]	-	$6.4 \cdot 10^{-3}$	Tested
Temperature in electrodes [K]	-	348	Tested

IV. RESULTS AND DISCUSSION

The graphs of polarization curves, showing the voltage dependence on current density based on the experimental and simulation results are shown in Figure 4.

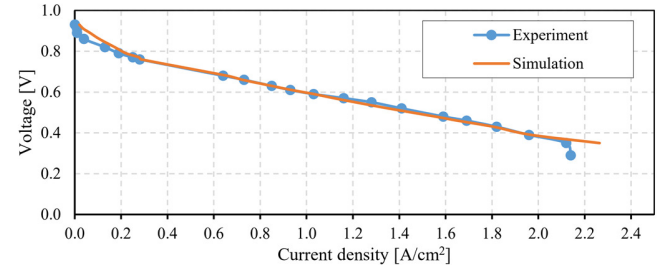
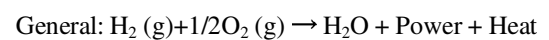
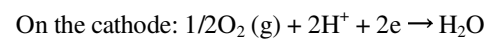
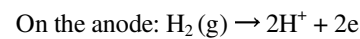


Fig. 4. Voltage-current density characteristics of PEMFC.

At lower current densities, the voltage decreases due to the activation energy required for the electrochemical reaction at the catalyst surface. The charge transfer coefficient and exchange current density mainly affect this region. After several adjustments based on the reference data in [15, 20, 22, 30], the studied model was appropriately adjusted. The voltage decreases linearly with current density at moderate current densities due to ohmic losses. The numerical and experimental result comparison shows that they are almost identical. Therefore, the input parameters have been calculated and selected correctly. At high current density, the rate of water formation at the cathode increases and causes flooding, and the formed water blocks the electrode holes, reducing the oxidant flow into the CL and causing the polarization curve to decrease sharply. The data in [19] determine the initial equilibrium water content at water saturation, adjusting the simulated polarization curve in this region to coincide with the experimental one.

Figure 4 shows a difference between the simulated and the experimental results at low current density, but the deviation is not more than 3%. At average current density, the simulation and experimental results coincide entirely. At the maximum current density, there is quite a difference between the results. However, they have high convergence overall within the research voltage range, from 0.29 to 0.94 V (as assumed in Section III), proving that the geometric model, input parameters, and boundary conditions are determined accurately and the built numerical model is highly reliable. In this study, we focus on the problem of PEMFC component flooding as a stated goal.

Figure 5 shows the water content of the main components, such as the cathode CL, membrane, and anode CL of the PEMFC with MPLs at a voltage of 0.39 V. Water is produced at the cathode as a result of the reaction in the PEMFC as follows:



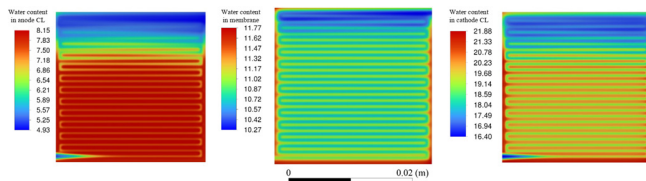


Fig. 5. Water contents on main PEMFC components

Water generated at the cathode will be discharged through grooves on the bipolar graphite plates of the cathode. Some of the water will penetrate the membrane, flow to the anode, and then exit through the groove on the bipolar graphite plates of the anode. If the water is not drained in time, it will cause flooding inside the PEMFC, which will hinder the reaction between oxygen and hydrogen and degrade the performance of the PEMFC. Good drainage is possible thanks to the groove structure of the graphite bipolar plates or the use of MPLs. Figure 6 illustrates water content in PEM for PEMFC with MPL and without MPL at 0.39 V. Without MPL, the amount of water stagnant in the membrane is more significant than when there is MPL. This shows that MPL improves water transport and increases the performance of PEMFCs (Figure 7).

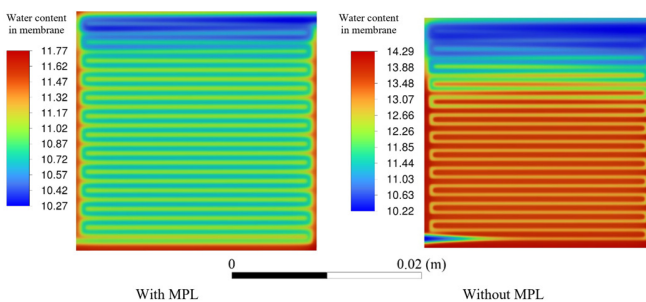


Fig. 6. Water contents in PEMs of the PEMFC with and without MPL.

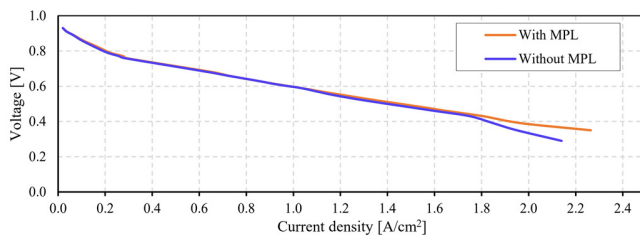


Fig. 7. Numerical polarization curves with and without MPL.

V. CONCLUSIONS

In this study, a PEMFC prototype with MPLs was fabricated. The polarization curve of this PEMFC was developed during testing on the test station with calibrated equipment. The first novelty of this research was that a fuel cell prototype was built with good performance using suitable commercial components, considering parameters such as assembly force, pressing time, and pressing temperature. The second novelty is the provision of an accurate experimental data set that serves as a carefully calibrated input data set for calibrating similar models in practical fuel cell design and manufacturing.

The ANSYS software was applied to develop the numerical model of the PEMFC prototype. The numerical polarization curve obtained from the calibrated model agrees with the experimental data. The numerical model simulates the PEMFC without MPLs to evaluate the flooding phenomenon inside the PEMFC that is difficult to measure experimentally. The combination of experimental and numerical studies is critical in fuel cell development. Experiments are used to create samples and measure their characteristics. In contrast, numerical simulation reduces the number of samples that need to be manufactured, reduces the cost and time of complex experiments, and allows for observation and analysis of phenomena within the fuel cell, such as water transport and flooding. Observing these phenomena during experiments requires expensive measuring equipment, is difficult to implement, and is often impossible.

Another point of interest is that the published works of other authors [20-24] often consider GDL without MPLs when modeling with the ANSYS PEM Fuel Cell module. Our study highlighted this issue by adding MPLs. That is an inevitable matter in modern fuel cell technology.

A final novelty is that oxygen was used in this study as an oxidizing agent when modeling with the ANSYS PEM Fuel Cell module, which typically uses air as an oxidizer [20-24]. The use of oxygen requires certain representations of the input data.

In future studies, the numerical model can optimize the fuel cell properties, including Pt/C loading, Nafion loading in CLs, the porosity and permeability of GDL with or without MPLs, and the membrane thickness for developing and manufacturing high-performance fuel cells.

ACKNOWLEDGMENT

This research is funded by Le Quy Don Technical University Research Fund under the grand number 2021.Q.04.

REFERENCES

- [1] "COP28 Agreement Signals 'Beginning of the End' of the Fossil Fuel Era," UNFCCC. https://unfccc.int/news/cop28-agreement-signals-beginning-of-the-end-of-the-fossil-fuel-era?gad_source=1&gclid=Cj0KCQiAxOauBhCaARIsAEBUSQRWASTPzA5aFB6zP-sEO_gwJE2Kg-3YAaLVnNk4Hynw2uGUT4TyIYgaAlcGEALw_wcB.
- [2] M. Marappan *et al.*, "Performance Studies of Proton Exchange Membrane Fuel Cells with Different Flow Field Designs – Review," *The Chemical Record*, vol. 21, no. 4, pp. 663–714, 2021, <https://doi.org/10.1002/tcr.202000138>.
- [3] C. Spiegel, *Designing and Building Fuel Cells*. New York, NY, USA: McGraw Hill, 2007.
- [4] A. Khadhraoui, T. Selmi, and A. Cherif, "Energy Management of a Hybrid Electric Vehicle," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 8916–8921, Aug. 2022, <https://doi.org/10.48084/etasr.5058>.
- [5] S. N. M., O. Tremblay, and L.-A. Dessaint, "A generic fuel cell model for the simulation of fuel cell vehicles," in *IEEE Vehicle Power and Propulsion Conference*, Dearborn, MI, USA, Sep. 2009, pp. 1722–1729, <https://doi.org/10.1109/VPPC.2009.5289692>.
- [6] S. N. Motapon, O. Tremblay, and L.-A. Dessaint, "Development of a generic fuel cell model: application to a fuel cell vehicle simulation," *International Journal of Power Electronics*, vol. 4, no. 6, pp. 505–522, Jan. 2012, <https://doi.org/10.1504/IJPELEC.2012.052427>.

- [7] C. Spiegel, *PEM Fuel Cell Modeling and Simulation Using MATLAB*. London, UK: Academic Press, 2008.
- [8] S. Javadpoor and D. Nazarpour, "Modeling a PV-FC-Hydrogen Hybrid Power Generation System," *Engineering, Technology & Applied Science Research*, vol. 7, no. 2, pp. 1455–1459, Apr. 2017, <https://doi.org/10.48084/etasr.760>.
- [9] J. Chakravorty, J. Saraswat, and V. Bhatia, "Modeling a Distributed Power Flow Controller with a PEM Fuel Cell for Power Quality Improvement," *Engineering, Technology & Applied Science Research*, vol. 8, no. 1, pp. 2585–2589, Feb. 2018, <https://doi.org/10.48084/etasr.1807>.
- [10] S. Barhate, R. Mudhalwadkar, and S. Madhe, "Fault Detection Methods Suitable for Automotive Applications in Proton Exchange Fuel Cells," *Engineering, Technology & Applied Science Research*, vol. 12, no. 6, pp. 9607–9613, Dec. 2022, <https://doi.org/10.48084/etasr.5262>.
- [11] M. A. Biberici and M. B. Celik, "Dynamic Modeling and Simulation of a PEM Fuel Cell (PEMFC) during an Automotive Vehicle's Driving Cycle," *Engineering, Technology & Applied Science Research*, vol. 10, no. 3, pp. 5796–5802, Jun. 2020, <https://doi.org/10.48084/etasr.3352>.
- [12] H. Bouzidi, L. Otmani, R. Doufnoune, L. Zerroual, and D. Benachour, "Influence of Membrane Type on Some Electrical Properties of a Single Microbial Fuel Cell," *Engineering, Technology & Applied Science Research*, vol. 12, no. 3, pp. 8492–8499, Jun. 2022, <https://doi.org/10.48084/etasr.4813>.
- [13] G. Sasikumar, J. W. Ihm, and H. Ryu, "Dependence of optimum Nafion content in catalyst layer and microporous layer on the performance of PEMFC," *Journal of Power Sources*, vol. 132, no. 1, pp. 11–17, May 2004, <https://doi.org/10.1016/j.jpowsour.2003.12.060>.
- [14] C.-J. Tseng and S.-K. Lo, "Effects of microstructure characteristics of gas diffusion layer and microporous layer on the performance of PEMFC," *Energy Conversion and Management*, vol. 51, no. 4, pp. 677–684, Apr. 2010, <https://doi.org/10.1016/j.enconman.2009.11.011>.
- [15] N. Khajeh-Hosseini-Dalasm, M. J. Kermani, D. G. Moghaddam, and J. M. Stockie, "A parametric study of cathode catalyst layer structural parameters on the performance of a PEM fuel cell," *International Journal of Hydrogen Energy*, vol. 35, no. 6, pp. 2417–2427, Mar. 2010, <https://doi.org/10.1016/j.ijhydene.2009.12.111>.
- [16] N. H. Hiep, N. Q. Quan, G. H. Thai, and P. T. San, "Numerical Modeling and Experimental Validation of a Hydrogen/Oxygen Fuel Cell for Underwater Vehicle Applications," *SAE International*, Warrendale, PA, USA, SAE Technical Paper 2023-01-5053, Aug. 2023, <https://doi.org/10.4271/2023-01-5053>.
- [17] A. Pfrang, S. Didas, and G. Tsotridis, "X-ray computed tomography of gas diffusion layers of PEM fuel cells: Segmentation of the microporous layer," *Journal of Power Sources*, vol. 235, pp. 81–86, Aug. 2013, <https://doi.org/10.1016/j.jpowsour.2013.01.179>.
- [18] F. Aldakheel *et al.*, "Gas permeability, wettability and morphology of gas diffusion layers before and after performing a realistic ex-situ compression test," *Renewable Energy*, vol. 151, pp. 1082–1091, May 2020, <https://doi.org/10.1016/j.renene.2019.11.109>.
- [19] *ANSYS FLUENT Theory Guide 17.2*. Canonsburg, PA, USA: ANSYS, 2022.
- [20] M. Arif, S. C. P. Cheung, and J. Andrews, "A systematic approach for matching simulated and experimental polarization curves for a PEM fuel cell," *International Journal of Hydrogen Energy*, vol. 45, no. 3, pp. 2206–2223, Jan. 2020, <https://doi.org/10.1016/j.ijhydene.2019.11.057>.
- [21] I. Nitta, S. Karvonen, O. Himanen, and M. Mikkola, "Modelling the Effect of Inhomogeneous Compression of GDL on Local Transport Phenomena in a PEM Fuel Cell," *Fuel Cells*, vol. 8, no. 6, pp. 410–421, 2008, <https://doi.org/10.1002/fuce.200700058>.
- [22] M. Arif, S. C. P. Cheung, and J. Andrews, "Numerical investigation of effects of different flow channel configurations on the 100 cm² PEM fuel cell performance under different operating conditions," *Catalysis Today*, vol. 397–399, pp. 449–462, Aug. 2022, <https://doi.org/10.1016/j.cattod.2021.07.016>.
- [23] M. Arif, S. C. P. Cheung, and J. Andrews, "Influence of Hydrophobicity and Porosity of the Gas Diffusion Layer on Mass Transport Losses in PEM Fuel Cells: A Simulation Study Supported by Experiments," *Energy & Fuels*, vol. 34, no. 10, pp. 13010–13022, Oct. 2020, <https://doi.org/10.1021/acs.energyfuels.0c02596>.
- [24] W. Xiaofei, Q. Yang, Z. Zhigang, and X. Liusheng, "Investigation of Ship Vibration Effects on the Gas Distribution and Output Voltage of a Proton Exchange Membrane Fuel Cell," *ACS Omega*, vol. 7, no. 24, pp. 20569–20583, Jun. 2022, <https://doi.org/10.1021/acsomega.2c00273>.
- [25] J. Zhao, S. Shahgaldi, I. Alaefour, Q. Xu, and X. Li, "Gas permeability of catalyzed electrodes in polymer electrolyte membrane fuel cells," *Applied Energy*, vol. 209, pp. 203–210, Jan. 2018, <https://doi.org/10.1016/j.apenergy.2017.10.087>.
- [26] J. T. Gostick, M. W. Fowler, M. D. Pritzker, M. A. Ioannidis, and L. M. Behra, "In-plane and through-plane gas permeability of carbon fiber electrode backing layers," *Journal of Power Sources*, vol. 162, no. 1, pp. 228–238, Nov. 2006, <https://doi.org/10.1016/j.jpowsour.2006.06.096>.
- [27] A. A. Shah, G.-S. Kim, P. C. Sui, and D. Harvey, "Transient non-isothermal model of a polymer electrolyte fuel cell," *Journal of Power Sources*, vol. 163, no. 2, pp. 793–806, Jan. 2007, <https://doi.org/10.1016/j.jpowsour.2006.09.022>.
- [28] M. Secanell, K. Karan, A. Suleman, and N. Djilali, "Multi-variable optimization of PEMFC cathodes using an agglomerate model," *Electrochimica Acta*, vol. 52, no. 22, pp. 6318–6337, Jun. 2007, <https://doi.org/10.1016/j.electacta.2007.04.028>.
- [29] G. Adilbish and Y.-T. Yu, "Effect of the Nafion content in the MPL on the catalytic activity of the Pt/C-Nafion electrode prepared by pulsed electrophoresis deposition," *International Journal of Hydrogen Energy*, vol. 42, no. 2, pp. 1181–1188, Jan. 2017, <https://doi.org/10.1016/j.ijhydene.2016.09.143>.
- [30] S. Li, J. Yuan, M. Andersson, G. Xie, and B. Sundén, "Influence of anisotropic gas diffusion layers on transport phenomena in a proton exchange membrane fuel cell," *International Journal of Energy Research*, vol. 41, no. 14, pp. 2034–2050, 2017, <https://doi.org/10.1002/er.3763>.
- [31] S. Li, J. Yuan, G. Xie, and B. Sundén, "Effects of agglomerate model parameters on transport characterization and performance of PEM fuel cells," *International Journal of Hydrogen Energy*, vol. 43, no. 17, pp. 8451–8463, Apr. 2018, <https://doi.org/10.1016/j.ijhydene.2018.03.106>.

Thermal Potential of a Twin-Screw Compressor as Thermoelectric Energy Harvesting Source

Claudia Savescu

Romanian Research and Development Institute for Gas Turbines COMOTI, Romania | Politehnica Bucharest, Romania
claudia.borzea@comoti.ro

Valentin Petrescu

Romanian Research and Development Institute for Gas Turbines COMOTI, Romania | Politehnica Bucharest, Romania
valentin.petrescu@comoti.ro

Daniel Comeaga

National University of Science and Technology, Politehnica Bucharest, Romania
daniel.comeaga@upb.ro

Razvan Carlanescu

Romanian Research and Development Institute for Gas Turbines COMOTI, Romania
razvan.carlanescu@comoti.ro

Mihaela Roman

Romanian Research and Development Institute for Gas Turbines COMOTI, Romania | Politehnica Bucharest, Romania
mihaela.roman@comoti.ro

Daniel Lale

Romanian Research and Development Institute for Gas Turbines COMOTI, Romania | Politehnica Bucharest, Romania
daniel.lale@comoti.ro

Andrei Mitru

Romanian Research and Development Institute for Gas Turbines COMOTI, Romania
andrei.mitru@comoti.ro (corresponding author)

Received: 19 October 2023 | Revised: 25 January 2024 | Accepted: 28 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6417>

ABSTRACT

This study evaluates the potential of a twin-screw compressor as a heat source to harness thermal energy. Thermoelectric generators are a feasible solution for microenergy harvesting from waste heat based on the Seebeck effect. Thermographic infrared images of the compressor were used to assess potential installation spots. The physical mounting of the thermoelectric modules must consider certain hindering aspects. At first, the compressor skid is subject to standards and authorizations for its components, leaving only a couple of spots for screw-mounted module installations. Another inconvenience is the bonds in any thermoelectric material causing them not to withstand lateral mechanical stress in other directions except the *c*-axis perpendicular to the layers. Therefore, vibration measurements have to be performed beforehand. Numerical simulations were conducted, relying on the acquired thermoelectric modules as well as on the temperature and vibration data measured on the compressor. The thermoelectric generators

studied are part of a multisource piezoelectric and thermoelectric energy harvesting system under research and development.

Keywords-thermoelectric generator; energy harvesting; waste heat; twin-screw compressor; thermographic imaging

I. INTRODUCTION

Waste heat is inherent in any industrial process and machinery that produces mechanical work. Almost half of the heat generated is wasted and often dissipated in the surrounding environment, thus limiting the efficiency of the system [1, 2]. Thermoelectric Generators (TEGs) [3] are considered a promising and feasible solution for energy harvesting applications from waste heat, under the condition that a temperature gradient is provided across the Thermoelectric Module (TEM). TEGs can generate electric power by converting temperature gradients (or differences) into electric energy, relying on the Seebeck effect, as opposed to the inverse Peltier effect that is used in Thermoelectric Coolers (TEC) and limits TEG efficiency. This alternative microenergy harvesting source has received increased research interest in the last two decades [4-11]. The maximum power output provided in [11] corresponds to approximately 200°C temperature differences, which is impossible to achieve in practical applications with industrial machinery and difficult to maintain even in the laboratory for a long time. To maintain the temperature difference for a longer period, it is necessary to ensure forced convection cooling of the cold side, which implies devices such as air fans or water circulation pumps, leading to higher power consumption than the TEGs would be able to produce. Most studies dealt with a theoretical analysis or assessed only one thermocouple, not a whole module with more coupled n-p semiconductor pairs [5, 6, 9, 12-14]. Simulations were usually conducted over a more or less arbitrary temperature range. In [15], the thermo-mechanical integrity and power-generation performance of BiTe TEGs were investigated with two pairs of thermocouples with rectangular, prism, and cylindrical legs at a temperature difference of 100°C in unconstrained mechanical conditions, numerically analyzing power output, conversion efficiencies, and thermal stresses on the legs. In [16], a TEG module with 127 thermocouples with different leg geometries was simulated in Ansys and analyzed under convective and isoflux boundary conditions, providing a copper heat sink to cool the cold side. In [17], an equivalent circuit model was developed in a SPICE program for the thermal and electrical design of thermoelectric-based systems. The model was exploited to obtain the electrical power output when the heat exchange between the hot and cold sources is not ideal, mainly because it is maximum with a load resistance matching the TEG's equivalent resistance, which is not just the electrical resistance but also depends on the thermal contact resistance. In [18], a coupled multiphysics simulation model was developed in Modelica/Dymola to ensure the TEGs' optimal thermal connection and find an optimal design and electrical connection for the energy harvesting system. The aim was to establish an optimal thermoelectric system for a given heat source and objectives. Finally, in [19], emergent thermoelectric materials were summarized and recent advances in energy harvesting from waste heat were discussed.

This study proposes a realistic approach on the thermal potential of screw compressors for thermoelectric energy harvesting. Using Finite Element Method (FEM) simulations, conducted in COMSOL Multiphysics, the expected voltage and power outputs of TEGs were evaluated by introducing the temperature ranges measured via infrared thermographic scanning. The vibration amplitudes of the compressor were also considered in a frequency domain study conducted to reveal the von Mises stress exerted on the TEG module. As far as is known, there is no study reporting mechanical stress induced by vibrating machinery.

II. COMPRESSOR THERMAL IMAGES

Thermographic images were taken with an infrared camera on a CU90G compressor on a test bench [20]. Figure 1 presents the skid with its main components. The thermal scans were obtained in November 2021, when atmospheric temperatures in Bucharest, Romania ranged from 10 to 18°C. This is confirmed by the lower limit captured by the thermal imager in the figures hereinafter. In the image processing software, the emissivity of 0.95 was set and the background temperature was considered 18°C as the ambient maximum. Figure 2 shows the temperatures recorded on the cast iron housing of the compressor. As can be observed, the maximum temperature (hot point) reached in the target area was 67.3°C on the compressor housing. The cold point was 17.3°C.



Fig. 1. Compressor skid within the test bed.

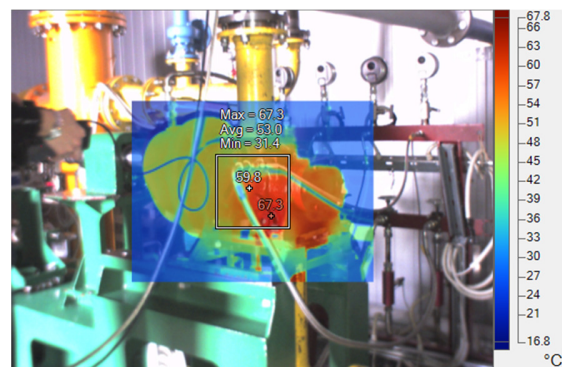


Fig. 2. Thermographic image of the compressor unit.

The compressor skid is subjected to regulations and standardizations, and drilling holes can render void its certifications. However, screw mounting with an applied torque of ~ 1.25 Nm per screw (0.128 kgm) is recommended [21] for satisfactory heat transfer, leaving only a couple of potential mounting places. The best option was the cover of the multiplier gearbox between the motor and the compressor (Figure 3 left), which is splashed with hot oil from within, making it ideal for this study's purposes. Figure 4 displays a thermographic image of the gearbox cover, showing a maximum temperature of 41.8°C .



Fig. 3. Optimal screw mounting spot for the TEGs.

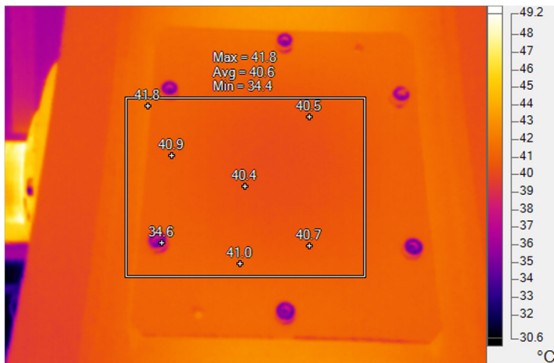


Fig. 4. Surface temperature range of the gearbox cover.

III. THERMOELECTRIC CONSIDERATIONS

Thermoelectric generators' efficiency depends on the temperature difference between the hot and cold sides, being related to the material's so-called figure of merit (zT), which depends on the material's conductivity according to:

$$zT = \frac{\varepsilon^2 \cdot \kappa_e \cdot T}{\kappa_d} \quad (1)$$

where zT [ND] is the figure of merit, ε [V/K] is the Seebeck coefficient, κ_e [$1/\Omega\text{m}$] is the electrical conductivity, T [K] is the temperature, and κ_d [W/mK] is the thermal conductivity. The Seebeck effect supposes a voltage being created across the TEM's sides, due to the temperature difference across a conductor or semiconductor. Two dissimilar (semi)conductors joined together represent a circuit's junctions. The A and B (semi)conductors are electrically connected in series and thermally in parallel. A junction is hot, with temperature T_{hot} ,

while the other one is cold, with temperature T_{cold} . The Seebeck voltage at the junctions of an open circuit is given by:

$$V = (\varepsilon_A - \varepsilon_B) \cdot (T_{hot} - T_{cold}) \Leftrightarrow \varepsilon_{AB} = \frac{V}{\Delta T} \quad (2)$$

where V [V] is the voltage, and ΔT [K] is the temperature difference. The thermoelectric modules require a very good contact for thermal transfer, most preferably with screw mounting. A spare cover was physically used to make holes and mount the thermoelectric harvesting components, as depicted in Figure 5. The original cover will be replaced in the upcoming experiments.

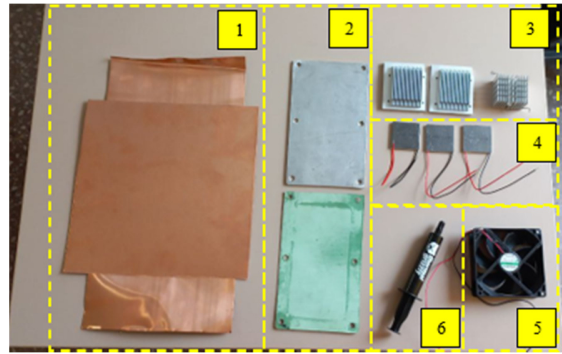


Fig. 5. Materials required for experiment: (1) Copper sheets for hot side contact; (2) Executed aluminum cover and sheet for sealing gasket; (3) Heat sinks; (4) Thermoelectric modules; (5) Fan; (6) Thermally conductive paste.

IV. SIMULATION MODEL

FEM simulations were performed in the COMSOL Multiphysics® software [22] using a multiphysics approach for a complete evaluation of both heat transfer and electric field problems. The simulation model was designed in an application builder, fully parametrized for thermoelectric coolers (TECs) [23], easing the geometry design for thermoelectric devices in general. The thermoelectric legs were assigned bismuth telluride (Bi_2Te_3) material of n-type and p-type, copper was used for the connection electrodes, and alumina ceramic was utilized for the plates acting as hot and cold sides. The overall dimensions of the chosen and subsequently acquired devices, of $40 \times 40 \times 5$ [mm] [11] were considered in the simulation model. Detailed constructive elements and parameters are provided in [13]. The following physics were coupled in the simulation model:

- **Heat Transfer in Solids** to numerically model the heat transfer. In solid domains, the temperature equation corresponds to the differential form of the Fourier law and may present additional factors, such as heat sources [12]. The general formulas employed in the simulation software are:

$$\rho C_p \mathbf{u} \cdot \nabla T + \nabla \cdot \mathbf{q} = Q + Q_{ted} \quad (3)$$

$$\mathbf{q} = -k \nabla T$$

where ρ [kg/m^3] is the density, C_p [J/(kgK)] is the specific heat, \mathbf{u} is the temperature T [K] or the electric potential V [V], Q [J] is the heat source, $k \mid \kappa_d$ [W/(m·K)] is the thermal conductivity, Q_{ted} [ND] is the thermoelastic dissipation

quality factor, and q [W/m²] is the conduction heat flux. A temperature T1 of 30°C was declared for the cold side, and a temperature T2 of 80°C was initially set for the hot side, as exhibited in Figure 6.

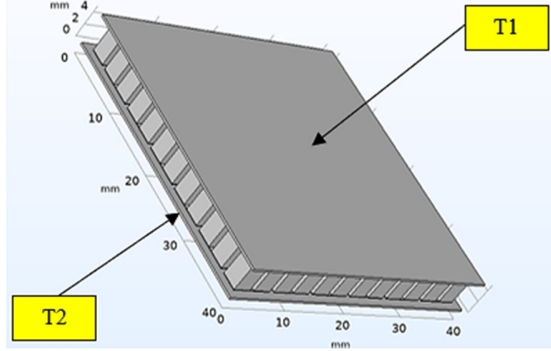


Fig. 6. TEG geometry and boundary conditions in heat transfer.

- **Electric Currents:** this physics interface, relying on Ohm's law, solves the equations for current conservation, considering the scalar value of the electric potential as a dependent variable. It is used for computing electric field, current, and electric potential distributions within conducting materials [12]. A reference impedance of 50 Ω is set by default in the simulation program, and the current conservation condition was applied to the Bi₂Te₃ elements with the initial value of voltage $V = 0$ V.

$$\nabla \cdot \mathbf{J} = Q_{j,v}$$

$$\mathbf{J} = \sigma \mathbf{E} + \mathbf{J}_e \quad (4)$$

$$\mathbf{E} = -\nabla V$$

where \mathbf{J} [A] is the electric current, $Q_{j,v}$ [A] is the current source, σ [S/m] is the electric conductivity, \mathbf{E} [V/m] is the electric field, \mathbf{J}_e [A] is the external current source, and V [V] is the electric potential. Ground (GND) and Floating Potential (FP) boundary conditions were applied, as in Figure 7, on the back surfaces in contact with the hot side of the module.

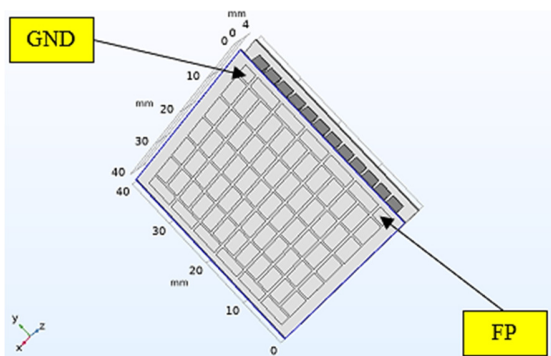


Fig. 7. TEG geometry showing the electric boundary conditions.

- **Thermoelectric Effect** multiphysics interface links together heat transfer in solids and electric currents' physics interfaces for a complete modeling of Peltier-Seebeck-

Thomson thermoelectric effects, solving the equation system for a complete evaluation of both thermal and electrical parameters describing the multiphysics interrelations [12]:

$$P = ST$$

$$q = PJ \quad (5)$$

$$J_e = -\sigma \nabla T$$

where P [V] is the Peltier coefficient, S [V/K] is the Seebeck coefficient, T [K] is the temperature, q [W/m²] is the conduction heat flux, J [A] is the induced electric current, J_e [A] is the external current source, and σ [S/m] is the electric conductivity. It is worth providing the main properties of the bismuth telluride material used by the simulation program (assumed the same for n- and p-type), as shown in Table I.

TABLE I. PROPERTIES OF BISMUTH TELLURIDE MATERIAL

T [K]	S [V/K]	k [W/m/K]	σ [S/m]
200	168e-6	24e-1	1.4286e5
250	192e-6	19e-1	1.1111e5
300	210e-6	16e-1	0.86957e5
350	225e-6	16e-1	0.71429e5
400	237e-6	17.5e-1	0.58824e5

- **Electromagnetic Heating** multiphysics coupling adds the source term for resistive (ohmic) heating, Q_e [W/m³] as:

$$\rho C_p \frac{\partial T}{\partial t} - \nabla \cdot (k \nabla T) = Q_e \quad (6)$$

where ρ [kg/m³] is density, C_p [J/(kgK)] is the specific heat, T [K] is the temperature, t [s] is the time, k [W/(m·K)] is the thermal conductivity, and Q_e [W/m³] is the resistive heating.

- **Electrical Circuit** is an extra physics interface that needs to be added when connecting external circuit components. A 2.4 k Ω load resistor was connected in parallel to obtain current and power information besides voltage. This value was not randomly chosen but represents the power required by the wireless sensor nodes [24], which need to be powered within the practical application using microenergy harvesting devices.
- **Solid Mechanics** was added to account for the vibration speeds on all axes, as measured on the CU90G compressor driven at 2500 rpm [25]. The boundary selected was the exterior hot side surface, the same as T2 (Figure 6). The input values are provided in Table II.

TABLE II. PRESCRIBED VELOCITY

Direction	Symbol	Value [U.M.]
Prescribed on x direction	v_x	0.998 [mm/s]
Prescribed on y direction	v_y	2.985 [mm/s]
Prescribed on z direction	v_z	1.692 [mm/s]

Solid Mechanics physics interface requires filling in additional material properties for Bismuth Telluride, namely Young's modulus (E) and Poisson's ratio (ν). Since these

characteristics are not given by the manufacturer, E was set to 32 GPa and ν was set to 0.25, as found in [26].

V. RESULTS AND DISCUSSION

A. Stationary Study - Thermal and Electrical Evaluation

A stationary study was conducted to assess the thermoelectric parameters. A parametric sweep for T_{hot} , range(40,5,80) degC, was set in to grasp the influence of the temperature difference on the electric output. Figure 8 shows the temperature gradient, from 30 to 80°C, across the thermoelectric module. The direction of the conductive heat flux is represented by arrows (from the hot side towards the cold side). Figure 9 portrays the electric potential. TEGs' voltage output is in VDC, hence the variation is from 0 V on the boundary where the ground condition was applied, to the maximum value of 1.55 V a temperature difference of 50°C, on the boundary declared with floating potential. If the floating potential is connected to a circuit or a value for electric current is set, this one acts like a terminal.

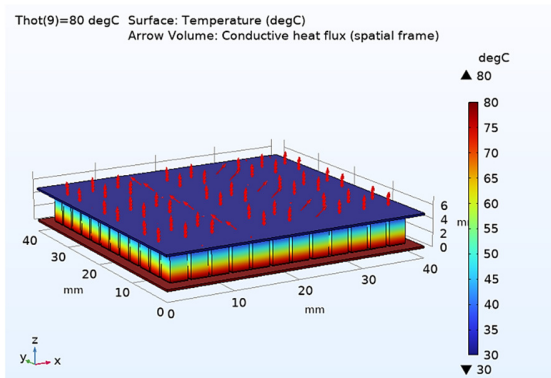


Fig. 8. Temperature gradient across the thermoelectric module.

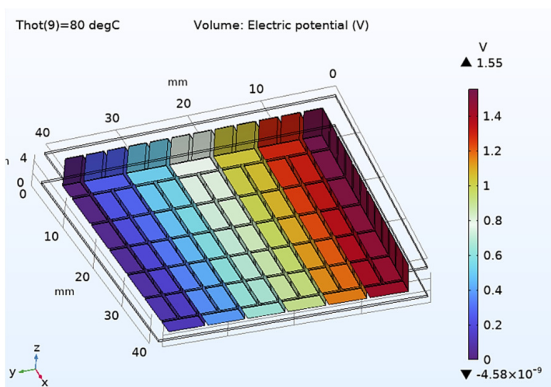


Fig. 9. Electric potential evaluation for hot side temperature of 80 °C.

Figure 10 demonstrates the voltage, current, and power curves for the hot side temperature. Secondary (right) axis values correspond to current [mA] and power [mW]. The evaluation was carried out on the load resistance of 2.4 kΩ. Figure 11 displays the current-voltage and current-power curves plotted in the simulation software. A linear current-voltage dependence indicates the correctness of the result trends, as well as the current-power curve.

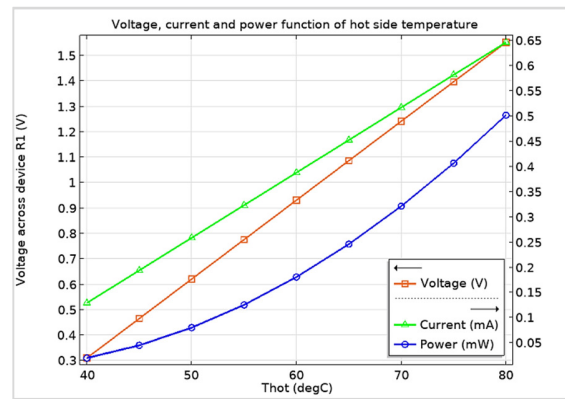


Fig. 10. Voltage, current, and power to load resistor, function of T_{hot} .

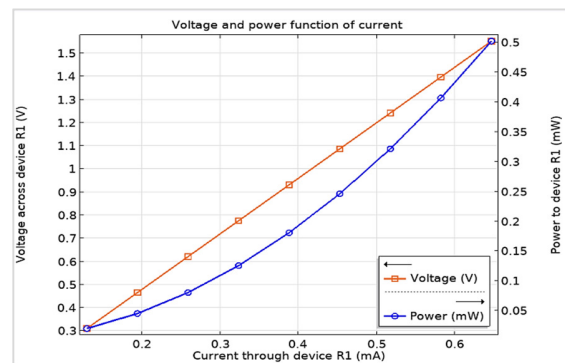


Fig. 11. Voltage and power function of current through resistor.

B. Frequency Domain Study – Stress Evaluation

Regarding the following stress evaluation [27], it should be noted that the crystal structure of Bi_2Te_3 consists of a hexahedral-layered structure with five atomic layers, a so-called quintuple ($\text{Te}^{(1)}\text{-Bi-Te}^{(2)}\text{-Bi-Te}^{(1)}$). Bonds within the quintuple are covalent, but between the $\text{Te}^{(1)}$ tellurium planes of adjacent quintuples there are weak van der Waals interactions along the c -axis of the unit cell, thus presenting an increased susceptibility to breakdown [23-24]. Figure 12 shows that the maximum stress value is 0.759 kPa, much less than the supported compression stress of 62 MPa declared in [26] for the same Bi_2Te_3 material.

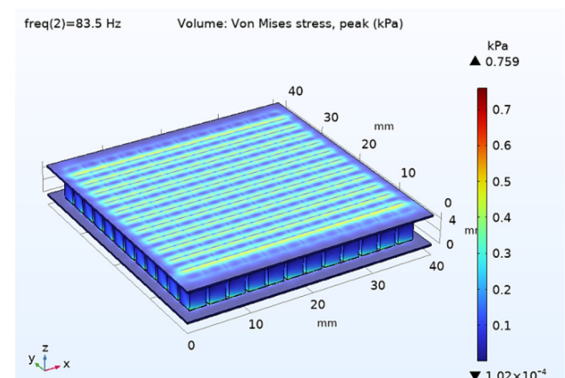


Fig. 12. Von Mises stress peak evaluation at 83.5 Hz.

The physical application involves sinusoidal vibration waves, hence the stress evaluation addresses the von Mises stresses in a frequency domain study, where everything is considered sinusoidal. The compressor does not yield any lateral linear forces on the TEG modules within its normal operation. The vibrational behavior of the rotary-bladed machinery does not present any risk to the thermoelectric modules' yield strength. The measured male rotor frequency of 83.5 Hz was considered for this purpose, due to the highest vibration amplitudes exhibited [25].

VI. CONCLUSIONS

This paper presents a realistic simulation model and results for the application of harvesting waste heat from a twin-screw compressor. All boundary loads considered in the numerical simulations are real data measured on the compressor and the model relied on physical TEGs. A frequency domain study revealed that the compressor's vibrations on all axes do not present a risk to damage the TEM and it is therefore safe to use in this application, since the compressor only produces sinusoidal vibration waves. The stationary study showed very pertinent results, comparable to the plots in the product specifications. The curves obtained for current, voltage, and power validated the correctness of the simulation model. This study did not aim to model an ideal heat source to obtain high values of thermoelectric power, as most studies in the literature, but to assess what could be practically expected from the harvesting of industrial waste heat. Consequently, the values of the electric parameters are insufficient to power wireless sensor nodes in terms of voltage output. Thermoelectric generators have been envisaged as a backup source for a multisource energy harvesting system, primarily based on piezoelectric harvesting from vibrations. Multiple harvesters must be electrically connected to meet the power requirements of the sensor nodes.

So far, no studies have been reported to apply mechanical vibrations in the simulation to assess von Mises stress in a frequency domain study along with classical TEG stationary studies. Furthermore, both the vibration and temperature data introduced in the simulation model were measured on an industrial compressor test bench. Future research will pursue conducting experimental tests, and adding a heatsink in the simulation model to compare to the physical model for proper result validation.

ACKNOWLEDGMENT

This study is part of the "Nucleu" program within the Romanian Research, Development, and Innovation Plan 2022-2027, conducted with the support of the Romanian Ministry of Research, Innovation, and Digitalization, contract 31N/2023, project number PN 23.12.09.

REFERENCES

- [1] H. Jouhara, N. Khordehgah, S. Almahmoud, B. Delpech, A. Chauhan, and S. A. Tassou, "Waste heat recovery technologies and applications," *Thermal Science and Engineering Progress*, vol. 6, pp. 268–289, Jun. 2018, <https://doi.org/10.1016/j.tsep.2018.04.017>.
- [2] K. R. Artoni, Ang, I. Yamazaki, K. Hirata, S. Singh, M. Matsunami, and T. Takeuchi, "Development of Cu₂Se/Ag₂(S,Se)-Based Monolithic Thermoelectric Generators for Low-Grade Waste Heat Energy Harvesting," *ACS Applied Materials & Interfaces*, vol. 15, no. 40, pp. 46962–46970, Oct. 2023, <https://doi.org/10.1021/acsami.3c09823>.
- [3] B. Safaei, S. Erdem, M. Karimzadeh Kolaroudi, and S. Arman, "State-of-the-art review of energy harvesting applications by using thermoelectric generators," *Mechanics of Advanced Materials and Structures*, 2023, <https://doi.org/10.1080/15376494.2023.2217660>.
- [4] K. S. Ong, L. Jiang, and K. C. Lai, "Thermoelectric Energy Conversion," vol. 4, I. Dincer, Ed. Oxford, UK: Elsevier, 2018, pp. 794–815.
- [5] B. Buonomo, F. Cascetta, A. di Pasqua, C. Fiorito, and O. Manca, "Numerical investigation on a thermoelectric generator in an exhaust automotive line with convergent metal foam," *Journal of Physics: Conference Series*, vol. 2385, no. 1, Sep. 2022, Art. no. 012057, <https://doi.org/10.1088/1742-6596/2385/1/012057>.
- [6] Y. J. Cui, B. L. Wang, and K. F. Wang, "Thermally induced vibration and strength failure analysis of thermoelectric generators," *Applied Thermal Engineering*, vol. 160, Sep. 2019, Art. no. 113991, <https://doi.org/10.1016/j.applthermaleng.2019.113991>.
- [7] A. Gürcan and G. Yakar, "Investigation of the performance of a thermoelectric generator system utilizing the thermal energy of air compressed in a compressor," *Journal of the Korean Physical Society*, vol. 80, no. 6, pp. 467–483, Mar. 2022, <https://doi.org/10.1007/s40042-022-00425-x>.
- [8] Z. Varga and E. Rácz, "Experimental Investigation of the Performance of a Thermoelectric Generator," in *2022 IEEE 20th Jubilee World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, Poprad, Slovakia, Mar. 2022, pp. 000159–000164, <https://doi.org/10.1109/SAMI54271.2022.9780741>.
- [9] J. H. Meng, X. X. Zhang, and X. D. Wang, "Multi-objective and multi-parameter optimization of a thermoelectric generator module," *Energy*, vol. 71, pp. 367–376, Jul. 2014, <https://doi.org/10.1016/j.energy.2014.04.082>.
- [10] M. N. Hanani, J. Sampe, J. Jaffar, and N. H. M. Yunus, "Development of a Hybrid Solar and Waste Heat Thermal Energy Harvesting System," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10680–10684, Jun. 2023, <https://doi.org/10.48084/etasr.5561>.
- [11] "TEGpro TE-MOD10W4V-40," TEGpro, Sep. 2014. [Online]. Available: <https://www.tegmart.com/datasheets/TGPR-10W4V-40S.pdf>.
- [12] A. Prasad and R. C. N. Thiagarajan, "Multiphysics Modelling and Multilevel Optimization of Thermoelectric Generator for Waste Heat Recovery," 2018. [Online]. Available: <https://www.comsol.com/paper/multiphysics-modeling-and-development-of-thermoelectric-generator-for-waste-heat-61281>.
- [13] C. Săvescu, A. Morega, Y. Veli, and V. Petrescu, "Numerical Modelling of Thermoelectric Energy Harvesting from Industrial Compressor Waste Heat," in *2023 13th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*, Bucharest, Romania, Mar. 2023, pp. 1–6, <https://doi.org/10.1109/ATEE58038.2023.10108390>.
- [14] N. Jaswanth and G. RaamDheep, "Thermoelectric maximum power point tracking by artificial neural networks," *Soft Computing*, vol. 27, no. 7, pp. 4041–4050, Apr. 2023, <https://doi.org/10.1007/s00500-023-07948-w>.
- [15] U. Erturun, K. Erermis, and K. Mossi, "Influence of leg sizing and spacing on power generation and thermal stresses of thermoelectric devices," *Applied Energy*, vol. 159, pp. 19–27, Dec. 2015, <https://doi.org/10.1016/j.apenergy.2015.08.112>.
- [16] C. Maduabuchi and R. Lamba, "Photovoltaic-Thermoelectric Power Generation: Effects of Photovoltaic Cell Type, Thermoelectric Leg Geometry, and Multistaging on System Performance," in *Sustainable Energy Storage for Furthering Renewable Energy*, Begell House, 2022.
- [17] G. Pennelli, E. Dimaggio, and M. Macucci, "Electrical and thermal optimization of energy-conversion systems based on thermoelectric generators," *Energy*, vol. 240, Feb. 2022, Art. no. 122494, <https://doi.org/10.1016/j.energy.2021.122494>.
- [18] M. Nesarajah and G. Frey, "Multiphysics Simulation in the Development of Thermoelectric Energy Harvesting Systems," *Journal of Electronic Materials*, vol. 45, no. 3, pp. 1408–1411, Mar. 2016, <https://doi.org/10.1007/s11664-015-4049-1>.

- [19] S. Singh, K. Hirata, S. K. Pandey, and T. Takeuchi, "Recent Advances in Energy Harvesting from Waste Heat Using Emergent Thermoelectric Materials," in *Emerging Materials: Design, Characterization and Applications*, L. R. Thoutam, S. Tayal, and J. Ajayan, Eds. Singapore: Springer Nature, 2022, pp. 155–184.
- [20] "COMOTI – Institutul National de Cercetare Dezvoltare Turbomotoare." <https://comoti.ro/en/home-2/>.
- [21] "TEGpro: Module Installation Notes," TEGpro. [Online]. Available: <https://www.tegmart.com/datasheets/TGPR-MOD-INST.pdf>.
- [22] "COMSOL Multiphysics® Software - Understand, Predict, and Optimize," COMSOL. <https://www.comsol.com/comsol-multiphysics>.
- [23] "Thermoelectric Cooler," COMSOL. <https://www.comsol.com/model/thermoelectric-cooler-30611>.
- [24] "LTC3588-2 - Nanopower Energy Harvesting Power Supply with 14V Minimum VIN," Linear Technology. [Online]. Available: <https://www.analog.com/media/en/technical-documentation/data-sheets/35882fc.pdf>.
- [25] C. Borzea, V. Petrescu, I. Vlăduță, M. Roman, and G. Badea, "Potential of Twin-Screw Compressor as Vibration Source for Energy Harvesting Applications," *Electrical Machines, Materials and Drives*, vol. 17, no. 1, pp. 91–96, 2021.
- [26] G. M. Guttman, Y. Gelbstein, G. M. Guttman, and Y. Gelbstein, "Mechanical Properties of Thermoelectric Materials for Practical Applications," in *Bringing Thermoelectricity into Reality*, IntechOpen, 2018.
- [27] F. Khelil, M. Belhouari, N. Benseddiq, and A. Talha, "A Numerical Approach for the Determination of Mode I Stress Intensity Factors in PMMA Materials," *Engineering, Technology & Applied Science Research*, vol. 4, no. 3, pp. 644–648, Jun. 2014, <https://doi.org/10.48084/etasr.442>.
- [28] D. Teweldebrhan, V. Goyal, and A. A. Balandin, "Exfoliation and Characterization of Bismuth Telluride Atomic Quintuples and Quasi-Two-Dimensional Crystals," *Nano Letters*, vol. 10, no. 4, pp. 1209–1218, Apr. 2010, <https://doi.org/10.1021/nl903590b>.
- [29] T. Nehari, A. Ziadi, D. Ouinas, and B. Boutabout, "Numerical Study of the Effect of the Penetration of a Crack in the Matrix of a Composite," *Engineering, Technology & Applied Science Research*, vol. 4, no. 3, pp. 649–655, Jun. 2014, <https://doi.org/10.48084/etasr.447>.

Experimental Analysis of Twin Screw Compressor's Energetic Efficiency Depending on Volume Ratio

Valentin Petrescu

Romanian Research and Development Institute for Gas Turbines COMOTI, Romania | Politehnica Bucharest, Romania
valentin.petrescu@comoti.ro

Claudia Savescu

Romanian Research and Development Institute for Gas Turbines COMOTI, Romania | Politehnica Bucharest, Romania
claudia.borzea@comoti.ro (corresponding author)

Teodor Stanescu

Romanian Research and Development Institute for Gas Turbines COMOTI, Romania | Politehnica Bucharest, Romania
teodor.stanescu@comoti.ro

Cristian Nechifor

Romanian Research and Development Institute for Gas Turbines COMOTI, Romania | Politehnica Bucharest, Romania
cristian.nechifor@comoti.ro

Filip Niculescu

Romanian Research and Development Institute for Gas Turbines COMOTI, Romania
filip.niculescu@comoti.ro

Sorin Tomescu

Romanian Research and Development Institute for Gas Turbines COMOTI, Romania | Politehnica Bucharest, Romania
sorin.tomescu@comoti.ro

Eduard Vasile

Romanian Research and Development Institute for Gas Turbines COMOTI, Romania | Politehnica Bucharest, Romania
eduard.vasile@comoti.ro

Received: 3 October 2023 | Revised: 25 January 2024 | Accepted: 28 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6425>

ABSTRACT

The current paper presents the results of the experimental analysis to assess and optimize the twin-screw compressor's efficiency by varying the volume ratio. The experimental tests are conducted on the compressor's test bench, with a dedicated automation system inside the control console. The control and monitoring software allows parameter recording for subsequent visualization, data curation, and post-processing. The evaluation of screw compressor's performance requires a simultaneous analysis of the

thermodynamic and flow processes, both of which depend on the compressor's geometry. The obtained volumetric and adiabatic efficiencies are good, with values over 0.88 and 0.69, respectively.

Keywords-*twin-screw compressor; volume ratio; energetic efficiency optimization; industrial automation*

I. INTRODUCTION

Twin-screw compressors belong to the category of positive displacement machines [1], and are very common in industrial applications [2], due to their robustness, reduced dimensions and relatively low maintenance requirements. Twin-screw compressors are commonly manufactured with helically grooved rotors of uniform pitch. The main components include a set of male and female rotors, a set of axial and radial bearings, and a slide valve. All these components are encased in a housing [3]. Oil-flooded twin-screw compressors have proven very reliable for applications in oil and gas industry. Unlike the oil-free version, they have direct contact between rotors. Even though the temperature range during the operation of a compressor is not that large and does not involve very high maximum temperatures, the effect of thermal expansion is significant and the small clearances required between rotors and between rotors and housing should be maintained over all working regime conditions [4, 5]. Screw compressors also generate noise and vibration during their operation [6]. The inlet and discharge ports' shape and size may influence to a great extent the screw compressor's performance, as well as the amplitude of flow oscillations in the discharge chamber. The characteristic pulsations generated in the suction and discharge areas are created by the working fluid and have a large influence on the mechanical noise and on the noise generated due to rotor rattle [5]. There are many approaches for analyzing and improving compressors' performance and efficiency [7, 8], from the helical lobes and rotor mechanical design and minimizing the clearances between rotors and casing to energetic aspects and increasing the efficiency of the driving motor regarding its electric consumption [9-13].

Authors in [14] analyzed and compared two profile types for asymmetrical rotors, demonstrating through CFD analysis that the new methods of generating rotor profiles do not bring a major improvement on screw compressors' energy efficiency, with the differences being at maximum 0.3-0.5%. Authors in [15] analyzed experimentally a screw compressor with rotors on which an anti-friction material was deposited, with the role of both accidental touch protection and clearance reduction. The authors obtained a 3% increase of the air flow vehiculated by the compressor, at constant energy consumption. Authors in [16] simulated with CFD and experimentally verified how the closing time of a compressor's suction port improves the volumetric efficiency. They found that the optimal suction closing angle is 30° behind the male rotor when the working chamber is at maximum volume. The volumetric efficiency increase reported is 1.5% compared to the conventional closure angle of 15° behind. Author in [17] reported a 26% performance increase by narrowing the discharge port. The computations, however, were only done for volume ratios of 4 and 8. The paper discusses the experimental works conducted on compressors with intermediate ratios of 2.6, 3.5 and 4.8, but on different discharge pressures. Authors in [18] showed how a gas compression skid with an oil-injected screw compressor

can be optimized regarding the energy consumption by adopting a lubrication system equipped with a geared pump with magnetic coupling. The compressor's power consumption estimations, based on the experimental operation of a compressor, is minimized when using a revamped lubrication system.

The literature survey shows that one cannot easily find comparisons between many compressors of the same type with different prebuilt volume ratios. The existent literature tends to present improvements of a single compressor, not addressing if the improvements perpetuate over the entire product series. The current paper deals with an experimental analysis of compressors' efficiency [19-21] by varying the volume ratio. The experimental validations outline the efficiency results. We aim to demonstrate the repeatability of the results over more compressors of the same type, presenting the results from an entire range of products, thus validating the efficiency improvements.

II. COMPRESSOR TEST BENCH

The screw compressor skid is presented in Figure 1. An oil-flooded compressor uses oil for lubricating and sealing the rotors on the air-end. This enables a quick generation of high pressure, as well as delivering compressed air on a single stage. Oil separation from air is carried out in the separator vessel, before the air leaves the system. The air is then recirculated into the air-end for another use [22].

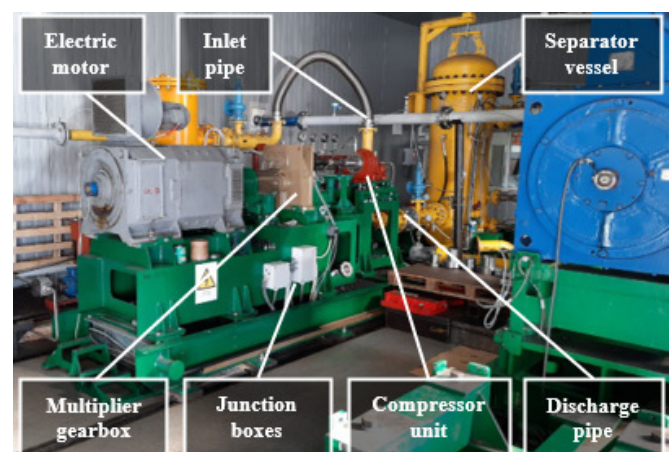


Fig. 1. Test bed with compressor skid.

The compressor is controlled and monitored by a Programmable Logic Controller (PLC) that also has the role of protecting the compression equipment, with warning limits set below the alarm limits in its custom programmed software. The warning limits are set to ensure an enhanced protection by letting the operators know that the parameters exceed the preestablished range of normal operation. These limits precede the alarm thresholds, which lead to the execution of emergency

shutdown sequences. The parameters are gathered in the PLC through analogue and digital inputs, and the execution elements are commanded through digital outputs via relays. Normally, the compressors manufactured by the Romanian Research and Development Institute for Gas Turbines COMOTI are provided with automation cabinets. In this case, however, since we are dealing with a fixed facility intended solely for this test bench, the automation elements and the PLC are located inside the control console presented in Figure 2. The important parameters of the compressor (pressures, temperatures, torque, speed) and of the driving electric motor (voltage, current, electric power) are monitored and their real-time values are displayed on the screens (Figure 4).



Fig. 2. Command and control console of the compressors test bench.

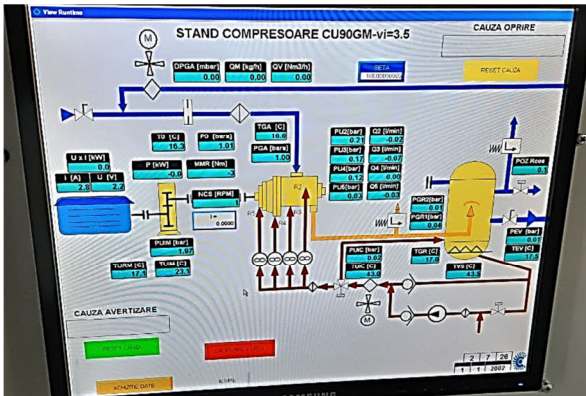


Fig. 3. The left side screen of Figure 2 displaying the parameters monitored by the PLC.

III. THE ANALYTICAL MATHEMATICAL MODEL

The compression capacity of a twin-screw compressor depends on the volume ratio and can be expressed as (1):

$$\Pi_c = V_i^k \quad (1)$$

where Π_c is the air compression ratio, V_i is the volume ratio, and k is the adiabatic coefficient, typically $1.26 \div 1.3$.

The volume ratio V_i of a twin-screw compressor has a direct effect on the internal compression ratio, P_i . A low V_i compressor exhibits a low compression ratio, while high V_i machines are used in systems where a higher compression ratio is desired. A low volume ratio compressor has a higher gas

volume at discharge than a machine with high volume ratio. A lower volume ratio will result in this case, by dividing the suction volume by the discharge volume [3]. The relation between the volume ratio and the compression ratio is expressed as [3]:

$$V_i = p_i^{\frac{1}{k}} \quad \text{or} \quad p_i = V_i^k \quad (2)$$

where V_i is the volume ratio, P_i is the compression ratio, k is the adiabatic coefficient corresponding to air ($k = 1.3$, it is being used herein).

The volume ratio can be calculated by:

$$V_i = \frac{V_s}{V_d} \quad (3)$$

where V_i is the internal volume ratio, V_s [m³] the internal suction volume, and V_d [m³] the internal discharge volume.

Alternatively, the compression ratio is calculated by:

$$p_i = \frac{p_d}{p_s} \quad (4)$$

where p_i is the internal compression ratio, p_d [bar] the internal discharge pressure, and p_s [bar] the internal suction pressure.

Depending on the manufacturer [23], different volume ratio machines can be available. Usually, the volume ratio of a twin-screw compressor lies within the range from 2.2 to 5.0 [3]. Compressors' efficiency can be assessed by calculating the volumetric, aerodynamic, and mechanical efficiencies. In each case the comparison is made to a theoretical optimum efficiency [24]. Adiabatic efficiency is a measure of the compression process efficiency with respect to the power consumption of the process. The theoretical power used in the compression of gases is defined by:

$$P_t = p_1 V_1 \frac{k}{k-1} \left[\left(\frac{p_2}{p_1} \right)^{\frac{k-1}{k}} - 1 \right] \quad [\text{W}] \quad (5)$$

where P_t [W] represents the theoretical power, p_1 [bar] is the suction pressure, V_1 [m³] is the suction gas volume, p_2 [bar] is the discharge pressure, and k is the isentropic coefficient of the gas.

The compressor's adiabatic efficiency is given by:

$$\eta_{ad} = \frac{P_a}{P_t} \quad (6)$$

where η_{ad} is the adiabatic efficiency, P_a [W] the absorbed power, and P_t [W] the theoretical power.

The absorbed power is calculated with (7), within the software, relying on the values given by a torque transducer and a tachometer measuring the rotational speed:

$$P_a = T \cdot \omega \Rightarrow P_a = T \cdot N \cdot \frac{\pi}{30} \quad (7)$$

where P_a [W] is the absorbed power, T [N·m] is the torque, ω [rad/s] is the angular velocity, and N [rpm] is the rotational speed.

IV. EXPERIMENTAL TESTS

The compressors were tested at the optimal pressure P_{optim} for each volume ratio V_i . The optimal pressures are provided in Table I. It is worth mentioning that bar is the standard unit for pressure when dealing with compressors, either bara (bar absolute) or barg (bar gauge), depending on the measuring reference.

TABLE I. OPTIMAL PRESSURES FOR COMPRESSORS' V_i

V_i	P_{optim}
2.6	6.1 bara
3.5	10 bara
4.8	13.5 bara

For a screw compressor with a volume ratio of $V_i = 3.5$, experimental tests were conducted, covering the entire operating range at different pressures and working speeds. Figure 4 illustrates the change in volumetric efficiency with respect to speed for different pressures. The volumetric efficiencies of the tested compressors are higher at lower speeds than the estimated values and the minimum efficiency obtained is 0.84, which denotes a significant improvement over 0.65. A decrease in the volumetric efficiency is observed with discharge pressure increase, due to the occurrence of more losses between the discharge chambers. A lower measured power of the screw compressors than the theoretical power calculated using the computation program provided by the original manufacturer [25] was recorded during testing. Possible reasons for this are the mechanical losses occurring due to the frictions between the mechanical parts, vibrations due to misalignments, bearings wear out, or damages of the rotor profile. All these reduce the compressor's efficiency. On the other hand, internal leakages are prone to take place, due to the clearances between the rotors and the compressor housing. During the compressor operation, a share of the pressurized air can leak back into the suction side, reducing the volumetric efficiency. This can lead to a reduction in compressor efficiency, and therefore the measured powers are lower than the calculated powers.

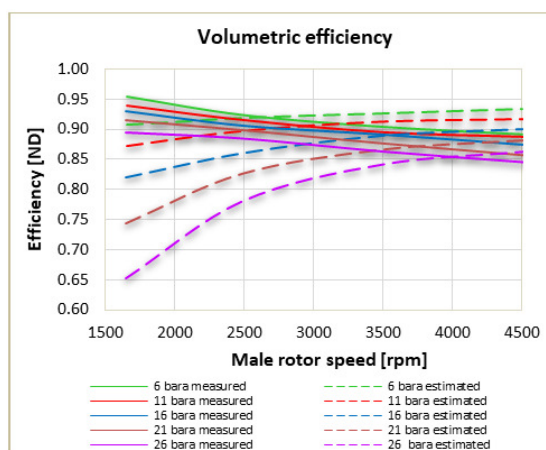


Fig. 4. Volumetric efficiency of CU90 compressors.

In Figure 5, the variation of the consumed power at various pressures is shown as a function of speed. The data obtained

show a decrease of the power consumption compared to the data obtained from the theoretical prediction program [25], this decrease being more significant at high discharge pressures. At the discharge pressure of 6 bar the differences are small, up to 2 kW, while at a discharge pressure of 26 bar, the difference is significant (25 kW).

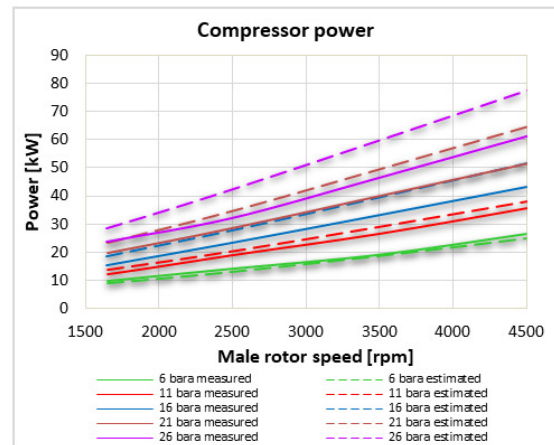


Fig. 5. Power of CU90 compressors.

The specific power is maintained quasi-constant regardless of speed. The specific power on each volume ratio did not record significantly high variations between the tested compressors with the same volume ratio. The specific power increases with the discharge pressure, considering that the volumetric and the adiabatic efficiencies are similar. The losses are generally of mechanical and thermodynamic nature, due to the friction between rotors, the friction within the multiplier gearbox, oil and gas leakages between the compressor chambers, and heat exchange between the gas and the compressor casing (making the process to be non-adiabatic). Figure 6 presents an undesirable situation when the rotors touch the housing during operation, leading to friction and potential mechanical gripping of the compressor. In this case, critical parameters monitored by the PLC (measured absorbed electric power, vibrations, or speed) exceed the alarm thresholds, triggering the execution of the automatic emergency shutdown sequence programmed.



Fig. 6. Traces of rotors friction with the housing.

Figure 7 shows the variation of adiabatic efficiencies with respect to speed for various working pressures. The graph shows that the highest efficiencies are achieved when the compression ratio is close to the optimal value of 10 bar.

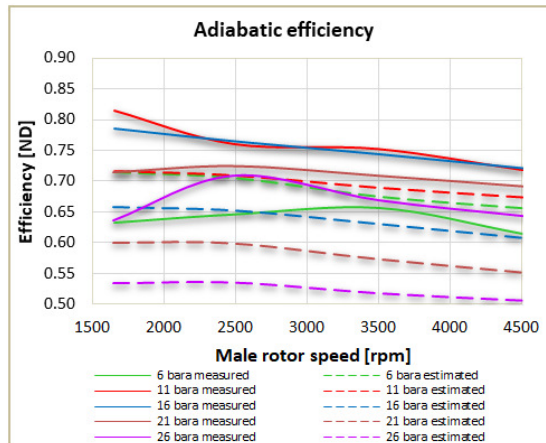


Fig. 7. Adiabatic efficiency of CU90 compressors.

The adiabatic efficiency of a twin-screw compressor decreases with the increase of speed and discharge pressure. This occurs because at higher speeds and pressures, the compression process becomes more difficult to control. At higher speeds, there is less time for gas compression, which can lead to higher losses. Moreover, higher gas speeds and pressures lead to a more turbulent gas flow and can also cause energy losses due to friction and heat transfer. At higher discharge pressures, the gas is compressed to a higher final temperature, which increases heat transfer. The temperature increase is proportional to the compression ratio, so at higher pressures, the temperature rises and energetic losses due to heat transfer can become significant, leading to a decrease in the adiabatic efficiency.

Hereinafter, we aim to show the changes in the specific power with respect to speed at various discharge pressures. A considerable improvement of the specific power consumed by the compressor can be noticed, especially at higher discharge pressures. The specific power of a screw compressor varies with compressor speed and discharge pressure. Typically, the specific power decreases with the increase of the compressor speed and with the decrease of the discharge pressure. At lower compressor speeds, the specific power is higher due to increased leakage losses and reduced efficiency. As compressor speed increases, leakage losses decrease and efficiency improves, resulting in a lower specific power. However, at higher speeds, internal losses increase due to friction and other factors, leading to a gradual increase of the specific power. Overall, the specific power of a screw compressor can be optimized by adjusting the compressor speed, the discharge pressure, and other operating parameters to obtain the optimal desired balance between efficiency and energy consumption. In the graph in Figure 8, one can observe that the specific power is quasi-constant regardless of speed. The compressor power represents the consumption in kW absorbed at a certain

operating condition. The specific power represents the energy consumption to compress 1 m³ of gas (air) in 1 min.

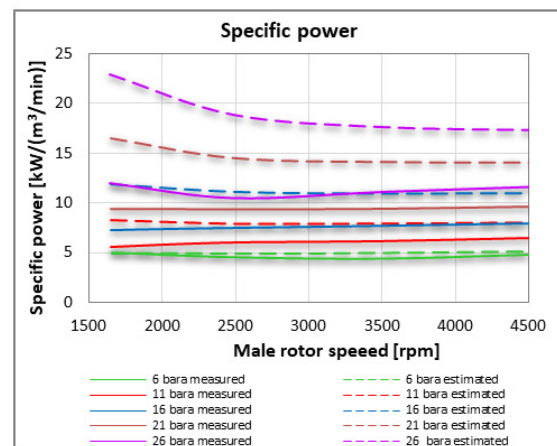


Fig. 8. Specific power of CU90 compressors.

V. RESULTS AND DISCUSSION

For the series manufacturing, a sample of 30 different compressors was considered for testing. The experimental parameters for these compressors are presented below, as 3 sets of 10 compressors designed with the same volume ratio ($V_i = 2.6$, $V_i = 3.5$, $V_i = 4.8$). Figure 9 shows the specific power derived experimentally for the tested compressors. It can be seen that the experiments show the specific power increase when the volume ratio is higher.

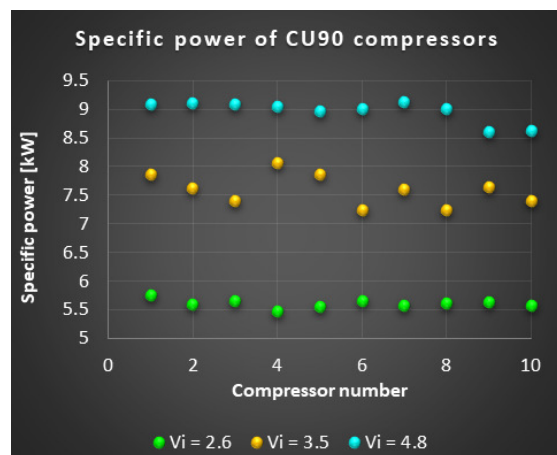


Fig. 9. Specific power for 30 CU90 compressors with different V_i .

Figure 10 shows the experimental volumetric efficiency for the 30 tested compressors. We can notice that volumetric efficiency increases when the volume ratio is lower. Compressors with an intermediate volume ratio of $V_i = 3.5$ show a rather unpredictable efficiency, some of them being situated both above the values for $V_i = 2.6$ and below for $V_i = 4.8$ in terms of volumetric efficiency. Figure 11 shows the adiabatic efficiency resulted from the bench tests with the 30 CU90 compressors. The compressors were tested at the same

speed of 25 m/s at the tip of the male rotor. The somehow random distribution of the adiabatic efficiency is caused by the variable clearances between the rotors and between rotors and housing. The clearances between rotors range from 1 to 84 μm .

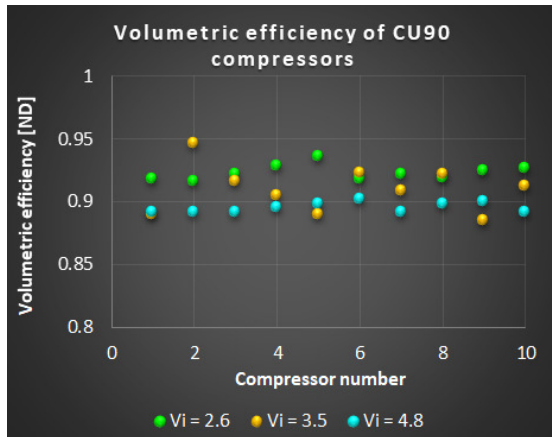


Fig. 10. Volumetric efficiency of 30 CU90 compressors with different V_i .

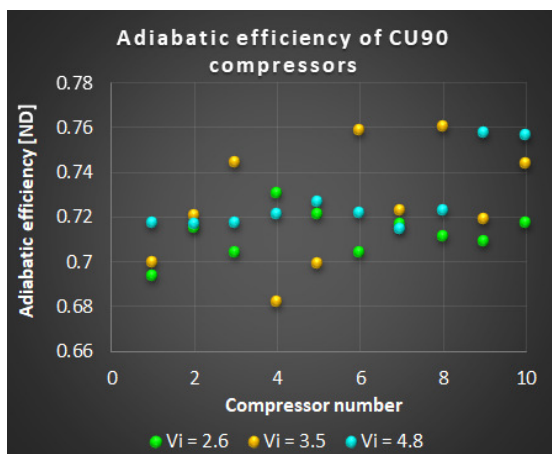


Fig. 11. Adiabatic efficiency of 30 CU90 compressors with different V_i .

VI. CONCLUSIONS

The current paper aimed to demonstrate the repeatability of the energy efficiency improvements over an entire series of compressors of the same type, presenting the results from a whole range of products. The findings have been validated experimentally. The differences between the tested compressors are due to the different internal clearances between the parts. The graphs illustrating the experimental comparisons present a sample of 30 compressors of the same type.

Achieving high volumetric efficiency at low speeds in a screw compressor can be done from multiple causes. One possible explanation is that at low speeds, the compressor can operate in a range where there are fewer leaks and more time for the gas to be compressed in the compression chambers. This may lead to higher volumetric efficiencies since more gas is compressed per unit time. Another possible explanation is

that the low speed can maintain an improved stability and operation efficiency due to lower internal losses, such as friction and pressure drops between compressor components. Additionally, the compression process may be more adiabatic at low speeds, meaning less heat transfer between the working fluid and compressor walls, resulting in less energy losses and a more efficient compression process.

As a general conclusion, the volumetric efficiencies obtained in our compressors are good, with values over 0.88. The adiabatic efficiencies obtained are also good, with values over 0.69. Nevertheless, it is worth mentioning that high volumetric efficiency at low speeds is related to the specific design and operating conditions of the twin-screw compressor, as well as to the compressed gas. Therefore, further analysis and experiments are needed to be carried out in order to fully understand the reasons behind the observed behavior.

ACKNOWLEDGMENT

This paper was realized as a part of the "Nucleu" Program within the Romanian Research, Development and Innovation Plan 2022-2027, conducted with the support of the Romanian Ministry of Research, Innovation and Digitalization, as part of contract 31N/2023, project number PN 23.12.05.

REFERENCES

- [1] A. Kovacevic, S. Rane, and N. Stosic, "Screw compressor with variable geometry rotors - analysis of designs by CFD," in *Fluid Machinery Congress 6-7 October 2014*, Imeche, Ed. Boston, MA, USA: Woodhead Publishing, 2014, pp. 91–101.
- [2] S. Tomescu and I. O. Bucur, "Numerical Investigation of Oil Gas Separation with the Use of VOF CFD", *Engineering, Technology & Applied Science Research*, vol. 11, no. 6, pp. 7841–7845, Dec. 2021, <https://doi.org/10.48084/etasr.4446>.
- [3] T. Bruce, *Screw Compressors: Misconception or Reality?* Rocky View, AL, Canada: NEXT Compression Corp.
- [4] E. Husak, A. Kovacevic, and I. Karabegovic, "Calculation of clearances in twin screw compressors", *IOP Conference Series: Materials Science and Engineering*, vol. 604, no. 1, 2019, Art. no. 012001, <https://doi.org/10.1088/1757-899x/604/1/012001>.
- [5] N. Stosic, I. K. Smith, and A. Kovacevic, "Chapter 5 - Screw compressors and expanders," in *Positive Displacement Machines*, I. A. Sultan and T. H. Phung, Eds. Academic Press, 2019, pp. 115–142, <https://doi.org/10.1016/B978-0-12-816998-8.00005-4>.
- [6] C. Borzea, V. Petrescu, I. Vlăduță, M. Roman, and G. Badea, "Potential of Twin-Screw Compressor as Vibration Source for Energy Harvesting Applications", *Electric Machines, Materials and Drives - Present and Trends*, vol. 2021, no. 1, pp. 1–6, 2021, <https://doi.org/10.36801/apme.2021.1.12>.
- [7] J. L. Rojas-Renteria, T. D. Espinoza-Huerta, F. S. Tovar-Pacheco, J. L. Gonzalez-Perez, and R. Lozano-Dorantes, "An Electrical Energy Consumption Monitoring and Forecasting System," *Engineering, Technology & Applied Science Research*, vol. 6, no. 5, pp. 1130–1132, Oct. 2016, <https://doi.org/10.48084/etasr.776>.
- [8] V. Drăgan, I. Mălăel, and B. Gherman, "A Comparative Analysis Between Optimized and Baseline High Pressure Compressor Stages Using Tridimensional Computational Fluid Dynamics," *Engineering, Technology & Applied Science Research*, vol. 6, no. 4, pp. 1103–1108, Aug. 2016, <https://doi.org/10.48084/etasr.696>.
- [9] S. Abdan, N. Basha, A. Kovacevic, N. Stosic, A. Birari, and N. Asati, "Development and Design of Energy Efficient Oil-Flooded Screw Compressors", *IOP Conference Series: Materials Science and Engineering*, vol. 604, 2019, Art. no. 012015, <https://doi.org/10.1088/1757-899x/604/1/012015>.

- [10] A. Bahadori, "Chapter 5 - Gas Compressors," in *Natural Gas Processing*, A. Bahadori, Ed. Boston, MA, USA: Gulf Professional Publishing, 2014, pp. 223–273, <https://doi.org/10.1016/B978-0-08-099971-5.00005-2>.
- [11] S. Z. Kassab, I. G. Adam, and H. A. Elgohary, "Performance of Oil-Injected Twin-Rotors Screw Compressor," in *International Conference on Aerospace Sciences and Aviation Technology*, Cairo, Egypt, May 2011, vol. 11, pp. 199–219, <https://doi.org/10.21608/asat.2011.27147>.
- [12] H. R. Takleh and V. Zare, "Employing thermoelectric generator and booster compressor for performance improvement of a geothermal driven combined power and ejector-refrigeration cycle", *Energy Conversion and Management*, vol. 186, pp. 120–130, 2019, <https://doi.org/10.1016/j.enconman.2019.02.047>.
- [13] S. Tomescu, V. Petrescu, A. Serban, and S. Voicu, "Energy Efficiency of an Oil Injected Screw Compressor Operating at Various Discharge Pressures," in *2021 10th International Conference on Energy and Environment (CIEM)*, Bucharest, Romania, Jul. 2021, pp. 1–5, <https://doi.org/10.1109/CIEM52821.2021.9614754>.
- [14] S. Patil, M. Davis, N. Stosic, A. Kovacevic, and N. Asati, "Contribution of modern rotor profiles to energy efficiency of screw compressors", *IOP Conference Series: Materials Science and Engineering*, vol. 1267, no. 1, 2022, Art. no. 012006, <https://doi.org/10.1088/1757-899x/1267/1/012006>.
- [15] M. Nitulescu, C. Slujitoru, V. Petrescu, V. Silivestru, G. Fetea, and S. Tomescu, "Reducing rotors clearance - a way to increase the performance of a screw compressor", *IOP Conference Series: Materials Science and Engineering*, vol. 1180, no. 1, 2021, Art. no. 012007, <https://doi.org/10.1088/1757-899x/1180/1/012007>.
- [16] H. Kameya, M. Ishikawa, and T. Saito, "Improvement of volumetric efficiency for screw compressors using inertial charging," in *7th International Conference on Compressors and their Systems 2011*, Boston, MA, USA: Woodhead Publishing, 2011, pp. 455–465.
- [17] N. Stosic, "Operation and performance of screw machines with high built-in volume ratio", *IOP Conference Series: Materials Science and Engineering*, vol. 604, Sep. 2019, Art. no. 012014, <https://doi.org/10.1088/1757-899x/604/1/012014>.
- [18] S. A. Serban, S. G. Tomescu, I. Vladuca, and S. Voicu, "Energy Improvement of an Oil Injected Screw Compressor Skid," *EMERG*, vol. VII, no. 1, pp. 51–59, 2021.
- [19] H. H. Patel and V. J. Lakhera, "A critical review of the experimental studies related to twin screw compressors", *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*, vol. 234, no. 1, pp. 157–170, 2019, <https://doi.org/10.1177/0954408919869534>.
- [20] N. Sessaiah, R. K. Sahoo, and S. K. Sarangi, "Theoretical and experimental studies on oil injected twin-screw air compressor when compressing different light and heavy gases", *Applied Thermal Engineering*, vol. 30, no. 4, pp. 327–339, 2010, <https://doi.org/10.1016/j.applthermaleng.2009.09.010>.
- [21] X. Liang and W. Wu, "Theoretical and Experimental Studies on Oil Injected Twin Screw Air Compressor under Unload Conditions", *IOP Conference Series: Earth and Environmental Science*, vol. 252, 2019, Art. no. 032118, <https://doi.org/10.1088/1755-1315/252/3/032118>.
- [22] "Oil Flooded Air Compressors," *Ingersoll Rand*. <https://www.ingersollrand.com/en-us/air-compressor/oil-flooded-ac>.
- [23] "Ingersoll Rand EPE50SE-GHH Air End Rebuild," *Industrial Air Power*. <https://www.industrialairpower.com/epe50se-ghh-air-end-rebuild.html>.
- [24] H. Wu, J. Liu, Y. Shen, M. Liang, and B. Zhang, "Research on Performance of Variable-Lead Rotor Twin Screw Compressor", *Energies*, vol. 14, no. 21, 2021, Art. no. 6970, <https://doi.org/10.3390/en14216970>.
- [25] N. Piccardo, "Optimize Your Compressor System," *Ingersoll Rand*. <https://www.ingersollrand.com/en-us/air-compressor/compressor-blog/optimize-a-compressed-air-system>.

Efficient Hardware Accelerator and Implementation of JPEG 2000 MQ Decoder Architecture

Layla Horrigue

Electronics and Micro-Electronics Laboratory, Faculty of Sciences, Monastir University, Tunisia
layla.k-12@hotmail.com

Refka Ghodhbani

Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
refka.ghodhbani@nbu.edu.sa (corresponding author)

Albia Maqbool

Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
albia.alam@nbu.edu.sa

Eman H. Abd-Elkawy

Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia | Department of Mathematics and Computer Science, Faculty of Science, Beni-Suef University, Egypt
eman.hassan@nbu.edu.sa

Jihane Ben Slimane

Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
jehan.saleh@nbu.edu.sa

Taoufik Saidani

Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
taoufik.saidan@nbu.edu.sa

Faheed A. F. Alrslani

Department of Information Technology, Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
f.alrslani@nbu.edu.sa

Amjad A Alsawaylimi

Department of Information Technology, College of Computing and Information Technology, Northern Border University, Saudi Arabia
amjad.alsawaylimi@nbu.edu.sa

Marouan Kouki

Department of Information Systems, Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
marouan.kouki@nbu.edu.sa

Amani Kachoukh

Department of Information Systems Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
amani.khasookh@nbu.edu.sa

Received: 13 February 2024 | Revised: 23 February 2024 | Accepted: 24 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7065>

ABSTRACT

Due to the extensive use of multimedia technologies, there is a pressing need for advancements and enhanced efficiency in picture compression. JPEG 2000 standard aims to meet the needs for encoding still pictures. JPEG 2000 is an internationally recognized standard for compressing still images. It provides a wide range of features and offers superior compression ratios and interesting possibilities when compared to traditional JPEG approaches. Nevertheless, the MQ decoder in the JPEG 2000 standard presents a substantial obstacle for real-time applications. In order to fulfill the demands of real-time processing, it is imperative to meticulously devise a high-speed MQ decoder architecture. This work presents a novel MQ decoder architecture that is both high-speed and area-efficient, making it comparable to previous designs and well-suited for chip implementation. The design is implemented using the VHDL hardware description language and is synthesized with Xilinx ISE 14.7 and Vivado 2015.1. The implementation findings show that the design functions at a frequency of 438.5 MHz on Virtex-6 and 757.5 MHz on Zync7000. For these particular frequencies, the calculated frame rate is 63.1 frames per second.

Keywords-JPEG2000; MQ decoder; FPGA; EBCOT; implementation

I. INTRODUCTION

The International Standard Organization/International Electrotechnical Commission (ISO/IEC) and the Joint Photographic Experts Group introduced the JPEG 2000 standard, which is a complex and high-performance digital image coding standard [1-4]. Positioned as the leading image compression standard, its goal is to surpass widely adopted standards such as JPEG. Numerous capabilities are included in JPEG 2000, including compressed domain processing, error resilience, continuous tone and bi-level compression, lossy and lossless compression, region of interest coding, and progressive transmission with pixel precision and resolution [2, 3]. The incorporation of these characteristics improves the efficiency of systems utilizing JPEG 2000 for picture compression. Although JPEG 2000 outperforms JPEG in terms of features and efficiency, it is notable for its considerably greater processing complexity. The improved functionality is accompanied by higher computational complexity and memory demands, mainly due to the utilization of the Embedded Block Coding with Optimized Truncation (EBCOT) technique, which is a fundamental component of JPEG2000 [5]. Multiple studies examining the JPEG2000 decoder algorithm [5, 6] have consistently found that entropy decoding is the operation that requires the most memory and processing time. Thus, the entropy decoder is the optimal choice for hardware acceleration. While the MQ decoding process inherently follows a sequential approach, achieving acceleration for this module is extremely difficult. Prior studies [7-12] have demonstrated that MQ decoder designs can achieve a one-clock-cycle decision, hence enhancing their efficiency.

Nevertheless, these architectures exhibit sluggish performance and need a substantial amount of silicon space. As a result, the throughput of these MQ decoders is lower than anticipated.

This study presents a cost-effective and efficient architecture for a high-speed MQ decoder, implemented using the VHDL hardware description language. The outcomes of the proposed design are compared with those of multiple FPGA implementations.

II. JPEG2000 STANDARD**A. Block Diagram – Algorithmic Chain**

Six distinct units are involved in the JPEG 2000 decoding process (Figure 1): a bitstream parser, an EBCOT, a dequantization unit, an Inverse Discrete Wavelet Transform (IDWT) unit, a color transform unit, and a tile combiner. The Tier 2 bitstream parser unit is specifically accountable for obtaining the compressed bitstream for every code block and transmitting it to the entropy block decoder by extracting pertinent data from image headers. Subsequently, the code stream is decoded into the wavelet domain by the EBCOT (Tier 1). The image is then arranged in raster order [3, 13] by the dequantization unit, which additionally increases the bit count of every coefficient [14]. The wavelet coefficients are then transformed into pixels using the IDWT, and these pixels are subsequently shifted from the YCbCr color system to the RGB color space using the color transform. Finally, to assemble the image in raster order, the tile combiner aligns the rows of every tile [3].

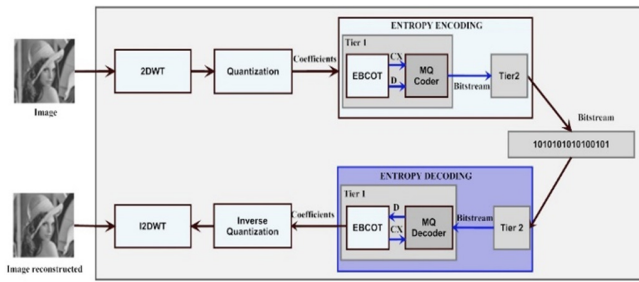


Fig. 1. Block diagram of the JPEG 2000 compression algorithm.

B. Entropy Block Decoder Overview

A simplified functional diagram of the entropy block decoder is illustrated in Figure 2. Two processing units comprise Tier 1: the decoding passes and the MQ decoder.

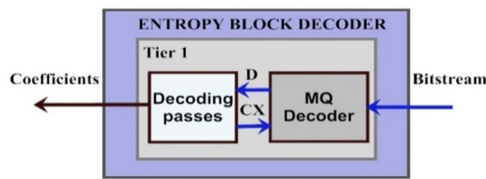


Fig. 2. The entropy block decoder.

a) Figure Labels

The decoding process utilizes the data from neighboring bits and the decoded judgments made by the MQ decoder to determine the importance of each bit in the image. During the decoding process, if a bit is found to be significant, the surrounding bits are used to offer contextual information to the MQ decoder for that particular bit. After obtaining the binary decision from the MQ decoder, the decoding operation must pause before processing the adjacent bits [15]. The decoding process has three separate phases: the Magnitude Refinement Pass (MRP), the Clean-Up Pass (CUP), and the Significance Propagation Pass (SPP) [1, 8]. SPP detects coefficients that are not labeled as significant but have at least one neighboring coefficient that is tagged as significant. The sign of the coefficient is established if the decoded bit is one. MRP focuses on coefficients that were previously determined to be significant in a previous bit plane and decodes the remaining bits of those coefficients. CUP decodes bits that are deemed inconsequential by the preceding two rounds.

b) The Binary Arithmetic Decoder

The binary arithmetic decoder used in JPEG 2000 is widely referred to as the MQ decoder. The likelihood estimate between the Most and the Least Likely Symbol (MPS and LPS, respectively) is considered, along with the context provided by the decoding steps, to calculate the output bit value. The resulting output value is subsequently utilized in the decoding process to assist in determining the importance of neighboring bits and their correct positioning in the final decoded image.

C. Binary Arithmetic Coding (BAC)

BAC executes entropy coding to construct the code stream based on the context data pairs from the bit plan coding. The BAC in JPEG 2000, frequently referred to as the MQ coder, is

based on the QM adaptive arithmetic encoder in JBIG but employs the Q coder's byte emission [7, 12]. Like in the Q coder, 0 and 1 are dynamically assigned to MPS and LPS. The MQ coder, as depicted in Figure 3, is based on the recursive probability subdivision of Elias coding. With each binary decision, the current probability interval is divided into two smaller intervals. The code string is modified if needed to indicate to the lower bound of the probability sub-interval for the symbol that occurred. Although the coding method includes the addition of binary fractions rather than the concatenation of integer code words, more probable binary decisions can frequently be coded for less than one bit per decision [11].

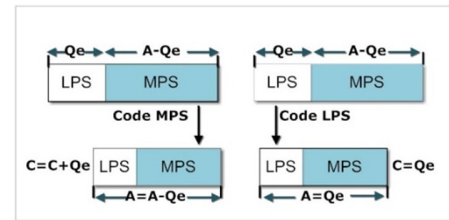


Fig. 3. MQ coder procedure.

D. MQ Decoder's Algorithm

The MQ decoder's algorithm is initialized using the INITDEC procedure, as shown in Figure 4. When all contexts have been read, the context (CX) is read and sent on to the DECODE procedure. The DECODE algorithm decodes the binary decision D and outputs a value of either 0 or 1. It is composed of four procedures: LPS_EXCHANGE, MPS_EXCHANGE, RENORMED, and BYTEIN. DECODE includes a probability estimation table that gives adaptive estimates of probability for every scenario. The compressed data have been decompressed once all CXs have been read, see [3, 5, 6, 11, 16] for a more detailed description of the algorithm.

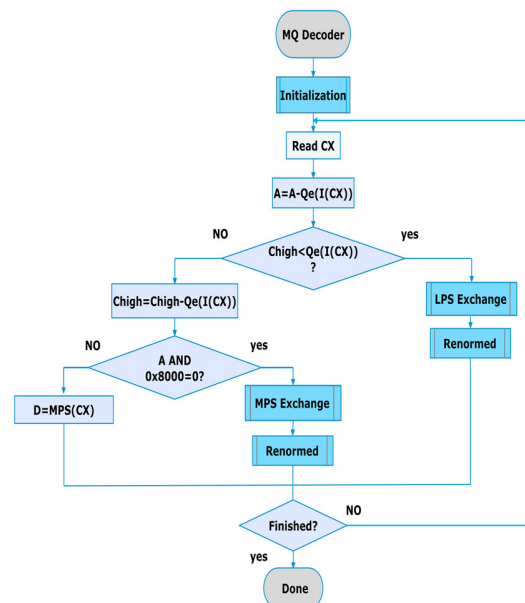


Fig. 4. The MQ-decoder's flowchart.

Figure 5 describes the arithmetic decoder's register configurations:

- In order to save the current interval value, the 16-bit A register is used.
- During the decoding procedure, the partially coded bits are stored in the C register, also known as the code register.
- By combining the Chigh register and the Clow register, we create a 32-bit C register. When we renormalize C, we move one bit of new data from Clow's bit 15 to Chigh's bit 0.

In the A register, the "a" bits stand for the fractional bits of the current interval value, and in the C register, the "x" bits imply the same. The "s" bits represent spacer bits that enforce crucial limitations on carry-over. The "b" bits indicate the specific locations from which entire bytes of data are taken from the C register. The letter "c" represents a carry bit.

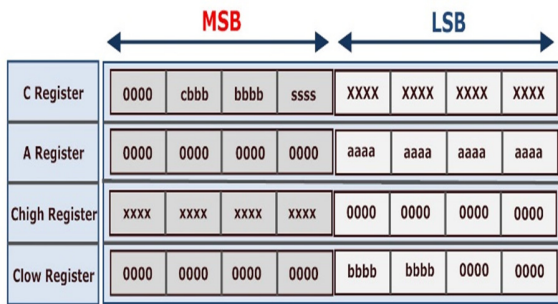


Fig. 5. Register structures for the MQ decoder.

III. PROPOSED MQ DECODER ARCHITECTURE

The main approach to reduce the execution time of the MQ decoder is to minimize the number of clock cycles needed to generate all the contexts of a picture. Given this, we propose a low-power, efficient hardware solution that obtains a bitstream and a context, then renders a decision in 4 clock cycles. This section provides a comprehensive description of the hardware architecture of our recently launched MQ decoder. Figure 6 depicts a simplified and functional diagram of the MQ decoder.

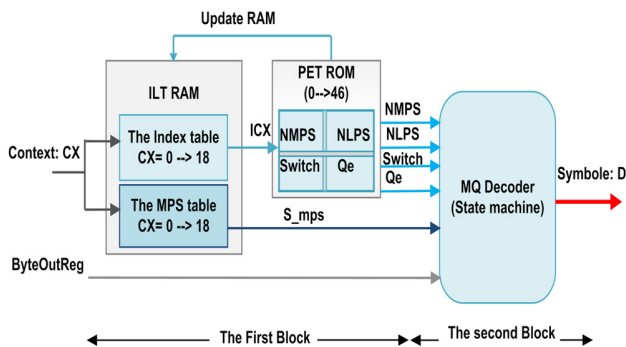


Fig. 6. MQ decoder functional diagram.

For every given CX value, the Bit Plane Code (BPC) can produce one of nineteen possible outcomes [1]. The CX values

are entered into a MQ decoder, where each context is associated with a corresponding item in the Index Look-Up Table (ILT). One field, ICX, is an index, and the other, S_mps, is the MPS value field. These two compose the ILT. While the ICX is used as a reference for the Probability Estimation Table (PET), the S_mps value decides if the symbol '0' or '1' is regarded as an MPS symbol. The initial values of the ILT are determined according to the standard JPEG 2000 and are later adjusted throughout the decoding process. The Positron Emission Tomography (PET) scan provides a detailed visualization of the decoder's internal structure [11, 16]. This architecture is comprised of two components: the probability estimator and the decoding block.

A. Probability Estimator

The context block prediction consists of the subsequent components: the ILT_RAM and the PET_ROM. These two principal components are composed, respectively, of 2 RAMs and 4 ROMs, as seen in Table I. The initial phase of the proposed architecture, as seen in Figure 7, commences with the retrieval of the context CX.

TABLE I. THE DIFFERENT COMPONENTS USED IN THE PREDICTION BLOCK

	RAM/ROM	Description/Use
ILT_RAM	RAM_ICX	It comprises 19 indexes, which correspond to the current index in the look-up table. Every index is encoded using 6 bits. In this scenario, the I (CX) functions as a pointer to the 4 ROMs.
	RAM_MPS	It has 19 MPSs (current Most Probable Symbols), and each MPS has a single bit of coding.
PET_ROM	ROM_NMPS	It has 47 NLPs (Next Most Probable Symbols). Every NLPs has a 6-bit code.
	ROM_NLPS	It contains 47 NLPs. Each NLPs is coded on 6 bits.
	ROM_Switch	It contains 47 elements (Switch). Each element is coded on 1 bit
	ROM_Qe	It consists of 47 components. Every element is encoded using 16 bits.

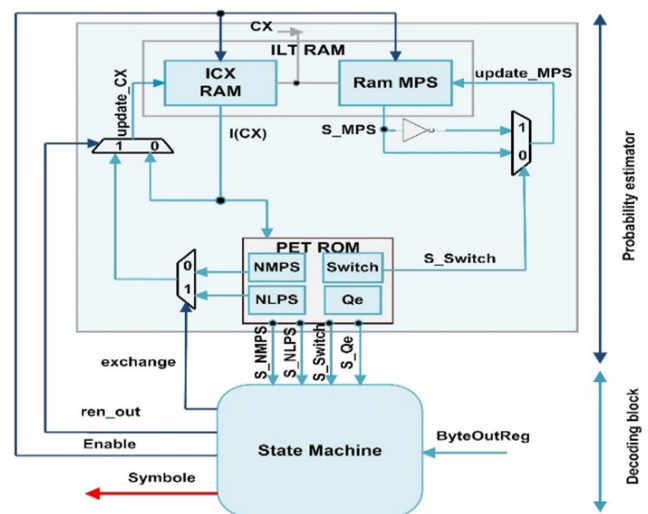


Fig. 7. Internal architecture of the MQ decoder.

To obtain the values of I (CX) and mps_D, the address bus is utilized to transfer the CX value to two RAMs, specifically RAM I (CX) and RAM MPS. Once the provided I(CX) value has been fed into each of the four ROMs (ROM_NMPS, ROM_NLPS, ROM_Switch, and ROM_Qe), the matching entries can be accessed. In cases where renormalization is required, the NMPS or NLPS value is copied and stored in the RAM (CX). Additionally, the S_MPS bit in RAM_MPS is modified when the S_Switch signal is very high.

B. The Decoding Block

In order to explain the functioning of the MQ decoder, we developed a state machine consisting of 11 states. We have created a hardware description of the process that is synchronized by the state machine, as depicted in Figure 8.

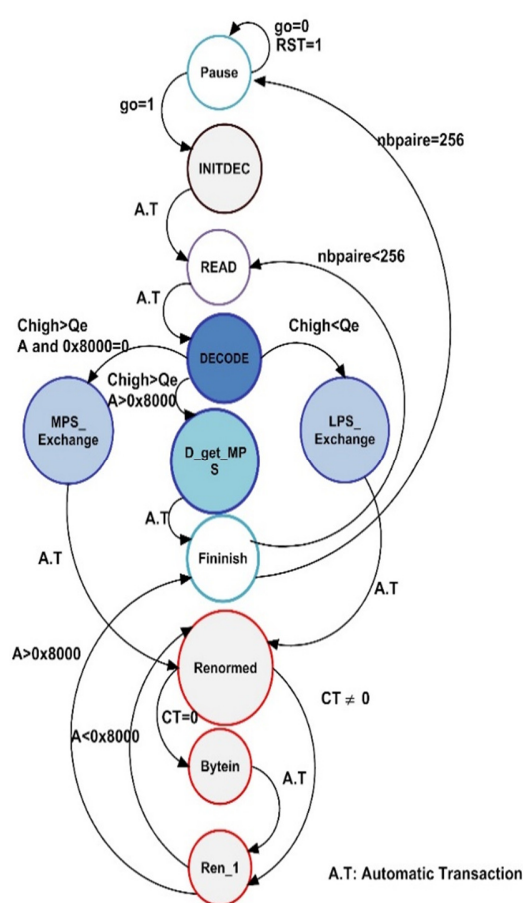


Fig. 8. The proposed state machine.

The state machine operates using a sequential pseudocode. Its primary states are:

- **INITDEC:** This state is utilized to initiate the MQ decoder. The compressed data pointer, BP (Byte Pointer B), is initialized to BPST (the starting value of BP). The initial byte of the compressed data is transferred to the least significant byte of Chigh, and subsequently, a new byte is retrieved. Ultimately, the register C undergoes a shift of 7

bits, the value of CT is decremented by 7, and the register A is set to the hexadecimal value 0x8000.

- **BYTEIN:** As seen in Figure 8, the bytein state is called by the RENORMD state when CT equals zero. This state handles the processing of a single byte of input, taking into consideration any extra bits that may come after the hexadecimal FF byte. The Clow and Chigh registers are concatenated to generate the Reg_C in Bytin. B is updated and put in the highest 8 bits of the Clow_register if it is not a hexadecimal FF byte. In this case, the value of BP is decremented. Indeed, B represents the byte specified by the compressed data buffer address BP. The byte B1, situated at the memory address BP+1, is examined to ascertain its equivalence to the hexadecimal value FF. When B1's value is greater than 8F in hexadecimal notation, the data byte in register C is incremented by adding the hexadecimal value FF, and bit counter CT is set to 8.
- **Ren_1:** During this state, both Reg_A and Reg_C are shifted to the left, and the counter CT is decremented by 1. If the value of reg_A is less than the hexadecimal value 8000, then the Renormed state is executed. Otherwise, the final state was executed successfully.
- **FINISH:** in this state if the number of pairs (nbpaire) is equal to the total number of CX, we will transition back to the pause state. Otherwise, the read operation will be repeated.

IV. IMPLEMENTATION AND RESULTS

A. Implementation

The proposed design was successfully tested on three platforms, namely Virtex5, Virtex4, and Spartan3A. This implementation utilized the default configurations of Xilinx ISE 14.7. Table II displays the maximum clock frequency and device consumption overview for the suggested MQ decoder design. It is possible for the proposed design to operate at 394.3 MHz, 320 MHz, and 176.8 MHz, respectively. The average decoding time for one decision is 4.10 cycles.

TABLE II. SYNTHESIS RESULTS FOR THE SUGGESTED DESIGN USING XILINX ISE 14.7

Used FPGA	XC5VLX50T	XC4VLX15	XC3SD3400a
Total slices LUTs	420	605	654
Total slices registers	288	322	345
Total FF pairs	197	289	290
Max. Frequency (MHz)	394.3	320	176.8

Using the default settings of Xilinx Vivado 15.1, the architecture's performance was assessed in terms of operating frequency and hardware requirements. The proposed MQ decoder architecture has been effectively implemented on three platforms: Zync7000, KINTEX7, and VIRTEX7, while adhering to all user-imposed constraints for implementation. Table III displays the summary of hardware requirements, Worst Negative Slack (WNS), data path delay, and requirements. Slack is determined by subtracting the arrival time from the required time [17]. WNS represents the amount of free time available once the timing constraints have been

met. The time period for 1 clock cycle at a frequency of 100 MHz is 10 ns, as specified in the constraints file for the purpose of completing the execution.

TABLE III. SYNTHESIS RESULTS FOR SUGGESTED ARCHITECTURE USING VIVADO 15.1

Used FPGA	Zync7000 xc7z045	KINTEX7 XC7K70T	VIRTEX7 XC7VX415T
Total FF slices	260	260	260
LUT	372	384	386
Memory LUT	14	14	14
WNS (ns)	3.686	3.612	3.502
Data path delay (ns)	0.817	0.94	0.803
Requirements (ns)	5	5	5
Frequency (MHz)	757.7	720	667.5
Throughput (MB/s)	118.35	112.47	104.27

In order to ascertain the maximum clock frequency, it is necessary to initially establish the required clock period [17, 18]. Upon implementation of the proposed design on Zynq 7000, the required clock period is calculated as 1.32 ns (by subtracting 3.68 ns from 5 ns), which gives a frequency of 757.7 MHz and a throughput of 118 MB/s.

B. Comparison

In order to assess the efficiency of our architecture in terms of hardware demands and operational speed, we conducted comparative analysis with other pre-existing architectures that employed the identical FPGA. The maximum frequency values reached by the proposed design and existing architectures using the same FPGA are shown in Table IV.

TABLE IV. COMPARISON OF THE MQ DECODER WITH THE ARCHITECTURES FROM [8, 9]

Used FPGA	Architecture [8]	Architecture [9]		Proposed	
	Frequency (MHz)	Frequency (MHz)	FPS	Frequency (MHz)	FPS
Virtex-6 XC6VLX75T-3FF68	437.1	-	-	438.5	63.1
Virtex-4 XC4VLX15-12FF68	-	195.3	28.07	321.1	46.2
Virtex-5 XC5VLX30-3FF324	-	222.8	32.02	394.1	56.7

From the comparison, it can be inferred that the proposed MQ decoder runs at a greater frequency. For example, the operating frequency of the suggested design, when targeted at Virtex 5, is 1.76 times greater than that of current architecture [9]. The frame rate at 321.1 MHz is predicted to be 46.2 Frames Per Second (FPS). However, at 394.1 MHz and 438.5 MHz, the frame rates are 56.7 FPS and 63.1 FPS, respectively. These values are specific to high-definition TVs with a resolution of 1920p.

The frequency achieved with the Virtex2 XC2V6000-6 platform and the average clock cycle obtained by using a counter in state machine HDL to build the design are used to calculate the theoretical throughput. Equation (1) is utilized to ascertain the throughput of sources [11, 19]. The throughput

and the average number of clock cycles from references [10, 11, 19, 20] are shown in Table V. Equation (1) calculates the throughput of the binary arithmetic decoder in MB/s. The terms w_{CB} and h_{CB} denote the width and height of each code block, measured in pixels, d_{CB} and f denote the pixel depth and design frequency in MHz, respectively, whereas μ represents the number of cycles needed per code block [11]. The design comparison uses (64×64) for the height and width per code block, and a byte for the depth of every each pixel. The processed data has a cumulative size of 4.096 bytes.

$$\text{Throughput (MB/s)} = w_{CB} \times h_{CB} \times d_{CB} \times f / \mu \quad (1)$$

TABLE V. MQ DECODER THROUGHPUT

Used FPGA	Virtex-2 XC2V6000-6						
Design	[9]	[10]	Proposed	[11]	[19]	[20]	[21]
No slice registers	328	498	313	313	-	315	-
No slice LUTs	579	944	640	630	-	586	-
Frequency (MHz)	142	140.4	200.2	205.5	45.6	157.2	127.9
Frame rate supported (FPS)	20.41	-	350	29.56	-	-	-
Avg. no of clock cycles	-	74.23	26.22	28.44	32.7	68.53	-
Throughput (MB/s)	-	7.74	31.27	35.74	5.7	9.4	5.3

Table V presents the expected frame rate, maximum cycle count, maximum clock frequency, and logic utilization of the proposed MQ decoder design compared to other existing architectures implemented in Virtex 2. Each code block in the recommended design requires 26,223 clock cycles in total. The proposed design produces a throughput of 31.27 based on the outcomes of FPGA synthesis. Compared to the architecture indicated in [18], it has a throughput that is 5.48 times higher. Table V demonstrates a significant decrease of 36% in memory requirements for the proposed architecture in comparison to the architecture mentioned in [10]. The predicted frame rate of the suggested MQ decoder design is 1.41 times higher than that of the MQ decoder architecture in [9].

V. CONCLUSION

This paper presents a detailed explanation of the advanced design of an MQ arithmetic decoder, which is built upon the MQ decoder technique used in JPEG 2000. The VHDL hardware description language was used to implement the proposed design, which was then synthesized for FPGA devices. The zynq7000 device achieves a maximum working speed of 757.7 MHz. For high-definition TV with a resolution of 1920 pixels, the estimated frame rate is 63.1 FPS. Furthermore, the memory requirement for the proposed architecture is diminished by 37.1% in comparison to the other existing architectures, and the proposed design achieves the highest maximum frequency among the compared designs. Hence, the proposed design can be effortlessly transferred to Xilinx architectures and has the ability to decode 63.1 FPS of high definition (HD, 1920 × 1080 pixels). This makes it a promising option for deployment as a high-speed real-time

JPEG 2000 decoder in diverse applications such as medical imaging, satellite imagery, digital cinema, and mobile applications.

ACKNOWLEDGMENT

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number "NBU-FFR-2024-1662-02."

REFERENCES

- [1] *JPEG 2000 Part 1: Final Draft International Standard (ISO/IEC JTC1/SC29/WG1 N1855, 2000.*
- [2] *ISO/IEC JTC 1 / SC 29 / WG 1, (ITU-T SG8) Coding of Still Pictures.* JBIG, 1999.
- [3] D. S. Taubman and M. W. Marcellin, *JPEG2000 Image Compression Fundamentals, Standards and Practice.* Boston, MA, USA: Springer US, 2002.
- [4] D. Santa-Cruz, R. Grosbois, and T. Ebrahimi, "JPEG 2000 performance evaluation and assessment," *Signal Processing: Image Communication*, vol. 17, no. 1, pp. 113–130, Jan. 2002, [https://doi.org/10.1016/S0923-5965\(01\)00025-X](https://doi.org/10.1016/S0923-5965(01)00025-X).
- [5] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Transactions on Image Processing*, vol. 9, no. 7, pp. 1158–1170, Jul. 2000, <https://doi.org/10.1109/83.847830>.
- [6] A. Samet, M. B. Ayed, M. Loulou, and N. Masmoudi, "Comparison between JPEG and JPEG2000 still image compression standard," in *Proc. Visualization, Imaging, and Image Processing*, 2002.
- [7] K. Sarawadekar and S. Banerjee, "VLSI design of memory-efficient, high-speed baseline MQ coder for JPEG 2000," *Integration*, vol. 45, no. 1, pp. 1–8, Jan. 2012, <https://doi.org/10.1016/j.vlsi.2011.07.004>.
- [8] D. J. Lucking, E. J. Balster, K. L. Hill, and F. A. Scarpino, "FPGA implementation of the JPEG2000 binary arithmetic (MQ) decoder," *Journal of Real-Time Image Processing*, vol. 8, no. 4, pp. 411–419, Dec. 2013, <https://doi.org/10.1007/s11554-011-0214-9>.
- [9] O. C. Kulkarni, K. Sarawadekar, and S. Banerjee, "VLSI implementation of MQ decoder in JPEG2000," in *IEEE Technology Students' Symposium*, Kharagpur, India, Jan. 2011, pp. 193–197, <https://doi.org/10.1109/TECHSYM.2011.5783844>.
- [10] A. Descampe, F.-O. Devaux, G. Rouvroy, J.-D. Legat, J.-J. Quisquater, and B. Macq, "A Flexible Hardware JPEG 2000 Decoder for Digital Cinema," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 11, pp. 1397–1410, Nov. 2006, <https://doi.org/10.1109/TCSVT.2006.884573>.
- [11] L. Horrigue, T. Saidani, R. Ghodhbani, J. Dubois, J. Miteran, and M. Atri, "An efficient hardware implementation of MQ decoder of the JPEG2000," *Microprocessors and Microsystems*, vol. 38, no. 7, pp. 659–668, Oct. 2014, <https://doi.org/10.1016/j.micpro.2014.06.005>.
- [12] S. D. Jayavathi and A. Shenbagavalli, "FPGA-based Auxiliary Minutest MQ-coder architecture of JPEG2000," *Journal of Real-Time Image Processing*, vol. 16, no. 5, pp. 1765–1779, Oct. 2019, <https://doi.org/10.1007/s11554-017-0683-6>.
- [13] K. Liu, Y. Zhou, Y. Song Li, and J. F. Ma, "A high performance MQ encoder architecture in JPEG2000," *Integration*, vol. 43, no. 3, pp. 305–317, Jun. 2010, <https://doi.org/10.1016/j.vlsi.2010.01.001>.
- [14] T. Acharya and P.-S. Tsai, *JPEG2000 Standard for Image Compression: Concepts, Algorithms and VLSI Architectures*, 1st ed. Hoboken, NJ, USA: Wiley-Interscience, 2004.
- [15] R. Ghodhbani, T. Saidani, L. Horrigue, A. M. Algarni, and M. Alshammari, "An FPGA Accelerator for Real Time Hyperspectral Images Compression based on JPEG2000 Standard," *Engineering, Technology & Applied Science Research*, vol. 14, no. 2, pp. 13118–13123, Apr. 2024, <https://doi.org/10.48084/etasr.6853>.
- [16] N. Ramesh Kumar, W. Xiang, and Y. Wang, "Two-Symbol FPGA Architecture for Fast Arithmetic Encoding in JPEG 2000," *Journal of Signal Processing Systems*, vol. 69, no. 2, pp. 213–224, Nov. 2012, <https://doi.org/10.1007/s11265-011-0655-1>.
- [17] T. Saidani and R. Ghodhbani, "Hardware Acceleration of Video Edge Detection with High Level Synthesis on the Xilinx Zynq Platform," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 8007–8012, Feb. 2022, <https://doi.org/10.48084/etasr.4615>.
- [18] T. Saidani, R. Ghodhbani, A. Alhomoud, A. Alshammari, H. Zayani, and M. B. Ammar, "Hardware Acceleration for Object Detection using YOLOv5 Deep Learning Algorithm on Xilinx Zynq FPGA Platform," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 13066–13071, Feb. 2024, <https://doi.org/10.48084/etasr.6761>.
- [19] H.-H. Chen, C.-J. Lian, T.-H. Chang, and L.-G. Chen, "Analysis of EBCOT decoding algorithm and its VLSI implementation for JPEG 2000," in *2002 IEEE International Symposium on Circuits and Systems (ISCAS)*, Feb. 2002, vol. 4, <https://doi.org/10.1109/ISCAS.2002.1010457>.
- [20] D. J. Lucking, E. J. Balster, K. L. Hill, and F. A. Scarpino, "FPGA implementation of the JPEG2000 binary arithmetic (MQ) decoder," *Journal of Real-Time Image Processing*, vol. 8, no. 4, pp. 411–419, Dec. 2013, <https://doi.org/10.1007/s11554-011-0214-9>.
- [21] T. Zhu, J. Zhou, and S. Liu, "Design and implementation of JPEG2000 arithmetic decoder based on Handel-C," in *2009 3rd International Conference on Anti-counterfeiting, Security, and Identification in Communication*, Hong Kong, China, Aug. 2009, pp. 505–508, <https://doi.org/10.1109/ICASID.2009.5276988>.

Smart Grid 2.0: Modeling Peer-to-Peer Trading Community and Incentives for Prosumers in the Transactive Energy Grid

Manal Mahmoud Khayyat

Department of Computer Science and Artificial Intelligence, College of Computing, Umm Al-Qura University, Makkah, Saudi Arabia
mmkhayat@uqu.edu.sa (corresponding author)

Sami Ben Slama

The Applied College, King Abdulaziz University, Jeddah, Saudi Arabia
sabdullah1@kau.edu.sa

Received: 2 February 2024 | Revised: 23 February 2024 | Accepted: 24 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7001>

ABSTRACT

Smart Grid 2.0 (SG 2.0) implementation constitutes an additional challenge in the industry and research fields. Energy consumption decreases when producers exchange excess energy consumers, including intelligent consumers, Distributed Generation (DG), such as wind and solar, and Electric Vehicles (EVs). By utilizing Demand Response (DR) based on Real-Time Pricing (RTP), the operation of every device in a smart home can be scheduled. Allowing users to trade energy directly with other energy producers (prosumers) rather than exclusively relying on the grid, peer-to-peer (P2P) energy trading in smart homes lowers energy prices for users. This article focuses on how the DR P2P energy trading affects consumers. The study conducted utilizes a two-stage scheduling technique to reduce consumers' electricity expenses. The initial stage involves arranging each device in the smart home based on RTP employing a deep learning method. The P2P energy trading between consumers in the second phase is made more accessible by the DR and the simulation results exhibit that energy trading decreases electricity bills in smart homes. Utility companies can reduce load during peak hours using DR-based P2P energy trading.

Keywords-artificial intelligence; deep reinforcement learning; peer-to-peer energy trading; smart community; photovoltaic-array; energy market

I. INTRODUCTION

SG 2.0 is an innovative technology that integrates local grids with distributed energy to reduce energy consumption. The former employs advanced A.I. technologies to combine local grids and distributed energy sources, making it an attractive and promising solution for the future. SG 2.0 is self-processing and can work with both power transmission and high-end solutions. The energy industry has undergone significant changes in energy demand, resource utilization, management, distribution, and exchange. These changes have led to a new energy-consuming process known as Prosumer [1, 2]. The industry is also experiencing a massive transformation involving the use of renewable energy sources, increased operational efficiency, and the implementation of smart infrastructure and services. These ecosystem actors continuously improve their work by establishing appropriate legal systems and experimenting with new marketing methods. P2P energy trading is emerging as a viable alternative, allowing end-users to buy or sell electricity to or from other customers instead of solely relying on electrical service providers.

However, P2P energy trading in distribution systems presents several challenges, including network limitations, increased communication needs, reduced end-user privacy, and financial stability for utility companies. Numerous experimental and research projects are underway worldwide to face these issues [3]. According to [4], Positive Energy Regions (PERs) can be utilized for overgeneration to produce more energy despite their limited Loss of Earning Power (LEP). Photovoltaic (PV) panels are a suitable renewable energy source for PEDs due to their ease of installation, increased capacity, and low maintenance costs. Distributed PV systems are prevalent in the household PV market due to their affordability and the multiple household control they provide. They are, therefore, better suited to densely populated urban areas. Local energy sharing is expected to increase the use of electricity generated from PV systems. Individuals and small to medium-sized private or public businesses will drive this, depending on the market setup and available support. However, there are more effective strategies than relying on contributions for long-term PV stabilization. Due to the difficulty in accessing capital, current business models, such as power purchase agreements and net

metering, are limited to small-scale PV systems. It may be necessary to restructure business practices to utilize distributed energy supply, demand, and sharing. Therefore, consumers with small PV systems must sell their excess power to the grid at market prices without assistance. PV customers may experience financial losses, further damaging the network's stability and reliability [5].

In [6], the authors reported that the EU Commission's proposed budget for renewable energy includes energy communities that allow for local energy trading on a European scale. The energy sector has adopted a new business strategy known as P2P trading. In this economic model, prosumers and consumers organize into energy societies, where they can offer their surplus production to other participants. The infrastructure of the grid should be designed to accommodate this new model. To comply with new regulations to increase the use of solar power while reducing the need for incentives, optimal planning and modeling of PV P2P business models are required. When analyzing the situation and discussing ways to trade power in the future, it is essential to consider the three ways to own a PV system: consumer, community equity, and third-party liability [7]. In [8], the authors describe three solar PV consumer business idea specifications. The first category is the single direct user group, where one customer generates PV power on-site. The second category is the local collective user group, where several users share the PV electricity produced without a public grid. The final group comprises a range of energy models in areas with numerous buildings fitted with PV cells. In these areas, users directly consume locally produced PV energy and the excess energy is shared via a public or private microgrid. The business process involves various properties in each set of boundary conditions, leading to many possibilities and uncertainties. It is crucial to comprehend and test the different potential designs and combinations. Currently, there is limited regulatory and modeling research on P2P energy trading. The community-owned PV system is gaining popularity as a developing business model in areas where people share energy, as it could be a profitable way to promote the use of PV systems [9]. DSM was only considered in early studies on the smart grid to decrease energy costs in the smart home. This plan reduces energy costs by shifting loads when energy prices are low. Authors in [10, 11] conducted a research supporting this strategy. Recent studies have also shown that combining DSM with energy storage and renewable energy can reduce costs [12, 13]. These contributions have provided optimization models with timelines for optimized power sources and loads. Authors in [14] focused on energy trading between microgrids in the smart grid. However, none of these methods can lower energy costs in intelligent homes because they do not consider trading energy with nearby houses. P2P energy trading is a relatively new idea for families. P2P service providers offer metering and billing services and manage the distribution network [15, 16]. These projects worked with DSM to develop business models instead of exclusively concentrating on lowering energy costs. In [17], the authors created a community mini grid with an online market where each house could trade energy. Two-level programming was used to describe an optimization problem, and the relationship between mini-grid prices and the Supply-Demand Ratio (SDR)

was examined. Authors in [18] emphasized the importance of presenting customers with accurate information when comparing prices. In [19], the authors used an SDR method to demonstrate that customer costs are comparable. Authors in [20] reported on the impact of energy storage systems on consumer-to-consumer trade. They compared the total energy held by all users to the total energy sent by all users. Nevertheless, none of the above technologies integrates DSM with P2P trading. In [21], the authors suggested that a P2P-based grid approach can decrease the energy costs of all microgrids and increase the value of locally Distributed Energy Resources (DERs). In [22], the authors suggest that a smart grid can reduce energy costs using an integrated demand management system that works with P2P energy trading. Authors in [23] recommend a distributed DSM system integrated with community energy trading to lower household electricity prices in a small community grid with a BT storage unit, establishing a good relationship between buyers and sellers. Authors in [24] proposed an energy-sharing scheme that responds to price and demand.

During the recent years, there has been considerable discussion about the role of multifactor power management systems in microgrids. In [25], the authors discussed some multifactorial schemes that control network power. Game theory can be employed as a mathematical method to examine how consumers and market participants interact. Research has primarily focused on the interaction between consumers and customers within the context of a contract. Support methods were employed to address contracts that provided secure electricity supply against unfavorable low market prices and financial losses resulting from imbalances between supply and demand in the spot market [26, 27]. In [28], the authors proposed a robust supply chain coordination model based on a quantity discount contract that considers only fixed demand loads

To maximize the benefits of P2P trading, it is essential to establish multiple homes and RESs competing with one another. The optimization of energy costs is typically described using linear or nonlinear programming. Linear programming is commonly used to address the concepts of linear optimization. It is important to note that solving a linear model requires less time than solving a nonlinear model. Although the optimization model is nonlinear, it can still be solved quickly. However, nonlinear optimization models take much longer than the linear ones, so such optimization techniques cannot be used to solve these problems. Therefore, this article discusses setting up a sustainable hybrid PV battery storage system to manage household energy while considering grid outages and Demand Response (DR). An optimization design is proposed to decrease energy costs while maintaining system reliability. DR is achieved through incentives, such as tariffs based on usage, real-time pricing, and the scheduling of widely used appliances that can or cannot be moved. Additional case studies were simulated to provide a comparison, considering different battery power rates and the amount of renewable energy used in lower and higher cases. Furthermore, this study analyzes various limitations concerning grid dependability, including the effect on project expenses and the impact of demand response methods.

This study suggests the use of P2P energy trading with hardware scheduling in smart homes to tackle the limitations. The report emphasizes the subsequent significant contributions: Firstly, smart homes provide users with various distinctive technologies, including loads, DGs, and EVs. Additionally, there are two different categories of smart home appliances: those that use temperature control and those that use electricity control. Real-Time Pricing (RTP) employs the DL algorithm to arrange the timetable for each intelligent home device. The adoption of DR allows for P2P power trading between two entities, namely the masters and consumers, resulting in a reduction in electricity costs for both parties involved.

II. METHODOLOGY

A. Distributed Power Sharing

The term Distributed Power Sharing describes a residential neighborhood community consisting of an Energy Pool Unit (EPU), Smart Homes (SHAs), and Traditional Homes (THAs). The SHAs are equipped with energy storage, such as Battery (BT), which enables them to sell excess energy to the EPU and purchase stock at market prices in real time. Non-market households can purchase electricity from the EPU at lower prices than the retail market when electricity is insufficient. The solar panels installed by THAs can sell their excess energy to the energy collector at a price higher than the Feed-in Tariff (FinT). Decentralized agents negotiate the buying and selling process through a pool as illustrated in Figure 1. This section describes how P2P energy exchange can efficiently operate in six families with variable output and consumption patterns. Figure 1 depicts the energy conservation process among customers through exchanging energy and optimizing the distribution network. If residents consume more energy locally than the P2P market produces, the grid sets the price for its sale.

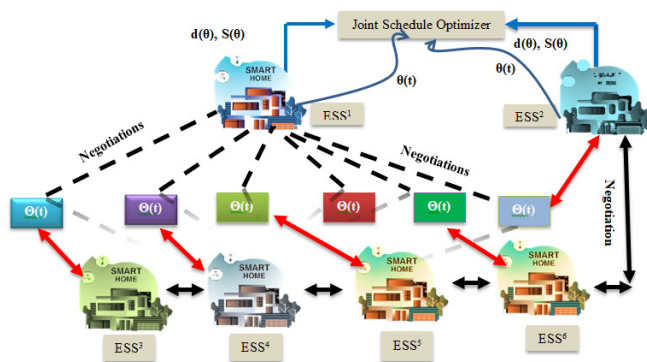


Fig. 1. P2P energy system design.

Following a surge in demand at the local exchange, the distribution business procures electricity at the prevailing grid prices. The P2P market structure seeks to minimize energy expenses and enhance societal well-being by optimizing pricing among nearby households. The model was specifically developed to aid six individual homes in four distinct settings. Users can monitor their Return on Investment (ROI) and expenses using the model interface. Home PV systems, Energy

Storage Systems (ESSs), and EVs are used for both energy generation and consumption. This technique empowers consumers to actively engage in the market by identifying the most economically efficient energy consumption. We introduce a precise approach for calculating exchange expenses. The sophisticated exchange algorithms, utilizing double auctions, consider the interests of both consumers and producers. This reduces the commercial motivation for a P2P house. According to popular belief, the surplus-to-demand ratio facilitates P2P trading. A series of minor blackouts, where immediate demand must always be met even if the utility network fails, are modeled. Starting time and duration of each event are determined arbitrarily. It is recommended to use files containing blackout data to create power blackout scenarios. The generated files can be used with the reverse conversion approach to create domestic power that collects and resells excess electricity from prosumers, consumers, and DGs at a discount. The energy in the pool comes from smart homes and DGs. All participants in the energy pool market must adhere to the principle of voluntarism and accept surplus-to-demand market prices.

Authors in [29, 30] proposed a methodology that reflects current retail prices and the home energy profile. The cost may exceed the retail market price when supply exceeds the demand. The authors discuss network blackout cases based on various constraints. During a power blackout, the entire network is inaccessible, resulting in no energy flow between homes and UGs. In this case, resources such as the PV-BT system must be provided to the domestic system during a power outage to prevent domestic power blackouts that would avert the domestic system from selling electricity to the UGs. Figure 1 demonstrates potential scenarios for distributed energy-sharing villages. The energy scenarios are:

- Scenario A is achieved when the community's SHAs are connected with BTs. SHAs can sell any excess energy to the EPU in real time. In a power outage, SHAs purchase additional energy from the EPU.
- Scenario B: In the event of a power outage, SHAs purchase additional energy from the EPU, considering the real-time electricity price.
- Scenario C: THAs purchase energy from the EPU at a discounted rate during periods of oversupply.
- Scenario D occurs when there is insufficient energy available during peak hours. In this scenario, THAs purchase the energy required to meet the demand at a lower cost than the V2G/retail market price. However, the EPU cannot supply the THAs with the power deficit needed.
- Scenario E occurs when THAs have excess power and the EPU is deficient in strength. In this scenario, THAs sell their extra energy to the energy pool in higher prices than the FinT.

B. Energy Pool Unit

It gets additional power from consumers, Distributed Generators (DGs), and renewable energy sources and sells it at a lower price in the retail market or a Real-Time Rate (RTP).

SHAs sell power to the pool, and DG sells the excess power. The surplus-to-demand ratio determines the market price of an energy pool unit [31]. Equation (1) estimates the neighborhood's retail market price of electricity and energy consumption. If the overall load profile is high, the price may be greater than the retail price (supply is higher than demand). The surplus/demand ratio enhances P2P commerce.

$$\begin{cases} f(t, p(t)) = m(t) + n(t, p(t)) \\ n(t, p(t)) = \alpha(t) \cdot (p(t))^2 + \beta(t) \cdot p(t) \\ p(t) = \frac{h_m(t) + K_{pv}(t) + g_m(t)}{g_n(t)} \\ n(t, p(t)) + \varepsilon < m(t) \end{cases} \quad (1)$$

C. Smart Home Agent

1) Reinforcement Q-learning

It is one of the most important ways for machine learning to make optimal choices in a world where nothing is known. The reward and the changed state of the environment are returned to the agent by the environment. The agent continues to learn until the latter provides greater advantages. The agent's main goal is to choose the optimal policy that offers the most rewards. The policy is how the agent behaves in a given scenario. This paper hypothesizes that a Markov decision-making process defines the environment. The agent state transition depends only on the current state and the action set in the current state without considering all the previous conditions and actions.

2) Home Agent Model

The SHA charges and discharges the BT unit based on the energy pool price and the home village demand. The BT is the most vital, controllable piece of equipment in the proposed P2P trading system, enabling users to become involved. The BT model is illustrated in (2) [32, 33]. The capacity is the main terminology when studying battery modeling. The battery functions in discharge current and electrolyte temperature, which affect the state of battery charge.

$$C^{BT}(I_c, \theta_c) = \frac{k_c C_0}{1 + (K_c - 1) \left(\frac{I_c}{I_c^*} \right)} \left(1 + \frac{\theta_c}{-\theta_f} \right)^\varepsilon \quad (2.1)$$

$$C^{BT}(t, \theta_c) = \begin{cases} C^{BT}(t) + P^r(t) \cdot \eta^c \cdot \theta_c \rightarrow P^r(t) > 0 \\ C^{BT}(t) + \frac{P^r(t)}{\eta^d} \cdot \theta_c \rightarrow P^r(t) < 0 \end{cases} \quad (2.2)$$

$$C^{BT}(t), P^r(t) : \begin{cases} 0 < C^{BT}(t) < C^{BT} \\ \overline{P_{\max}^c} \leq P^r(t) \leq 0 \rightarrow (P^r(t) \geq 0) \\ P^r(t) \leq 0 \leq \overline{P_{\max}^d} \rightarrow (\leq P^r(t)) \end{cases} \quad (2.3)$$

In household BT, the SHA must manage power in real time for the benefit of the customers. The BT's decision will affect battery capacity, public retail, and neighborhood pricing.

- Discharge case: The SHA would buy energy from the retail market or the local energy pool if the remaining energy did not fulfill the user's needs (Scenario D). The BT may discharge more electricity than needed to trade energy.

Traded energy reduces the supply-to-demand ratio, depreciating the power of the local energy pool.

- Charge Case: The agent must purchase electricity from the retail market or the local energy pool.

If there is no energy storage in the energy pool, the agent must buy electricity from the retail market. The Markov Decision Process (MDP) may be an excellent alternative for optimal decision-making with numerous associated stages.

D. Single Home Sharing Energy

1) PV Supply

It makes controlled power near a PV module with a Maximum Power Point Tracking (MPPT) controller and a DC/DC converter. The production of PV energy on rooftops is described by [34]:

$$\begin{cases} I_{GEN} = I_{ph} - I_s e^{\left(\frac{N_s V_{PV} + N_p I_{PV} R_s}{V_T} \right)} - \frac{N_s V_{PV} + N_p I_{PV} R_s}{R_{sh}} \\ P^{pv}(\forall t) = P^{pv}(s, t) * D(s, t) \end{cases} \quad (3)$$

2) Household Load Consumption

An agent is allocated to follow up the total home energy demand defined by the set of appliances consumption (lighting, heating, leisure, baking, etc.) [35]:

$$I_{DEM} = \sum_{i=0}^n I_{AP_i} = \sum_{\forall k} \Delta T_k(0,1) P_k^{App}(s, t) \quad (4)$$

3) BT Energy Storage Unit

Equation (5) defines how much power a BT can provide. Charging and discharging its two activities. The same equation describes the BT state of charge, which is limited by its capacity. It depicts the initial and final loading conditions of the system. Finally, (5) shows how much energy is stored in a storage unit over a typical period and under specific conditions [37].

$$\begin{cases} SoC^{BT}(t, s) = SoC^{BT}(t-1, s) \\ \quad + \Delta v(\eta^c * p^c(t, s) - \frac{p^d(t, s)}{\eta^d}) \\ SoC^{BT}(t) = SoC_0^{BT} + SoC_T^{BT}(1 - D^{BT}) \\ SoC_0^{BT} = P^{BT}(1, s) = P^{BT}(T, s) \\ C^{BT}(1 - D^{BT}) \leq P^{BT}(t, s) \leq C_0^{BT}(t) \\ SoC_1^{BT}(t, s) = C_0^{BT}(0, t) = SoC_T^{BT}(t, s) \end{cases} \quad (5)$$

4) Objective Function

An objective function is a mathematical equation for maximizing production benefits. The outcome depends on interdependent factors. It is a formula designed to meet profit and production targets. The goal of this study is to keep the cost of the THA (PV/BT) for a detached house as low as possible over its lifetime (see (6)). The capital cost is given by:

$$y_1 = \zeta^{BT} \cdot C^{BT}(t) + \zeta^{PV} P^{pv}(s, t) \quad (6.1)$$

The annual cost is given by:

$$Y_2 = C^{BT}(t, s) (\Psi^{BT} \delta^{BT} + z^{BT}) + P^{pv}(t, s) (\Psi^{pv} \delta^{pv} + z^{pv}) \quad (6.2)$$

The daily cost is given by:

$$Y_3 = \sum_{t=1}^{M'} \Psi \cdot \left[\sum_{i=1}^T \Delta v (\zeta^1(t, s) \cdot P^{GH}(t) - \zeta^2(t, s) \cdot P^{GH}(t)) \right] \quad (6.3)$$

The objective function is given by:

$$\min_a Y_1 + \sum_{i=1}^m (1+w)^{(1-i)} Y_2 + Y_3 \quad (6.4)$$

5) Energy Balance

The energy balance is centered on the basic thermodynamic principle according to which energy cannot be destroyed, but can be gained, lost, or conserved by a system. The condition achieved when the energy consumption is equal to the energy expenditure is called an energy balance. To reach this standard, it is anticipated that the energy management system will be able to program devices in a flexible manner. As a result, the constraint (7) must be fulfilled at all times [38].

$$\begin{aligned} &P^{GH}(s, t) + P^{pv}(s, t) + P_d^{BT}(s, t) = \\ &P_r^{BT}(s, t) + P^{HG}(s, t) + \sum_{v,k} \underline{\Delta}_k(0,1) P_k^{App}(s, t) \end{aligned} \quad (7)$$

III. MULTI-HOUSE CASE STUDY

A. Energy Trading Community

The MDP contains a set of states' $S(i)$, including the SOC_i , price, and the community price. A set of procedures for each state is set, involving buying or selling different amounts of electricity to the retail market or local power pool. The Q-value $Q(s, a)$ for the various actions when the agent is in the S state is recorded. Q-value learning uses a weighted average of the previous and new data. The agent can choose the actions with the highest Q value. The optimum active combination is provided by (8) [40, 41]. Equation (8.1) seeks to find the actual state(s) and tends to provide the best combination of action for all rules. Equation (8.2) aims to connect the output groups ($N(s)$) to the Q value. The Q-learning Trad-Algorithm-1 provides the Energy Trading community, including both smart and Trad users.

$$\begin{aligned} &\overline{Q}(s, a) \leftarrow Q(s, a) + aV(s, a) + \phi \max_{Q_s} Q_s(r_i()) \\ &a = \max_{Q_s} Q_s \left(r_i() \xrightarrow{\text{step1}} (8.1) \right) \\ &Q(s, a) = \frac{\min(N(s)) \cdot q(i, a)}{\sum \min(N(s))} \xrightarrow{\text{step2}} (8.2) \end{aligned} \quad (8)$$

where $\min(N(s))$ signifies the application of the operator to the output groups $N(s)$, $q(i, a)$ is the q value that corresponds to the aggregated rule i for the chosen action a . Next, the action sequence is applied and the new states are acquired by calculating $H(s', a)$ and Q value changes.

$$\begin{aligned} &H(s', a) = \frac{\min(N(s')), \max_{action} (q(i, a))}{\sum \min(N(s'))} \xrightarrow{\text{step3}} (9.1) \\ &\Delta Q(s', a) = V(s, a, s') + \phi H(s', a) \xrightarrow{\text{step4}} (9.2) \\ &q(i, a) \xleftarrow{\text{step5}} q(i, a) + \beta \cdot \Delta Q(i, a) + w(i, a) \xrightarrow{\text{step5}} (9.3) \end{aligned} \quad (9)$$

where $H(s', a)$ is the maximum function, and the Q value for the new case s' , $V(s, a, s)$ and Φ are identical to the ones in (9). Ultimately, the q value will be changed during each operation following (9.3). Where β is the learning rate and $w(i, a)$ refers to the value of truth deformed to the chosen action a according to rule i .

Q-learning Trad-Algorithm 1: Energy Trading community

Input: Solar output, households power demand, Temperature, electricity price (RTP, TOU), Agent ID status, SoCBS

Output: Maximum function $H(s', a)$, new case Q -value $s'(V(s, a, s'))$, output groups $N(s)$, Q -value records, optimal action a

Initialize memory G of size N ;

Initialize preprocess function $Q(s)$

Initialize target networks $Q(s', a)$

For iteration in $[1, \text{Max}+1]$:

Episode: $s=1, 2, \dots, M$ ($s=\Sigma Mi$)

Get the initial state $s0$

Compute output groups $N(s)$

End

For q -value in $N(s)$ do:

Convolute $N(s) \rightarrow w[i, a]$

$q(i, a) \xleftarrow{\text{step5}} q(i, a) + \beta \cdot \Delta Q(a) + w(i, a) \xrightarrow{\text{step5}} (8.3)$

End

For $i=1, 2, 3 \dots n-1$ do:

$\text{Max}(ai, aj) + q(i * ai * aj) * \Phi * \text{max}[Qs(ri, a)] + V \rightarrow V$;

$\Phi \text{max}[Qs(ri, a)] + aV(s, a) + Q(s, a) \rightarrow Q$

$Q \xleftarrow{\text{from } i \text{ to } n} Q / \sum w(i, a)$

$V \xleftarrow{\text{from } i \text{ to } n} V / \sum w(i, a)$

//Execute operation at in smart home environments and

observe $st+1$ (s')

$\Delta Q(s', a) = V(s, a, s') + \phi H(s', a, s, a)$

$q(i * ai, aj) \xleftarrow{\text{step5}} q[i * ai][aj] + \beta \cdot \Delta Q(a) + w(i)$

// Select a limited set of K occurrences, $1 \leq \varepsilon \leq K$;

for $i=1, 2, 3 \dots n-1$ do:

If $M(s=\Sigma Mi) < \varepsilon$

// Update target

$A[ai, aj] = \max A[ai, aj] * q(i * ai * aj)$

$S' \rightarrow S$

End

Else

$ai, aj = \varepsilon$

End

End

IV. RESULTS

A. System Initialization

The proposed approach suggests that integrating a storage system would be beneficial only if the existing electrical system functions perfectly. The installation of a PV array could save money. According to the optimization problem, a storage system is unnecessary because the power system is perfectly stable. Solar panels have reduced user expenses in two ways. Firstly, during peak hours, smart homes can operate autonomously. Secondly, excess energy can be sold to the power company for a profit. Figures 2(a) and 2(b) illustrate the sequencing of the PV system and the total number of devices required for a typical day when ESS batteries are used. The PV and ESS aim to meet the energy demand of homes during

lunchtime and do not affect savings. Energy savings seldom exceed 3%.

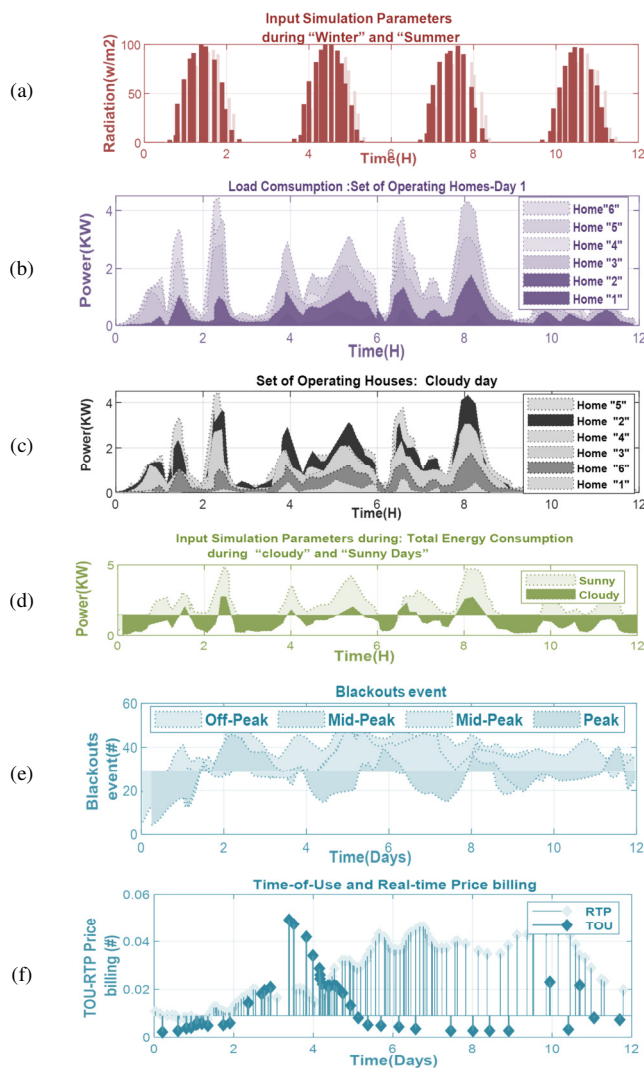


Fig. 2. (a) Solar radiation (w/m^2) during winter and summer, (b) total load consumption and (c)-(d) load consumption during cloudy and sunny days, (e) data available for main grid breakdowns occurring in KSA, (f) TOU and RTP prices taken into account in the simulation during 2022.

In peak periods, the grid can provide any deficit power (Figures 2(c) and 2(d)). This section analyzes the effect of renewable energy storage on cost reduction. Anything between 0% and 100% of domestic renewable energy can be stored anywhere. The relationship between storage capacity and cost savings is also examined through the employment of three scenarios: inconsiderable (less than 50%), medium, and large (up to 50%). The total cost of each system is calculated implementing the single-house model when the houses do not exchange energy. Additionally, the proposed algorithm calculates the cost of trading between homes and the utility. Figure 2(f) displays the energy savings to the grid resulting from the price disparity between these two systems. It portrays the relationship between grid savings and homes during peak

hours (non-RES). Each house has a minimum amount of energy stored. The life of the storage device is affected by falling below this minimum energy level. To save money and keep energy levels above this minimum, households should charge their energy storage devices in the early slots, even if unnecessary. Inefficient charging of storage devices increases energy costs. Therefore, storage penetration rates above a certain threshold. Adding energy storage to a grid is initially advantageous due to lower energy costs, but this advantage diminishes as storage capacity increases. The energy savings from storing energy on the grid are negligible. This article analyzes grid trading and the impact of solar panel penetration on costs and savings. The expenses and protection of energy storage capacity are also discussed. Figure 2(f) demonstrates how the percentage of households with storage devices affects the overall cost of the microgrid. The inefficiency of the load increases energy costs for homes that use storage devices. The cost curves increase with storage penetration, and smaller energy storage reduces overhead. The cost curves increase with storage penetration, and smaller energy storage reduces overhead (see Appendix B). A regional meteorological station provided the most recent measurements of the specified data for 12 hours [1].

B. Case Study: Multi-Houses

1) Results considering Grid without Blackouts

The proposed approach suggests that the integration of an ESS would only be beneficial if the existing electrical system is functioning perfectly. The installation of a PV array could save money. According to the optimization problem, a storage system is not necessary because the power system is perfectly stable. Solar panels have reduced user expenses in two ways. During peak hours, smart homes can operate autonomously. Excess energy can be sold to the power company for a profit. Figures 3(a) and 3(b) depict the sequencing of the PV system and the total number of devices for a typical day when ESSs are used. The PVs and ESSs strive to meet the lunchtime household need. During peak periods the grid can provide the deficit power (Figures 3(c) and 3(d)). This section examines the impact of RES on reducing expenses. The relationship between storage capacity and cost savings is explored using inconsiderable (less than 50%), medium, and large (up to 50%) scenarios. The single-house model calculates the total cost of each system when the houses do not exchange energy. Similarly, the proposed algorithm calculates the cost of trading between homes and the utility. Figures 5(e) and 5(f) show the energy savings to the grid from the price disparity between these two systems. They exhibit the relationship between grid savings and home usage during peak hours (non-RES). Figures 3(e) and 3(f) show how the percentage of households with storage devices affects the overall cost of the microgrid. It demonstrates that load inefficiency increases energy costs when using storage devices (cost curves increase with storage penetration). Smaller energy storage reduces overhead. Therefore, storage only saves money with renewable energy sources or microgrid connections.

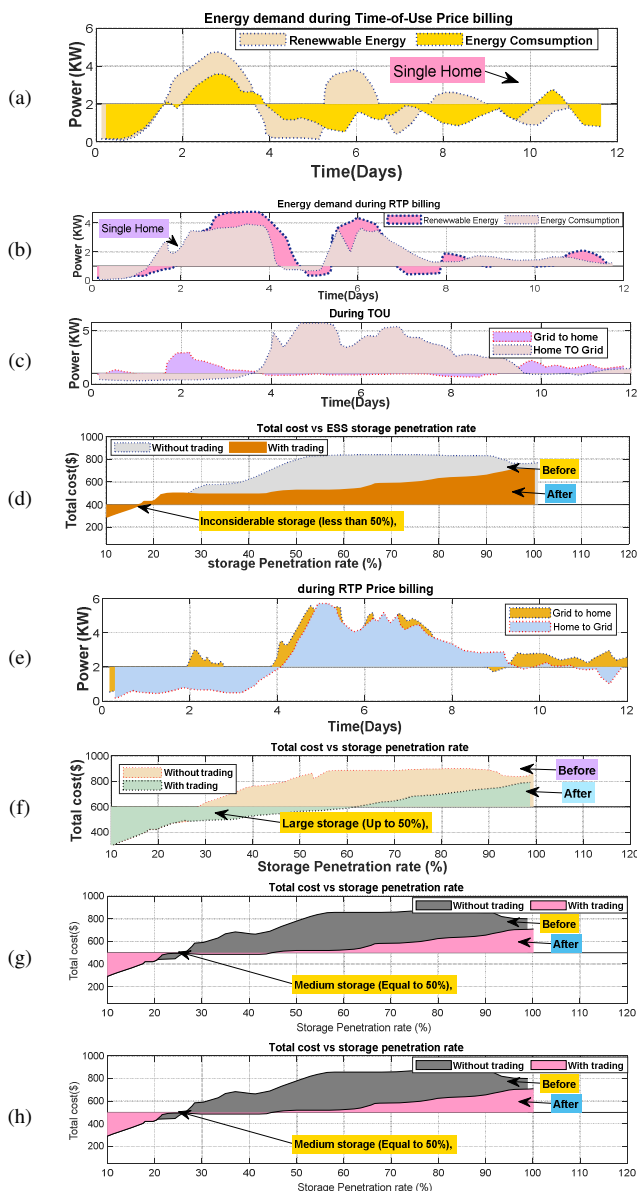


Fig. 3. A typical day without blackouts.

2) Daily Performance Results

Figure 4 reveals how each house in the TOU and RTP cases, benefits from the excess solar energy. Based on the idea behind the proposed method, the results indicate the behavior of each node λ -Connectⁱ and the impact of TOU and RTP parameters. Each home manages its energy consumption during RTP and TOU. Figure 4 displays the best P2P EMS implementation parameters compared to home $\# \lambda$ -Connect1 running independently. House #2 exports PV electricity to house $\# \lambda$ -Connect1 from 11:00 am. At 3:00 pm, the additional PV energy from house #2 is stored in the HBSS of house $\# \lambda$ -Connect1. To reduce duty cycles and degradation costs, the EMS requires the ESS in House $\# \lambda$ -Connect2 to operate less efficiently and more expensively than the HBSS in-house #1. Figure 4(e) discloses that the in-house ESS discharged energy

to supply the consumption of House $\# \lambda$ -Connect2 and stored the excess energy. This procedure increases the deterioration cost of the ESS. The reduction in daily energy expenses in house 1 compensates for the rise in battery degradation expenses. Figure 4(h) illustrates that house $\# \lambda$ -Connect2 imports energy from house 2 at no energy sharing costs. House #1 imports energy for free from its surplus from 6:00 am to 11:00 am and from 4:00 pm to 7:00 pm to minimize its reliance on power from the supplier at high rates. At 7:30 am, the rated capacity of the homes) exceeded the surplus (Table I). The PV array is used to meet this additional load. If the energy limit is managed efficiently, the energy consumption of many households (λ -Connect2, λ -Connect6) will not exceed it. This will prevent customers and prosumers from having to purchase expensive energy. Node Connect has been verified in both TOU and RTP modes.

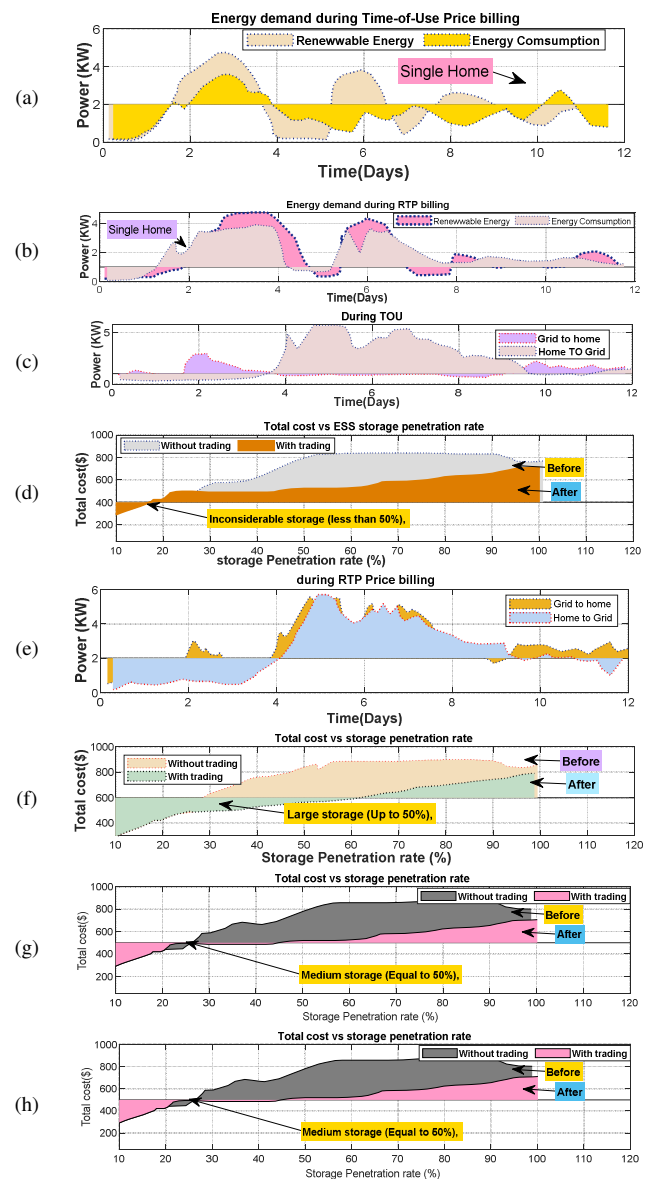


Fig. 4. A typical day without blackouts.

TABLE I. OBTAINED RESULTS WITHOUT BLACKOUTS

Cases			Objective function	PV (kW)	P^{ESS} (kW)
With PV/ESS	TOU	ESS	2631.00	3.899	0.000
	RTP	ESS	3987.00	3.897	0.000
Without PV/ESS	TOU	----	5998.00	---	---
	RTP	----	6876.00	---	---

3) Annual Performance Results

Figure 5 demonstrates the way the change in Home # λ -Connect1 annual consumption affects the percentage reduction in the annual cost of EMS for all nodes in the community. It shows that as node λ -Connect1 annual consumption increases, its annual percentage reduction increases. As household #1's annual consumption increases, (a) its annual percentage reduction decreases from 8.2% to 6.0%, (b) its percentage reduction decreases from 11.0% to 9.789% and then remains the same, (c) its percentage reduction decreases to 0.25%, and (d) its percentage reduction increases to 6.0% (see Table IV). Figure 5 indicates that as a node's load in an EMS community increases, most excess generation is used internally, leaving little energy to share with other homes. As in the case of household #1, the percentage reduction in the household's annual energy cost decreases. As a result, other nodes in the neighbourhood (e.g. home # λ -Connect2 and # λ -Connect4) will reduce their annual energy costs due to the lack of energy. Figures 5(c) and 5(d) demonstrate the annual consumption effects with and without outage events. Figure 5(d) presents two smart customer handling and storage processes from $T = 0$ to ΔT_d . Figure 5(e) compares retail and community prices with limited solar penetration. The community price is higher because the local power pool has acquired less solar power. This will limit the return on battery storage devices employed by smart users and will participate in the societal energy trade. Contrary to the idea, only some people will buy electricity from the power pool. The electricity costs for the smart user are also higher than in the first example. Similarly in (Figure 5(f)), the presented algorithm can benefit consumers. The trade actions of both agents are comparable, which indicates the algorithm's reliability. According to Figure 5(d), the proposed approach may lead to cost savings for both users of the energy community. The proposed methodology is solid and pervasive. Fuzzy Q-learning seeks to increase renewable energy usage. For example, the surplus-to-demand ratio is high when solar energy is plentiful in the middle of the day.

4) Low Solar Penetration

To prove that the proposed algorithm increases the use of renewable energy, it is implemented for the case of the addition of renewable energy penetration. In this scenario, the local energy pool collects only 30% of solar energy, which could increase the price of the community and reduce participation in the energy exchange (see Figure 6). Intelligent THA agents, trade with the neighborhood's energy pool when there is low solar penetration, as observed in Figure 6(f). The power pool and smart user do not exchange electricity from ΔT_c to ΔT_d . Community prices increase while solar energy penetration is low. Users then sell or store electricity. The phenomena mentioned above demonstrate the algorithm's accuracy. After empirically learning to maximize energy management, the

agent chooses the best version based on the q-value table. Although it is less effective than global optimization, continuous form-free online processes are nevertheless possible. Smart consumers can reduce their electricity costs and make the use of solar energy easier with the help of the proposed methodology. Utilizing the suggested fuzzy Q-learning algorithm, the augmented consumer revenue from the use of renewable energy can encourage renewable energy growth. Neighborhoods' monthly energy costs can be reduced in various ways. An intelligent, energetic community with a community energy pool is thus provided.

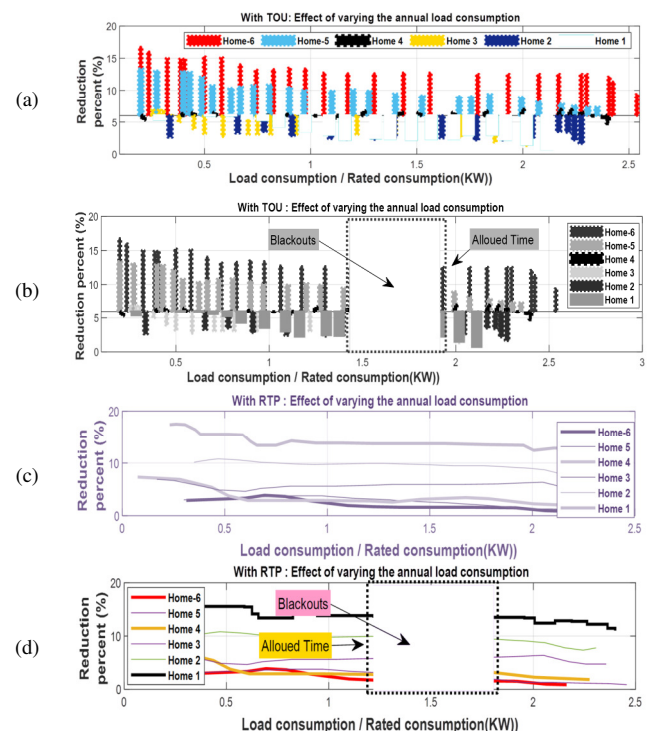


Fig. 5. Effect of changing the annual consumption of House No. 1 on the % reduction of the community pool while participating in the home energy management system with RTP and TOU and with/without outages.

V. DISCUSSION

The price will remain between the finite and retail costs, thanks to the proposed pricing methodology. Statistics showcase that this pricing scheme can help pure energy users reduce their monthly energy expenditure while boosting energy customers' profits. Reinforcement Q-learning can be applied to SG 2.0, e.g. in trading community energy. This ability allows SG 2.0 to constantly improve its methods and options for managing its energy demand. The proposed case study allowed SG 2.0 and renewable home energy generation to be more efficient for both consumers and prosumers. The community power pool will provide cheaper electricity, and smart users can sell the additional power to the community. SG 2.0 can change the user from a consumer to a prosumer. Through sophisticated storage and control systems, users can engage in the energy market, influencing electricity pricing and boosting profits. The results revealed that reinforcement Q-learning

solved the energy demand cost with high and low PV penetration. In this vein, the Q-Learning approach can be used to solve ongoing concerns related to SG 2.0.

findings derived from these two cases, both with and without the PV and BT can be pinpointed.

VI. CONCLUSION

This article proposes combining Demand Response (DR) using Real-Time Pricing (RTP) with peer-to-peer (P2P) energy trading to lower electricity expenses for prosumers and consumers. An accurate algorithm was initially employed to arrange the timing of smart home gadgets. A complex algorithm was used to maximize photovoltaic (PV), wind, storage, and Electric Vehicle (EV) resources in prosumer areas. In the second stage, users exchange excess electricity with other consumers who source it from the grid. Prosumers can effectively supply power to customers at a fair price by monitoring and identifying surplus power. Implementing P2P energy trading in smart homes successfully reduced costs for users and providers. Consumers 1 and 2 reduced their electricity expenses by 9.023 and 2.689 units, respectively, by obtaining power directly from customers instead of the grid. Prosumers 1 and 2 can reduce their electricity costs by 20.205 and 15.898 units by participating in P2P energy trading. The simulation results demonstrate the efficacy of the proposed method in decreasing power costs for both prosumers and consumers.

ACKNOWLEDGEMENT

The authors would like to thank the Deanship of Scientific Research at Umm Al-Qura University for supporting this work by Grant Code: (22UQU4400271DSR03).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Model Scope at <https://www.sciencedirect.com/science/article/abs/pii/S0959652621020102> (Sami Benslama 2021).

APPENDIX A: NOMENCLATURE

$f(t, p(t))$	Real-time market price in the community
$K_{pv}(t)$	THA solar energy at time t
$p(t)$	Energy ratio in the pool unit
$n(t, p(t))$	Total retail electricity market
$\alpha(t)$	Time dependent parameter
$\beta(t)$	Time dependent parameter
$g_m(t)$	Energy amount that SHA sells to the EPU at time t
C^{BT}_{max}	BT maximum capacity
$p'(t)$	BT charging/discharging rate at time t
$C^{BT}(t)$	BT capacity system
η^c	BT charging efficiency
η^d	BT discharging efficiency
$p^{c,max}$	BT maximum charging rate
$p^{d,max}$	BT maximum discharging rate
$P^{App}(t, s)$	App: Appliance rated power (kW)
$\Delta T_k(0, 1)$	Equal to 1 if the appliance is ON at scenario s and time t
UO^n	App upper-band operations time slot (s)
$P^{UG}(t, s)$	UG power delivered at time t and scenario s (kW)
ψ^{BT}	Equal to 1 if the BT has been replaced at year i
ψ^{pv}	Equal to 1 if the PV has been replaced at year i
z^{BT}	BT maintenance costs (\$/kWh/yr)
z^{pv}	PV maintenance costs (\$/kWh/yr)
a	Vector variable choices
W^b	The daily blackouts
ζ^o	Average cluster day

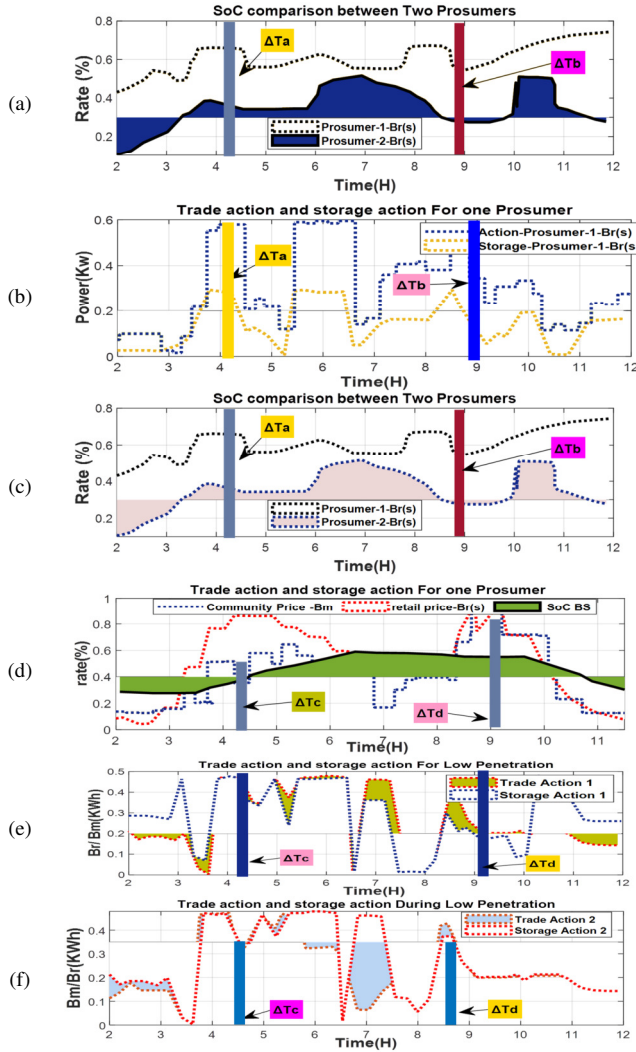


Fig. 6. A typical day with blackouts.

TABLE II. OBTAINED RESULTS WITH BLACKOUTS

Cases			Objective function	PV (kW)	PBT (kW)
With PV and BT	TOU	Li-Ion	2631.00	4	7.86
	RTP	Li-Ion	3987.00	3.997	7.99
Without PV and BT	TOU	-----	5998.00	3.997	8.01
	RTP	-----	6876.00	4	8.01

Table II displays the results when grid blackouts are considered in the problem. The obtained outcomes are only considered in PV and battery cases. It is noticed that under similar electricity use, the consumer's energy costs are lower than the customers'. Two users and three clients were engaged to assess the algorithm's consistency and accuracy. Likewise, the electricity usage for all consumers is comparable. Prosumer 1 has a SOC of 0.45, while Prosumer 2 has a SOC of 0.29. The

ζ^b	Average cluster breakout scenario
$\Sigma \Delta t^r$	BT rated power (kW)
$\Sigma \Delta t^s$	BT rated power (kW)
$SoC^{BT}(t,s)$	SOC at time t and scenario s
$g_a(t)$	THA/SHA aggregated-demand amount requesting energy from the pool
$h_m(t)$	Rest of the energy in the pool at time t
P^{GH}_{max}	Maximum grid-to-home power (kW)
P^{HG}_{max}	Maximum home-to-grid power (kW)
$P^{GH}(t)$	Grid-to-home power (kW)
$P^{HG}(t)$	Home-to-grid power (kW)
b_1	Is equal to 1 if the utility grid is delivering power to the home at representative scenario s and time t (binary)
b_2	Is equal to 1 if the home is delivering power to the grid at representative scenario s and time t (binary)
$O^r(t,s)$	Grid outage at time t and scenario s
$P_d^{BT}(t,s)$	BT power delivered to the home at scenario s and time t (kW)
$P_r^{BT}(t,s)$	Received power by the BT from the THA at time t and scenario s (kW)
ζ^{BT}	BT capital-cost (\$/kW)
ζ^{PV}	PV array capital-cost (\$/kW)
B_{limit}	Fixed budget limit (\$)
$B_{initial}$	Initial budget (\$)
χ^n	App operation time slots (s)
LO^n	App lower-band operations time slot (s)
$P^{OT}(t,s)$	Outage power required at time t and scenario s (kW)
σ^{BT}	BT replacement cost (\$/kWh)
σ^{PV}	PV replacement cost (\$/kWh)
ψ	Weight scenario s
Δv	Time step variation (h)
$\zeta^r(t)$	TOU or RTP tariff at scenario s and time t (\$/kWh)
$\zeta^s(t)$	Cost of the selling energy at scenario s and time t (\$/kWh)
$k^r(daily)$	is the total number of first days (clusters)
$k^r(year)$	the anticipated number of outages per year
$p^d(t,s)$	Power delivered by the BT to the THA at time t and scenario s (kW)
$p^c(t,s)$	BT power received from the THA at time t and scenario s (kW)
D^{BT}	BT depth discharge (pu)

APPENDIX B: INPUT PARAMETER OF THE HCPV SYSTEM

Parameter	Value
System technical parameters	
PV related power	1.0 kW
Interest rate	4.80%
PV system lifetime	25.0
Rated capacity (kW)	8.02 kW
Investment cost (δ_{PV}) (\$/KW)	769.0 \$/kW
PV cell number	N_s 3; N_p 6
$P_{Grid,max}$ (Kw)	9,725 kW
Maximum G_2H/H_2G - (P^{HG} , P^{HG})	10 kW
PV efficiency η_{PV} (pu)	0.13%
Max rated PV array power (kW)	4.2 kW
BT depth of discharge (pu)	0.6
PV array specification costs	
Whole capital	1130 \$/kW
Total maintenance cost per year	5.001 \$/kW
Replacement	398.31 \$/kWh
Expected-lifetime per year	21
BS array specification costs	
Whole capital	280 \$/kW
Total maintenance cost per year	14.2 \$/kW
Replacement	305. \$/kWh
Expected-lifetime per year	11
Whole capital	1130 \$/kW
BT charge efficiency (pu)	0.97
BT discharge efficiency (pu)	0.98

REFERENCES

- [1] S. Ben Slama, "Prosumer in smart grids based on intelligent edge computing: A review on Artificial Intelligence Scheduling Techniques," *Ain Shams Engineering Journal*, vol. 13, no. 1, Jan. 2022, Art. no. 101504, <https://doi.org/10.1016/j.asej.2021.05.018>.
- [2] R. Rodriguez *et al.*, "Sizing of a fuel cell-battery backup system for a university building based on the probability of the power outages length," *Energy Reports*, vol. 8, pp. 708–722, Nov. 2022, <https://doi.org/10.1016/j.egy.2022.07.108>.
- [3] G. K. Jabash Samuel and J. Jasper, "MANFIS based SMART home energy management system to support SMART grid," *Peer-to-Peer Networking and Applications*, vol. 13, no. 6, pp. 2177–2188, Nov. 2020, <https://doi.org/10.1007/s12083-020-00884-8>.
- [4] B. M. Manjunatha, S. N. Rao, A. S. Kumar, V. L. Devi, P. R. Mohan, and K. Brahmanandam, "An Enhanced Z-Source Switched MLI Capacitor for Integrated Micro-Grid with Advanced Switching Pattern Scheme," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 8936–8941, Aug. 2022, <https://doi.org/10.48084/etasr.4909>.
- [5] T. Khan, M. Yu, and M. Waseem, "Review on recent optimization strategies for hybrid renewable energy system with hydrogen technologies: State of the art, trends and future directions," *International Journal of Hydrogen Energy*, vol. 47, no. 60, pp. 25155–25201, Jul. 2022, <https://doi.org/10.1016/j.ijhydene.2022.05.263>.
- [6] Y. Liu, L. Wu, and J. Li, "Peer-to-peer (P2P) electricity trading in distribution systems of the future," *The Electricity Journal*, vol. 32, no. 4, pp. 2–6, May 2019, <https://doi.org/10.1016/j.tej.2019.03.002>.
- [7] L. Novoa, R. Flores, and J. Brouwer, "Optimal renewable generation and battery storage sizing and siting considering local transformer limits," *Applied Energy*, vol. 256, Dec. 2019, Art. no. 113926, <https://doi.org/10.1016/j.apenergy.2019.113926>.
- [8] J. L. Rojas-Renteria, T. D. Espinoza-Huerta, F. S. Tovar-Pacheco, J. L. Gonzalez-Perez, and R. Lozano-Dorantes, "An Electrical Energy Consumption Monitoring and Forecasting System," *Engineering, Technology & Applied Science Research*, vol. 6, no. 5, pp. 1130–1132, Oct. 2016, <https://doi.org/10.48084/etasr.776>.
- [9] A. F. Moreno Jaramillo, D. M. Laverty, D. J. Morrow, J. Martinez del Rincon, and A. M. Foley, "Load modelling and non-intrusive load monitoring to integrate distributed energy resources in low and medium voltage networks," *Renewable Energy*, vol. 179, pp. 445–466, Dec. 2021, <https://doi.org/10.1016/j.renene.2021.07.056>.
- [10] S. Nebili, I. Benabdallah, and A. Cherif, "Decoupling Control Applied to the Smart Grid Power Dispatching Problem," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 8960–8966, Aug. 2022, <https://doi.org/10.48084/etasr.5083>.
- [11] Y. Dai, Y. Gao, H. Gao, and H. Zhu, "Real-time pricing scheme based on Stackelberg game in smart grid with multiple power retailers," *Neurocomputing*, vol. 260, pp. 149–156, Oct. 2017, <https://doi.org/10.1016/j.neucom.2017.04.027>.
- [12] J. Shu, R. Guan, L. Wu, and B. Han, "A Bi-Level Approach for Determining Optimal Dynamic Retail Electricity Pricing of Large Industrial Customers," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2267–2277, Mar. 2019, <https://doi.org/10.1109/TSG.2018.2794329>.
- [13] H. Thomas, H. Sun, and B. Kazemtabrizi, "Closest Energy Matching: Improving peer-to-peer energy trading auctions for EV owners," *IET Smart Grid*, vol. 4, no. 4, pp. 445–460, 2021, <https://doi.org/10.1049/stg2.12016>.
- [14] S. Mohammadi, F. Eliassen, and Y. Zhang, "Effects of false data injection attacks on a local P2P energy trading market with prosumers," in *IEEE PES Innovative Smart Grid Technologies Europe*, The Hague, Netherlands, Oct. 2020, pp. 31–35, <https://doi.org/10.1109/ISGT-Europe47291.2020.9248761>.
- [15] A. Al-Sorour, M. Fazeli, M. Monfared, A. Fahmy, J. R. Searle, and R. P. Lewis, "Enhancing PV Self-Consumption Within an Energy Community Using MILP-Based P2P Trading," *IEEE Access*, vol. 10, pp. 93760–93772, 2022, <https://doi.org/10.1109/ACCESS.2022.3202649>.
- [16] M. I. Azim, W. Tushar, T. K. Saha, C. Yuen, and D. Smith, "Peer-to-peer kilowatt and negawatt trading: A review of challenges and recent advances in distribution networks," *Renewable and Sustainable Energy*

- Reviews, vol. 169, Nov. 2022, Art. no. 112908, <https://doi.org/10.1016/j.rser.2022.112908>.
- [17] H. Zang and J. Kim, "Reinforcement Learning Based Peer-to-Peer Energy Trade Management Using Community Energy Storage in Local Energy Market," *Energies*, vol. 14, no. 14, Jan. 2021, Art. no. 4131, <https://doi.org/10.3390/en14144131>.
- [18] Z. He, K. P. Tran, S. Thomassey, X. Zeng, J. Xu, and C. Yi, "Multi-objective optimization of the textile manufacturing process using deep-Q-network based multi-agent reinforcement learning," *Journal of Manufacturing Systems*, vol. 62, pp. 939–949, Jan. 2022, <https://doi.org/10.1016/j.jmsy.2021.03.017>.
- [19] A. Sheffrin, "Empirical Evidence of Strategic Bidding in the California ISO Real-time Market," in *Electricity Pricing in Transition*, A. Faruqui and B. K. Eakin, Eds. Boston, MA, USA: Springer, 2002, pp. 267–281.
- [20] Y. Wu, Z. Liu, B. Li, J. Liu, and L. Zhang, "Energy management strategy and optimal battery capacity for flexible PV-battery system under time-of-use tariff," *Renewable Energy*, vol. 200, pp. 558–570, Nov. 2022, <https://doi.org/10.1016/j.renene.2022.09.118>.
- [21] M. I. Azim, W. Tushar, and T. K. Saha, "Coalition Graph Game-Based P2P Energy Trading With Local Voltage Management," *IEEE Transactions on Smart Grid*, vol. 12, no. 5, pp. 4389–4402, Sep. 2021, <https://doi.org/10.1109/TSG.2021.3070160>.
- [22] K. Chaurasia and H. R. Kamath, "New Approach using Artificial Intelligence-Machine Learning in Demand Side Management of Renewable Energy integrated Smart Grid for Smart City," in *4th International Conference on Innovative Computing and Communication*, Delhi, India, Feb. 2021, pp. 1–5, <https://doi.org/10.2139/ssrn.3833753>.
- [23] Y. Wang, X. Wang, C. Shao, and N. Gong, "Distributed energy trading for an integrated energy system and electric vehicle charging stations: A Nash bargaining game approach," *Renewable Energy*, vol. 155, pp. 513–530, Aug. 2020, <https://doi.org/10.1016/j.renene.2020.03.006>.
- [24] F. Zeng, Y. Chen, L. Yao, and J. Wu, "A novel reputation incentive mechanism and game theory analysis for service caching in software-defined vehicle edge computing," *Peer-to-Peer Networking and Applications*, vol. 14, no. 2, pp. 467–481, Mar. 2021, <https://doi.org/10.1007/s12083-020-00985-4>.
- [25] M. B. Roberts, A. Bruce, and I. MacGill, "Impact of shared battery energy storage systems on photovoltaic self-consumption and electricity bills in apartment buildings," *Applied Energy*, vol. 245, pp. 78–95, Jul. 2019, <https://doi.org/10.1016/j.apenergy.2019.04.001>.
- [26] B. D. Raj, A. Sarkar, and D. Goswami, "An efficient framework for brownout based appliance scheduling in microgrids," *Sustainable Cities and Society*, vol. 83, Aug. 2022, Art. no. 103936, <https://doi.org/10.1016/j.scs.2022.103936>.
- [27] J. Wang, H. Zhong, Q. Xia, G. Li, and M. Zhou, "Sharing Economy for Renewable Energy Aggregation," in *Sharing Economy in Energy Markets: Modeling, Analysis and Mechanism Design*, J. Wang, H. Zhong, Q. Xia, G. Li, and M. Zhou, Eds. New York, NY, USA: Springer, 2022, pp. 107–142.
- [28] S. Benjaafar, G. Kong, X. Li, and C. Courcoubetis, "Peer-to-Peer Product Sharing," in *Sharing Economy: Making Supply Meet Demand*, M. Hu, Ed. New York, NY, USA: Springer, 2019, pp. 11–36.
- [29] P. R. Padghan, S. Arul Daniel, and R. Pitchaimuthu, "Grid-tied energy cooperative trading framework between Prosumer to Prosumer based on Ethereum smart contracts," *Sustainable Energy, Grids and Networks*, vol. 32, Dec. 2022, Art. no. 100860, <https://doi.org/10.1016/j.segan.2022.100860>.
- [30] P. Pediaditis, D. Papadaskalopoulos, A. Papavasiliou, and N. Hatziaargyriou, "Bilevel Optimization Model for the Design of Distribution Use-of-System Tariffs," *IEEE Access*, vol. 9, pp. 132928–132939, 2021, <https://doi.org/10.1109/ACCESS.2021.3114768>.
- [31] M. Mallet *et al.*, "Trends in local electricity market design: Regulatory barriers and the role of grid tariffs," *Journal of Cleaner Production*, vol. 358, Jul. 2022, Art. no. 131805, <https://doi.org/10.1016/j.jclepro.2022.131805>.
- [32] Y. Takeda and K. Tanaka, "Bidding Agent Model for P2P Energy Trading," *IEEE Transactions on Industry Applications*, vol. 140, no. 10, pp. 738–745, Oct. 2020, <https://doi.org/10.1541/ieejias.140.738>.
- [33] M. Aloud, "Adaptive GP agent-based trading system under intraday seasonality model," *Intelligent Decision Technologies*, vol. 11, no. 2, pp. 235–251, Jan. 2017, <https://doi.org/10.3233/IDT-170291>.
- [34] S. Ben Slama, "Design and implementation of home energy management system using vehicle to home (H2V) approach," *Journal of Cleaner Production*, vol. 312, Aug. 2021, Art. no. 127792, <https://doi.org/10.1016/j.jclepro.2021.127792>.
- [35] B. S. Sami, N. Sihem, and Z. Bassam, "Design and implementation of an intelligent home energy management system: A realistic autonomous hybrid system using energy storage," *International Journal of Hydrogen Energy*, vol. 43, no. 42, pp. 19352–19365, Oct. 2018, <https://doi.org/10.1016/j.ijhydene.2018.09.001>.
- [36] S. Zamanloo, H. Askarian Abyaneh, H. Nafisi, and M. Azizi, "Optimal two-level active and reactive energy management of residential appliances in smart homes," *Sustainable Cities and Society*, vol. 71, Aug. 2021, Art. no. 102972, <https://doi.org/10.1016/j.scs.2021.102972>.
- [37] J. Liu, H. Yang, and Y. Zhou, "Peer-to-peer energy trading of net-zero energy communities with renewable energy systems integrating hydrogen vehicle storage," *Applied Energy*, vol. 298, Sep. 2021, Art. no. 117206, <https://doi.org/10.1016/j.apenergy.2021.117206>.
- [38] I. Hammou Ou Ali, M. Ouassaid, and M. Maaroufi, "Optimal appliance management system with renewable energy integration for smart homes," in *Renewable Energy Systems*, A. T. Azar and N. A. Kamal, Eds. Cambridge, MA, USA: Academic Press, 2021, pp. 533–552.
- [39] D. Hemkumar, S. Ravichandra, and D. V. L. N. Somayajulu, "Impact of data correlation on privacy budget allocation in continuous publication of location statistics," *Peer-to-Peer Networking and Applications*, vol. 14, no. 3, pp. 1650–1665, May 2021, <https://doi.org/10.1007/s12083-021-01078-6>.
- [40] L. Ma, L. Wang, and Z. Liu, "Multi-level trading community formation and hybrid trading network construction in local energy market," *Applied Energy*, vol. 285, Mar. 2021, Art. no. 116399, <https://doi.org/10.1016/j.apenergy.2020.116399>.
- [41] N. E. H. Bourebia and C. Li, "A greedy energy efficient clustering scheme based reinforcement learning for WSNs," *Peer-to-Peer Networking and Applications*, vol. 15, no. 6, pp. 2572–2588, Nov. 2022, <https://doi.org/10.1007/s12083-022-01368-7>.

Exploring the Mechanical Behavior of Concrete enhanced with Fibers derived from recycled Plastic Bottles

Lana Ayad Abdulateef

Civil Engineering Department, Nawroz University, Kurdish Region, Iraq
lana.ayad@nawroz.edu.krd (corresponding author)

Sara Hikmat Hassan

Department of Civil and Environmental Engineering, University of Zakho, Iraq
sara.hassan@uoz.edu.krd

Ahmed Mohammed Ahmed

Department of Civil and Environmental Engineering, University of Zakho, Iraq
eng.ahmedahmed1998@gmail.com

Received: 12 January 2024 | Revised: 1 February 2024 | Accepted: 3 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6895>

ABSTRACT

The increasing issue of plastic waste has become detrimental to human society, particularly with the increase in disposable plastic bottles in many countries. This study investigates the impact of incorporating plastic bottle waste fibers on the slump, density, compressive strength, split tensile strength, and flexural strength of concrete. This material was selected for its cost-effectiveness and wide availability, addressing the prevalent global concern of environmental pollution resulting from inadequate waste management practices. This study describes a systematic plan to fabricate and test cubes, cylinders, and beams using Fiber-Reinforced Concrete (FRC). A comparative analysis was performed between concrete reinforced with plastic bottle waste fibers, in varying ratios of 1, 2, and 3%, and plain concrete. The results showed a positive impact on concrete strength with fiber addition, although at the expense of reduced workability and decreased concrete density. In particular, a significant improvement in the ductility of the concrete was observed. The analysis shows that a fiber ratio of 2% emerges as the most optimal dosage to achieve improved concrete properties. This study provides valuable insights into the imperative pursuit of sustainable concrete production and the environmental challenges posed by plastic waste.

Keywords-concrete; Fiber-Reinforced Concrete (FRC); plastic waste; concrete strength

I. INTRODUCTION

Concrete stands out as a highly adaptable material in the realm of civil engineering construction, capable of being molded into various forms and shapes [1]. Its fundamental components are derived from natural sources [2]. However, the addition of specific natural or artificial ingredients allows customization of concrete properties [3]. This versatile material has numerous advantageous characteristics, including commendable compressive strength, durability, impermeability, specific gravity, and resistance to fire [4-5]. However, concrete exhibits certain unfavorable properties, such as tension weakness, brittleness, susceptibility to cracking, lower impact strength, and considerable weight [6]. Fortunately, there are remedial measures to mitigate these drawbacks and improve the overall performance of concrete [7]. Several adverse properties of concrete are due to

microcracks at the mortar-aggregate interface [8]. A potential solution to address this problem involves incorporating fibers as an additional ingredient in the concrete mix. The inclusion of fibers in the cement-based matrix serves as a deterrent to unwanted microcrack formation [9]. By impeding the propagation of cracks under load, this approach can lead to improvements in both the static and dynamic properties of the cement-based matrix [10].

Waste poses a substantial challenge in terms of disposal and management and presents significant environmental, economic, and social issues [11]. Industrial activities contribute significantly to the generation of non-biodegradable solid waste [12]. An illustrative example is the widespread use of plastic bottles, specifically those made from polyethylene terephthalate (PET), to package mineral water in many countries [13]. PET bottles exhibit non-biodegradable properties, but efforts are made to address their environmental

impact through recycling [14]. These bottles are recycled and reused in various industries, contributing to sustainable practices and reducing the overall environmental footprint [15]. In civil engineering construction, waste management has become an appealing and essential disposal option, contributing significantly to environmental conservation [3, 10]. The industry recognizes the importance of sustainable practices in handling construction waste to minimize its impact on the environment [16]. Efficient waste management involves strategies such as recycling, reusing materials, and adopting responsible disposal methods [17]. By integrating these practices into construction processes, civil engineering efforts can not only reduce the environmental footprint but also promote a more sustainable and environmentally friendly approach to construction projects [7].

Architectural concrete stands out as a preferred choice for critical infrastructure projects such as bridges, dams, and tunnels, primarily due to its exceptional strength and durability [18]. This material exhibits remarkable resistance to wear and tear, ensuring prolonged structural integrity even in challenging environments. Beyond its utilitarian attributes, architectural concrete brings an aesthetic dimension to these essential structures, enhancing their visual appeal [19-20]. In the realm of architectural engineering, the use of polymer-modified concrete (PTE) has attracted attention, particularly for designing facades of buildings and bridges, as well as external walls [21]. Several studies have emphasized the significant benefits associated with the incorporation of waste plastic bottle fibers into concrete formulations. This innovative approach not only contributes to sustainability by recycling plastic waste but also imparts advantageous properties to concrete [22]. One notable advantage is the ability of concrete reinforced with waste plastic bottle fibers to reduce the material's weight, making it lighter and more manageable during construction [23]. The synergistic combination of structural strength, aesthetic appeal, and environmentally friendly practices makes PTE concrete with plastic waste fibers a compelling choice for the design and construction of modern infrastructure. Several studies investigated this type of Fiber Reinforced Concrete (FRC), observing that the incorporation of PET fibers in varying ratios led to a reduction in the workability of the concrete produced. Additionally, other studies showed that the inclusion of fiber content has an impact on the flow properties of concrete, with a reduction in slump observed as fiber content increased. This study aims to mitigate pollution and investigate cost-effective methods for producing FRC, incorporating plastic bottle fibers into reinforced concrete, and systematically evaluating the alterations in its mechanical behavior.

II. MATERIALS AND METHOD

A. Materials

All materials were tested according to ASTM standards to ensure accurate characterization. Cement serves as a crucial binding material to enhance the strength of concrete. This study used Type I Ordinary Portland Cement (OPC), adhering to ASTM C150 specifications [24]. Table I presents the properties of the cement used, which has been deemed acceptable by ASTM requirements. Potable tap water was used, meeting the

standards for drinkability and satisfying the criteria outlined in [25]. The water used was confirmed to be free of organic impurities. Locally sourced fine aggregates (Table II) were used. Rigorous tests, including sieve analysis, specific gravity, and additional evaluations per ASTM standards, were conducted to verify the properties of the sand used. The results of these tests were in accordance with the ASTM specifications. Table II also details the characteristics of the coarse aggregates used. To validate the properties of the aggregates, various tests, including sieve analysis, specific gravity, density, and additional sieve analysis tests, were performed following ASTM standards. These tests confirmed that the properties of the coarse aggregate met ASTM standards.

Recycled plastic fibers were derived from the cutting of PET water bottles. A uniform collection of plastic bottles of the same type was collected, followed by a thorough cleaning and drying process to eliminate any impurities. Subsequently, the bottles were cut longitudinally, introducing corrugations into the fiber profile to enhance structural integrity. The length of the cut ranged from 4 to 6 cm, as shown in Figure 1. The fiber content was chosen as a percentage of the weight of the concrete mix, with three specific percentages used for each case: 1, 2, and 3%. This study also used Sika Plastocrete admixture, a liquid plasticizer designed for use in concrete and mortar that serves a dual purpose, functioning as a highly efficient plasticizer and a waterproofing agent.

TABLE I. CEMENT PROPERTIES

Property	Value
Specific gravity	3.11
Consistency (%)	31
Initial setting time	200
Final setting time	230
Compressive strength (2 days) (Mpa)	24.8
Compressive strength (7 days) (Mpa)	40.3
Compressive strength (28 days) (Mpa)	46.4

TABLE II. PROPERTIES OF FINE AND COARSE AGGREGATES

Property/ value	Fine aggregates	Coarse aggregates
Density (Kg/m ³)	-	1709
Specific gravity	2.632	2.710
Water absorption (%)	1.37	0.78
Bulk specific gravity (Dry)	2.540	2.618
Bulk specific gravity (SSD)	2.581	2.576



Fig. 1. PET fibers prepared in this study.

B. Mix Design

The concrete mix used in this study adheres to a grade of M40 and its composition was meticulously produced following the ACI design. The mixture included 500 kg/m³ cement, 746 kg/m³ sand, 1270 kg/m³ gravel, and 249 kg/m³ water. Concrete components, including coarse and fine aggregates, cement, and fibers, were manually dry-mixed for about 4-5 min. Subsequently, the Sika Plastocrete admixture was incorporated into the mixture, as shown in Figure 2. The resulting concrete mix was poured into the mold in three layers and manually compacted according to ASTM standards. Cubes and beams adhered to the specified compaction standards (36 and 60 times, respectively) and a standard square steel rod (16 mm diameter, 60 mm length) with one end rounded was utilized. The cylinders were compacted 25 times using a small, round, straight steel tamping rod (10 mm diameter, 305 mm length). The immersion method was chosen for curing. Initially, the specimens were covered with nylon sheets for 24 h to prevent water loss through evaporation. Subsequently, the molds were removed, and the specimens were placed inside a water tank for continuous immersion curing. The specimens remained submerged until each testing age date, i.e. 3, 7, and 28 days. Figure 3 shows the cured samples prepared in three types.



Fig. 2. Concrete mixing.



Fig. 3. Cured samples.

C. Experiments

The concrete slump test [26] is a method to measure the workability of fresh concrete, a crucial characteristic in determining its usability. Concrete, after being compacted into three layers 25 times with a rounded rod, was subjected to a slump test to gauge its workability. The bulk density test involved the calculation of the density of each cube to evaluate the impact of fibers on its density. The density was determined by dividing the mass of a cube by its volume, employing the following formula:

$$\rho = M/V \quad (1)$$

where M is the mass in kg and V is the volume of the cube in m³. The compressive strength test, aligned with ASTM C39 specifications, used standard cubes (150×150×150 mm) instead of typical cylinders. Testing was carried out at three different ages, specifically after 3, 7, and 28 days [27]. Three specimens were used for each age, and the average value was calculated from the test results.

$$\sigma = P/A \quad (2)$$

where σ is the compressive strength, P is the maximum applied load in N, and A is the area of the loaded face of the cube in square mm.

The splitting tensile strength test was performed following ASTM C496 [28]. Cylindrical specimens of standard size, with a diameter of 150 mm and a height of 300 mm, were used in this test. After the curing process, three specimens for each age were surface-dried and positioned horizontally on the machine plate, with the diameter oriented. Wooden strips were placed at the top and bottom to prevent the concrete specimen from being crushed at those points during the test.

$$T = \frac{2 \times P}{\pi \times D \times L} \quad (3)$$

where P is the maximum applied load in N, D is the diameter of the specimen in mm, and L is the length of the specimen in mm. The resulting splitting tensile strength, denoted by T , is measured in MPa.

The flexural strength test, performed under ASTM C 293 [29], involved the use of beams of 560 mm length, 150 mm width, and 150 mm depth. The center-point loading method was used, where the entire load was applied at the center span of the beams during the test. The following formula was used to calculate the flexural strength:

$$R = \frac{3 \times P \times L}{2 \times B \times D^2} \quad (4)$$

where P is the maximum load in N, L is the span length in mm, B is the average sample width at fracture, and D is the average specimen depth at fracture. R is measured in MPa.

III. RESULTS AND DISCUSSION

A. Bulk Density Test

The results of the density test, shown in Figure 4, reveal a notable distinction between FRC and plain concrete. The density of FRC is lower than that of plain concrete. This divergence can be attributed to the specific gravity of the PET

fibers used in the concrete mix. In PET-FRC, a portion of the concrete volume was replaced with PET fibers, and because of their inherently lower density compared to the concrete matrix, an overall reduction in density was observed. This decrease in density underscores the influence of using PET fibers on the overall density characteristics of the concrete.

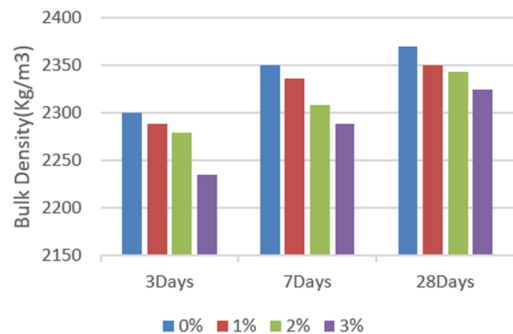


Fig. 4. Results of bulk density test.

B. •Slump Test

Figure 5 shows the slump test results, revealing a notable decrease in the workability of PET-FRC compared to plain concrete. This reduction can be attributed to the formation of a fibrous structure within the concrete, which leads to a decrease in the slump value. Furthermore, it can be observed that as the polymer fiber content increased, workability experienced a progressive reduction. The reduced workability in PET-FRC is a consequence of using fibers, as they alter the rheological properties of concrete. The presence of these fibers introduces a structural framework that hinders the free movement of particles, resulting in a decreased ability of the concrete mix to flow and deform. This effect is more pronounced at higher PF values, indicating that the amount of polymer fibers directly influences the workability of the concrete mix.

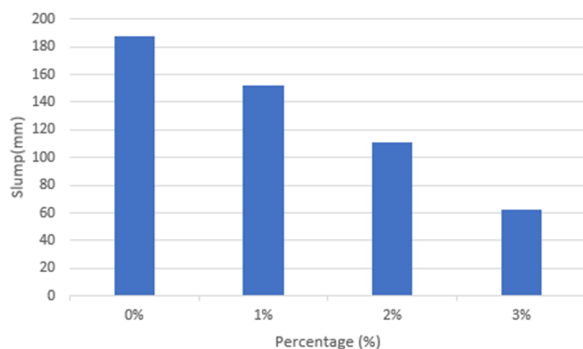


Fig. 5. Results of the slump test.

C. •Compressive Strength Test

Table III presents the results of the compressive strength test, showing a notable improvement in the compressive strength of PET-FRC compared to plain concrete. Strength improvement ranged from 2.4 to 8.7%, with variations depending on the curing age and the percentage of fibers

added. Specifically, at the 28-day mark, PET-FRC mixes containing 1, 2, and 3% fiber exhibited increases in compressive strength by 5.5, 8.7, and 7.3%, respectively. The significant increase in compressive strength observed in PET-FRC is attributed to the constructive role of PET fibers. These fibers delay the appearance and extent of cracks within the concrete structure. In particular, the maximum increase in compressive strength was achieved at a fiber content of 1%. The reinforcing effect of PET-FRC is instrumental in improving the overall compressive strength of concrete. By impeding the initiation and propagation of cracks, fibers contribute to a more robust and resilient concrete matrix. The results underscore the positive influence of using PET fibers on compressive strength, highlighting the potential of PET FRC as a promising material for applications where improved compressive strength is a critical requirement.

TABLE III. RESULTS OF COMPRESSIVE STRENGTH TEST

Fiber (%) in mix	3 days (MPa)	7 days (MPa)	28 days (MPa)
0%	25.860	35.837	42.923
1%	26.520	37.315	45.448
2%	27.243	38.302	46.993
3%	26.643	37.517	46.265

D. Splitting Tensile Strength

Table IV shows the results obtained from the tensile strength test of PET-FRC. The results show a substantial impact of fibers on the splitting tensile strength of the concrete, as an increase in fiber content corresponds to an increase in splitting tensile strength. The mixtures containing fibers exhibited a notable improvement in the splitting tensile strength, ranging from 5.1% to 19%, compared to control concrete and depending on the curing age. The most significant increase was observed with 2% fiber content. PET fibers play a crucial role in improving the concrete's resistance to crack propagation, as evidenced by the considerable boost in tensile strength. This mechanism involves PET fibers acting as a reinforcing agent that impedes the development and spread of cracks within the concrete matrix. This reinforcing effect is most pronounced at higher fiber content, such as the 2% level, indicating a positive correlation with the tensile strength of PET FRC. These results underscore the advantageous impact of PET fiber reinforcement in improving the mechanical properties of concrete and its potential to mitigate the adverse effects of cracking.

TABLE IV. RESULTS OF SPLITTING TENSILE STRENGTH TEST

Fiber (%) in mix	3 days (MPa)	7 days (MPa)	28 days (MPa)
0 %	1.742	2.162	3.601
1%	1.887	2.353	4.121
2%	1.922	1.452	4.299
3%	1.901	2.389	4.208

E. Flexural Strength Test

Table V presents the flexure test results, indicating a clear improvement in the flexural strength of PE-FRC compared to plain concrete and other variants. The improvement ranged from 6.3% to 18.6% compared to plain concrete, depending on

the age of the specimen. The maximum increase occurred with a fiber content of 2%. This improvement is attributed to the role of fibers in preventing and limiting crack extension, coupled with the improved compressive and tensile strength of PET-FRC. The flexural strength increased with increasing fiber content as a result of improved resistance to pulling. In summary, PE-FRC, especially with 2.0% fiber content, significantly improves flexural strength through crack prevention and improved material strength.

TABLE V. RESULTS OF THE FLEXURAL STRENGTH TEST

Fiber (%) in mix	3 days (MPa)	7 days (MPa)	28 days (MPa)
0 %	3.788	5.143	6.397
1%	4.042	5.739	7.418
2%	4.154	5.937	7.868
3%	4.053	5.798	7.591

IV. CONCLUSION

This comprehensive study investigated the use of PET fibers in concrete as a sustainable solution to address both economic and environmental challenges. The primary objective was to evaluate the impact of PET fibers on the behavior and mechanical properties of concrete, with a focus on reducing material costs and addressing solid waste issues associated with plastics. This study exhibited the reduced workability of FRC with PET fibers, revealing insights into their intricate impact on aggregate movement. The lower density of PET-FRC, compared to plain concrete, offers a novel application in lightweight structures, especially as an alternative for retaining structures. This expands the design possibilities and shows the versatility of PET-FRC. The results of the slump test unveiled a reduction in workability for PET-FRC, emphasizing the influence of fibers on impeding aggregate movement. The incorporation of PET fibers into concrete demonstrated a substantial improvement in its compressive, tensile, and flexural strength. This improvement can be primarily attributed to the reinforcing effect induced by the fibers, along with their ability to delay the initiation and extension of cracks. This positive impact is especially pronounced when the fiber proportion reaches 2%, indicating the potential to optimize the mix for specific structural requirements. This study provides valuable information on the multifaceted benefits of integrating PET fibers into concrete, supporting sustainable construction practices, and offering innovative solutions to contemporary challenges.

RECOMMENDATION

Future studies could investigate diverse plastic sources, optimize fiber ratios, investigate workability improvement, conduct long-term durability assessments, and perform comprehensive environmental impact analyses for a more holistic understanding and advancement of environmentally friendly concrete production.

ACKNOWLEDGMENT

The authors express their gratitude to the Office of Research and Development of Nawroz University for their support in this study.

REFERENCES

- [1] E. L. Pereira, A. L. de Oliveira Junior, and A. G. Fineza, "Optimization of mechanical properties in concrete reinforced with fibers from solid urban wastes (PET bottles) for the production of ecological concrete," *Construction and Building Materials*, vol. 149, pp. 837–848, Sep. 2017, <https://doi.org/10.1016/j.conbuildmat.2017.05.148>.
- [2] A. R. Sandhu, M. T. Lakhia, A. A. Jhatial, H. Karira, and Q. B. Jamali, "Effect of River Indus Sand and Recycled Concrete Aggregates as Fine and Coarse Replacement on Properties of Concrete," *Engineering, Technology & Applied Science Research*, vol. 9, no. 1, pp. 3832–3835, Feb. 2019, <https://doi.org/10.48084/etasr.2558>.
- [3] A. I. Al-Hadithi and M. A. Abbas, "The Effects of Adding Waste Plastic Fibers on the Mechanical Properties and Shear Strength of Reinforced Concrete Beams," *Iraqi Journal of Civil Engineering*, vol. 12, no. 1, 2018.
- [4] J. M. Irwan, R. M. Asyraf, N. Othman, K. H. Koh, M. M. K. Annas, and S. K. Faisal, "The Mechanical Properties of PET Fiber Reinforced Concrete from Recycled Bottle Wastes," *Advanced Materials Research*, vol. 795, pp. 347–351, 2013, <https://doi.org/10.4028/www.scientific.net/AMR.795.347>.
- [5] N. A. Memon, M. A. Memon, N. A. Lakho, F. A. Memon, M. A. Keerio, and A. N. Memon, "A Review on Self Compacting Concrete with Cementitious Materials and Fibers," *Engineering, Technology & Applied Science Research*, vol. 8, no. 3, pp. 2969–2974, Jun. 2018, <https://doi.org/10.48084/etasr.2006>.
- [6] I. M. Nikbin *et al.*, "Life cycle assessment and mechanical properties of high strength steel fiber reinforced concrete containing waste PET bottle," *Construction and Building Materials*, vol. 337, Jun. 2022, Art. no. 127553, <https://doi.org/10.1016/j.conbuildmat.2022.127553>.
- [7] N. Bheel, R. A. Abbasi, S. Sohu, S. A. Abbasi, A. W. Abro, and Z. H. Shaikh, "Effect of Tile Powder Used as a Cementitious Material on the Mechanical Properties of Concrete," *Engineering, Technology & Applied Science Research*, vol. 9, no. 5, pp. 4596–4599, Oct. 2019, <https://doi.org/10.48084/etasr.2994>.
- [8] A. I. Al-Hadithi, A. T. Noaman, and W. K. Mosleh, "Mechanical properties and impact behavior of PET fiber reinforced self-compacting concrete (SCC)," *Composite Structures*, vol. 224, Sep. 2019, Art. no. 111021, <https://doi.org/10.1016/j.compstruct.2019.111021>.
- [9] D. Foti, "Use of recycled waste pet bottles fibers for the reinforcement of concrete," *Composite Structures*, vol. 96, pp. 396–404, Feb. 2013, <https://doi.org/10.1016/j.compstruct.2012.09.019>.
- [10] F. Pelisser, O. R. K. Montedo, P. J. P. Gleize, and H. R. Roman, "Mechanical properties of recycled PET fibers in concrete," *Materials Research*, vol. 15, pp. 679–686, Aug. 2012, <https://doi.org/10.1590/S1516-14392012005000088>.
- [11] H. M. Adnan and A. O. Dawood, "Strength behavior of reinforced concrete beam using re-cycle of PET wastes as synthetic fibers," *Case Studies in Construction Materials*, vol. 13, Dec. 2020, Art. no. e00367, <https://doi.org/10.1016/j.cscm.2020.e00367>.
- [12] A. C. Bhogayata and N. K. Arora, "Fresh and strength properties of concrete reinforced with metalized plastic waste fibers," *Construction and Building Materials*, vol. 146, pp. 455–463, Aug. 2017, <https://doi.org/10.1016/j.conbuildmat.2017.04.095>.
- [13] A. K. K. Reddy and C. A. Kumar, "Study of Mechanical Properties of Concrete using Waste Pet Bottle Fibres for the Reinforcement," *International Journal of Engineering and Management Research (IJEMR)*, vol. 4, no. 2, pp. 232–240, 2014.
- [14] K. Ramadevi and R. Manju, "Experimental investigation on the properties of concrete with plastic PET (bottle) fibres as fine aggregates," *International Journal of Emerging Technology and Advanced Engineering*, vol. 2, no. 6, pp. 42–46, 2012.
- [15] S. B. Kim, N. H. Yi, H. Y. Kim, J.-H. J. Kim, and Y. C. Song, "Material and structural performance evaluation of recycled PET fiber reinforced concrete," *Cement and Concrete Composites*, vol. 32, no. 3, pp. 232–240, Mar. 2010, <https://doi.org/10.1016/j.cemconcomp.2009.11.002>.
- [16] C. R. Gagg, "Cement and concrete as an engineering material: An historic appraisal and case study analysis," *Engineering Failure*

- Analysis*, vol. 40, pp. 114–140, May 2014, <https://doi.org/10.1016/j.engfailanal.2014.02.004>.
- [17] T. S. Vadivel and M. Doddurani, "An experimental study on mechanical properties of waste plastic fiber reinforced concrete," *International Journal of Emerging Trends in Engineering and Development*, vol. 3, no. 2, pp. 395–401, 2013.
- [18] M. N. Nurbayeva *et al.*, "Influence of Fiber on the Strength Characteristics of Fine-Grained Concrete," in *Proceedings of 2021 4th International Conference on Civil Engineering and Architecture*, Singapore, 2022, pp. 151–161, https://doi.org/10.1007/978-981-16-6932-3_13.
- [19] B. Panda, S. Chandra Paul, and M. Jen Tan, "Anisotropic mechanical performance of 3D printed fiber reinforced sustainable construction material," *Materials Letters*, vol. 209, pp. 146–149, Dec. 2017, <https://doi.org/10.1016/j.matlet.2017.07.123>.
- [20] J. Quek and R. Rajeev, "Composite System for Structural Enhancement and Efficient Heritage Conservation: A Case Study," in *Proceedings of the Third International Conference on Sustainable Civil Engineering and Architecture*, Da Nang City, Vietnam, 2024, pp. 419–428, https://doi.org/10.1007/978-981-99-7434-4_44.
- [21] R. T. Durai Prabhakaran *et al.*, "Plants and architecture: the role of biology and biomimetics in materials development for buildings," *Intelligent Buildings International*, vol. 11, no. 3–4, pp. 178–211, Oct. 2019, <https://doi.org/10.1080/17508975.2019.1669134>.
- [22] S. Bahij, S. Omary, F. Feugeas, and A. Faqiri, "Fresh and hardened properties of concrete containing different forms of plastic waste – A review," *Waste Management*, vol. 113, pp. 157–175, Jul. 2020, <https://doi.org/10.1016/j.wasman.2020.05.048>.
- [23] D. K. Dadzie, A. K. Kaliluthin, and D. R. Kumar, "Exploration of Waste Plastic Bottles Use in Construction," *Civil Engineering Journal*, vol. 6, no. 11, pp. 2262–2272, Nov. 2020, <https://doi.org/10.28991/cej-2020-03091616>.
- [24] "ASTM C150-07: Standard Specification for Portland Cement." 2007, [Online]. Available: <https://www.astm.org/standards/c150>.
- [25] "ASTM C1602/C1602M-22: Standard Specification for Mixing Water Used in the Production of Hydraulic Cement Concrete." [Online]. Available: https://www.astm.org/c1602_c1602m-22.html.
- [26] "ASTM C143/C143M-12: Standard Test Method for Slump of Hydraulic-Cement Concrete." [Online]. Available: https://www.astm.org/c0143_c0143m-12.html.
- [27] "ASTM C39/C39M-21: Standard Test Method for Compressive Strength of Cylindrical Concrete Specimens." [Online]. Available: https://www.astm.org/c0039_c0039m-21.html.
- [28] "ASTM C496-96: Standard Test Method for Splitting Tensile Strength of Cylindrical Concrete Specimens." [Online]. Available: <https://www.astm.org/c0496-96.html>.
- [29] "ASTM C293/C293M-16: Standard Test Method for Flexural Strength of Concrete (Using Simple Beam With Center-Point Loading)." [Online]. Available: https://www.astm.org/c0293_c0293m-16.html.

A Study on the Influence of FDM Parameters on the Tensile Behavior of Samples made of PET-G

Dragos Gabriel Zisopol

Mechanical Engineering Department, Petroleum-Gas University Ploiesti, Romania
zisopold@upg-ploiesti.ro

Mihail Minescu

Mechanical Engineering Department, Petroleum-Gas University Ploiesti, Romania
mminescu@upg-ploiesti.ro

Dragos Valentin Iacob

Production Department, Marelli Ploiesti, Romania
dragoshich@gmail.com (corresponding author)

Received: 22 January 2024 | Revised: 19 February 2024 | Accepted: 24 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6949>

ABSTRACT

This experimental study investigated the influence of FDM 3D printing parameters on the tensile behavior of PET-G-made parts. In this context, 27 test specimens were produced using FDM on the Anycubic 4 Max Pro 2.0 printer with layer heights applied in one pass $L_h = 0.10/0.15/0.20$ mm and filling percentages $I_d = 50/75/100$ %. All these samples were tensile tested on the Barrus White 20 kN universal testing machine. The experimental results determined maximum tensile strength, elongation percentage at break, and Young's modulus. The two parameters considered, I_d and L_h , influence the maximum tensile strength, the elongation percentage at break, and Young's modulus. The findings demonstrated that the filling percentage has a strong influence on the maximum tensile strength and the elongation percentage at break of the PET-G samples, and L_h has a decisive influence on Young's modulus.

Keywords: FDM parameters; PET-G; tensile strength; percent elongation; modulus of elasticity

I. INTRODUCTION

Additive manufacturing technologies have revolutionized a part of the manufacturing processes, introducing a new concept of adding material in overlapping layers. The former have been adopted worldwide by key industries, such as aerospace, energy, medical, automotive, and manufacturing [1-3]. Additive manufacturing technologies emerge over conventional manufacturing technologies (subtractive and formative technologies) with the following main advantages: the ability to change the shape of the finished product without constraints, ease of use, energy efficiency, and environmental impact [2, 4-10]. However, additive manufacturing technologies have certain limitations, namely printing speed, part surface quality, and mechanical characteristics that are inferior to parts made with conventional technologies [11-13].

The mechanical features of the parts are of great importance and it is necessary to optimize the printing parameters to obtain their maximum values. The main FDM parameters that contribute to the mechanical characteristics of parts are the height of the layer applied in one pass (L_h), the filling

percentage (I_d), the printing speed (P_s), the filling pattern (I_p), the part orientation, the extruder temperature (E_t), and the table temperature (B) [14-26].

In [14], the influence of the FDM parameters (height of the layer applied in one pass and filling percentage) on the tensile strength and hardness of PLA parts was studied. The outcomes showed that better results were obtained with an increase in the filling percentage. In [15], the influence of the height of the layer applied in one pass and the color of the filament on the dimensional accuracy and tensile strength of FDM PLA parts was studied. According to the findings, the tensile strength decreases with increasing L_h and the color of the filament has an impact on tensile strength, as the FDM-made specimens from gray and red colored filaments had better mechanical characteristics compared to those made with black and natural colored filaments. In [16], a study on the optimization of FDM printing parameters to improve the tensile properties of parts produced from ABS and nylon was presented, varying three parameters: I_p (tri-hexagonal, triangular, octet), I_d (10/50/100%) and P_s (60/65/70 mm/s). The results exhibited

that among the three parameters studied, I_d had a strong influence on tensile strength since it increased with a higher I_d . In [27], the mechanical and durability performance of concrete with different doses of bacteria were investigated. The results disclosed that bacteria can improve the physical properties of cement concrete. In [28], the effect of different gradations of coarse aggregates on the mechanical properties of No-Fines Concrete (NFC) was studied, showing that the aggregate gradation significantly affects the compressive, splitting tensile, and flexural strength of NFC. In [29], an experimental study was carried out on NFC with a fixed proportion of cement-to-aggregate (c-a) of 1:6 with a 0.40 water/cement (w/c) ratio. It was discovered that the aggregate classification, the c-a ratio, and the w/c ratio have a significant impact on compressive strength. In [30], the influence of FDM process parameters (air gap, extruder temperature, layer thickness, infill density, and raster angle) on the tensile strength of parts printed with PLA material was investigated. The findings demonstrated that layer thickness and infill density have a direct impact on tensile strength and percentual elongation, and the best result for tensile strength was for the sample with the highest infill density.

This study aims to determine the influence of FDM on the tensile behavior of samples made of PET-G and to present the optimal combination of parameters for the best tensile strength results. Statistical analysis was performed on the findings using Minitab for a better understanding of the effects of FDM parameters on the tensile behavior of PET-G-made samples. The samples were produced and tested in laboratories of the Faculty of Mechanical and Electrical Engineering of the Petroleum-Gas University of Ploiești.

II. EXPERIMENTAL METHOD

Figure 1 illustrates the steps of the method used to investigate the influence of FDM parameters on the tensile behavior of PET-G parts. Solidworks 2023 [31] was used to create the 2D and then the 3D model of the tensile specimen, shown in Figure 2. The 3D model of the specimen was converted from SLD to STL format. The STL file, which corresponds to the specimen depicted in Figure 2, was processed in Cura Slicer software [32] by choosing the PET-G material and entering the FDM printing parameters from Table I, followed by the generation of the G-Code file.

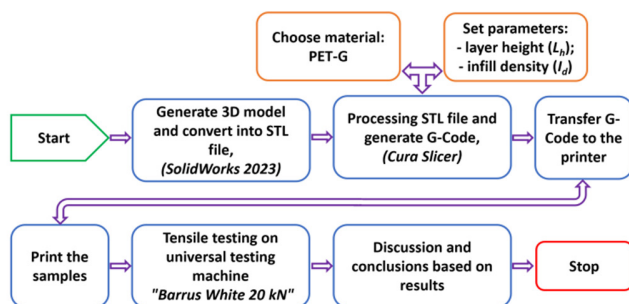


Fig. 1. The steps of the method used to investigate the influence of FDM parameters on the tensile behavior of PET-G parts.

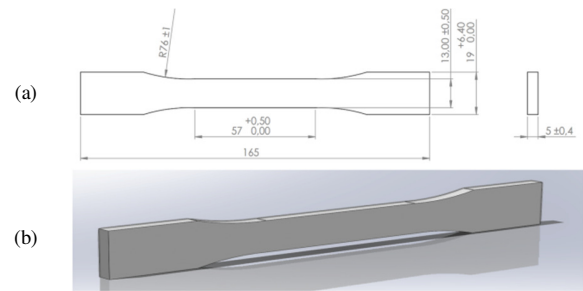


Fig. 2. Tensile test specimen models in SolidWorks 2023: (a) 2D, (b) 3D.

TABLE I. FDM PRINTING PARAMETERS FOR TENSILE SPECIMENS

Constant parameters	Variable parameters		Material
	Layer height L_h (mm)	Infill density I_d (%)	PET-G (pieces)
Part Orientation: X-Y Extruder Temperature, $E_t = 250\text{ }^\circ\text{C}$ Table Temperature, $B_t = 70\text{ }^\circ\text{C}$ Printing Speed, $P_s = 30\text{ mm/s}$ Filling Pattern I_p : Grid	0.10	100	3
		75	3
		50	3
	0.15	100	3
		75	3
		50	3
	0.20	100	3
		75	3
		50	3

Figure 3 displays the PET-G tensile specimen made in Cura Slicer software with constant parameters from Table I and $L_h = 0.20\text{ mm}$ and $I_d = 50\%$. The G-Code file was transferred to the Anycubic 4 Max Pro 2.0 3D printer. The latter has a print volume of $270 \times 210 \times 190\text{ mm}$, on which the 27 tensile specimens presented in Figure 4 were fabricated using Everfill brand PET-G filament with 1.75 mm diameter. The 27 were tensile tested on the "Barrus White 20 kN" universal testing machine, according to the ASTM D638-14 standard, utilizing a speed of 5 mm/min [33].

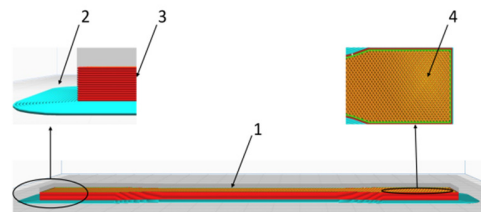


Fig. 3. Tensile test specimen in Cura Slicer software: (1) specimen, (2) specimen holder, (3) overlapping layers, (4) filling pattern.

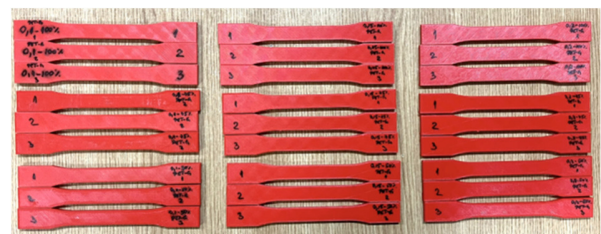


Fig. 4. PET-G tensile test specimens manufactured by FDM on Anycubic 4 Max Pro 2.0 printer.

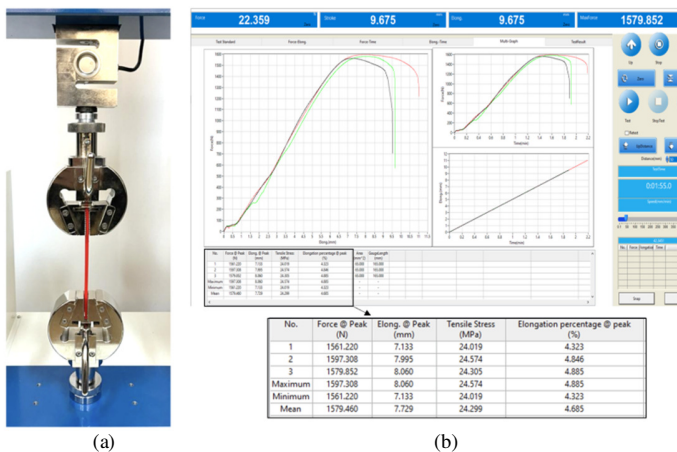


Fig. 5. Tensile test on Barrus White 20 kN machine: (a) Tensile test, (b) results of tensile test for $L_h = 0.20$ and $I_d = 100\%$.

III. RESULTS AND DISCUSSION

Tensile strength, percentage elongation at break, and Young modulus are the mechanical characteristics determined from tensile tests on the 27 specimens portrayed in Figure 4. Figure 5 presents the 27 specimens of PET-G after tensile testing.

A. Influence of FDM Parameters on the Tensile Strength of PET-G Tensile Specimens

Figure 7 illustrates the mean tensile strength values obtained from testing the specimens. As can be seen, I_d has a major influence on the tensile strength of PET-G specimens. The best results, 24.29-28.25 MPa, were obtained for specimens having filling percentage $I_d = 100\%$. By increasing I_d from 50 to 75%, the tensile strength of the specimens increased by 19.55-20.93%, while by rising I_d from 75 to 100%, tensile strength increased by 24.64-29.75%. The Pareto chart observed in Figure 8 was plotted using Minitab [30] and indicates that the filling percentage ($B = I_d$) has a decisive influence on the tensile strength of PET-G FDM specimens.

By utilizing Minitab software [34] and the FDM parameters in Table I, $L_h = 0.10/0.15/0.20$ mm and $I_d = 50/75/100\%$, the contour plot of the tensile strength of PET-G tensile specimens, pinpointed in Figure 10, was plotted [5]. Analysing the contour plot in Figure 10 led to an observation of how the L_h and I_d parameters of the FDM affect the specimen tensile strength. By increasing the filling percentage (I_d) it can be observed that this augmentation has a significant impact on the increase in tensile strength of PET-G tensile specimens.

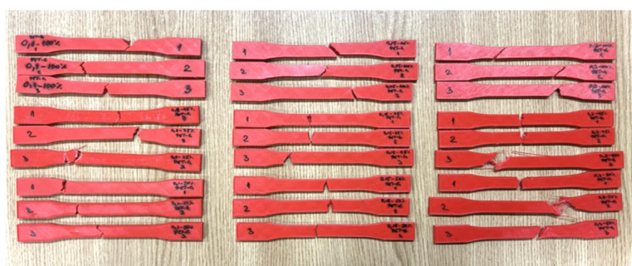


Fig. 6. PET-G specimens after tensile testing.

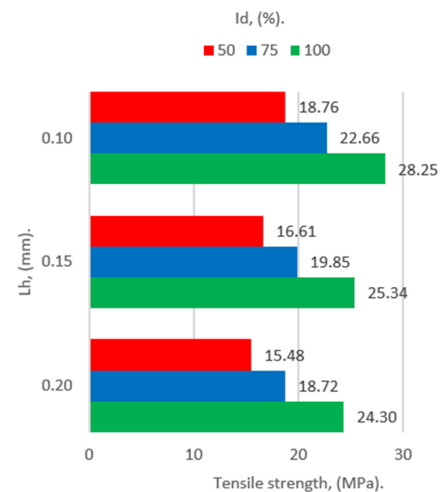


Fig. 7. Mean values of the tensile strength of PET-G specimens.

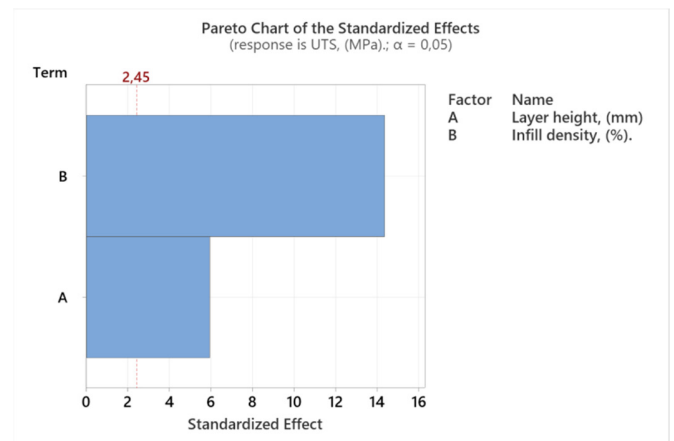


Fig. 8. Pareto chart of the influence of the parameters $A = L_h$ and $B = I_d$ on the tensile strength of PET-G tensile specimens.

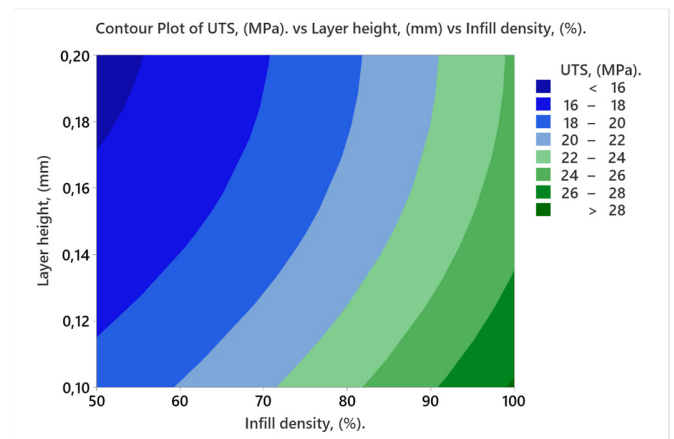


Fig. 9. Contour plot of the tensile strength of PETG tensile specimens.

B. Influence of FDM Parameters on the Elongation Percentage at Break of PET-G Tensile Specimens

Figure 10 presents the average values of the elongation percentage at break acquired from the tensile testing of the

specimens. Filling percentage (I_d) influences the elongation percentage at break of PET-G tensile specimens. The best results, 4.59-4.68%, were obtained for specimens with $I_d = 100\%$. By increasing I_d from 50 to 75%, the elongations at break of PET-G tensile specimens rose by 0.53-17.51%, while by increasing I_d from 75 to 100% the elongations at break rose by 15.91-33.43%.

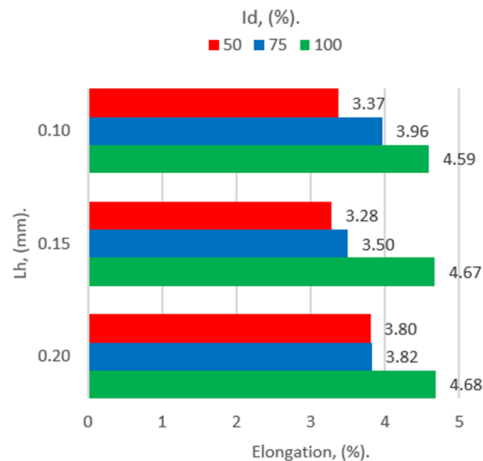


Fig. 10. Mean values of elongation percentage at break of PET-G specimens.

The Pareto chart shown in Figure 11 was plotted using Minitab and demonstrates that the filling percentage ($B = I_d$) strongly influences the elongation percentage at break of PET-G-made specimens.

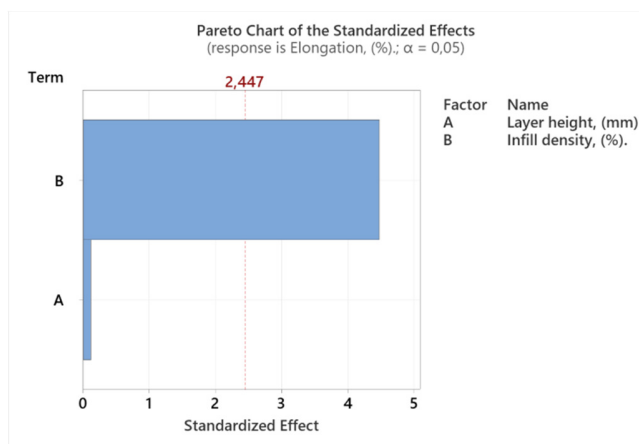


Fig. 11. Pareto chart showing the influence of the parameters $A = L_h$ and $B = I_d$ on the percentage elongation at break of PETG tensile specimens.

Figure 12 exhibits a contour plot of the elongation percentage of the specimens. It can be observed how the L_h and I_d of the FDM influence the elongation percentage at specimen break. Increasing I_d has the main influence on increasing the elongation percentage at break of PET-G tensile specimens.

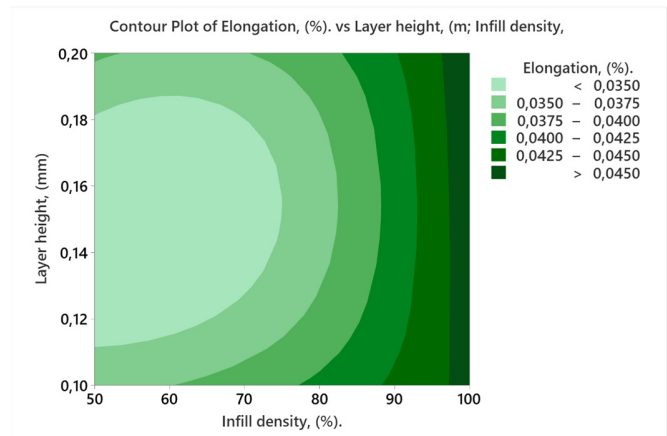


Fig. 12. Contour plot of percentage elongations at break of PET-G tensile specimens.

C. Influence of FDM Parameters on Young Modulus Values Obtained in the Tensile Test of PET-G Specimens

Figure 13 displays the average Young modulus values obtained from tensile testing of the specimens. The FDM parameters affect the Young modulus values. The best result, 0.62 GPa, was acquired for specimens with $I_d = 100\%$ and the height of the layer applied at one pass $L_h = 0.10$ mm. By increasing I_d from 50 to 75%, the Young modulus values increased by 11.93-21.08 %, and by increasing I_d from 75 to 100%, the Young modulus values increased by 5.31-7.67%. When decreasing the L_d applied from 0.20 to 0.15 mm, the Young modulus values increased by 4.33-24.04%, and when decreasing from 0.15 to 0.10 mm, the Young modulus values increased by 0.56-13.65%.

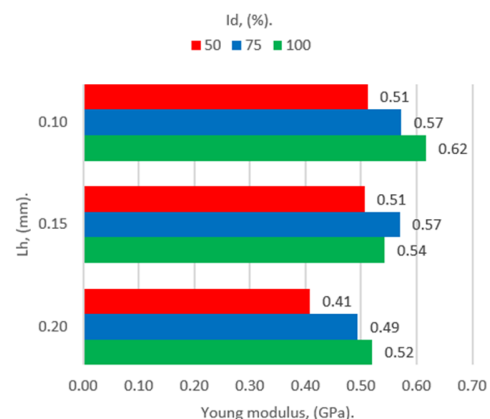


Fig. 13. Average values of Young modulus obtained in tensile test of PET-G specimens.

Figure 14 depicts a Pareto chart, indicating that the height of the layer applied at one pass ($A = L_h$) has the greatest influence on the Young modulus values obtained in the tensile test of PET-G specimens. Figure 15 portrays a contour plot of the Young modulus values acquired in the tensile test of PET-G specimens. It can be observed how increasing I_d and decreasing L_h determine an increase in the Young modulus values obtained in the tensile test of PET-G specimens.

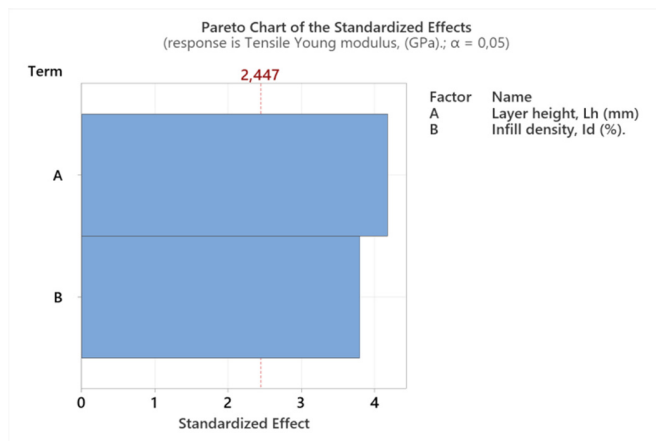


Fig. 14. Pareto chart of the influence of the parameters $A = L_h$ and $B = I_d$ on the Young modulus values obtained in the tensile test of PET-G specimens.

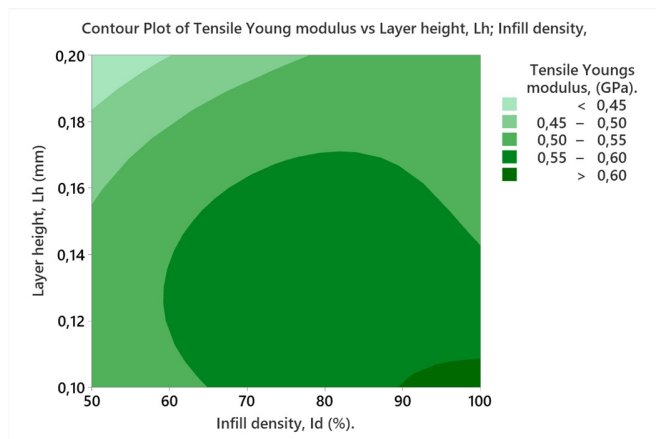


Fig. 15. Contour plot of Young modulus obtained in tensile test of PET-G specimens.

IV. CONCLUSIONS

The study investigated the influence of 3D FDM printing parameters on the tensile behavior of parts made of Everfill brand PET-G. In this context, 27 FDM-made specimens on the Anycubic 4 Max Pro 2.0 printer with layer heights applied at a pass $L_h = 0.10/0.15/0.20$ mm and filling percentage $I_d = 50/75/100\%$ were tensile tested on a Barrus White 20 kN universal testing machine. The parameters I_d and L_h influence the tensile behavior (tensile strength, percentage elongation at break, and Young modulus) of 3D-printed PET-G specimens. Increasing the filling percentage (I_d), increases the break strength of the specimens. This agrees with the results of [14, 16, 30]. By comparing the outcomes of this study to [14], the tensile strength of PET-G for $I_d = 50/75/100\%$ is less than the tensile strength of PLA with the same fill density with 31.10-56.22/16.50-44.05/18/60-58.14%. The minimum value of the tensile strength of the PET-G FDM specimens, 15.48 MPa, was recorded when using $L_h = 0.20$ mm and $I_d = 50\%$, and the maximum value of the tensile strength, 28.25 MPa, was recorded for $L_h = 0.10$ mm and $I_d = 100\%$.

Increasing the filling percentage (I_d) increased the elongation percentage of the specimens made by PET-G FDM.

The minimum value of elongation percentage at break of PET-G FDM specimens (3.28%) was recorded when using $L_h = 0.15$ mm and $I_d = 50\%$, while the maximum value (4.68%) was recorded for $L_h = 0.20$ mm and $I_d = 100\%$. Increasing the filling percentage (I_d) and decreasing the height of the layer applied in one pass (L_h) increased the tensile Young modulus values obtained in the tensile test of the PET-G FDM specimens. The minimum value of the Young modulus (0.41 GPa) was recorded when using $L_h = 0.20$ mm and $I_d = 50\%$, and the maximum value of the Young modulus (0.62 GPa) was documented in the case of $L_h = 0.10$ mm and $I_d = 100\%$. Minitab was used to analyze the results from a statistical point of view, to better understand the effects of FDM parameters on the behavior of tensile strength, percentage elongation, and Young modulus. The applicability of this study is for creating 3D printed pieces by FDM from PET-G with the optimal parameters for the best mechanical proprieties and printing time. Future studies could investigate other materials, such as ASA and rPET-G, with different percentages of recycled material.

REFERENCES

- [1] C. Jianu, S. I. Spiridon, and G. Ionita, "Tehnologia de fabricație aditivă 3D: Inovație și aplicabilitate în domeniul ingineriei," *Buletinul AGIR*, vol. 3-4, pp. 102–107, 2020.
- [2] "Innovation trends in additive manufacturing," European Patent Office, Munich, Germany, 2023.
- [3] D. G. Zisopol, M. Tănase, and A. I. Portoacă, "Innovative Strategies for Technical-Economical Optimization of FDM Production," *Polymers*, vol. 15, no. 18, Jan. 2023, Art. no. 3787, <https://doi.org/10.3390/polym15183787>.
- [4] D. G. Zisopol and A. Dumitrescu, *Materiale și tehnologii primare. Aplicații practice și studii de caz*. Ploiesti, Romania: Universității Petrol-Gaze din Ploiești, 2005.
- [5] D. G. Zisopol and A. Dumitrescu, *Ecotehnologie. Studii de caz*. Ploiesti, Romania: Editura Universității Petrol-Gaze din Ploiești, 2020.
- [6] D. G. Zisopol, A. Dumitrescu, and C. N. Trifan, *Ecotehnologie: Noțiuni teoretice, aplicații și studii de caz*. Ploiesti, Romania: Editura Universității Petrol-Gaze din Ploiești, 2010.
- [7] D. G. Zisopol and M. J. Săvulescu, *Bazele tehnologiei*. Ploiesti, Romania: Editura Universității din Ploiești, 2003.
- [8] M. J. Săvulescu and D. G. Zisopol, *Tehnologii industriale și de construcții*. Ploiesti, Romania: UPG Ploiesti Publishing House, 2002.
- [9] D. Zisopol, *Tehnologii industriale și de construcții. Aplicații practice și studii de caz*. Ploiesti, Romania: Publisher: Editura Universității din Ploiești, 2003.
- [10] M. J. Săvulescu, D. G. Zisopol, and I. Nae, *Bazele tehnologiei materialelor: Îndrumar de lucrări practice*. Ploiești, Romania: Editura Premier Ploiești, 1997.
- [11] D. G. Zisopol, N. Ion, and A. I. Portoaca, "Comparison of the Charpy Resilience of Two 3D Printed Materials: A Study on the Impact Resistance of Plastic Parts," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10781–10784, Jun. 2023, <https://doi.org/10.48084/etasr.5876>.
- [12] D. G. Zisopol, A. I. Portoaca, and M. Tanase, "Improving the Impact Resistance through Annealing in PLA 3D Printed Parts," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11768–11772, Oct. 2023, <https://doi.org/10.48084/etasr.6281>.
- [13] D. G. Zisopol, I. Nae, A. I. Portoaca, and I. Ramadan, "A Statistical Approach of the Flexural Strength of PLA and ABS 3D Printed Parts," *Engineering, Technology & Applied Science Research*, vol. 12, no. 2, pp. 8248–8252, Apr. 2022, <https://doi.org/10.48084/etasr.4739>.
- [14] D. G. Zisopol, I. Nae, A. I. Portoaca, and I. Ramadan, "A Theoretical and Experimental Research on the Influence of FDM Parameters on

- Tensile Strength and Hardness of Parts Made of Polylactic Acid," *Engineering, Technology & Applied Science Research*, vol. 11, no. 4, pp. 7458–7463, Aug. 2021, <https://doi.org/10.48084/etasr.4311>.
- [15] D. Frunzaverde *et al.*, "The Influence of the Layer Height and the Filament Color on the Dimensional Accuracy and the Tensile Strength of FDM-Printed PLA Specimens," *Polymers*, vol. 15, no. 10, 2023, <https://doi.org/10.3390/polym15102377>.
- [16] C. Vălean, L. Marşavina, M. Mărghitaş, E. Linul, N. Razavi, and F. Berto, "Effect of manufacturing parameters on tensile properties of FDM printed specimens," *Procedia Structural Integrity*, vol. 26, pp. 313–320, Jan. 2020, <https://doi.org/10.1016/j.prostr.2020.06.040>.
- [17] D. G. Zisopol, A. I. Portoaca, and M. Tanase, "Dimensional Accuracy of 3D Printed Dog-bone Tensile Samples: A Case Study," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11400–11405, Aug. 2023, <https://doi.org/10.48084/etasr.6060>.
- [18] D. G. Zisopol, I. Nae, A. I. Portoaca, and I. Ramadan, "A Theoretical and Experimental Research on the Influence of FDM Parameters on Tensile Strength and Hardness of Parts Made of Polylactic Acid," *Engineering, Technology & Applied Science Research*, vol. 11, no. 4, pp. 7458–7463, Aug. 2021, <https://doi.org/10.48084/etasr.4311>.
- [19] D. G. Zisopol, D. V. Iacob, and A. I. Portoaca, "A Theoretical-Experimental Study of the Influence of FDM Parameters on PLA Spur Gear Stiffness," *Engineering, Technology & Applied Science Research*, vol. 12, no. 5, pp. 9329–9335, Oct. 2022, <https://doi.org/10.48084/etasr.5183>.
- [20] D. G. Zisopol, M. Minescu, and D. V. Iacob, "A Theoretical-Experimental Study on the Influence of FDM Parameters on the Dimensions of Cylindrical Spur Gears Made of PLA," *Engineering, Technology & Applied Science Research*, vol. 13, no. 2, pp. 10471–10477, Apr. 2023, <https://doi.org/10.48084/etasr.5733>.
- [21] D. G. Zisopol, M. Minescu, and D. V. Iacob, "A Study on the Evaluation of the Compression Behavior of PLA Lattice Structures Manufactured by FDM," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11801–11806, Oct. 2023, <https://doi.org/10.48084/etasr.6262>.
- [22] D. G. Zisopol, M. Minescu, and D. V. Iacob, "A Study on the Influence of aging of the Butt-welded PE100 SDR11 on Shore A Hardness and Tensile Strength," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12722–12727, Feb. 2024, <https://doi.org/10.48084/etasr.6635>.
- [23] S. Subramonian *et al.*, "Artificial Neural Network Performance Modeling and Evaluation of Additive Manufacturing 3D Printed Parts," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11677–11684, Oct. 2023, <https://doi.org/10.48084/etasr.6185>.
- [24] B. Belarbi, M. E. A. Ghernaout, and T. Benabdallah, "Implementation of a New Geometrical Qualification (DQ) Method for an Open Access Fused Filament Fabrication 3D Printer," *Engineering, Technology & Applied Science Research*, vol. 9, no. 3, pp. 4182–4187, Jun. 2019, <https://doi.org/10.48084/etasr.2689>.
- [25] D. G. Zisopol, I. Nae, and A. I. Portoaca, "Compression Behavior of FFF Printed Parts Obtained by Varying Layer Height and Infill Percentage," *Engineering, Technology & Applied Science Research*, vol. 12, no. 6, pp. 9747–9751, Dec. 2022, <https://doi.org/10.48084/etasr.5488>.
- [26] D. G. Zisopol, A. I. Portoaca, I. Nae, and I. Ramadan, "A Comparative Analysis of the Mechanical Properties of Annealed PLA," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 8978–8981, Aug. 2022, <https://doi.org/10.48084/etasr.5123>.
- [27] A. S. Buller, A. M. Buller, T. Ali, Z. A. Tunio, S. Shabbir, and M. A. Malik, "Experimental Characterization of Bacterial Concrete Against Mechanical and Durability Performance," *Engineering, Technology & Applied Science Research*, vol. 11, no. 1, pp. 6703–6707, Feb. 2021, <https://doi.org/10.48084/etasr.3983>.
- [28] A. S. Buller, Z. A. Tunio, F. U. R. Abro, T. Ali, and K. A. Jamali, "Influence of Coarse Aggregate Gradation on the Mechanical Properties of Concrete, Part II: No-Fines Vs. Ordinary Concrete," *Engineering, Technology & Applied Science Research*, vol. 9, no. 5, pp. 4623–4626, Oct. 2019, <https://doi.org/10.48084/etasr.3021>.
- [29] Z. A. Tunio, F. U. R. Abro, T. Ali, A. S. Buller, and M. A. Abbasi, "Influence of Coarse Aggregate Gradation on the Mechanical Properties of Concrete, Part I: No-Fines Concrete," *Engineering, Technology & Applied Science Research*, vol. 9, no. 5, pp. 4612–4615, Oct. 2019, <https://doi.org/10.48084/etasr.3046>.
- [30] V. Ambade, S. Rajurkar, G. Awari, B. Yelamasetti, and S. Shelare, "Influence of FDM process parameters on tensile strength of parts printed by PLA material," *International Journal on Interactive Design and Manufacturing (IJIDeM)*, Aug. 2023, <https://doi.org/10.1007/s12008-023-01490-7>.
- [31] "3D CAD Design Software", *Solidworks*, <https://www.solidworks.com>.
- [32] "UltiMaker Cura," *UltiMaker*. <https://ultimaker.com/software/ultimaker-cura/>.
- [33] "ASTM D638-14: Standard Test Method for Tensile Properties of Plastics." <https://doi.org/10.1520/D0638-14>.
- [34] "Data Analysis, Statistical & Process Improvement Tools." <https://www.minitab.com/en-us/>.

Shoreline Changes and Sediment Transport along Nhat Le Coast, Vietnam

Vu Dinh Cuong

Key Laboratory of River and Coastal Engineering (KLORCE) - Vietnam Academy for Water Resources, Vietnam
cuongvd.vkhtlvn@gmail.com

Nguyen Thanh Hung

Key Laboratory of River and Coastal Engineering (KLORCE) - Vietnam Academy for Water Resources, Vietnam
nthung@vawr.org.vn (corresponding author)

Tran Dinh Hoa

Vietnam Academy for Water Resources, Vietnam
tranhhoa08@gmail.com

Nguyen Tien Thanh

Thuyloi University, Vietnam
thanht@tlu.edu.vn

Received: 27 December 2023 | Revised: 12 January 2024 | Accepted: 18 January 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6822>

ABSTRACT

One of the most beautiful beaches in Northern Vietnam, Nhat Le, has recently experienced severe erosion as a result of the ensemble interaction of natural factors, such as tropical cyclones, extreme weather events, and human activities. Consequently, negative impacts on tourism and social and economic development have been recorded. This paper aims to provide a deep understanding of the changes in shoreline and longshore sediment transport at Nhat Le estuary based on two modules of LITDRIFT and LITLINE of the LITPACK software package combined with geospatial analysis. The rate of change statistics is calculated using the Digital Shoreline Analysis System (DSAS) from 30-year multi-temporal satellite data (1989-2019) for multiple historical shoreline positions. The Module of LITDRIFT is employed to estimate sediment transport and the shoreline position calculated from the LITLINE module. These data are then compared with measured topographic data and satellite images. Wave climate conditions are incorporated into the LITDRIFT module to identify the volume of sediment transport along the coast on seasonal and annual bases. The results illustrate that a mean erosion rate of about 2 m per year was observed in the southern sandspit of Nhat Le from 1989 to 2019. This rate reaches 4.5 m per year during 2009-2019.

Keywords-Nhat Le beach; shoreline change; littoral sediment transport; LITPACK model; DSAS

I. INTRODUCTION

Vietnam is one of the countries affected by global climate change, especially in coastal zone and river deltas. With over 3,260 km and about more than 20 million people living on the Vietnamese coastline, it is one of the most densely populated areas in Southeast Asia. Furthermore, Vietnam is experiencing rapid urban and economic growth, and faces a high risk of coastal hazards related to tropical cyclones, floods, or extreme heavy rainfall. Natural coastal erosion may be caused by nearshore currents, wind, waves, weathering, or as a consequence of a reduction in river discharge through decreasing rainfall or snowmelt [1-2]. In other words, accretion

could also relate to river discharge during the rainy season or through increasing rainfall and the coasts are likely supplied by river sediment. River discharge and sediments from the river can be controlled by dam systems or weather patterns [3-5]. Many researches state that coastal erosion induced by human is often associated with the construction of river dams, sand mining, and the construction of coastal infrastructure (e.g. groynes, ports, inlet jetties, breakwaters, and sea walls) [6-10]. The important role of non-climatic anthropogenic drivers for the increasing vulnerability of low-lying coastal communities besides long-term and very slow nature impacts of sea level rise is studied in [11]. Accelerated sea level rise could lead to the permanent or frequent inundation of unprotected low-lying

land. Combined with climate conditions and storm surge activities, such areas are at a high risk of land subsidence or coastline retreat. All these factors constitute big challenges for understanding the shoreline changes and sediment transport in coastal regions generally and in Vietnam particularly. In the Central coast of Vietnam, more than one fifth of the coastline is being eroded, whereas coastal erosion has become one of the most serious environmental issues in Vietnam, affecting the national economy and social development [12]. For example, severe erosion is observed in Cua Dai beach, Hoi An City [13] and Hai Hau [14-15]. Specifically, Nhat Le and Bao Ninh are long sandy beaches (16 km long) located around Nhat Le estuary, and Nhat Le beach is one of the most beautiful beaches on the North Central coast of Vietnam. In recent years, there has been coastal erosion, particularly in the northern commune of Hai Thanh and the southern commune of Bao Ninh in Nhat Le mouth. The erosion has caused direct damage to coastal houses and infrastructure, and has had a significant impact on coastal tourism. The appearance of a coastal sand dune system is a multifaceted unfavorable topographical factor. For instance, under the influence of the wind, the phenomena of flying and flowing sands have caused sand dunes to gradually move towards the continent, narrowing the already small coastal plain and increasing flooding occurrences, especially in Nhat Le estuary. This may prevent individuals from insightfully understanding the regime of sediment transport, accretion-erosion and shoreline changes in the specific area. Within the coastal zone management, the prediction of coastal evolution with numerical models has proven to be a powerful technique to assist in the comprehension of the processes involved in the performance of necessary interventions and the selection of the most appropriate project design [16]. One-dimensional coastal morphology models (one-line models) have demonstrated practical capability in predicting long-term shoreline change [17]. One-line theory implies that all contour lines have similar shapes and move landward and seaward together up to a limiting offshore depth, the depth of closure, as if there were only one contour line [18]. Thus, one contour line can be used to describe changes in the beach plan shape and volume as the beach erodes and accretes. Many numerical models based on the one-line theory have been implemented for predicting coastal line changes. A well-known numerical model for shoreline change phenomena is LITPACK. Parameters such as wind, wave, current, and water level crucially affect sediment transport phenomena and they should be thus supplied as inputs for the numerical models which are taken from a two-dimensional (2D) hydrodynamics model (MIKE 21FM for this study).

Monitoring coastal areas in developing countries generally and in Vietnam particularly has always been an issue in that it could not provide an accurate identification of the location, the direction, and the geometric shape of the shoreline for a long time period. This information is vital to monitor and model the coastal erosion/accretion processes. To conduct the shoreline monitoring, the traditional methods of ground survey are not only time-consuming, but also not practical enough for large area monitoring. So, the remote sensing analysis method has been deployed in many studies regarding different countries [19-22]. In Vietnam, many researches applied the Digital

Shoreline Analysis System (DSAS) to detect the shoreline changes. Based on this technique, authors in [23] employed statistical measures of Net Shoreline Movement (NSM) and End Point Rate (EPR) involved in DSAS for monitoring the shoreline changes in the district of Ky Anh, Ha Tinh province from 1989 to 2013 with only five satellite images. Authors in [24] used DSAS to detect changes in shoreline without estimating the volume of sediment transport for Nhat Le estuary [24]. Utilizing a series of 10 optical satellite images (Landsat and Sentinel-2) from 2010 to 2020 and the statistical models of NSM, EPR, and Linear Regression Rate (LRR) are deployed for 5-year and 50 m intervals. The study pointed out the highest coastal erosion rate for the periods of 2010-2015 and 2015-2020, which is -5.94 m and -2.37 m per year, respectively. Furthermore, the highest accretion rate for these periods was found to be 2.05 m and 3.43 m per year. With four satellite images and DSAS, the shoreline changes for Dong Thap province, Vietnam from 2005 to 2019 were detected [25]. Shoreline changes in other provinces in Vietnam have been estimated using DSAS [26, 27]. Generally, DSAS technique has been widely applied, but its use in combination with LITPACK for understanding the shoreline changes and sediment transport in Nhat Le estuary is limited.

Therefore, the goal of this study is to deeply identify (1) the erosion-accretion evolution along the coast of Nhat Le and (2) the volume and direction of sediment transport along the coast in a seasonal and an annual scale. In order to achieve these targets, the DSAS technique combined with LITPACK software are applied to identify the rate of change and sediment transport for multiple historical coastline positions along the Nhat Le coast. The satellite images of Landsat and Sentinel from 1989 to 2019 were extracted and utilized. The coastal dynamics were digitized, visualized, and compared by two statistical parameters provided in DSAS, namely Shoreline Change Envelope (SCE) and LRR. The Linear Regression (LR) method of determining shoreline position change rate is an important technique, as it minimizes potential random error and short-term variability (cyclical changes) through the use of a statistical approach [28].

II. STUDY AREA

Nhat Le is the estuary of Nhat Le river originating from the Truong Son Cordillera with two major tributaries, Dai Giang and Kien Giang, and discharges into the sea at Dong Hoi city. Located in the tropical monsoon climate regime, the hydrodynamic regime in the estuary of Nhat Le is controlled by factors like waves, currents, tides and the river. Notably, there are two distinguishing seasons, one rainy (from September to March) and one dry (from April to August). Very heavy rains are mainly recorded in September, October, and November. Torrential rains may occur from July to October causing considerable rainfall and river floods due to severe typhoons coming from the Western Pacific basin. The total flows of four months of the flood season (September – December) produce 76% of the annual flow. The tidal regime in the estuary of Nhat Le is mainly characterized by semidiurnal tidal patterns with tidal amplitudes varying averagely from 0.7 to 0.8 m. The maximum and minimum values are over 1.61 m and 0.05 m, respectively [29].



Fig. 1. Sketch map of the study area.

III. MATERIALS AND METHODS

A. Materials

Landsat satellite images from 1989 to 2019 were acquired and freely downloaded from the United States Geological Survey (USGS) web-site (<https://earthexplorer.usgs.gov>). Images of Landsat satellite and Sentinel-2 images have a spatial resolution of 30 m and 10 m, respectively. Sentinel-2 image snapshotted on November 2nd, 2018 was used to verify the simulations of the changes in the shoreline from the module LITLINE. For the study area, the selected images were downloaded on the basis of several criteria: (1) One image per year and from the dry season (from February to June), (2) cloud cover less than 10% to limit unwanted atmospheric effects, and (3) the images should be at the same sea level (around mean sea level - national datum). A total of 30 images were selected according to these criteria.

B. Detection of Shoreline Changes using DSAS

DSAS is an add-in to Esri ArcGIS (a free available ArcGIS extension) designed by the USGS. It enables a user to detect shoreline movements in tempo-spatial scales and to capture the expansion and regression in coastline morphology. Furthermore, it facilitates calculating rate-of-change statistics from multiple historical shoreline positions. The shoreline is precisely identified through the process of handling remote sensing images, including the following steps:

- Calculating NDWI (Normalized Difference Water Index) to separate a water object from other objects [30].
- Dividing and separating shorelines based on the NDWI index. The DSAS creates straight lines perpendicular to the baseline and inter-sects the shoreline, thereby calculating the variability of shoreline based on the distance from the shoreline point to the baseline.

DSAS provides various methods in defining the rate-of-change statistics (e.g. SCE, EPR, and LRR). For this study, SCE and LLR were selected to calculate the coastline change. SCE is the distance between the shoreline furthest from and closest to the baseline at each transect. It represents the total change in shoreline movement for all shoreline positions [31]. The LRR is calculated using the least square regression lines from all shoreline positions along the transect. The rate of change in shoreline is defined by the slope of the regression

line. LRR has typically the advantage of using all the time-series data and discarding the impact of spurious values on the overall accuracy of change rate [32].

C. Modeling of Sediment Transportation

LITPACK model, developed by DHI is a software package for simulating non-cohesive sediment transport in wave and currents, littoral drift, coastline evolution, and profile development along quasi-uniform beaches [33]. The model employs a fully deterministic approach to estimate longshore transport. It consists of the modules of Longshore current and littoral drift (LITDRIFT), coastline evolution (LITLINE), cross-shore profile evolution (LITPROF), and sedimentation in trenches (LITTREN). The applied LITDRIFT and LITLINE modules calculate the amount of sediment transport on the coast and simulate the coastal response to gradients in the alongshore sediment transport under the conditions of waves, water level, and topography. Both models are based on one-line theory in which the cross section of shoreline is assumed to be unchanged during erosion/accretion for shoreline change modeling. It is worth noting that LITLINE calculates the coastline evolution. Then, LITDRIFT module is used for the estimation of longshore Sediment Transport (ST) rates. The module includes important ST mechanisms, such as non-linear wave motion, turbulent bottom boundary layer, wave breaking, and sediment grading. LITDRIFT computes longshore sediment transport from a time series of waves, water level parameters, and cross-shore profile characteristics. Other parameters, like graded sediments, currents, wind, and local roughness can be included as input data.

D. Model Setup and Validation

1) Model Setup

The LITPACK model is setup for two coastline segments around Nhat Le estuary. The first segment is located in the north site of Nhat Le beach with a length of 8 km and the other in the south site of Bao Ninh beach with a length of 6 km. Each segment is divided into 9 cross-sections labeled from CR_L1 to CR_L9 for the north coastline and CR_R1 to CR_R9 for the south coastline. Further divisions are performed between the cross-sections to create 800 and 600 sub-sections with a length of 10 m. Everything is selected as an input for the model. Locations of shoreline and cross-shore profile are displayed in Figure 2. The data used for setting up the models were utilized as follows:

- The cross-shore profiles were extracted from the topographic map from the baseline to the seabed points at elevation of around -12m, producing cross-shore profiles of 1935 m length. This length was set to 5 m resolution to the model with 387 points.
- A sediment characteristics' database on the analysis of the bottom sediment samples collected at the sites entailed in the survey was constructed. Average particle diameter D50 value ranged from 0.3 mm to 0.4 mm, while sediment sorting varies from 1.6 to 2.0. The calculation set the bottom roughness equal to 0.035 uniformly for all sections.
- Wave and water level characteristics at the cross-shore profiles were extracted from the results of Mike21FM

model [34]. Figure 2 shows a wave rose in the offshore and 6 wave roses in the near shore at the study area. The tidal water level in the study ranged from -0.9 m to 0.7 m.

- Shoreline data were extracted from the topographic map and Sentinel images in April 2018 and October 2018

2) Model Validation

There are two steps for validating the LIPPACK model. As the first step, LITDRIFT module is validated with the volume of longshore sediment transport analyzed from topographic data obtained in April 2018 and October 2018 for cross-shore profiles. The sediment volumes in Table I reveal that the calculated sediment volume from LITDRIFT is reasonably consistent with the measured sediment volume from the survey. A difference of less than 40% between the calculated and the measured sediment volume is observed in Table I. In other words, the model can simulate real-world conditions with an accuracy degree of more than 60%. This could be acceptable for sediment transport modeling. Based on this, LITDRIFT module can be used for the calculations of longshore sediment transport.

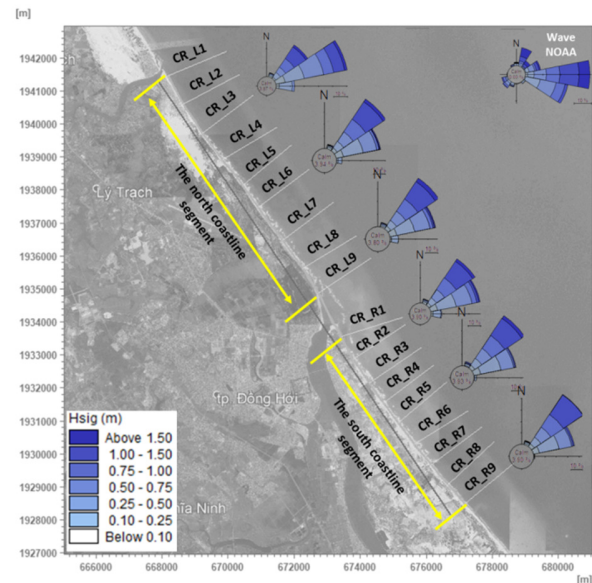


Fig. 2. Sketch map of the study area.

TABLE I. COMPARISON BETWEEN MODELED AND MEASURED SEDIMENT VOLUME (m^3)

Sediment volume from cross-section	CR_L6-CR_L7	CR_L7-CR_L8	CR_L8-CR_L9	CR_R1-CR_R2	CR_R2-CR_R3	CR_R3-CR_R4
Modeled	24.07	28.46	26.23	25.65	18.31	21.22
Measured	38.98	28.54	36.41	30.24	30.54	34.73
Relative error (%)	38.2	0.03	27.9	15.2	39.9	38.9

As the second step, LITLINE module was setup using the cross-shore profiles and sediment characteristics from the first step. The outputs of LITLINE were validated with the shoreline data analyzed from satellite images and topographic data. The shoreline positions calculated using LITLINE module from April 2018 to October 2018 were considered as the initial shoreline position. The calculated shoreline positions were calibrated with the observed shoreline data as the results of field surveys and of the Sentinel image analysis. Figure 3 demonstrates a close correlation between modeled and observed shoreline with a correlation coefficient greater than 0.9. Generally, the modeled shoreline is highly consistent with the measured shoreline for both north and south coast-lines.

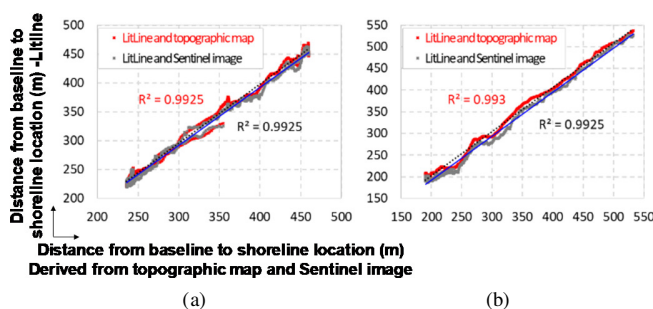


Fig. 3. Correlation between modeled and observed data. (a) North coastline, (b) south coastline.

IV. RESULTS AND DISCUSSION

A. Shoreline Changes from DSAS Analysis

In this study, the coastline was divided into 160 cross sections with 100 m equidistance from each other, as depicted in Figure 4. It is worth to note that this division is different from the division using the LITPACK model. The time periods are divided into 10-year intervals to detect the rate of shoreline changes. From that, the three periods of 1989-1999, 1999-2009, and 2009-2019 were analyzed within the overall 1989-2019 period. An important point for this selection is that Nhat Le estuary is strongly affected by sediment transporting from the river under the extreme flood events in 1990, 1999, 2009, and 2020.

The indicators of SCE [35] and LRR [36-37] were used to assess the rate of change statistics in the coastline of Nhat Le beach. A typical example of the regression line for 18 transects located in the north coastline is illustrated in Figure 5. The distances from the shoreline points to the baseline are represented by circle points and the dashed line is a linear regression line in the form $y = ax + b$ with coefficient a representing the slope of the line. The accretion rate is described as the shifting of the shoreline towards the seaward side with positive values of LRR, whereas the erosion rate is documented as the shoreline shift landward with the negative values of LRR. Figures 6 and 7 portray a variation in the mean rate of accretion and erosion in the selected four periods in Nhat Le coast using SCE and LRR, respectively. Especially, it is illustrated that during 1989-1999, the sandspit in the south of Nhat Le estuary is strongly eroded with a rate of 3.2 m/year,

whereas the sandspit in the north of Nhat Le estuary is strongly accreted with a rate of 4.5 m/year. A mean erosion rate of 3.5 m/year is found in the south coastline at the position of 100 m from the river mouth. During 1999-2009, the mean accretion rate of 3 m/year is smaller than the mean erosion rate in the previous period in the southern sandspit.

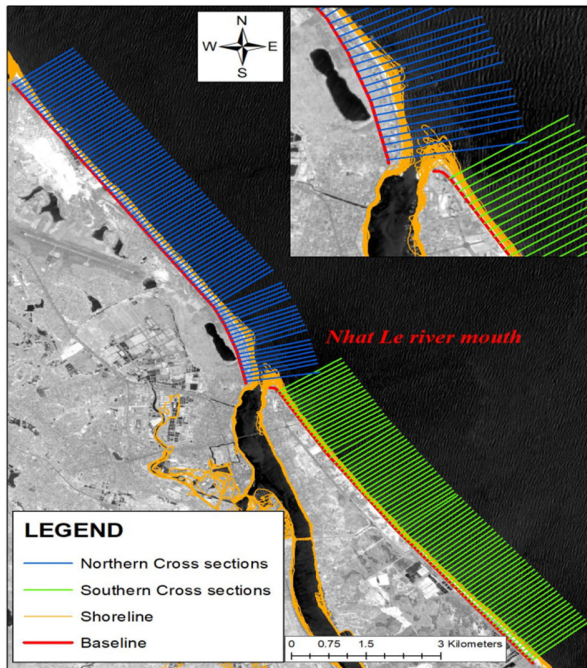


Fig. 4. Position of calculated and observed shorelines.

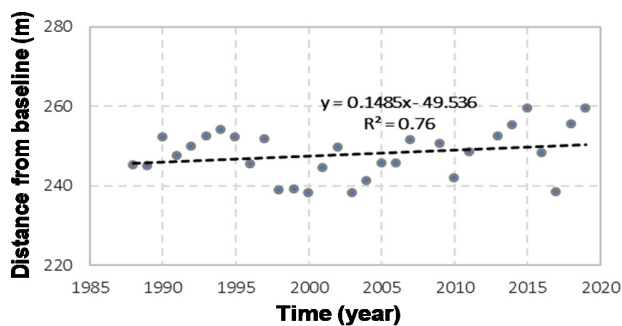


Fig. 5. Linear regression graph at cross section No 18.

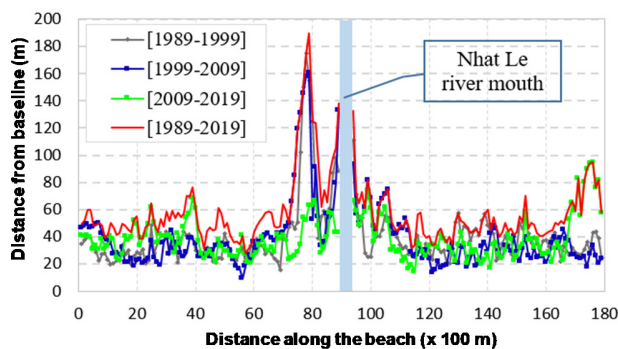


Fig. 6. SCE indicator to evaluate shoreline variability.

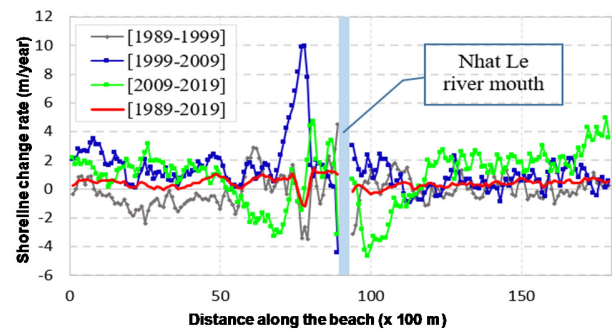


Fig. 7. LRR indicator to evaluate shoreline variability.

The maximum mean accretion rate of 10.1 m/year is discovered at the position of 100 m towards the north side from the river mouth. The sandspit in the north of Nhat Le estuary is the greatest one eroded with a rate of 4.5 m/year. During 2009-2019, there is a strong erosion in the southern sandspit with a rate of 4.5 m/year. Over the entire 1989-2019 period, the mean erosion rate 2 m/year and 1.2 m/year is recorded in the southern sandspit of Nhat Le estuary and at the position of 120 m from the Nhat Le estuary towards north, respectively. Using the SCE and LRR indicators reveals that the most significant shoreline changes are mainly concentrated in the vicinity of Nhat Le river mouth for all the three considered 10-year periods and the entire 1989-2019 period. These changes progressively decrease towards the sides of the mouth. A detailed analysis of shoreline changes, with a particular focus on the area near the Nhat Le estuary is shown in Figure 8. The large change rate of shoreline is definitely related to big sediment movement.

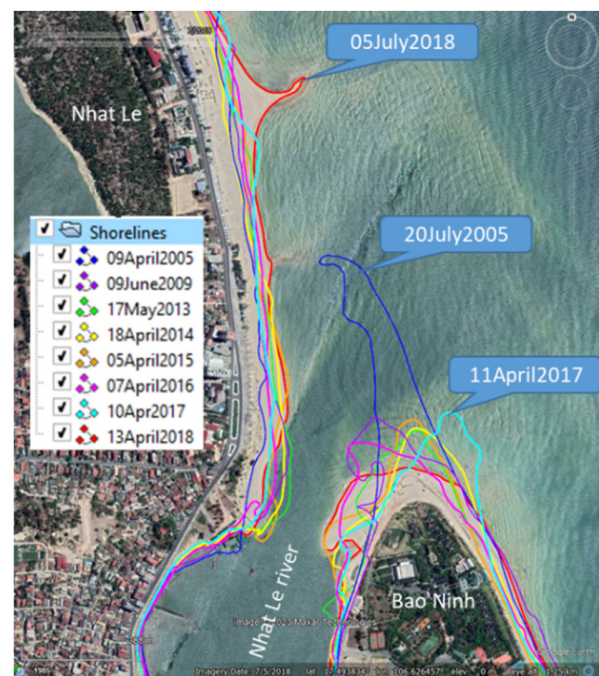


Fig. 8. Shoreline changes along the Nhat Le coast.

B. Longshore Sediment Transport

Obviously, understanding the sediment transport around estuaries and the longshore is crucial in formulating effective solutions to stabilize the shoreline. The longshore sediment transport can be calculated to determine the volume and direction in Nhat Le beach using the modules of the LITPACK model. LITDRIFT is applied in order to estimate the sediment transport along the coast of Nhat Le beach. Furthermore, the wave regime in the sea is calculated for a representative year, considering two distinct seasons, namely the northeast monsoon from October to March and the southeast monsoon from April to September. To measure the long-shore sediment transport, the wave and wind data from the WAVEWATCH III model (point at 17.5°N and 107°E) (<http://www.noaa.gov/>) with 2 parameters of direction and magnitude in 2013 are used. The selection of this year is due to the landfall of WUTIP tropical storm and the caused power loss in Quang Binh province. The wave data for boundary conditions of the LITDRIFT module were computed by the Mike21FM model

[33]. The sediment transport results for the typical year 2013 calculated from the LITDRIFT module are depicted in Figure 9 and Table II.

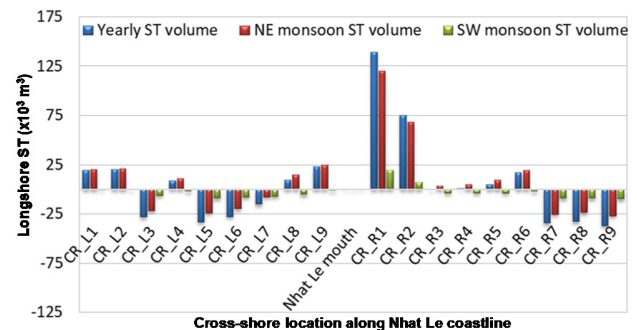


Fig. 9. Volumes of longshore sediment transport in annual and seasonal scales.

TABLE II. VOLUMES OF SEDIMENT TRANSPORT AT CROSS SECTIONS IN AN ANNUAL AND SEASONAL SCALE OF NORTHWEST AND SOUTHEAST MONSOON

Cross sections	Annual ($\times 10^3 \text{ m}^3/\text{year}$)		Northeast monsoon ($\times 10^3 \text{ m}^3/\text{year}$)		Southwest monsoon ($\times 10^3 \text{ m}^3/\text{year}$)	
	Net sediment transport from north to south	Total sediment transport	Net sediment transport from north to south	Total sediment transport	Net sediment transport from north to south	Total sediment transport
CR_L1	19.45	49.9	20.4	36.47	-0.94	13.43
CR_L2	20.41	47.26	20.87	34.73	-0.46	12.53
CR_L3	-28.26	42.77	-21.9	32.14	-6.36	10.63
CR_L4	8.73	44.2	10.94	31.58	-2.21	12.62
CR_L5	-33.29	55.15	-24.12	39.87	-9.17	15.28
CR_L6	-27.99	54.96	-19.97	40.04	-8.02	14.92
CR_L7	-15.12	55.66	-7.81	39.17	-7.31	16.49
CR_L8	9.2	76.37	14.39	54.67	-5.19	21.7
CR_L9	22.81	59.35	24.33	43.33	-1.52	16.02
Nhat Le river mouth						
CR_R1	139.1	149	120	122.9	19.1	26.1
CR_R2	75.1	98.57	68.03	77.55	7.07	21.02
CR_R3	-0.45	47.36	3.57	33.97	-4.02	13.39
CR_R4	0.79	45.04	4.85	31.85	-4.06	13.19
CR_R5	4.84	52.49	9.06	37.06	-4.23	15.43
CR_R6	16.72	59.67	19.09	42.89	-2.37	16.78
CR_R7	-34.43	52.98	-25.48	38.88	-8.95	14.1
CR_R8	-32.73	50.8	-23.69	36.69	-9.04	14.11
CR_R9	-36.91	53.18	-27.13	38.78	-9.78	14.4

As displayed in Figure 9, volumes of longshore sediment transport for 18 cross-shore profiles along the Nhat Le coastline are exposed to the annual and seasonal scales of northeast and southwest monsoons. Generally, the volume of sediment transport during the northeast monsoon is greater than that during the southwest monsoon. There is a shift from the south towards the north during the southwest monsoon. For cross sections CR_L1, CR_L2, CR_L4, and CR_L8 to CR_L9 in the north coastline of Nhat Le river mouth, the movement of net sediment transport is from the north to the south with the approximate volume of $20 \times 10^3 \text{ m}^3/\text{year}$. On the contrary, the movement of net sediment transport from the south to the north with the volume of over $20 \times 10^3 \text{ m}^3/\text{year}$ is observed for the other cross sections (CR_L3, CR_L5 to CR_L7). It is worth noting that the largest volume of sediment transport with over $70 \times 10^3 \text{ m}^3/\text{year}$ is observed in cross sections CR_R1 and

CR_R2, located near the river mouth. This volume is shifted from the north towards the south. There is a similar movement at the cross sections CR_R3 to CR_R6, but the volumes of sediment transport remarkably decrease by under $15 \times 10^3 \text{ m}^3/\text{year}$. The volume of sediment transport at the cross sections CR_R7 to CR_R9 is notably increased again by over $20 \times 10^3 \text{ m}^3/\text{year}$ from south to north. This reveals that sediment transport volumes substantially fluctuate in spatial-temporal scale.

Table II displays the volumes of net and total sediment transport at all cross sections. The negative values indicate the direction of sediment transport from south to north. In an annual scale, sediment transport volume of cross sections is mostly estimated from 40 to 60 strongly $10^3 \text{ m}^3/\text{year}$. This value is higher at several cross sections, e.g. CR_L8 (76×10^3

m^3/year), CR_R2 ($98 \times 10^3 \text{ m}^3/\text{year}$), and CR_R1 ($149 \times 10^3 \text{ m}^3/\text{year}$). Among that, total sediment transport volume of roughly 30 to over $40 \times 10^3 \text{ m}^3$ at these cross sections is observed during the northeast monsoon season. A larger amount of sediment transport is found out at cross sections CR_L8 (over $54 \times 10^3 \text{ m}^3$), CR_R2 ($77 \times 10^3 \text{ m}^3$), and CR_R1 ($122 \times 10^3 \text{ m}^3$). Notably, the volume of sediment transport at the cross sections is smaller with values from over 10 to under $17 \times 10^3 \text{ m}^3$. Some cross sections have larger volumes, such as CR_L8, CR_R2, and CR_R1 with over $21 \times 10^3 \text{ m}^3$, $21 \times 10^3 \text{ m}^3$, and $26 \times 10^3 \text{ m}^3$, respectively. There are large volumes of sediment transport at the CR_L4, CR_L8, and CR_R3 to CR_R5 cross sections but small net sediment transport. There is a similarity in the volume of sediment transport at these cross sections from north to south and south to north.

Figure 10 illustrates the results in the volume of sediment transport with its distribution at four typical cross sections along the coastline located in the north and south of Nhat Le estuary. CR_L6 and CR_L9 represent the north whereas CR_R6 and CR_R1 represent the south of Nhat Le estuary. Cross sections CR_L6 and CR_R6 are significantly far from the estuary, whereas the others are close. It can be observed that for the cross-sections along the straight coast, far from the river mouth, the morphology of the slope cross-section is relatively uniform. The distribution of sediment transport quantity along the shore is concentrated in a relatively narrow beach cross-section, approximately 300-400 m wide, coinciding with the coastal wave breaking zone. This distribution follows a curve with a single peak, as depicted in Figures 10(a) and 10(c). At cross section CR_L6, the total sediment transport (Net drift) is from the south to the north. So, the dominant sediment transport component is from the south to the north (Drift -ve), while the sediment transport component from the north to the south (Drift +ve) is less significant. The largest total amount of sand and mud transport reached $425 \text{ m}^3/\text{s/m}$. At cross section CR_R6, the total sediment transport volume is from the north to the south, because the dominant sediment transport component is from the north to the south, surpassing the sediment transport component from the south to the north. The largest total amount of sediment transport reached $175 \text{ m}^3/\text{s/m}$. Therefore, the total amount of sediment transport at section CR_L6 on the north coast and at section CR_R6 on the south coast is relatively small and shows minimal variation.

Furthermore, for the coast near the river mouth, the beach profile exhibits a gentle and irregular shape, characterized by the presence of a raised terrain area known as the shoal bar that obstructs the river mouth. The distribution of sediment transport along the coastline, across the beach cross-section, is considerably broader, spanning a width of approximately $200 \text{ m}^3/\text{s/m}$. This distribution takes the form of a curved pattern with two peaks, which is the result of repeated wave breaking occurrences (as depicted in Figures 10(b) and 10(d)). At cross section CR_L9, the total volume of sediment transport is from the north to the south, with the maximum amount reaching $145 \text{ m}^3/\text{s/m}$. Conversely, at section CR_R1, the total volume of sediment transport is also from the north to the south, with the largest quantity of sand and mud reaching $1130 \text{ m}^3/\text{s/m}$. Consequently, the total amount of sediment transport at cross

section CR_R1, located near the river mouth on the south bank, is considerably greater than that at cross section CR_L9, situated near the river mouth on the north bank.

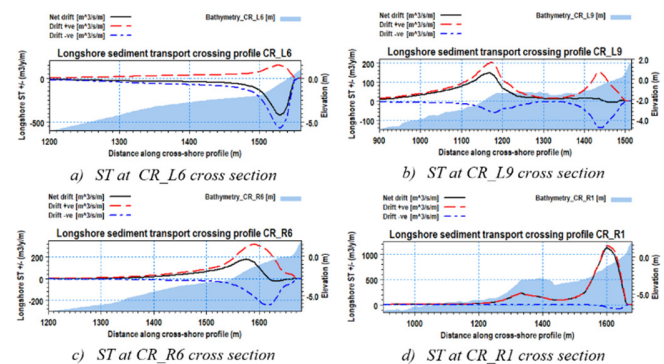


Fig. 10. Ranges of sediment transport across cross sections.

V. CONCLUSIONS

The Nhat Le coastline experienced significant changes over a 30-year period from 1989 to 2019, as observed through the analysis of multi-temporal remote sensing images. The coastline variability is particularly pronounced in the vicinity of the Nhat Le river mouth, gradually stabilizing as one moves further away from the mouth. The period from 2009 to 2019 witnessed more substantial shoreline changes compared to the other two periods. More specifically, the shoreline at Hai Thanh commune (located north of the river mouth) and Bao Ninh commune (located south of the river mouth) exhibited an average erosion rate of approximately 5 m per year. These areas experienced ongoing coastal erosion, resulting in the retreat of the shoreline over time.

The direction and volume of longshore sediment transport in Nhat Le beach were calculated with the help of the LITPACK software. The sediment transport exhibits significant variability among different cross-shore sections and between the two monsoon seasons. On average, the total volume of sediment transport through all cross-shore profiles is approximately $50 \times 10^3 \text{ m}^3$ per year. However, the highest volume of sediment transport occurs at the southern cross-shore profile close to the river mouth, reaching a value of $122 \times 10^3 \text{ m}^3$ per year. This indicates that the sediment transport is influenced by local factors such as the proximity to the river mouth, which can result in higher sediment transport rates in that specific area. The width of cross-shore sediment transport is around within 350 - 400 m along the coast and can be extended to 500 - 600 m for areas near the river mouth.

ACKNOWLEDGMENT

This paper was supported by the research projects: KC08.16/16-20 "Research on erosion and accretion along the coast and estuaries from Quang Binh to Thua Thien Hue provinces, considering impacts of the upper reservoirs and proposing solutions" and "Research and propose solutions to stabilize estuaries in the central coastal area for socio-economic development and disaster prevention and pilot application for Thuan An estuary, Thua Thien Hue province".

REFERENCES

- [1] E. C. F. Bird, *Coastal Geomorphology: An Introduction*, 2nd ed. Hoboken, NJ, USA: Wiley, 2008.
- [2] B.-L. Cui and X.-Y. Li, "Coastline change of the Yellow River estuary and its response to the sediment and runoff (1976–2005)," *Geomorphology*, vol. 127, no. 1, pp. 32–40, Apr. 2011, <https://doi.org/10.1016/j.geomorph.2010.12.001>.
- [3] A. Liaghat, A. Adib, and H. R. Gafouri, "Evaluating the Effects of Dam Construction on the Morphological Changes of Downstream Meandering Rivers (Case Study: Karkheh River)," *Engineering, Technology & Applied Science Research*, vol. 7, no. 2, pp. 1515–1522, Apr. 2017, <https://doi.org/10.48084/etasr.969>.
- [4] D. D. Cham, N. T. Son, N. Q. Minh, N. T. Hung, and N. T. Thanh, "Hydrodynamic Condition Modeling along the North-Central Coast of Vietnam," *Engineering, Technology & Applied Science Research*, vol. 10, no. 3, pp. 5648–5654, Jun. 2020, <https://doi.org/10.48084/etasr.3506>.
- [5] A. N. Laghari, W. Rauch, and M. A. Soomro, "A Hydrological Response Analysis Considering Climatic Variability: Case Study of Hunza Catchment," *Engineering, Technology & Applied Science Research*, vol. 8, no. 3, pp. 2981–2984, Jun. 2018, <https://doi.org/10.48084/etasr.2056>.
- [6] N. T. Hung, D. M. Duc, D. T. Quynh, and V. D. Cuong, "Nearshore Topographical Changes and Coastal Stability in Nam Dinh Province, Vietnam," *Journal of Marine Science and Engineering*, vol. 8, no. 10, Oct. 2020, Art. no. 755, <https://doi.org/10.3390/jmse8100755>.
- [7] J. A. Warrick, A. W. Stevens, I. M. Miller, S. R. Harrison, A. C. Ritchie, and G. Gelfenbaum, "World's largest dam removal reverses coastal erosion," *Scientific Reports*, vol. 9, no. 1, Sep. 2019, Art. no. 13968, <https://doi.org/10.1038/s41598-019-50387-7>.
- [8] J.-P. Bravard, M. Goichot, and S. Gaillot, "Geography of Sand and Gravel Mining in the Lower Mekong River," *EchoGeo*, no. 26, Dec. 2013, Art. no. 13659, <https://doi.org/10.4000/echogeo.13659>.
- [9] C. Jordan *et al.*, "Sand mining in the Mekong Delta revisited - current scales of local sediment deficits," *Scientific Reports*, vol. 9, no. 1, Nov. 2019, Art. no. 17823, <https://doi.org/10.1038/s41598-019-53804-z>.
- [10] K. F. Nordstrom, "Living with shore protection structures: A review," *Estuarine, Coastal and Shelf Science*, vol. 150, pp. 11–23, Oct. 2014, <https://doi.org/10.1016/j.ecss.2013.11.003>.
- [11] "Special Report on the Ocean and Cryosphere in a Changing Climate," *IPCC*. <https://www.ipcc.ch/srocc/>.
- [12] H. N. Thanh, *National Research project coded KC08.16/16-20. Research on erosion and accretion along the coast and estuaries from Quang Binh to Thua Thien Hue provinces, considering impacts of the reservoirs and proposing solutions*. Hanoi, Vietnam: Key Laboratory of River and Coastal Engineering, 2019.
- [13] D. D. Cham, N. Q. Minh, N. T. Lam, N. T. Son, and N. T. Thanh, "Identification of Erosion-Accretion Causes and Regimes Along the Quang Nam Coast, Vietnam," in *10th International Conference on Asian and Pacific Coasts*, Hanoi, Vietnam, Sep. 2019, pp. 809–814, https://doi.org/10.1007/978-981-15-0291-0_111.
- [14] D. M. Duc, K. Yasuhara, N. M. Hieu, and N. C. Lan, "Climate change impacts on a large-scale erosion coast of Hai Hau district, Vietnam and the adaptation," *Journal of Coastal Conservation*, vol. 21, no. 1, pp. 47–62, Feb. 2017, <https://doi.org/10.1007/s11852-016-0471-7>.
- [15] N. T. Thanh and N. H. Son, "Understanding Shoreline and Riverbank Changes Under the Effect of Meteorological Forcings," in *10th International Conference on Asian and Pacific Coasts*, Hanoi, Vietnam, Sep. 2019, pp. 1303–1310, https://doi.org/10.1007/978-981-15-0291-0_177.
- [16] H. Hanson *et al.*, "Modelling of coastal evolution on yearly to decadal time scales," *Journal of Coastal Research*, vol. 19, no. 4, pp. 790–811, 2003.
- [17] M. Dabees and J. W. Kamphuis, "Online, A Numerical Model for Shoreline Change," pp. 2668–2681, Dec. 2015, <https://doi.org/10.1061/9780784404119.202>.
- [18] J. W. Kamphuis, *Introduction To Coastal Engineering And Management*, 3rd ed. Hackensack, NJ, USA: World Scientific, 2020.
- [19] Y. Song, Y. Shen, R. Xie, and J. Li, "A DSAS-based study of central shoreline change in Jiangsu over 45 years," *Anthropocene Coasts*, vol. 4, no. 1, pp. 115–128, Jan. 2021, <https://doi.org/10.1139/anc-2020-0001>.
- [20] R. M. Abou Samra and R. R. Ali, "Applying DSAS tool to detect coastal changes along Nile Delta, Egypt," *The Egyptian Journal of Remote Sensing and Space Science*, vol. 24, no. 3, Part 1, pp. 463–470, Dec. 2021, <https://doi.org/10.1016/j.ejrs.2020.11.002>.
- [21] K. Nassar, W. E. Mahmod, H. Fath, A. Masria, K. Nadaoka, and A. Negm, "Shoreline change detection using DSAS technique: Case of North Sinai coast, Egypt," *Marine Georesources & Geotechnology*, vol. 37, no. 1, pp. 81–95, Jan. 2019, <https://doi.org/10.1080/1064119X.2018.1448912>.
- [22] M. M. Billah, "Mapping and Monitoring Erosion-Accretion in an Alluvial River Using Satellite Imagery – The River Bank Changes of the Padma River in Bangladesh," *Quaestiones Geographicae*, vol. 37, no. 3, pp. 87–95, Sep. 2018, <https://doi.org/10.2478/quageo-2018-0027>.
- [23] T. N. An, "An application of digital shoreline analysis system (dsas) to study shoreline change along ky anh coasts (ha tinh province) during 1989–2013," *Science Journal of Natural Resources and Environment*, vol. 21, pp. 66–72, Jun. 2019.
- [24] H. Le, N. Quang, V. Tuan, and M. H. Nguyen, "Analyzing River Estuary Changes Using Remote Sensing Images, a Case Study of Nhat Le River of Quang Binh Province," *TNU Journal of Science and Technology*, vol. 227, no. 3, pp. 15–24, Feb. 2022, <https://doi.org/10.34238/tnu-jst.5409>.
- [25] T. T. Kim, P. T. M. Diem, N. N. Trinh, N. K. Phung, and N. T. Bay, "Riverbank movement of the Mekong River in An Giang and Dong Thap Provinces, Vietnam in the period of 2005–2019," *Vietnam Journal of Hydrometeorology*, vol. 6, pp. 35–45, 2020.
- [26] D. V. To and P. T. P. Thao, "A shoreline analysis using DSAS in Nam Dinh coastal area," *International Journal of Geoinformatics*, vol. 4, no. 1, pp. 37–42, Mar. 2008.
- [27] H.-H. Nguyen, C. McAlpine, D. Pullar, S. J. Leisz, and G. Galina, "Drivers of Coastal Shoreline Change: Case Study of Hon Dat Coast, Kien Giang, Vietnam," *Environmental Management*, vol. 55, no. 5, pp. 1093–1108, May 2015, <https://doi.org/10.1007/s00267-015-0455-7>.
- [28] B. C. Douglas and M. Crowell, "Long-Term Shoreline Position Prediction and Error Propagation," *Journal of Coastal Research*, vol. 16, no. 1, pp. 145–152, 2000.
- [29] N. Dai, *Collection and adjustment of meteorological and hydrological data of Quang Binh province from 1956 to 2005. Research report*. Department of Natural Resources and Environment of Quang Binh Province, 2006.
- [30] S. K. McFEETERS, "The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features," *International Journal of Remote Sensing*, vol. 17, no. 7, pp. 1425–1432, May 1996, <https://doi.org/10.1080/01431169608948714>.
- [31] E. A. Himmelstoss, "DSAS 4.0 Installation Instructions and User Guide," in *Digital Shoreline Analysis System (DSAS) version 4.0 – An ArcGIS extension for calculating shoreline change: U.S. Geological Survey Open-File Report 2008-1278*, E. R. Thieler, E. A. Himmelstoss, J. L. Zichichi, and A. Ergul, Eds. 2009.
- [32] M. G. Honeycutt, M. Crowell, and B. C. Douglas, "Shoreline-Position Forecasting: Impact of Storms, Rate-Calculation Methodologies, and Temporal Scales," *Journal of Coastal Research*, vol. 17, no. 3, pp. 721–730, 2001.
- [33] *An integrated modeling system for littoral processes and coastline kinetics, short introduction and tutorial*. Copenhagen, Denmark: DHI, 2009.
- [34] N. T. Hung, V. D. Cuong, N. V. Hung, and N. Q. Minh, "Research on Seasonal Variability of Hydrody-Namic Regime in Nhat Le Estuary Area, Quang Binh Province," *Journal on Science and Technology of Irrigation (Vietnam Academy for Water Resources)*, no. 48, pp. 91–104, Oct. 2018.
- [35] E. R. Thieler, D. Martin, and A. Ergul, *The Digital Shoreline Analysis System, version 2.0: Shoreline change measurement software extension for ArcView. USGS Open-File Report 03-076*. USGS, 2003.
- [36] E. R. Thieler, E. A. Himmelstoss, J. L. Zichichi, and A. Ergul, "The Digital Shoreline Analysis System (DSAS) Version 4.0 - An ArcGIS

extension for calculating shoreline change," U.S. Geological Survey, Reston, VA, USA, 2008–1278, 2009, <https://doi.org/10.3133/ofr20081278>.

- [37] W. A. Ali, M. F. Kaiser, S. Kholief, and M. El-Tahan, "Assessment of coastal changes along Baltim resort (Egypt) using remote sensing and DSAS method," *Egyptian Journal of Aquatic Biology and Fisheries*, vol. 21, no. 1, pp. 37–48, May 2017, <https://doi.org/10.21608/ejabf.2017.2380>.

Characterization of Pure and Doped ZnO Nanostructured Powders elaborated in Solar Reactor

Adriana-Gabriela Schiopu

Pitesti University Centre, Faculty of Mechanics and Technology, Politehnica Bucharest National University of Science and Technology, Romania
gabriela.schiopu@upb.ro

Mihai Oproescu

Pitesti University Centre, Faculty of Mechanics and Technology, Politehnica Bucharest National University of Science and Technology, Romania
mihai.oproescu@upb.ro (corresponding author)

Vasile Gabriela Iana

Pitesti University Centre, Faculty of Mechanics and Technology, Politehnica Bucharest National University of Science and Technology, Romania
vasile_gabriel.iana@upb.ro

Sorin Georgian Moga

Pitesti University Centre, Regional Center of Research and Development for Materials, Processes and Innovative Products Dedicated to the Automotive Industry, Politehnica Bucharest National University of Science and Technology, Romania
sorin_georgian.moga@upb.ro

Denis Aurelian Negrea

Pitesti University Centre, Regional Center of Research and Development for Materials, Processes and Innovative Products Dedicated to the Automotive Industry, Politehnica Bucharest National University of Science and Technology, Romania
aurelian.negrea@upb.ro

Denisa Stefania Vilcoci

Pitesti University Centre, Regional Center of Research and Development for Materials, Processes and Innovative Products Dedicated to the Automotive Industry, Politehnica Bucharest National University of Science and Technology, Romania
denisa.vilcoci@upb.ro

Georgiana Cirstea

Pitesti University Centre, Regional Center of Research and Development for Materials, Processes and Innovative Products Dedicated to the Automotive Industry, Politehnica Bucharest National University of Science and Technology, Romania
georgiana.cirstea93@upb.ro

Catalin Marian Ducu

Pitesti University Centre, Regional Center of Research and Development for Materials, Processes and Innovative Products Dedicated to the Automotive Industry, Politehnica Bucharest National University of

Science and Technology, Romania
marian_catalin.ducu@upb.ro

Miruna-Adriana Iota

Doctoral School Materials Science and Engineering, National University of Science and Technology Politehnica Bucharest | National R&D Institute for Non-Ferrous and Rare Metals, INCDMNR-IMNR, 102 Biruintei Blvd, 077145 Pantelimon, Romania
iota.miruna@imnr.ro

Received: 17 January 2024 | Revised: 5 February 2024 | Accepted: 18 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6923>

ABSTRACT

The synthesis of nano-oxides is an important field of nanotechnology, as these materials possess unique properties and applications. Several methods have been developed for synthesizing nano-oxides, each offering advantages and disadvantages depending on the desired material characteristics. Solar energy focused on solar reactors can be utilized for nano-oxide elaboration, offering a sustainable and environmentally friendly approach. The current article presents the research carried out for the elaboration of pure and doped nanostructured zinc oxides using solar energy. The morphostructural characteristics were determined by X-Ray Diffraction (XRD), Scanning Electron Microscopy (SEM), and the Brunauer-Emmett-Teller method. The attenuated total reflectance Fourier transform infrared spectroscopy confirmed the synthesis of pure and doped nanostructured ZnO. The optical properties were highlighted by UV-VIS Spectroscopy. The research points out that crystallite sizes vary between 37 and 51 nm due to the influence of doping metal. The morphology associated with these particles is predominantly whiskers with elongated parts between 0.18 and 1.4 μm . Doping with Fe, Si, Yb, and Ce causes a wider band gap compared to pure ZnO nanoparticles. As solar energy becomes more accessible and efficient, solar-driven synthesis of pure and doped ZnO is poised to be a crucial factor in shaping the future of material science and technology.

Keywords: zinc oxide; nanostructured; morphology; characterization; solar reactor

I. INTRODUCTION

Solar energy is a sustainable energy source, offering several advantages over conventional energy sources like fossil fuels. Solar energy is derived from the sun's radiant energy, a virtually limitless resource. Unlike fossil fuels, which are finite and will be eventually depleted, solar energy can be harnessed indefinitely [1]. Solar energy can be harnessed in various forms, from rooftop solar panels to large-scale solar farms, making it adaptable to different locations and power requirements [2]. This versatility allows for the integration of solar energy into various energy systems. The field of solar synthesis, which utilizes solar energy to produce materials, has witnessed significant advancements in recent years, leading to the development of novel materials with unique properties and broader applications [3, 4]. These advancements hold immense potential for enhancing the performance and sustainability of solar energy technologies [5]. One notable area of innovation lies in the synthesis of nanomaterials utilizing solar energy. Solar energy offers several advantages over traditional synthesis methods, including:

- **Sustainability:** Solar energy is a renewable, clean, and abundant energy source [6], making it a more sustainable approach to nanomaterial synthesis compared to energy-intensive methods that rely on fossil fuels or hazardous chemicals [7].

- **Efficiency:** Solar-driven synthesis processes can be more efficient than conventional methods in terms of energy consumption and waste generation [8, 9].
- **Control over size and shape:** Solar energy can be used to control the size and shape of nanomaterials more precisely, leading to enhanced properties and applications [10].
- **Green chemistry:** Solar-based synthesis methods can minimize the use of hazardous chemicals and solvents, reducing the environmental impact of nanomaterials production [11-14].

Solar synthesis techniques enable the production of these nanomaterials with high purity and crystallinity, enhancing their photocatalytic activity. Solar synthesis methods can be employed to synthesize materials with controlled size, morphology, and surface chemistry, tailoring their properties for specific applications [15-23]. In addition to the synthesis of nanomaterials, solar synthesis has also been explored to produce other materials such as $\text{Bi}_{1.7}\text{Pb}_{0.3}\text{Sr}_2\text{Ca}_{(n-1)}\text{Cu}_n\text{O}_y$ ($n = 3\div 5$) ceramics [4].

The choice of the synthesis method depends on the desired specific properties of the nano-oxide, such as particle size, morphology, composition, and crystallinity. ZnO doped nanoparticles can be synthesized using various methods, including sol-gel, co-precipitation, hydrothermal, green synthesis, and Chemical Vapor Deposition (CVD) [24-27]. The choice of the synthesis method depends on the desired particle

size, morphology, and doping level. For instance, sol-gel methods are well-suited for obtaining uniform and small-sized ZnO NP (nanoparticles), while hydrothermal synthesis can produce larger particles with high crystallinity.

The use of solar energy for pure and doped ZnO nanostructured powder synthesis is an emerging field with the potential to revolutionize the way we produce nanomaterials for various applications. Pure ZnO, $\text{Zn}_{1-x}\text{Al}_x\text{O}$, $\text{Zn}_{1-x}\text{In}_x\text{O}$, $\text{Zn}_{1-x}\text{Co}_x\text{O}$, $\text{Zn}_{1-x}\text{Bi}_x\text{O}$, metallic Zn nanophases, nanophases of Si and SiO_x , CuO and Mn-doped CuO, as well as pure TiO_2 and Fe, Co or Mn-doped TiO_2 have been prepared by this method [10, 18]. Processing in a solar reactor has the advantages of giving a physical approach rather a chemical one with the strength of the absence of solvent contamination and the uniformity of nanoparticle distribution.

II. MATERIALS AND METHODS

Solar energy can be used to evaporate metal oxide precursors, leading to the formation of nanoparticles. The method is simple, scalable, and energy efficient, making it suitable for production of nano-oxide powders. The sun can be used to heat solid precursors in specially designed solar reactors. Such reactors are built in the CNRS-PROMES laboratory, UPR 8521, a member of the CNRS (French National Centre for Scientific Research) [28]. The acquired research uses as precursors micrometric oxides, easily procured commercially (Sigma Aldrich, p.a.). Doping was achieved for each dopant element through mechanosynthesis starting from metal oxides M_xO_y ($\text{M}=\text{Mg}, \text{Fe}, \text{Si}, \text{Ce}, \text{Yb}$), in a proportion of 0.1%, in weight percentage. Inside the solar reactor, the pure and doped ZnO in the form of tablets, is placed on a continuously cooled support. The solar energy captured by mirrors, located on the laboratory's 1st level, is directed to a spherical mirror, on the 6th level of the laboratory [28].

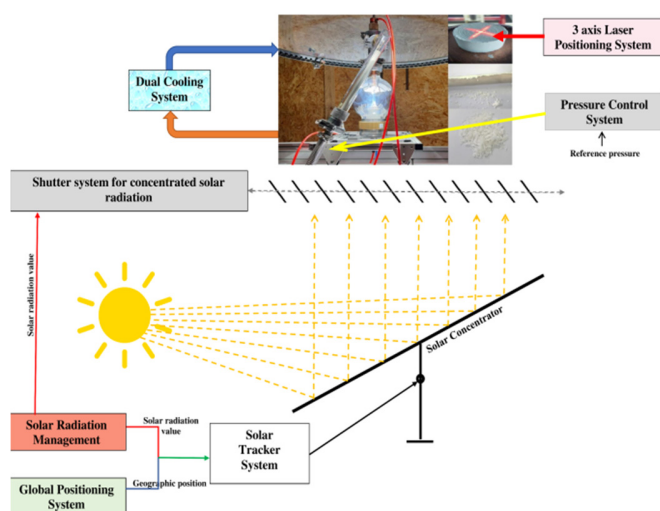


Fig. 1. Production of nanostructured powders in the reactor.

The role of the mirror is to concentrate the solar energy, directed towards the raw tablet. Before the experiment starts, the position of the tablet is verified with a laser beam.

Concentrated energy from mirror produces ultra-fast (2-3 min) evaporation. The evaporated particles are driven, due to the pressure in the reactor, to a nanoporous ceramic filter. Their rapid condensation on filter leads to the formation of nanostructured powders. The quantity of elaborated powders depends on the vapor pressure of the material. When the vapor pressure on the top of the melted material is low, the nanopowder quantity is less than 10 g/hour at flux higher than 1 kW/m^2). It is one of the reasons why when larger quantities are needed, a requirement is to associate ball milling with solar elaboration. The entire reactor is semi-automated. The position of the sun and its trajectory are obtained based on the Global Positioning System (GPS) and positioning maps. The Solar Tracker System positions the concentrator plane mirrors, on the laboratory level, to obtain the maximum available solar flux. The target in the center of the reactor core is positioned with the support of the 3 axis Laser Positioning System. The Pressure Control System (implemented with a controlled vacuum pump) maintains the desired pressure in the reactor. The Shutter system is activated by means of an electric motor to assure the solar flux on the spheric mirror. The process parameters are detailed in Table I.

TABLE I. PROCESS PARAMETERS

Sample	Solar flux (W/m^2)	Pressure (mbar)
NP/ZnO	950	300
NP/Mg/ZnO	924	300
NP/Fe/ZnO	750	300
NP/Si/ZnO	910	300
NP/Yb/ZnO	924	300
NP/Ce/ZnO	800	300

The purity of phase is determined by XRD using a Rigaku Ultima IV diffractometer, in Bragg-Brentano geometry. The acquisition parameters were: $\text{CuK}\alpha$ radiation (45 kV, $\lambda = 0.154 \text{ \AA}$, and 40 mA) and a D/teX Ultra one-dimensional detector with graphite monochromator. SEM by Hitachi SU 8230 Electron Microscope allowed identifying the morphologies. The ATR-FTIR spectra of all powders were recorded with a multireflexion ATR sensor of Bruker Fourier transmission infrared spectrometer, Tensor 27, in the range of 300 to 4000 cm^{-1} , at 4 cm^{-1} resolution. This method is particularly well-suited for analyzing nanoparticles, as it eliminates the need for sample preparation and allows for in situ studies. The powders were analyzed with Micromeritics® TriStar II Plus in a nitrogen atmosphere (absorption-desorption isotherms at 77 K) to determine the specific surface area. The moisture content of the samples was removed by degassing/drying at 300°C for approximately 3 hours before analysis. The optical properties of the particles were studied using UV-Visible spectroscopy by Ocean Optics HR2000+.

III. RESULTS AND DISCUSSION

A. Structural Characterization

The crystallinity and the phases of the as-elaborated pure and doped samples were investigated by XRD. Phase analysis of the nanoparticles was conducted in the 2θ range of $[25-101^\circ]$ at a step of 0.05° and a scan speed of $2^\circ/\text{min}$. The ICDD PDF4+ 2022 database was used for crystalline phase identification [29-34]. The patterns are shown in Figures 3-5.

The diffraction peaks indicate that the particles of pure and Mg, Fe, Si, and Yb-doped ZnO powders present only ZnO peaks corresponding to the P63mc space group [34]. No other impure phases were obtained.

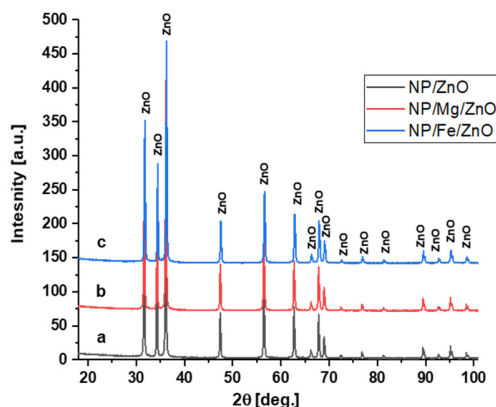


Fig. 2. XRD patterns of the: pure and Mg- and Fe-doped ZnO.

Only in the Ce-doped sample CeO_2 phase, the 225: Fm-3m space group is identified in the XRD spectrum, as shown in Figure 4. The crystallite mean size is determined with Whole-Powder-Pattern Fitting (WPPF) using Pdx12. Elaboration using solar energy leads to obtaining pure and doped nanometric particles with dimensions smaller than 60 nm.

The crystal structure parameters space group, average crystallite size D (nm), cell parameters a and c (Å), cell volume V (Å³), and unit cell volume u (Å) are also determined from XRD data (Table II).

The lattice parameters for commercial ZnO are 3.253 Å, 5.210 Å, and 47.73 Å³, respectively. The difference in lattice parameters of pure and doped ZnO nanostructured powders elaborated by solar energy is attributed to the strain induced by the nanoparticle's surface. The surface of a nanoparticle is curved, and this curvature creates a strain field that alters the bond lengths between atoms. As a result, the lattice parameters of ZnO nanoparticles are slightly smaller than those of bulk ZnO. This suggests that the crystal structure of pure and doped

ZnO nanostructured powders is relatively robust and does not change significantly under solar synthesis conditions when solar flow varies, and the pressure is constant [35]. The smaller particle size of ZnO nanostructures allows more active sites for pollutant adsorption and decomposition, making them attractive for water, soil, and air purification applications. It can be deduced that the synthesis of ZnO pure and doped nanostructured powders is successfully achieved using concentrated solar energy in the solar reactor.

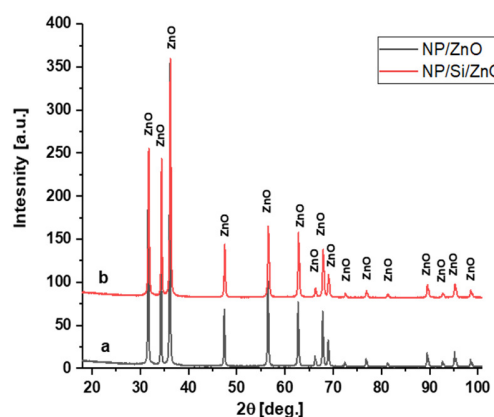


Fig. 3. XRD patterns of pure and Si-doped ZnO.

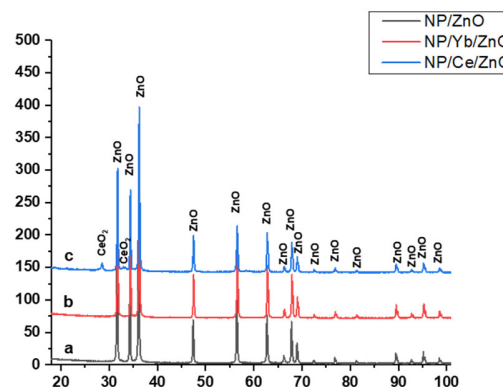


Fig. 4. XRD patterns of pure and Yb- and Ce-doped ZnO.

TABLE II. STRUCTURAL DATA FROM XRD PATTERNS

Sample	Phase	Space group	D (nm)	a (Å)	c (Å)	c/a	V (Å ³)	u (Å)
NP/ZnO	100% ZnO	186 : P63mc	51.16	3.2523	5.2106	1.6021	47.73	3.6274
NP/Mg/ZnO			44.14	3.2515	5.2093	1.6021	47.697	3.6266
NP/Fe/ZnO			47.94	3.2520	5.2105	1.6022	47.722	3.6272
NP/Si/ZnO			37.71	3.2517	5.2094	1.6021	47.703	3.6267
NP/Yb/ZnO			49.42	3.2517	5.2099	1.6022	47.708	3.6269
NP/Ce/ZnO	96.98% ZnO	186 : P63mc		3.2526	5.2109	1.6020	47.742	3.6277
	3.02% CeO ₂	225 : Fm-3m		5.425	5.425	NA	159.66	5.4250

B. Morphological Characterization

Polyhedral ZnO nanoparticles are observed in micrometric powders, as shown in Figure 5. After solar radiation, the morphology of the nanoparticles is influenced by the method of concentrating solar energy, becoming pellets and whiskers, illustrated in Figure 6. The graphical representation of the

distribution of particle sizes in ZnO micrometric particles exhibits a bimodal distribution while for ZnO nanoparticles is monomodal.

The morphology of Mg-ZnO doped powders from Figure 7 presents a nanorod form, densely packed together, with a length that varies between 0.18 and 1.1 μm. This micrograph

shows the highly crystalline and uniform nature of the Mg-doped ZnO nanoparticles elaborated with solar energy. The particle size histogram reveals that the most common length for Mg-ZnO doped nanoparticles lies between 0.38 and 0.61 μm .

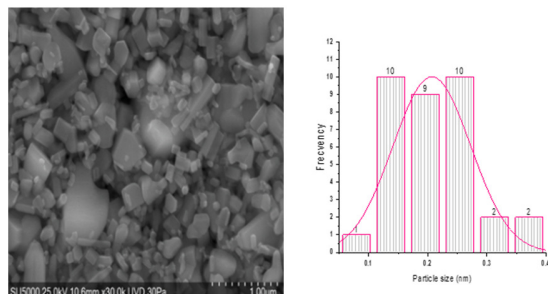


Fig. 5. SEM micrograph of ZnO nanoparticle commercial powders and particle size histogram.

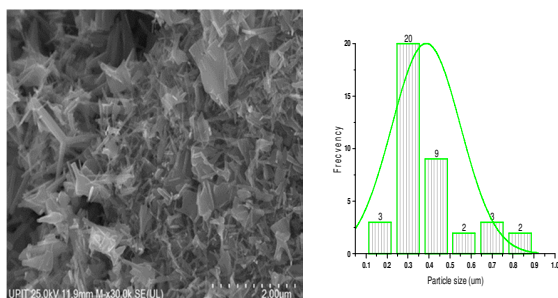


Fig. 6. SEM micrograph of ZnO nanopowders and particle size histogram.

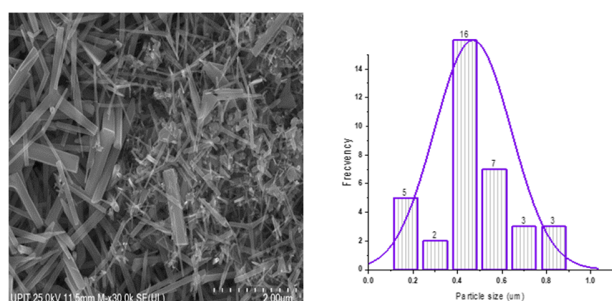


Fig. 7. SEM micrograph and particle size histogram of NP/Mg/ZnO nanopowders.

Fe-doped ZnO nanoparticles have a specific morphology for synthesis with solar energy. Figure 8 highlights the formation of tetrapods with a uniform distribution of the length of the whiskers up to 1 μm . The same way of formation of tetrapods is also observed in Figure 9, in which the morphologies of the NP/Si/ZnO particles are presented. In addition to these, there are also a few almost spherical particles not totally elongated in tetrapods. Figure 10 demonstrates the uniform formation of tetrapod Yb-doped ZnO particles with elongated whiskers, characterized by lengths between 0.06 and 1.4 μm . The length size histogram presents a monomodal distribution. The morphologies change in the case of NP/Ce/ZnO due to the presence of the 2 phases ZnO and CeO_2 .

Figure 11 shows a mixture of spherical and tetrapodic morphologies.

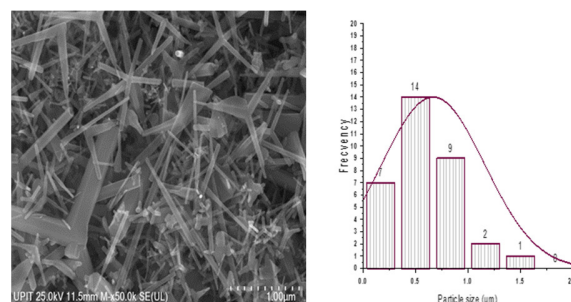


Fig. 8. SEM micrograph and particle size histogram of NP/Fe/ZnO nanopowders.

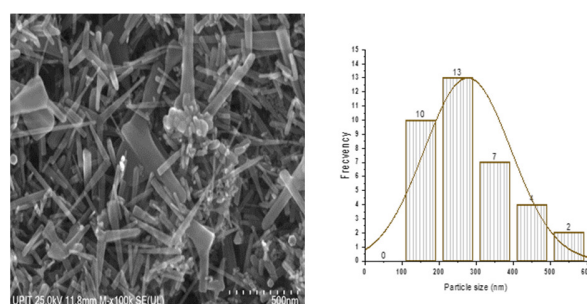


Fig. 9. SEM micrograph and particle size histogram of NP/Si/ZnO nanopowders.

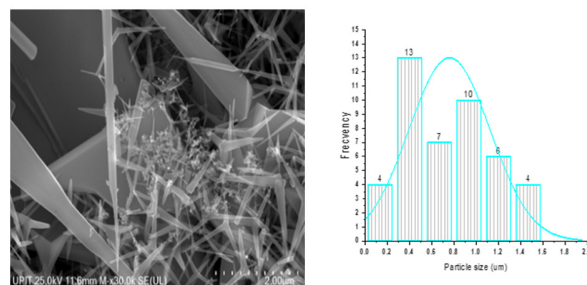


Fig. 10. SEM micrograph and particle size histogram of NP/Yb/ZnO nanopowders.

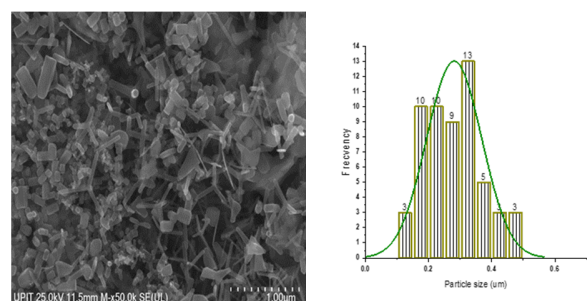


Fig. 11. SEM micrograph and particle size histogram of NP/Ce/ZnO nanopowders.

C. Surface Area

The surface area of nanoparticles, determined by BET method, and their lattice parameters are closely related. The specific surface area of the doped nanoparticles increases depending on the dopant element type, as shown in Table III. For dopants with a higher atomic radius the surface area increases accordingly. This trend can be explained by the fact that the atomic radius decreases from left to right across a period (Mg/160 pm; Fe/124 pm; Zn/133 pm; Si/114 pm; Ce/182 pm; Yb/160 pm). Also, the influence of lattice parameters on surface area is particularly evident in nanoparticles with high aspect ratios, such as nanowhiskers. These structures exhibit a large surface area due to their elongated shape and relatively small cross-sectional area.

TABLE III. SURFACE CHARACTERISTICS OF PURE AND DOPED ZnO NANOPARTICLES

Sample	Surface area (m ² /g)	Total pore volume (cm ³ /g)	Nanoparticle average size (BET) nm	Max. pore volume (cm ³ /g)
NP/ZnO	4.0618	0.003975	263.3127	0.002327
NP/Mg/ZnO	11.494	0.016832	93.0503	0.004769
NP/Fe/ZnO	12.5207	0.016345	85.4197	0.004530
NP/Si/ZnO	16.0099	0.026692	66.4331	0.005919
NP/Yb/ZnO	21.6544	0.025043	49.3903	0.005459
NP/Ce/ZnO	14.9353	0.020634	71.6102	0.004837

As the lattice parameters of pure and doped ZnO nanoparticles decrease (as shown in Table III), their surface area increases rapidly due to the large surface-to-volume ratio, because the atoms in the crystal lattice are packed more tightly together, which reduces the spacing between them and increases surface area. Smaller ZnO nanoparticles have a higher surface area and are more reactive. Additionally, the increased surface area allows for more efficient charge separation, which is essential for the photocatalytic process. Thus, ZnO nanostructures elaborated by solar energy can have enhanced photocatalytic activity compared to bulk ZnO. This increase in surface area can have a significant impact on the properties of nanoparticles, including their reactivity, catalytic activity, and optical properties [36].

D. Optical Characterization

Optical properties were determined for all developed nanoparticles. The UV-VIS spectra are shown in Figure 12. The Tauc plot method was used to energy bandgap determination. The Tauc plot is obtained by plotting the absorption coefficient vs. the square root of E (energy band gap). The bandgap was determined from the intercept of the linear fit with the absorption coefficient axis. The band gap of pure and doped ZnO nanopowders elaborated on solar reactor was determined by extrapolating the linear region of the energy gap from Figure 13. The determined energy gap of ZnO nanoparticles is typically 2.83 eV, depending on the size and morphology of the nanoparticles [39]. Doping ZnO with Mg and Si can slightly increase the band gap, resulting in a small blue shift of the absorption band. However, the effect is much smaller than for Fe doping. Doping ZnO with Fe increases the band gap at 3.14 eV, resulting in a small blue shift of the absorption band. This can be attributed to the formation of Fe-related defect states within the band gap, which emit more

photons. Doping ZnO with Yb increases the band gap at 3.06 eV while doping ZnO with Ce grows the band gap at 3.00 eV, resulting in a small blue shift of the absorption band. Absorbance depends on the type of nanoparticle and corresponds to each dopant.

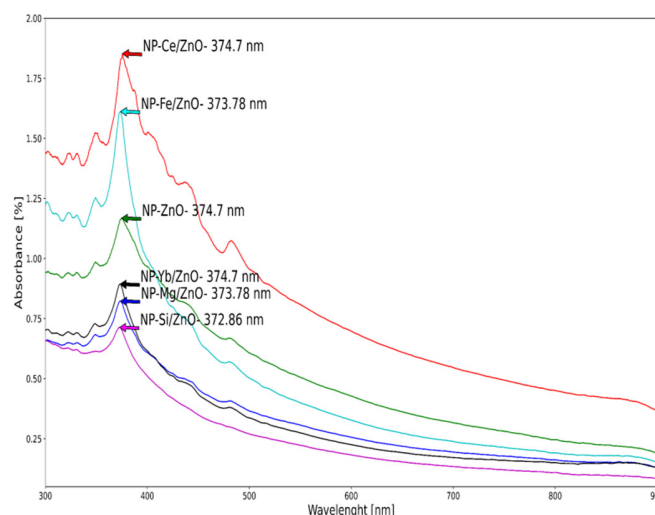


Fig. 12. UV-Vis spectra of pure and doped ZnO powders.

The values of optical energy, absorbance, and wavelengths, determined from the UV-VIS spectra, are summarized in Table IV.

TABLE IV. OPTICAL CHARACTERISTICS OF NP/ZNO

Sample	E _g (eV)	Wavelength (nm)	A max (%)	Dopant atomic radius (pm)
NP/ZnO	2.99	375.62	2.54	133
NP/Mg/ZnO	2.83	374.7	1.17	160
NP/Fe/ZnO	3.04	373.78	0.82	124
NP/Si/ZnO	3.14	373.78	1.61	114
NP/Yb/ZnO	3.02	372.86	0.71	160
NP/Ce/ZnO	3.06	373.32	0.89	182

A smaller lattice parameter of doped ZnO nanoparticles corresponds to a wider band gap, which means that solar elaborated nanoparticles absorb light at a shorter wavelength. When the ZnO particles become smaller, the quantum confinement restricts the movement of electrons between VB and CB, increasing the band gap. The surface effects and the elongated form of particles dominate the behavior. This can be beneficial for applications such as photocatalysis, where ZnO nanoparticles are used to break down organic molecules using light energy and for ultraviolet light detection [37-43]. The surface area and lattice parameters of solar synthesized ZnO nanoparticles are interconnected properties that have a profound impact on their overall properties. As research in this field continues to advance, we can expect to see even more innovative materials emerge, further propelling the growth of the solar energy industry.

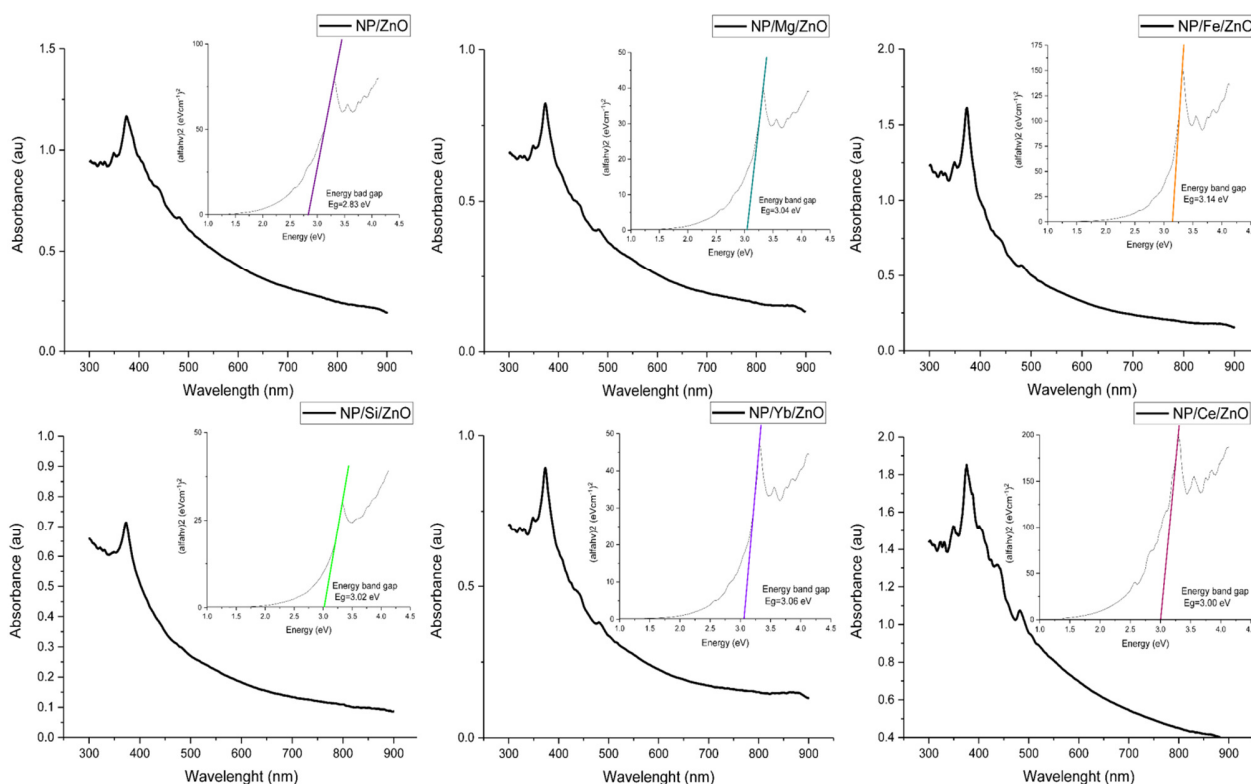


Fig. 13. Energy band gap spectra of pure and doped ZnO powders.

IV. CONCLUSIONS

Among all methods of ZnO nanoparticle elaboration, solar-evaporation offers advantages such as energy efficiency, environmental friendliness, and the ability to synthesize nano-oxides with tailored properties, making them promising approaches for sustainable nano-oxide production. The novelty of solar synthesis lies in its ability to address environmental concerns, reduce costs, enhance control over material properties, promote sustainable manufacturing practices, and enable the discovery of novel materials with transformative applications.

The tetrapod morphology (provided by SEM) of pure and doped ZnO nanometric particle (from XRD) synthesis by solar energy can significantly influence their properties and applications. For instance, these nanoparticles are suitable for gas sensors and field-effect transistors due to their high surface area (demonstrated by BET). Having a larger proportion of atoms or molecules on their surfaces leads to enhanced interactions with their surroundings. By analyzing the ATR-FTIR spectrum of Mg-, Fe-, Si-, Yb-, and Ce-doped ZnO nanoparticles, researchers can gain insights into the structural doping effects. Tetrapod type morphology observed to have been formed by solar synthesis exhibits unique optical properties demonstrated by UV-VIS analysis. Thus, we can state that these nano-oxides can be used in various sensing and photonic applications.

As solar energy becomes more accessible and efficient, solar-driven synthesis of pure and doped ZnO is poised to be a

crucial factor in shaping the future of materials science and technology. Therefore, understanding and controlling the structure and morphology of ZnO nanoparticles is crucial for optimizing their performance in various applications.

ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under grant agreement No 823802. We thank the CNRS-PROMES laboratory, UPR 8521, belonging to the French National Centre for Scientific Research (CNRS), for providing access to its installations, the support of its scientific and technical staff, and the financial support of the SFERA-III project (Grant Agreement No 823802). The characterization was supported by The National University of Science and Technology POLITEHNICA Bucharest, Pitești University Centre under grant agreement CIPCS-2022-6 (Enhanced the efficiency of solar cells by coating with nanostructured metal oxide layers).

REFERENCES

- [1] D. Tan, Y. Wu, Z. Zhang, Y. Jiao, L. Zeng, and Y. Meng, "Assessing the Life Cycle Sustainability of Solar Energy Production Systems: A Toolkit Review in the Context of Ensuring Environmental Performance Improvements," *Sustainability*, vol. 15, no. 15, Jan. 2023, Art. no. 11724, <https://doi.org/10.3390/su151511724>.
- [2] S. Kuskaya, F. Bilgili, E. Mugaloglu, K. Khan, M. E. Hoque, and N. Toguc, "The role of solar energy usage in environmental sustainability: Fresh evidence through time-frequency analyses," *Renewable Energy*, vol. 206, pp. 858–871, Apr. 2023, <https://doi.org/10.1016/j.renene.2023.02.063>.
- [3] N. Novas, R. M. Garcia, J. M. Camacho, and A. Alcayde, "Advances in Solar Energy towards Efficient and Sustainable Energy," *Sustainability*,

- vol. 13, no. 11, Jan. 2021, Art. no. 6295, <https://doi.org/10.3390/su13116295>.
- [4] G. I. Mamniashvili, G. Donadze, V. Tavkhelidze, and D. Gulamova, "Superconducting Precursors in Bi/Pb Multiphase Cuprates Fabricated by the Solar Technology and their Comparative Study by Torque Magnetometry Methods," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12390–12395, Dec. 2023, <https://doi.org/10.48084/etasr.6487>.
 - [5] N. S. M. N. Izam, Z. Itam, W. L. Sing, and A. Syamsir, "Sustainable Development Perspectives of Solar Energy Technologies with Focus on Solar Photovoltaic—A Review," *Energies*, vol. 15, no. 8, Jan. 2022, Art. no. 2790, <https://doi.org/10.3390/en15082790>.
 - [6] M. Oelgemoller, "Solar Photochemical Synthesis: From the Beginnings of Organic Photochemistry to the Solar Manufacturing of Commodity Chemicals," *Chemical Reviews*, vol. 116, no. 17, pp. 9664–9682, Sep. 2016, <https://doi.org/10.1021/acs.chemrev.5b00720>.
 - [7] L. G. Ceballos-Mendivil, Y. Carvajal-Campos, J. Tanori-Cordova, J. C. Luque-Ceballos, H. Villafan-Vidales, and C. A. Estrada, "Solar synthesis of nanostructured zirconia: microstructural and thermal characterization," *Materials Research Express*, vol. 7, no. 11, Aug. 2020, Art. no. 115014, <https://doi.org/10.1088/2053-1591/abcbb8>.
 - [8] N. Shohoji, L. Guerra Rosa, J. Cruz Fernandes, D. Martinez, and J. Rodriguez, "Catalytic acceleration of graphitisation of amorphous carbon during synthesis of tungsten carbide from tungsten and excess amorphous carbon in a solar furnace," *Materials Chemistry and Physics*, vol. 58, no. 2, pp. 172–176, Mar. 1999, [https://doi.org/10.1016/S0254-0584\(98\)00275-2](https://doi.org/10.1016/S0254-0584(98)00275-2).
 - [9] K. K. Brar *et al.*, "Green route for recycling of low-cost waste resources for the biosynthesis of nanoparticles (NPs) and nanomaterials (NMs)—A review," *Environmental Research*, vol. 207, May 2022, Art. no. 112202, <https://doi.org/10.1016/j.envres.2021.112202>.
 - [10] N. Shohoji *et al.*, "Synthesising Carbo-Nitrides of some D-Group Transition Metals Using a Solar Furnace at PSA," *Materials Science Forum*, vol. 730–732, pp. 153–158, 2013, <https://doi.org/10.4028/www.scientific.net/MSF.730-732.153>.
 - [11] G. Flamant, A. Ferriere, D. Laplace, and C. Monty, "Solar processing of materials: opportunities and new frontiers," *Solar Energy*, vol. 66, no. 2, pp. 117–132, Jun. 1999, [https://doi.org/10.1016/S0038-092X\(98\)00112-1](https://doi.org/10.1016/S0038-092X(98)00112-1).
 - [12] R. Roman, I. Canadas, J. Rodriguez, M. T. Hernandez, and M. Gonzalez, "Solar sintering of alumina ceramics: Microstructural development," *Solar Energy*, vol. 82, no. 10, pp. 893–902, Oct. 2008, <https://doi.org/10.1016/j.solener.2008.04.002>.
 - [13] Z. H. Wang, C. J. Choi, B. K. Kim, J. C. Kim, and Z. D. Zhang, "Characterization and magnetic properties of carbon-coated cobalt nanocapsules synthesized by the chemical vapor-condensation process," *Carbon*, vol. 41, no. 9, pp. 1751–1758, Jan. 2003, [https://doi.org/10.1016/S0008-6223\(03\)00127-1](https://doi.org/10.1016/S0008-6223(03)00127-1).
 - [14] J. Ahire and B. M. Bhanage, "Solar energy-controlled shape selective synthesis of zinc oxide nanomaterials and its catalytic application in synthesis of glycerol carbonate," *Journal of Solid State Chemistry*, vol. 295, Mar. 2021, Art. no. 121927, <https://doi.org/10.1016/j.jssc.2020.121927>.
 - [15] D. D. Dionysiou, S. C. Pillai, and S. Rtimi, "Editorial overview: Nanomaterials for energy and environmental applications: advances and recent trends," *Current Opinion in Chemical Engineering*, vol. 36, Jun. 2022, Art. no. 100805, <https://doi.org/10.1016/j.coche.2022.100805>.
 - [16] Y. Zhang, K. Poon, G. S. P. Masonsong, Y. Ramaswamy, and G. Singh, "Sustainable Nanomaterials for Biomedical Applications," *Pharmaceutics*, vol. 15, no. 3, Mar. 2023, Art. no. 922, <https://doi.org/10.3390/pharmaceutics15030922>.
 - [17] M. Parashar, V. K. Shukla, and R. Singh, "Metal oxides nanoparticles via sol-gel method: a review on synthesis, characterization and applications," *Journal of Materials Science: Materials in Electronics*, vol. 31, no. 5, pp. 3729–3749, Mar. 2020, <https://doi.org/10.1007/s10854-020-02994-8>.
 - [18] A. G. Plaiasu, C. M. Topala, A. Dinu, C. Sutan, and M. Abrudeanu, "Copper oxides nanopowders: Synthesis by S.P.V.D. And characterization," *Revista de Chimie*, vol. 66, no. 10, pp. 1636–1638, Oct. 2015.
 - [19] W. Jiang *et al.*, "Preparation and properties of superparamagnetic nanoparticles with narrow size distribution and biocompatible," *Journal of Magnetism and Magnetic Materials*, vol. 283, no. 2, pp. 210–214, Dec. 2004, <https://doi.org/10.1016/j.jmmm.2004.05.022>.
 - [20] Z. Sadowski and A. Pawlowska, "Synthesis of Metal Oxide Nanoparticles and Its Biomedical Applications," in *Nanotechnology Applied To Pharmaceutical Technology*, M. Rai and C. Alves dos Santos, Eds. New York, NY, USA: Springer, 2017, pp. 91–111.
 - [21] L. M. Cursaru, A. G. Plaiasu, C. M. Ducu, R. M. Piticescu, and I. A. Tudor, "Carbon Nanotube/Polyaniline Composite Films Prepared by Hydrothermal-Electrochemical Method for Biosensor Applications," in *International Semiconductor Conference*, Sinaia, Romania, Oct. 2018, pp. 249–252, <https://doi.org/10.1109/SMICND.2018.8539793>.
 - [22] J. Jeevanandam *et al.*, "Green approaches for the synthesis of metal and metal oxide nanoparticles using microbial and plant extracts," *Nanoscale*, vol. 14, no. 7, pp. 2534–2571, Feb. 2022, <https://doi.org/10.1039/D1NR08144F>.
 - [23] M.-A. Iota, L.-M. Cursaru, A.-G. Schiopu, I. A. Tudor, A.-M. Motoc, and R. M. Piticescu, "Fe₃O₄ Core-Shell Nanostructures with Anticancer and Antibacterial Properties: A Mini-Review," *Processes*, vol. 11, no. 7, Jul. 2023, Art. no. 1882, <https://doi.org/10.3390/pr11071882>.
 - [24] L. Yan, A. Uddin, and H. Wang, "ZnO Tetrapods: Synthesis and Applications in Solar Cells," *Nanomaterials and Nanotechnology*, vol. 5, Jan. 2015, Art. no. 19, <https://doi.org/10.5772/60939>.
 - [25] N. A. Hussien, J. S. Al Malki, F. A. R. Al Harthy, A. W. Mazi, and J. A. A. Al Shadadi, "Sustainable Eco-Friendly Synthesis of Zinc Oxide Nanoparticles Using Banana Peel and Date Seed Extracts, Characterization, and Cytotoxicity Evaluation," *Sustainability*, vol. 15, no. 13, Jan. 2023, Art. no. 9864, <https://doi.org/10.3390/su15139864>.
 - [26] A. C. Constandache *et al.*, "Morphological and structural investigations of ZnO resulted from green synthesis," *Scientific Bulletin-University Politehnica of Bucharest*, vol. 85, no. 2, pp. 275–283, 2023.
 - [27] A. Mir, N. Becheikh, L. Khezami, M. Bououdina, and A. Ouderni, "Synthesis, Characterization, and Study of the Photocatalytic Activity upon Polymeric-Surface Modification of ZnO Nanoparticles," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12047–12053, Dec. 2023, <https://doi.org/10.48084/etasr.6373>.
 - [28] E. Guillot, R. Rodriguez, N. Boullet, and J.-L. Sans, "Some details about the third rejuvenation of the 1000 kWth solar furnace in Odeillo: Extreme performance heliostats," *AIP Conference Proceedings*, vol. 2033, no. 1, Nov. 2018, Art. no. 040016, <https://doi.org/10.1063/1.5067052>.
 - [29] M. Quiros, S. Grazulis, S. Girdzijauskaitė, A. Merkys, and A. Vaitkus, "Using SMILES strings for the description of chemical connectivity in the Crystallography Open Database," *Journal of Cheminformatics*, vol. 10, no. 1, May 2018, Art. no. 23, <https://doi.org/10.1186/s13321-018-0279-6>.
 - [30] A. Merkys, A. Vaitkus, J. Butkus, M. Okulic-Kazarinas, V. Kairys, and S. Grazulis, "COD::CIF::Parser: an error-correcting CIF parser for the Perl language," *Journal of Applied Crystallography*, vol. 49, no. 1, pp. 292–301, Feb. 2016, <https://doi.org/10.1107/S1600576715022396>.
 - [31] S. Grazulis, A. Merkys, A. Vaitkus, and M. Okulic-Kazarinas, "Computing stoichiometric molecular composition from crystal structures," *Journal of Applied Crystallography*, vol. 48, no. 1, pp. 85–91, Feb. 2015, <https://doi.org/10.1107/S1600576714025904>.
 - [32] S. Grazulis *et al.*, "Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration," *Nucleic Acids Research*, vol. 40, no. D1, pp. D420–D427, Jan. 2012, <https://doi.org/10.1093/nar/gkr900>.
 - [33] S. Grazulis *et al.*, "Crystallography Open Database – an open-access collection of crystal structures," *Journal of Applied Crystallography*, vol. 42, no. 4, pp. 726–729, Aug. 2009, <https://doi.org/10.1107/S0021889809016690>.
 - [34] R. T. Downs and M. Hall-Wallace, "The American Mineralogist crystal structure database," *American Mineralogist*, vol. 88, no. 1, pp. 247–250, Jan. 2003.

- [35] N. L. Nguyen and N. N. Ha, "Understanding ZnO surface defects from first-principles simulation," *Surface Science*, vol. 732, Jun. 2023, Art. no. 122272, <https://doi.org/10.1016/j.susc.2023.122272>.
- [36] M. M. Nadareishvili, G. Mamniashvili, D. Jishiashvili, G. Abramishvili, C. Ramana, and J. Ramsden, "Investigation of the Visible Light-Sensitive ZnO Photocatalytic Thin Films," *Engineering, Technology & Applied Science Research*, vol. 10, no. 2, pp. 5524–5527, Apr. 2020, <https://doi.org/10.48084/etasr.3392>.
- [37] H. M. Pouran, F. L. Martin, and H. Zhang, "Measurement of ZnO Nanoparticles Using Diffusive Gradients in Thin Films: Binding and Diffusional Characteristics," *Analytical Chemistry*, vol. 86, no. 12, pp. 5906–5913, Jun. 2014, <https://doi.org/10.1021/ac500730s>.
- [38] T. M. Awwad, S. M. Shaaban, E. M. Ragab, and A. Mir, "Photo-Catalytic Activity Improvement for Organic Pollutant Removal in Wastewater using Zinc Oxide Quantum Dots: An Experimental and Modeling Study," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12138–12144, Dec. 2023, <https://doi.org/10.48084/etasr.6451>.
- [39] C. Wang, Z. Chen, Y. He, L. Li, and D. Zhang, "Structure, morphology and properties of Fe-doped ZnO films prepared by facing-target magnetron sputtering system," *Applied Surface Science*, vol. 255, no. 15, pp. 6881–6887, May 2009, <https://doi.org/10.1016/j.apsusc.2009.03.008>.
- [40] K. Kaviyarasu, X. Fuku, G. T. Mola, E. Manikandan, J. Kennedy, and M. Maaza, "Photoluminescence of well-aligned ZnO doped CeO₂ nanoplatelets by a solvothermal route," *Materials Letters*, vol. 183, pp. 351–354, Nov. 2016, <https://doi.org/10.1016/j.matlet.2016.07.143>.
- [41] S. K. Noukelag *et al.*, "Investigation of structural and optical properties of biosynthesized Zincite (ZnO) nanoparticles (NPs) via an aqueous extract of *Rosmarinus officinalis* (rosemary) leaves," *MRS Advances*, vol. 5, no. 45, pp. 2349–2358, Jan. 2020, <https://doi.org/10.1557/adv.2020.220>.
- [42] S. K. Dhoke, "Synthesis of nano-ZnO by chemical method and its characterization," *Results in Chemistry*, vol. 5, Jan. 2023, Art. no. 100771, <https://doi.org/10.1016/j.rechem.2023.100771>.
- [43] A. Anzlovar, Z. Crnjak Orel, and M. Zigon, "Nanocomposites with nano-to-sub-micrometer size zinc oxide as an effective UV absorber," *Polimeri : casopis za plastiku i gumu*, vol. 29, no. 2, pp. 84–87, Nov. 2008.

Quicklime-stabilized Tuff and Clayey Soils for Highway A3 Construction in Northern Tunisia

Nejib Ghazouani

Department of Civil Engineering, College of Engineering, Northern Border University, Saudi Arabia | Civil Engineering Laboratory (LGC), National Engineers School of Tunis (ENIT), University of Tunis El Manar, Tunisia

nejib.ghazouani@nbu.edu.sa (corresponding author)

Received: 23 January 2024 | Revised: 8 February 2024 and 17 February 2024 | Accepted: 26 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6955>

ABSTRACT

This study presents a comprehensive examination of the effects of quicklime (QL) addition on the stabilization of two distinct clayey soils with high (CH) and low plasticity (CL-tuff). The results showed that incorporating QL into the soils substantially improves their stabilization characteristics. Specifically, the addition of QL results in a notable decrease in the final water content of both soils, as shown by a reduction from 23.04 to 19.06% in CH and from 18.07 to 17.1% in CL-tuff at 4% QL addition. Furthermore, this study reveals a transformation in the plasticity properties of soils. Liquid Limit (LL) and Plasticity Index (PI) were reduced, with CH-tuff exhibiting a significant decrease in PI from 48 to 12 and an increase in Plastic Limit (PL) from 21.8 to 55 at 4% QL. CL-tuff also showed reduced plasticity, with PI decreasing to 8.33 at 4% QL. Additionally, the Immediate Bearing Index (IBI) was improved for both soil samples, indicating improved load-bearing capacities. For CH samples, IBI improved from 6.37 to 11.99 at 4% QL addition, while for CL-tuff, it increased dramatically from 4.5 to 23.6 for the same QL percentage. The findings underscore the effectiveness of QL in improving soil properties crucial for chemical stabilization, providing evidence that QL addition can be a key technique in soil stabilization, especially for soils with high plasticity or those requiring increased bearing strength.

Keywords-clayey soils; quicklime stabilization; soil plasticity; water content reduction; bearing capacity

I. INTRODUCTION

Clayey soils exhibit high sensitivity to variations in moisture content but are also subject to adverse impacts of environmental stressors, including climatic conditions, long-term static loading, and dynamic impacts. Therefore, stabilization of such soils, when used as road subgrade, requires remedial treatment to increase their strength properties and decrease permeability [1-10]. The use of quicklime (QL) for soil stabilization is a widely recognized technique [11]. Incorporating QL into fine-grained soils engenders a profound alteration in the soil's physicochemical properties. QL stabilization catalyzes the flocculation of cohesive soil particles, modifying the soil microstructure by inducing particle aggregation. The mechanism underlying this transformation is described in [12]. The increase of pore water-electrolyte concentration attenuates the diffused double layer upon lime addition, thus strengthening Van der Waals forces and initiating the formation of flocculent structures. This reorganization of the soil matrix fosters enhanced interparticle bonds, improving resistance to compaction efforts and leading to a higher Optimum Moisture Content (OMC).

The interaction between lime treatment and the permeability of expansive soils has been well-documented in recent studies. In [13], it was shown that lime modification initially increases the permeability of expansive soils, which

stabilizes or slightly decreases with time. This behavior is due to the early effects of lime on soil structure, while the subsequent decrease is attributed to long-term pozzolanic reactions that create cementitious bonds. In [14], it was shown that treating clayey soils with lime and fly ash significantly reduces permeability, by up to 95% after a 14-day period, which is related to the reduction in the soil plasticity. In [15], a threshold for lime treatment was observed, as soil permeability decreased when lime content exceeded 4%.

In [16], Mercury Intrusion Porosimetry (MIP) was used to reveal a dual-pore structure similar to that of compacted clays, supporting previous studies [17-18]. The impact of cyclic wetting and drying was analyzed, revealing minimal initial changes in microporosity, trending towards smaller pore sizes. However, a pronounced increase in macroporosity was observed after multiple cycles of controlled suction, particularly with a QL content of 2%, implying a disruption of interparticle bonds, while microporosity remained relatively constant. This study illuminated several phenomena consequent to lime treatment: a reduction in volumetric change that illustrates the swell potential, an increase in stress sensitivity and bond formation through pozzolanic reactions, and the degradation of soil properties under severe wetting and drying cycles compared to more moderate amplitude cycles. Severe cycles were observed to induce irreversible swelling, whereas controlled cycles led to irreversible shrinkage. Despite

exposure to varying wetting and drying cycles, the long-term mechanical property improvement of QL was not substantiated.

This paper investigates the stabilization effects of QL on soils with tuff content. The investigation delves into the macroscopic behavior of tuff-enriched clayey soils stabilized with QL, revealing substantial improvements in terms of moisture reduction, plasticity modification, and strength improvement. These advances are crucial for the reliable application of such soils in highway construction. By providing empirical evidence of the efficacy of QL stabilization in this context, this study opens new avenues for modeling and computational research in the field, empowering sustainable construction practices.

II. EXPERIMENTAL INVESTIGATION

This study investigated Proctor compaction, California Bearing Ratio (CBR), and Atterberg limits. The influence of initial parameters, such as water content and lime content, on macroscopic soil characteristics was examined.

A. Materials

The samples used in this study were extracted from the construction site of Highway A3 in Beja governorate, Northern Tunisia. The first sample was extracted from a subsurface stratum located between 3 to 5 m deep (a). The second was obtained from a shallower depth ranging from 1 to 2.5 m (b).

B. Physical Properties of Natural Soils

To obtain the Particle Size Distribution (PSD) of the two soil samples, sieve analysis (according to NF P94-056) and Hydrometer (according to NF P94-057) tests were carried out.

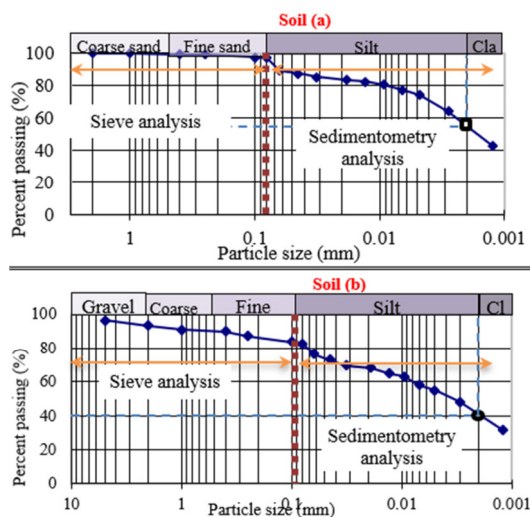


Fig. 1. PSD curves of the two soil samples.

Based on PSD results and the Unified Soil Classification System (USCS), soil 1 is identified as clay with High Plasticity (CH) with 55% of particles finer than 2 microns, and soil 2 as clay with Low Plasticity (CL) with tuff, with 40% of particles finer than 2 microns. The extension - tuff is purposefully added to remind the tuff fraction. From now on, the CH and CL-tuff notations will be used to designate the respective soils.

Methylene blue value (VBS) tests (NF P 94-068) were used to assess the swelling potential of the expansive clay. In [19], VBS tests carried out in two different clay fractions yielded a value close to 9, suggesting that both clay types exhibit expansive properties that require chemical stabilization. Similarly, in this study, chemical stabilization is required due to the high clay content and high plasticity of the two types of soils.

TABLE I. SUMMARY OF SOIL SAMPLES' PROPERTIES

Properties	CH	CL-Tuff
PSD properties		
Test sample weight (g)	398.71g	485.11g
Maximum grain size	2 mm	5 mm
Fraction passing #200	94%	80%
Atterberg limits		
Liquid limit (LL)	69.80	39.50
Plastic limit (PL)	21.80	18.50
Plasticity index (PI)	48.00	21.00
Methylene Blue Value	8.63	5.50

C. QL as Chemical Stabilizing Material

Lime, or calcium oxide (CaO), is a white, alkaline, crystalline substance produced by heating high-purity limestone (97-99% CaCO_3) in vertical kilns at 1100 to 1300°C. This process, which generates QL is exothermic, emitting significant heat and light while releasing carbon dioxide and water vapor. QL has a PSD ranging from 0.5 to 25 μm and is used in various industrial applications. Using locally sourced QL for soil stabilization promotes sustainability by reducing environmental impact and taking advantage of in situ soil improvement over material substitution. By improving the engineering properties of weak and problematic soils in situ, this study supports the sustainable practice of minimizing material transport and avoiding the ecological impacts associated with the extraction and import of alternative materials.

D. Experimental Method

The experimental program aimed to investigate the short-term effects of adding QL to natural silt and tuff soils. The study was conducted according to the following method:

- The humidification of 0/20 mm fraction of the samples at 3 or 4 water contents. The prepared samples were mixed with increasing QL content, varying from 1% to 4%.
- The determination of the Plasticity Index (PI) for each sample.
- Perform normal Proctor test to investigate the compaction properties and CBR test to determine the Immediate Bearing Index (IBI).

III. RESULTS AND DISCUSSION

A. Soil Initial Compaction Properties

Figure 2 shows the results of the normal Proctor test (NF P 94-093) for both soil samples before QL treatment. In road construction, compaction efforts typically aim to reach 95% of the maximum dry density ($g_{D,max}$), allowing for an optimal

moisture content (ω_{OMC}) within a specific range, indicated as $\Delta\omega = [\omega_1, \omega_2]$. $\Delta\omega$ serves as a gauge measure of the soil's reaction sensitivity to increases in water content. For instance, $\Delta\omega$ values for CL-tuff and CH were 6% and 13%, respectively, indicating that clayey CH is more sensitive to changes in water content than CL-tuff. This is evident from the compaction tests, where the tuff reaches higher dry densities with increasing water content. Despite clay's tendency to be water-sensitive, these results suggest counterintuitive behavior. When considering the wide variance in water content requirements as shown by the two Proctor compaction curves, it is crucial to precisely determine the optimal lime percentages to be added based on these findings.

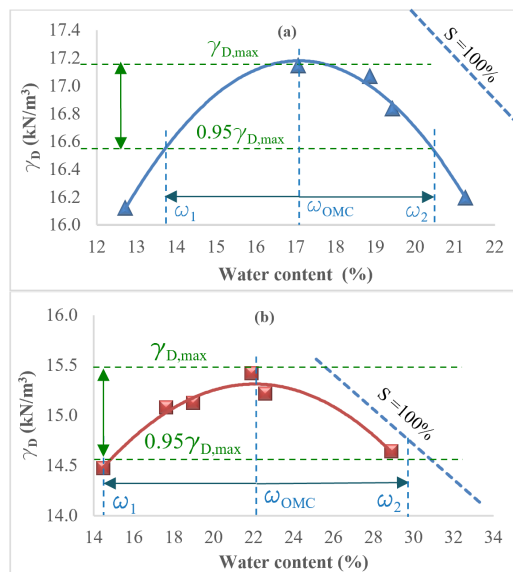


Fig. 2. Standard Proctor compaction curves for (a) CL-Tuff and (b) CH.

B. Soil Initial Swell Potential

The swelling potential is measured by recording the soaked specimen's volume change over a soaking period, with the peak increase observed after four days of immersion. For detailed design procedures for field soil-lime QL stabilization on roadbeds and railway embankments, the interested reader may refer to the relevant guidelines in [19-20]. Figure 3 shows that the swelling of the silty CH is 6%. However, as expected, the volume increase of CL-tuff is too low (0.66%, almost 1/10 of that of CH). This soil is not expansive, because the soil contains a tuff fraction and contains a lower fraction of clay than the CH, and consequently its plasticity is also lower.

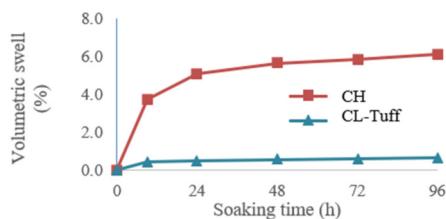


Fig. 3. Variation of volumetric swell as a function of the soaking time.

C. Soil Initial Bearing Properties

The CBR test was conducted according to NF P 94-078 and the values were obtained for the two soils on the wet side. This choice is supported by previous studies that reported that the expected compaction state in the field is performed nearly at the maximum dry density and then an increase in water content can occur by water infiltration, simulating in this manner the wetting path. However, the hydric cycle as wetting-drying was not investigated here, and some expected trends of CBR may be correlated with the uniaxial compressive strength given in [8, 16, 21]. Figure 4 provides IBI values that gradually decrease as the moisture content of soil samples increases. The results show a high sensitivity of the CBR to water content. For CH, a 6% increase in water content, from 18 to 24%, leads to a 81.7% decrease in CBR. It can also be noted that for the water content closest to the normal Proctor optimum, the CBR value is close to 5. For CL-tuff, an increase in the water content of 12% causes a decrease in CBR of 76.56%. Therefore, the CBR values of the CL-tuff soil decrease more rapidly than those of CH. This fact is explained by the potential dissolution of the tuff in the water which results from the interparticle bonds being broken.

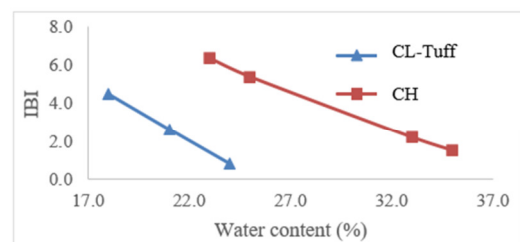


Fig. 4. Variation of CBR as a function of water content for both soils.

D. Hydric and Plasticity Properties of Treated Soils

Soil samples exhibiting a higher water content, identified as being on the wet side, were prepared and then treated with varying amounts of QL, 1% to 4% of the soil sample's weight. Figure 5 shows that the final water content in the soil decreases with increasing lime content. This trend is observed in both soil types and is due to the addition of dry QL particles and the exothermic nature of the QL-soil hydration reaction, which together promote moisture reduction in the soil. CL-tuff samples exhibit a significant decrease in water content. This can be attributed to the reduction of the soil's affinity for water, which is a consequence of the pozzolanic reaction. The pozzolanic reaction facilitates the binding of flocculated soil particles, thereby diminishing the soil's ability to retain water [22-25]. The pozzolanic reaction is also associated with the absorption of water by solid tuff particles.

For clay soils, the reduction in water content with the addition of QL is less pronounced compared to CL-tuff, which can be primarily attributed to the pozzolanic reactions that promote increased flocculation in soils with a high clay content. This phenomenon is supported by the findings in [26], where an experimental investigation in low-plasticity fine soils revealed that a 4% QL content is optimal for stabilizing pH at a constant value, indicating a saturation point in calcium

concentration. It is anticipated that the increase in pH would be more significant in CL-tuff because of a more energetic alkaline reaction (e.g. cation exchange and pozzolanic reactions). Regarding the plasticity properties, the addition of QL increases the plasticity limits of both soil types without significantly altering their liquid limits, as shown in Figure 6.

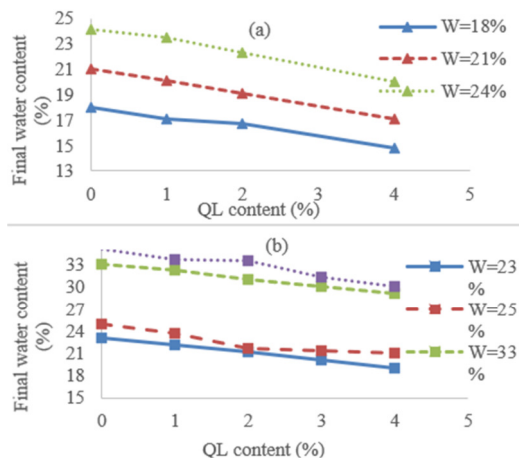


Fig. 5. (a) Tuff and (b) clayey silt hydric conditions after treatment.

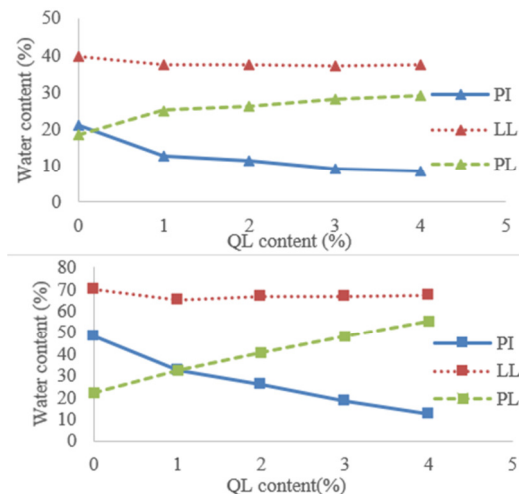


Fig. 6. Effect of QL addition on (a) tuff and (b) silt plasticity.

The increase in plasticity limit stems from QL-induced flocculation and cementation, which promote bond formation via pozzolanic reactions. With higher QL dosages, these reactions intensify over time, broadening the soil's solid range and consequently lowering the plasticity index. With a 2% QL dosage applied to CL-tuff, there is a discernible alteration in PI, with variations reaching up to 45%. In the case of CH, PI experiences a significant change of 62% at 3% QL dose. Such changes in the plasticity state significantly modify the soil's behavior, transitioning it from low to higher consistency. Figure 6 shows that both soil samples undergo a considerable decrease in their respective PI after QL treatment. This reduction reflects the transformative impact of QL on soil consistency, irrespective of the initial textural differences.

E. Effect of Lime Addition on Methylene Blue Value (VBS)

Figure 7 shows the results of a chemical analysis carried out on CH after QL addition to determine VBS. Soil VBS significantly decreased by 60% with 2% QL content and further decreased by 75% with 4% QL. This reduction in VBS underscores QL's effectiveness in mitigating the reactive clay fraction within the soil. At 2% QL content, the swelling potential of the soil was markedly reduced, leading to a decrease in volumetric changes. The optimal QL content for soil stabilization must consider the full range of parameters investigated in this study.

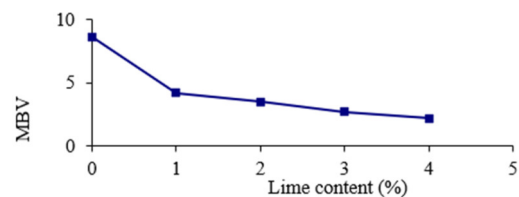


Fig. 7. Variation of VBS for lime content on the clayey silt.

F. Effect of QL on Immediate Bearing Index (IBI)

IBI testing, according to the NF P 94-078 standard, was carried out on both soils with moisture contents exceeding the optimum Proctor values of their untreated counterparts. These moisture contents were then used as a baseline for the stabilization study. As shown in Figure 8, the application of lime treatment significantly improved the IBI of the soils, demonstrating its positive effect on soil-bearing capacity. For both soils and at constant moisture content, IBI increases considerably with the addition of QL.

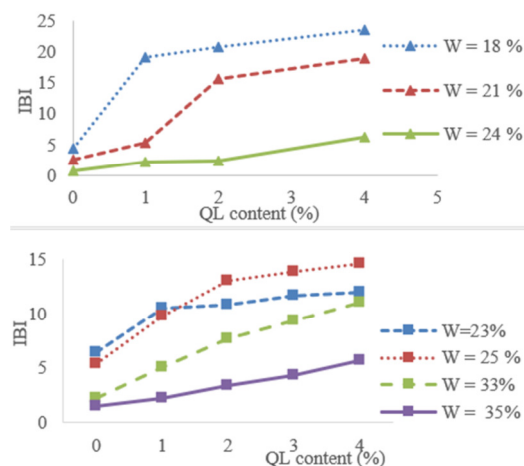


Fig. 8. Variations in IBI with lime addition for (a) tuff and (b) clay.

Without treatment, CH had a better IBI than CL-tuff soil. However, after treatment, CL-tuff achieved a higher IBI. Therefore, it seems that treating this soil can lead to better results. The presence of undesirable elements in CH impedes the kinetics of setting and hardening of the QL-soil mixture and, therefore, the performance of the treated soils. For CL-tuff and all QL content, IBI values decrease with increasing moisture content. This decrease is significant, reaching 89% for

a moisture content variation between 18 and 24% (moisture content greater than OMC). For CH, the IBI does not always decrease with increasing moisture content. Starting with a QL content of 2%, there is an increase in IBI between a moisture content of 23% and 25%. Beyond 25%, IBI decreases again. This can be explained by the fact that for a moisture content of 23% and a QL content of more than 2%, the final moisture content of the mixture is lower than the OMC. Therefore, for the same compaction energy, the mixture is less compact and has a lower load-bearing capacity than a mixture with an initial moisture content of 25% that reaches the OMC after treatment.

G. Effect of QL on Clayey Silt CBR

Figure 9 shows that CBR was improved by adding QL. According to [19-20], the CBR for soaked samples up to 7 days should be greater than 30 for embankments 2m deep or shorter. The CBR values increase with the width of the embankment. This means that when soaking for 4 days and CBR is around 25, a greater value can be reached for this type of soil with QL immersed for more than 4 days. A minimum value for low-lime content should be more than 2%. Therefore, a very small amount of QL, such as 2% or more, improves the strength of a low plasticity CL-tuff soil.

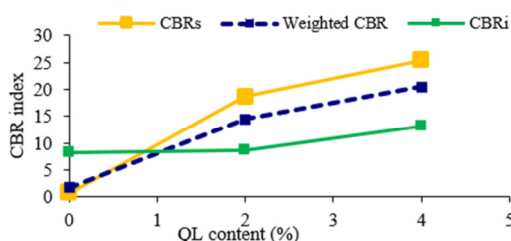


Fig. 9. CBR variation with QL content.

IV. CONCLUSION

The current study demonstrated that the addition of QL benefits soil stabilization, highlighting how it effectively reduces moisture content, modifies plasticity, and improves strength. The reduction in moisture content achieved through QL addition suggests a lower susceptibility of stabilized soil to weather-induced degradation, which is crucial to maintaining the integrity of highway foundations. The modification in soil plasticity, characterized by reduced liquid limit and plasticity index, results in a decreased potential for expansion and contraction, thus offering a more stable base for pavement layers. Finally, improved strength, as indicated by the increase in the immediate bearing index, implies a greater ability to support the heavy loads experienced by highways. These improvements directly correlate with the requirements for materials used in highway applications, thereby affirming the suitability of the treated soils for such purposes. Furthermore, they collectively indicate that QL stabilization makes the soil more resilient and reliable for highway construction, ensuring long-term performance and reduced maintenance costs.

ACKNOWLEDGMENT

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA

for funding this research work through the project number "NBU-FFR-2024-2105-02".

REFERENCES

- [1] H. Menaceur, O. Cuisinier, F. Masrouri, and H. Eslami, "Impact of monotonic and cyclic suction variations on the thermal properties of a stabilized compacted silty soil," *Transportation Geotechnics*, vol. 28, May 2021, Art. no. 100515, <https://doi.org/10.1016/j.trgeo.2021.100515>.
- [2] A. Mehenni, O. Cuisinier, and F. Masrouri, "Impact of Lime, Cement, and Clay Treatments on the Internal Erosion of Compacted Soils," *Journal of Materials in Civil Engineering*, vol. 28, no. 9, Sep. 2016, Art. no. 04016071, [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0001573](https://doi.org/10.1061/(ASCE)MT.1943-5533.0001573).
- [3] N. S. Ikhlef, M. S. Ghembaza, and M. Dadouch, "Effect of Cement and Compaction on the Physicochemical Behavior of a Material in the Region of Sidi Bel Abbas," *Engineering, Technology & Applied Science Research*, vol. 4, no. 4, pp. 677–680, Aug. 2014, <https://doi.org/10.48084/etasr.467>.
- [4] G. Stoltz, O. Cuisinier, and F. Masrouri, "Weathering of a lime-treated clayey soil by drying and wetting cycles," *Engineering Geology*, vol. 181, pp. 281–289, Oct. 2014, <https://doi.org/10.1016/j.enggeo.2014.08.013>.
- [5] H. Sellaf and B. Balegh, "An Experimental Study on the Effect of Plastic Waste Powder on the Strength Parameters of Tuff and Bentonite Soils Treated with Cement," *Engineering, Technology & Applied Science Research*, vol. 13, no. 2, pp. 10322–10327, Apr. 2023, <https://doi.org/10.48084/etasr.5580>.
- [6] V. Robin, A. A. Javadi, O. Cuisinier, and F. Masrouri, "An effective constitutive model for lime treated soils," *Computers and Geotechnics*, vol. 66, pp. 189–202, May 2015, <https://doi.org/10.1016/j.compgeo.2015.01.010>.
- [7] N. Chabrat, O. Cuisinier, and F. Masrouri, "In situ ageing of a lime/cement-treated expansive clayey soil," in *8th International Symposium on Deformation Characteristics Of Geomaterials*, Porto, Portugal, Sep. 2023.
- [8] O. Cuisinier, G. Stoltz, and F. Masrouri, "Long-Term Behavior of Lime-Treated Clayey Soil Exposed to Successive Drying and Wetting," pp. 4146–4155, Mar. 2014, <https://doi.org/10.1061/9780784413272.403>.
- [9] O. Cuisinier, F. Masrouri, and A. Mehenni, "Alteration of the Hydromechanical Performances of a Stabilized Compacted Soil Exposed to Successive Wetting–Drying Cycles," *Journal of Materials in Civil Engineering*, vol. 32, no. 11, Nov. 2020, Art. no. 04020349, [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0003270](https://doi.org/10.1061/(ASCE)MT.1943-5533.0003270).
- [10] M. Otieno, C. Kabubo, and Z. Gariy, "Mechanical and Structural Correlation of Lateritic Soil Road Base Stabilized with Cement and Selected Biochars," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11070–11077, Aug. 2023, <https://doi.org/10.48084/etasr.5973>.
- [11] A. A. E. Driss, K. Harichane, and M. Ghrichi, "Effect of lime on the stabilization of an expansive clay soil in Algeria," *Journal of Geomechanics and Geoengineering*, vol. 1, pp. 1–10, 2023, <https://doi.org/10.38208/jgg.v1i1.413>.
- [12] G. R. Pokkunuri, R. K. Sinha, and A. K. Verma, "Field Studies on Expansive Soil Stabilization with Nanomaterials and Lime for Flexible Pavement," *Sustainability*, vol. 15, no. 21, 2023, <https://doi.org/10.3390/su152115291>.
- [13] M. Awad, I. Al-Kiki, and A. A. Khalil, "Permeability of Expansive Soils Modified/Stabilized with lime (Review Paper)," *Diyala Journal of Engineering Sciences*, vol. 14, no. 2, 2021, [Online]. Available: <https://www.iasj.net/iasj/article/210287>.
- [14] M. S. Islam, T. Islam, and N. Khatun, "Permeability Alteration of Low Plastic Clay and Poorly Graded Sand Using Lime and Fly Ash," *Indian Geotechnical Journal*, vol. 51, no. 5, pp. 967–978, Oct. 2021, <https://doi.org/10.1007/s40098-020-00493-5>.
- [15] N. D. Quang and J. C. Chai, "Permeability of lime- and cement-treated clayey soils," *Canadian Geotechnical Journal*, vol. 52, no. 9, pp. 1221–1227, Sep. 2015, <https://doi.org/10.1139/cgj-2014-0134>.
- [16] G. Stoltz, O. Cuisinier, and F. Masrouri, "Weathering of a lime-treated clayey soil by drying and wetting cycles," *Engineering Geology*, vol.

- 181, pp. 281–289, Oct. 2014, <https://doi.org/10.1016/j.enggeo.2014.08.013>.
- [17] A. Lloret, M. V. Villar, M. Sánchez, A. Gens, X. Pintado, and E. E. Alonso, "Mechanical behaviour of heavily compacted bentonite under high suction changes," *Géotechnique*, vol. 53, no. 1, pp. 27–40, Feb. 2003, <https://doi.org/10.1680/geot.2003.53.1.27>.
- [18] P. Delage, D. Marcial, Y. J. Cui, and X. Ruiz, "Ageing effects in a compacted bentonite: a microstructure approach," *Géotechnique*, vol. 56, no. 5, pp. 291–304, Jun. 2006, <https://doi.org/10.1680/geot.2006.56.5.291>.
- [19] B. Celauro, A. Bevilacqua, D. Lo Bosco, and C. Celauro, "Design Procedures for Soil-Lime Stabilization for Road and Railway Embankments. Part 1-Review of Design Methods," *Procedia - Social and Behavioral Sciences*, vol. 53, pp. 754–763, Oct. 2012, <https://doi.org/10.1016/j.sbspro.2012.09.925>.
- [20] B. Celauro, A. Bevilacqua, D. Lo Bosco, and C. Celauro, "Design Procedures for Soil-Lime Stabilization for Road and Railway Embankments. Part 2-Experimental Validation," *Procedia - Social and Behavioral Sciences*, vol. 53, pp. 568–579, Oct. 2012, <https://doi.org/10.1016/j.sbspro.2012.09.907>.
- [21] O. Cuisinier and F. Masrouri, "Impact of wetting/drying cycles on the hydromechanical behaviour of a treated soil," *E3S Web of Conferences*, vol. 195, 2020, Art. no. 06008, <https://doi.org/10.1051/e3sconf/202019506008>.
- [22] P. V. Sivapullaiah, B. G. Prasad, and M. M. Allam, "Effect of Sulfuric Acid on Swelling Behavior of an Expansive Soil," *Soil and Sediment Contamination: An International Journal*, vol. 18, no. 2, pp. 121–135, Feb. 2009, <https://doi.org/10.1080/15320380802660289>.
- [23] M. H. Fasihnikoutalab, A. Asadi, C. Unluer, B. K. Huat, R. J. Ball, and S. Pourakbar, "Utilization of Alkali-Activated Olivine in Soil Stabilization and the Effect of Carbonation on Unconfined Compressive Strength and Microstructure," *Journal of Materials in Civil Engineering*, vol. 29, no. 6, Jun. 2017, Art. no. 06017002, [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0001833](https://doi.org/10.1061/(ASCE)MT.1943-5533.0001833).
- [24] S. Pourakbar, A. Asadi, B. B. K. Huat, and M. H. Fasihnikoutalab, "Soil stabilisation with alkali-activated agro-waste," *Environmental Geotechnics*, vol. 2, no. 6, pp. 359–370, Dec. 2015, <https://doi.org/10.1680/envgeo.15.00009>.
- [25] S. Pourakbar, B. B. K. Huat, A. Asadi, and M. H. Fasihnikoutalab, "Model Study of Alkali-Activated Waste Binder for Soil Stabilization," *International Journal of Geosynthetics and Ground Engineering*, vol. 2, no. 4, Nov. 2016, Art. no. 35, <https://doi.org/10.1007/s40891-016-0075-1>.
- [26] M. Karasahin, E. Keskin, and İ. Sahinoglou, "Effect of Lime on Unconfined Compressive Strength of a Low Plasticity Clayey Soil," *Eurasian Journal of Civil Engineering and Architecture*, vol. 3, no. 2, pp. 32–40, 2019.

Unbalanced Distribution Network Cross-Country Fault Diagnosis Method with Emphasis on High-Impedance Fault Syndrome

Balamurali Krishna Ponukumati

School of Electrical Engineering, KIIT Deemed to be University, India
pbmk20031993@gmail.com (corresponding author)

Anil Kumar Behera

School of Electrical Engineering, KIIT Deemed to be University, India
anil.beherafel@kiit.ac.in

Lipsa Subhadarshini

School of Electrical Engineering, KIIT Deemed to be University, India
lipsa.subhadarshinifel@kiit.ac.in

Pampa Sinha

School of Electrical Engineering, KIIT Deemed to be University, India
pampa.sinhafel@kiit.ac.in

Manoj Kumar Maharana

School of Electrical Engineering, KIIT Deemed to be University, India
mkmfel@kiit.ac.in

Arapirala Venkata Pavan Kumar

Department of EEE, Madanapalle Institute of Technology & Science, India
pawanrao82@gmail.com

Received: 16 January 2024 | Revised: 27 January 2024 and 25 January 2024 | Accepted: 7 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6917>

ABSTRACT

Unusual fault scenarios can occur on the utility grid in a power system network. Cross-Country Faults (CCFs) connected to the High-Impedance Fault (HIF) syndrome are more prone to occur in forested areas due to thunderstorms, cyclones, and improper vegetation management and tree pruning. Finding and categorizing CCFs associated with HIF syndrome is a great challenge. This study employed the cross-correlation method to reconstruct the signals produced by CCFs with HIF, which were shown to be complicated, aperiodic, asymmetric, and nonlinear. A decreased sensitivity to random noise means that a given modification might not affect equally all component peaks. This allows for more precise signal recovery. The maximum voltage cross-correlation coefficients were carefully evaluated as distinguishing elements in the development of a suggested fault detection technique. The proposed concept was evaluated on a modified imbalanced IEEE 240 bus system under different case studies. These case studies cover a wide range of scenarios, such as the switching of a capacitor bank, feeder energization, and the effects of nonlinear loads under noisy conditions.

Keywords-Cross Country Faults (CCF); High Impedance Fault (HIF); cross-correlation; slime mould optimization technique; non-negative matrix factorization; method of peak detection

I. INTRODUCTION

The Cross-Country High Impedance Fault (CCHIF) can be described as "high-impedance ground faults happening in separate phases of one circuit at different places at the same time as the fault inception time" in a typical scenario. The main challenges in cross-country faults in transmission/distribution systems are the detection of distance, direction, and phase selection. These faults are simultaneous or evolving problems in the network. The relay may experience failure to function correctly in cases where there is an internal issue within its designated zone or may operate erroneously when a fault occurs outside it. This complexity arises from the unpredictable distribution of zero- and negative-sequence components along the protected line, which is used by the above mentioned functions. This unpredictability affects both solidly grounded systems and those that are either isolated or impedance grounded [1].

Wavelet Transform (WT) is a signal-processing technique that has been widely used for HIF detection [2]. Using a collection of filters, WT can analyze HIF signals at various resolution levels. Artificial Neural Networks (ANNs) and fuzzy expert systems, along with WT, have been employed to detect HIF in distribution networks [3-6]. In [7], a graph theory-based zone identification method and Random Search Multilevel Support Vector Machine (RSMSVM) algorithm were used to accurately categorize the fault zone. The moving sum approach and mathematical morphology [8-9] are different signal processing tools, and conventional relays can help detect HIFs. Ongoing research is focused on precise identification and location estimation of HIFs using methods such as evidential reasoning [10], parameter estimation [11], and computation of fault resistances [12].

In Unified Power Flow Controller (UPFC) compensated transmission lines, a backup protection technique based on differential apparent power can help the traditional distance protection system with HIF detection [13]. In [14], a signal-processing-based method was proposed to extract the necessary data to detect and locate HIF in the situation of a single line to ground fault, where the voltage and current signal were decomposed into orthogonal components. In [15], a method for identifying malfunctioning feeders in HIF was presented, considering the Peterson coil's influence. In [16], a single-phase composite power-based HIF protection approach for a low-resistance grounded distribution network was proposed, without considering the polarity of the current transformers. In [17-18], new models for the identification of HIFs were introduced. Nowadays, wind farms are gradually integrated into grids at various voltage levels. When a wind farm is connected to the grid through a transmission line, the voltage of the system constantly fluctuates, and as a result, fault classification becomes more intricate.

Recent studies have highlighted the growing use of Support Vector Machines (SVM) for power system fault monitoring. The SVM approach focuses on recognizing system faults, particularly power system instability, using bus voltages, times before and after faults, and generator angles as training data. A trained SVM effectively identified abnormal conditions after

disturbance [19]. Other studies evaluated the impact of integrating wind power into the system, employing the Multi-Band Power System Stabilizer (MBPSS) to suppress dynamic oscillations. Simulations confirm MBPSS's ability to enhance system stability under severe faults and high integration of wind farms [20-21]. This study proposes a unique approach to identifying the location of an HIF fault zone in a distribution system. The innovative contributions of this study are:

- A cross-correlation-based scheme for identifying cross-country HIFs in the distribution system is proposed. The method involves analyzing the cross-correlation between the voltage under normal operating conditions and the voltage during a faulty state to identify distinctive characteristics in the signature.
- The correlogram as a useful tool for data analysis is validated by demonstrating its ability to reduce the effect of uncorrelated random noise. Feature extraction aims to create fresh features from the ones already present in a dataset, thereby decreasing the dataset's size. These refined features should succinctly encapsulate the meaningful information in the original features. Through the amalgamation of the initial features in this manner, a more concise rendition of the original set can be produced.

II. PROPOSED METHOD

A. Feature Extraction based on Time Cross-Correlation

A correlation describes how closely two variables are related. Correlation is helpful because it shows the connection between two variables, which enables users to forecast how a system will behave in the future. In simpler terms, cross-correlation serves as a statistical tool for assessing the extent to which two correlated signals resemble each other. The time series is expected to be organized chronologically with $I = 1, \dots, N$. The periodogram is conventionally defined as the squared modulus of the Discrete Fourier Transform (DFT) and serves as an approximation for Power Spectral Density (PSD):

$$P(v_k) = \frac{2T}{N^2} \left(\left[\sum_{i=1}^N f_i \cos(2\pi v_k t_i) \right]^2 + \left[\sum_{i=1}^N f_i \sin(2\pi v_k t_i) \right]^2 \right) \quad (1.a)$$

By applying normalization to the cross-correlation function, a Pearson correlation coefficient that is time-adjusted can be obtained:

$$\rho_{xx}(t_1, t_2) = \frac{K_{xx}(t_1, t_2)}{\sigma_x(t_1)\sigma_x(t_2)} = \frac{E[(X_{t_1} - \mu_{t_1})(X_{t_2} - \mu_{t_2})]}{\sigma_x(t_1)\sigma_x(t_2)} \quad (1.b)$$

Specifically, a stochastic process's normalized cross-correlation is defined as:

$$\rho_{xy}(\tau) = \frac{K_{xy}(\tau)}{\sigma_x \sigma_y} = \frac{E[(X_t - \mu_x)(Y_{t+\tau} - \mu_y)]}{\sigma_x \sigma_y} \quad (1.c)$$

$$Kx(\tau) = E[x(t_1)y(t_1 + \tau)] = E[x(t_1 - \tau)y(t_1)] \\ = E[y(t_1)x(t_1 - \tau)] = Ky(-\tau) \quad (1.d)$$

where μ_x is the mean and σ_y is the standard deviation of the processes X_t and Y_t , respectively, which are constant over time, and K_{xy} is the cross co-variance function respectively.

Frequencies are given by $v_k = k/T$, where k can be any integer from 1 to N , and $v_k = k/(N-1)/2$, where N is an odd number. The maximum frequency is the Nyquist frequency $v_{Nyq} = (N/2)(1/T)$, the minimum frequency is $v_{min} = 1/T$, and $T = N(t_N - t_1)/(N - 1)$. The cross-correlation between 2 signals, $x(n)$ and $y(n)$ is given by:

$$\hat{R}_{xy}(m) = \begin{cases} \sum_{n=0}^{N-m-1} x_{n+m} y_n & m \geq 0 \\ \hat{R}_{yx}(-m) & m < 0 \end{cases} \quad (2)$$

where m is equal to 2, 1, 0, 1, 2. Subscript xy denotes the order in which the two variables are associated, and index m denotes a parameter that shifts with time. The subscripts are arranged so that x precedes y to indicate how one sequence is moved concerning the other. Compared to the equal sampling scenario, the uneven sampling case's window function shapes can account for the significant red-noise leakage shown in the simulations and the increased noise. Using even sampling, standard Fourier analysis results can be recreated, complete with the well-established characteristics of window functions. In the case of representing noise, it can be written:

$$x(t) = A \cos(\omega t + \theta) \quad (3)$$

where θ is a random variable, $y(t)$ represents noise, and $x(t)$ and $y(t)$ are uncorrelated functions. Then, the autocorrelation coefficient of $x(t)$ is $R_x(\tau) = \frac{A^2}{2} \cos \omega \tau$. $R_y(\tau)$ is the autocorrelation function of noise $y(t)$. $R_y(\tau)$ should be decaying in nature. Hence correlograms are decaying in. Consider:

$$R_y(\tau) = Y_0^2 e^{-\tau/\tau} \quad (4)$$

After adding $x(t)$ and $y(t)$, it becomes:

$$z(t) = x(t) + y(t) \quad (5)$$

$$Z = A \cos(\omega t + \theta) + \text{noise}$$

$$R_z(\tau) = R_x(z) + R_y(\tau), \quad Y_0^2 \gg \frac{A^2}{2} \quad (6)$$

If $x(t)$ contains several frequencies or a small band of frequencies, this band can be recovered from the new signal $Z(t)$. In this study, the Source Separation Non-Negative Matrices (SSNMF) algorithm was used to remove the significant noise.

B. Reducing Noise

The SSNMF method is represented graphically by a sweep sequence of amplitude spectra. The Frequency-Following Response (FFR) can be more easily seen, and extra noise is eliminated from all recordings thanks to the SSNMF decomposition. The analysis involves examining the patterns of improvement in FFR and reduction in noise as the number of sweeps increases. A model was developed using an exponential curve fit. Pitch processing and neuroplasticity processes during signal interruptions are two possible uses of the SSNMF algorithm on the FFR signal [22]. The machine learning algorithm takes a non-negative input matrix A and uses it to learn and factorize a smaller matrix S , which serves as the spectral basis, and T , which serves as the information coding.

$$A_{ij} \approx (ST)_{ij} = \sum_{k=1}^n S_{ik} T_{kj} \quad (7)$$

Here, i and j stand for elements across matrix A 's initial (a flattened frequency-time vector) and secondary (a sequence of amplitude spectrograms) dimensions, respectively. Following this approach, FFR and noise were independently reconstructed, as shown in Figure 1, using the following formulas: multiply the input data matrix T by their respective S - T ratios, and then divide the result by the matrix T .

$$FFR = \frac{A(S_{FFR} T_{FFR})}{ST} \quad (8)$$

$$\text{Noise} = \frac{A(S_{noise} T_{noise})}{ST} \quad (9)$$

The following equation is used to outline the efficiency improvement in the FFR signal (i.e. working of the SSNMF method):

$$B(n) = B_{BS} e^{\left(-\frac{n}{\tau}\right)} + B_{DC} \quad (10)$$

In this context, B represents the performance index, specifically FFR improvement, n is the total amount of sweeps for each signal, B_{BS} is the asymptotic amplitude of the fitted curve minus the direct current component, and the constant e represents Euler's mathematical constant (approximately 2.7182) and denotes the time constant of the fitted curve, indicating the number of sweeps required to reach 63% of the asymptotic amplitude. B_{DC} refers to the direct current component of the fitted curve, representing the overall elevation of the fitted curve. For noise reduction, an alternative model was employed, showing results that exhibited an increasing trend as the number of sweeps increased.

$$B(n) = B_{BS} e^{\left(-\frac{n}{\tau}\right)} + B_{DC}$$

$$FFR_{\text{enhancement}} =$$

$$0.254 * \left(e^{\left(-\frac{n}{555}\right)}\right) + 0.005 \quad (11)$$

$$\text{NoiseReduction from signal} =$$

$$20.653 * \left(1 - e^{\left(-\frac{n}{290}\right)}\right) - 20.991 \quad (12)$$

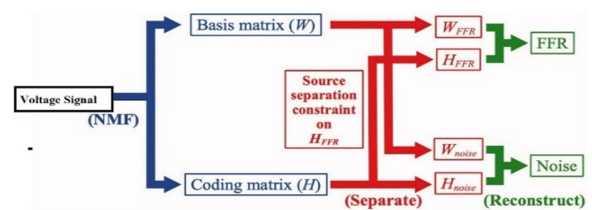


Fig. 1. Block diagram of source separation using SSNMF algorithm.

C. Method of Peak Detection (MOPD) for Determining QRS

Applying Normal-Exponential-Bernoulli (NEB) and mixture probability models, a brand-new peak identification algorithm was used to analyze extensive two-dimensional electrical signals:

$$X_i \sim ND(\theta_i + \mu, \sigma^2) \text{ and } \theta_i \sim \text{Exp}(\phi) \quad (13)$$

III. SYSTEM MODELING

A. Modified 240 Node System Modeling

This system consists of three feeders, as shown in Figure 2. Feeder A has 17 nodes, feeder B has 60 nodes, and feeder C has 162 nodes, all receiving power from a 69 kV substation. The major length of this distribution system is 23 miles, and it supplies electricity to more than 1100 people. Transformers are used for secondary supply clients. Real direct data on power consumption (kW) are available through smart meters at different locations based on the slime mould optimization technique. There are two capacitor banks in the 240-bus system. The substation features on-load tap changers at nodes 2038 and 3079. The introduction of photovoltaic (PV) systems at different locations alters this test system. As shown in Figure 2, a total of 20 PV systems were installed in the 240-node unbalanced distribution system. The size of the selected PV system size was 40 kVA, even if PV system sizes and locations are selected on a random basis using empirical data. Voltage readings are taken throughout the grid at various points.

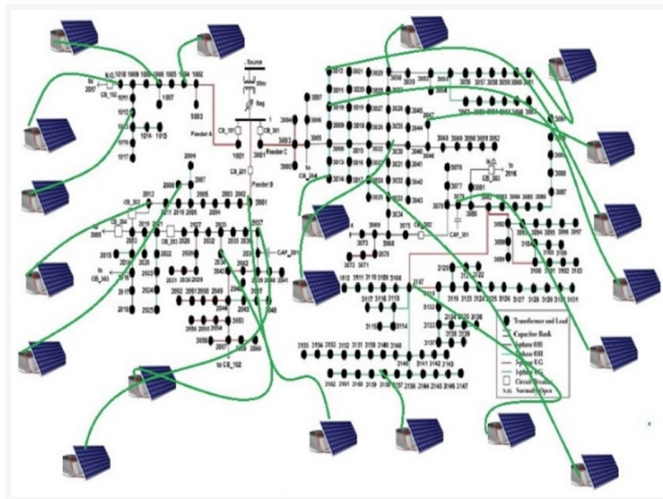


Fig. 2. The IEEE modified 240 bus system.

B. HIF Modeling for Cross Country

A HIF exhibits traits similar to an electric arc, including non-linearity, inconsistency, and recurrence in the fault current signals. Additionally, these signals exhibit the existence of high-frequency components and harmonics. Various studies have proposed various HIF model types. This study considered a 2-diode HIF model, similar to [5]. The model consists of two DC voltage sources (V_p and V_n), two diodes (D_p and D_n), and two resistances (R_p and R_n) that are coupled in an antiparallel fashion. Asymmetric fault currents were produced in the distribution network using several sets of fault resistances R_p and R_n that have different values. When the line voltage is higher than the positive DC voltage (V_p), the fault current begins to flow in the direction of the ground. On the contrary, it returns from the ground when the line voltage is less than the negative DC voltage (V_n). If the phase voltage falls within the range of V_p and V_n , there is no fault current. This study considered three diverse situations built on three different

voltage and resistance values, as shown in Table I. Variation is necessary to achieve the various HIF behaviors and their effects on the electrical power distribution system [6]. When HIFs occur in a feeder, both negative and zero sequence currents are introduced. HIFs are mostly line-to-ground faults.

TABLE I. HIF PARAMETERS

Sl. No.	HIF parameter values			
	V_p (V)	R_p (Ω)	V_n (V)	R_n (Ω)
Condition 1	3588	208	3847	212
Condition 2	6180	245	6155	245
Condition 3	8092	270	9358	290

IV. RESULTS AND DISCUSSION

To validate the proposed algorithm for the detection of cross-country HIFs, different capacitor bank switching feeder energization at various inception angles and locations were simulated, with both balanced and unbalanced loads as well as electronic power loads. The signal is generated at time $t = 0.1$ s with a sampling frequency of 2048 Hz, and the calculation is carried out one cycle after the fault event to determine the values of Q , R , and S . In this study cross-country-HIF and other transients were generated in the 240 bus distribution network on buses 1010, 1015, 2019, 2031, 3022, and 3035. Buses 1037, 1014, 1005, 1003, 1011, 1015, 2012, 2019, 2030, 2024, 2021, 2032, 2042, 2053, 2066, 2041, 2031, 3034, 3005, 3013, 3011, 3027, 3030, 3042, 3049, 3094, and 3067 are the best locations for the monitoring points, which were found using the slime mould optimization technique. At the measurement point, voltage signal correlation was tested under both normal and various transient conditions, such as cross-country faults. The two signals' cross-correlogram becomes their autocorrelation when everything is working properly. The cross-correlogram produces various waveforms for various types of transients. Figure 3 shows the extraction of features for CCHIF detection on bus 1010.

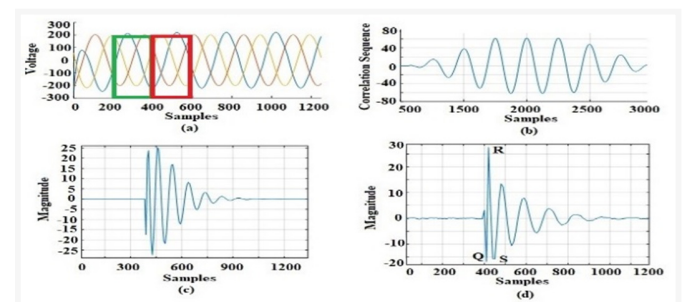


Fig. 3. Extraction of features for CCHIF detection at bus 1010: (a) voltage signal during CCHIF, (b) correlation with pure signal, (c) signal after correlation and MOPD operation, and (d) signal after filtering.

Since there are two periodic switching fingerprints in two consecutive windows that do not overlap, they can be dismissed by subtracting the matching data from both windows. A technique based on cross-correlation is used to guarantee that the difference is small or zero under typical conditions. This is performed by calculating the cross-correlation of the windows at varying delays. The windows'

similarity is at its highest when the estimated cross-correlation value is maximized, and the windows' length is recalculated after taking into account the appropriate latency. This method guarantees that the subtraction of similar samples yields the correct result.

Figure 4 shows the correlograms for some example studies. The CCHIF correlogram under condition 3 in Figure 5(a) was generated at 0.02 s on bus 1010 and observed from bus 1013. Figure 4 (b, c) shows the correlograms of the capacitor bank switching and the feeder energization on bus 1015 at 0.02 s, respectively. Figure 5 shows that the Q , R , and S points are indistinguishable in scenario c . Therefore, the associated signal is subjected to MOPD action, followed by SSNMF operation to produce the waveforms shown in Figure 5. From this figure, three separate values can be obtained for Q , R , and S , which were used for CCHIF detection.

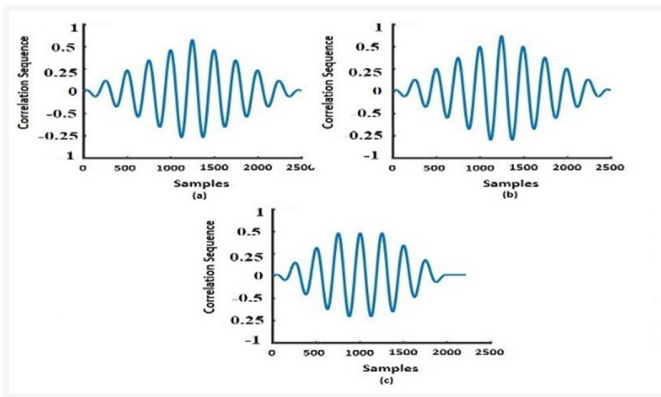


Fig. 4. Cross-correlogram of CCHIF: (a) CCHIF, (b) switching of the capacitor bank, and (c) energization of the feeder.

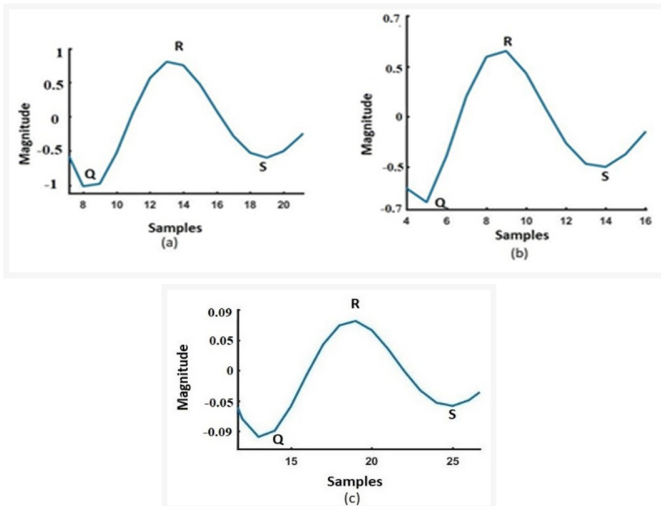


Fig. 5. Cross-correlogram post MOPD and filter operation to obtain Q , R , and S points: (a) CCHIF, (b) signal for switching of the capacitor bank.

V. METHOD VERIFICATION UNDER DIFFERENT CONDITIONS

A. Influence of Noise

The voltage and current signals in real-time systems are always contaminated with noise. Unwanted electrical impulses that interfere or distort with an original (or desired) signal are known as noise or interference. As a result, the effectiveness of the suggested approach for cross-country HIF detection was investigated in noisy environments. The entire recorded signal contains power system noise, which has a normal probability distribution. The Signal-to-Noise Ratio (SNR) that states this noise, is given by:

$$SNR_{dB} = 20 \cdot \log_{10} \left(\frac{A_{signal}}{A_{noise}} \right) \quad (14)$$

B. CCHIF Fault Detection Influence of Power Electronics Load

Power electronics interfaced nonlinear loads, such as time-varying harmonics in distribution networks, are now widely employed by both home and industrial customers. The proposed solution was tested with a non-linear load to assess its performance. A 6-pulse converter bridge that provides a DC load of 650 kW at 3.6 kV and is coupled at various points of the 240-bus distribution network was used to design the non-linear load.

VI. CONCLUSIONS

This study introduces a reliable correlation-based method for detecting cross-country HIFs in distribution networks. The proposed method was effective in identifying this type of fault under various conditions, including capacitor bank switching, load switching, feeder energization, and unbalanced load. Unlike complex approaches, this technique analyzes voltage signals at a single monitoring point and performs correlation operations with those signals in typical scenarios. The identification of the faulty phase during cross-country HIFs relies on just three features extracted from the cross-correlogram of voltage signals. To streamline the feature extraction process, a cross-correlation-based method was used, which helps mitigate the impact of random uncorrelated noise in the signal. The slime mould optimization technique was used to determine the optimal monitoring point locations to identify the faulty bus in the network. The efficiency of the proposed method was verified on a real-time platform, and its simplicity was highlighted by its lack of reliance on distributed parameters, artificial intelligence techniques, or time synchronization with monitoring devices.

REFERENCES

- [1] F. M. Gatta, A. Geri, S. Lauria, M. Maccioni, A. Cerretti, and L. D'Orazio, "Large Temporary Overvoltages in MV Network due to a Series Fault in the HV Subtransmission System," in *2021 IEEE Madrid PowerTech*, Madrid, Spain, Jun. 2021, pp. 1–6, <https://doi.org/10.1109/PowerTech46648.2021.9494822>.
- [2] K. Solak and W. Rebizant, "Analysis of differential protection response for cross-country faults in transmission lines," in *2010 Modern Electric Power Systems*, Wroclaw, Poland, Sep. 2010.
- [3] A. A. M. Zin, N. A. Omar, A. M. Yusof, and S. P. A. Karim, "Effect of 132kV Cross-Country Fault on Distance Protection System," in *2012*

- Sixth Asia Modelling Symposium, Bali, Indonesia, Feb. 2012, pp. 167–172, <https://doi.org/10.1109/AMS.2012.17>.
- [4] Z. Y. Xu, W. Li, T. S. Bi, G. Xu, and Q. X. Yang, "First-Zone Distance Relaying Algorithm of Parallel Transmission Lines for Cross-Country Nonearthed Faults," *IEEE Transactions on Power Delivery*, vol. 26, no. 4, pp. 2486–2494, Jul. 2011, <https://doi.org/10.1109/TPWRD.2011.2158455>.
 - [5] T. Bi, W. Li, Z. Xu, and Q. Yang, "First-Zone Distance Relaying Algorithm of Parallel Transmission Lines for Cross-Country Grounded Faults," *IEEE Transactions on Power Delivery*, vol. 27, no. 4, pp. 2185–2192, Jul. 2012, <https://doi.org/10.1109/TPWRD.2012.2210740>.
 - [6] S. R. Kumar Joga, P. Sinha, M. K. Maharana, C. Jena, A. Mishra, and A. Roy, "Stockwell Transform and Data Mining based Fault Diagnosis Method to protect Microgrid," in *2022 3rd International Conference for Emerging Technology (INCET)*, Belgaum, India, Feb. 2022, pp. 1–5, <https://doi.org/10.1109/INCET54531.2022.9824471>.
 - [7] S. R. K. Joga, P. Sinha, and M. K. Maharana, "A novel graph search and machine learning method to detect and locate high impedance fault zone in distribution system," *Engineering Reports*, vol. 5, no. 1, 2023, Art. no. e12556, <https://doi.org/10.1002/eng2.12556>.
 - [8] S. AsghariGovar and H. Seyed, "Adaptive CWT-based transmission line differential protection scheme considering cross-country faults and CT saturation," *IET Generation, Transmission & Distribution*, vol. 10, no. 9, pp. 2035–2041, 2016, <https://doi.org/10.1049/iet-gtd.2015.0847>.
 - [9] S. Singh and D. N. Vishwakarma, "A Novel Methodology for Identifying Cross-Country Faults in Series-Compensated Double Circuit Transmission Lines," *Procedia Computer Science*, vol. 125, pp. 427–433, Jan. 2018, <https://doi.org/10.1016/j.procs.2017.12.056>.
 - [10] A. Swetapadma and A. Yadav, "An artificial neural network-based solution to locate the multilocation faults in double circuit series capacitor compensated transmission lines," *International Transactions on Electrical Energy Systems*, vol. 28, no. 4, 2018, Art. no. e2517, <https://doi.org/10.1002/etep.2517>.
 - [11] V. Ashok and A. Yadav, "A Protection Scheme for Cross-Country Faults and Transforming Faults in Dual-Circuit Transmission Line Using Real-Time Digital Simulator: A Case Study of Chhattisgarh State Transmission Utility," *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, vol. 43, no. 4, pp. 941–967, Dec. 2019, <https://doi.org/10.1007/s40998-019-00202-w>.
 - [12] V. Ashok, A. Yadav, and A. Y. Abdelaziz, "MODWT-based fault detection and classification scheme for cross-country and evolving faults," *Electric Power Systems Research*, vol. 175, Oct. 2019, Art. no. 105897, <https://doi.org/10.1016/j.epsr.2019.105897>.
 - [13] A. Nikander and P. Järventausta, "Identification of High-Impedance Earth Faults in Neutral Isolated or Compensated MV Networks," *IEEE Transactions on Power Delivery*, vol. 32, no. 3, pp. 1187–1195, Jan. 2017, <https://doi.org/10.1109/TPWRD.2014.2346831>.
 - [14] B. Kumar and A. Yadav, "Backup protection scheme for transmission line compensated with UPFC during high impedance faults and dynamic situations," *IET Science, Measurement & Technology*, vol. 11, no. 6, pp. 703–712, 2017, <https://doi.org/10.1049/iet-smt.2016.0458>.
 - [15] O. E. Batista, R. A. Flauzino, M. A. de Araujo, L. A. de Moraes, and I. N. da Silva, "Methodology for information extraction from oscillograms and its application for high-impedance faults analysis," *International Journal of Electrical Power & Energy Systems*, vol. 76, pp. 23–34, Mar. 2016, <https://doi.org/10.1016/j.ijepes.2015.09.019>.
 - [16] B. K. Ponukumati, P. Sinha, M. K. Maharana, A. V. P. Kumar, and A. Karthik, "An Intelligent Fault Detection and Classification Scheme for Distribution Lines Using Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 8972–8977, Aug. 2022, <https://doi.org/10.48084/etasr.5107>.
 - [17] B. K. Ponukumati, P. Sinha, M. K. Maharana, C. Jenab, A. V. Pavan Kumar, and K. Akkenaguntla, "Pattern Recognition Technique Based Fault Detection in Multi-Microgrid," in *2021 IEEE 2nd International Conference on Applied Electromagnetics, Signal Processing, & Communication (AESPC)*, Bhubaneswar, India, Nov. 2021, pp. 1–6, <https://doi.org/10.1109/AESPC52704.2021.9708541>.
 - [18] B. K. P. Sinha, M. K. Maharana, C. Jena, A. V. Pavan Kumar, and K. Akkenaguntla, "Power System Fault Detection Using Image Processing And Pattern Recognition," in *2021 IEEE 2nd International Conference on Applied Electromagnetics, Signal Processing, & Communication (AESPC)*, Bhubaneswar, India, Nov. 2021, pp. 1–5, <https://doi.org/10.1109/AESPC52704.2021.9708475>.
 - [19] P. Ilius, M. Almuhami, M. Javaid, and M. Abido, "A Machine Learning-Based Approach for Fault Detection in Power Systems," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11216–11221, Aug. 2023, <https://doi.org/10.48084/etasr.5995>.
 - [20] B. Dhoubi, M. A. Zdiri, Z. Alaas, and H. H. Abdallah, "Analyzing the Effects of MBPSS on the Transit Stability and High-Level Integration of Wind Farms during Fault Conditions," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10652–10658, Jun. 2023, <https://doi.org/10.48084/etasr.5838>.
 - [21] P. Balamurali Krishna and P. Sinha, "Detection of Power System Harmonics Using NBPSO Based Optimally Placed Harmonic Measurement Analyser Units," in *2018 Second International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, Feb. 2018, pp. 369–373, <https://doi.org/10.1109/ICCMC.2018.8488114>.
 - [22] Y. Xue, X. Chen, H. Song, and B. Xu, "Resonance Analysis and Faulty Feeder Identification of High-Impedance Faults in a Resonant Grounding System," *IEEE Transactions on Power Delivery*, vol. 32, no. 3, pp. 1545–1555, Jun. 2017, <https://doi.org/10.1109/TPWRD.2016.2641045>.

A Real Test and Simulation Result Comparison of Selected Properties of Hybrid Composite Materials

Naqib Daneshjo

Faculty of Commerce, University of Economics in Bratislava, Slovak Republic
daneshjo47@gmail.com

Dusan Sabadka

Faculty of Mechanical Engineering, Technical University of Kosice, Slovak Republic
dusan.sabadka@tuke.sk

Peter Malega

Faculty of Mechanical Engineering, Technical University of Kosice, Slovak Republic
peter.malega@tuke.sk (corresponding author)

Received: 17 January 2024 | Revised: 8 February 2024 | Accepted: 18 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6922>

ABSTRACT

In this study, the notion of composite materials is thoroughly assessed. Actual and simulation in a specific computer software stress testing of hybrid composites are investigated. The paper deals with the mechanics of rigid bodies, their elasticity, strength, and stiffness. In addition to a general overview of the former's behavior and properties, this paper presents the possibility of calculating the bearing capacity of various materials in relevant computer programs. The production and testing process of the composite samples are described. The latter are then subjected to simulated tests in computer software. The main objective of this study is to compare real test results of hybrid composites, namely combined carbon fibers, glass fibers, aramid-carbon fibers, aramid honeycomb, and metal mesh with the simulation findings.

Keywords-simulation; composite materials; hybrid composites; tensile test; simulation of tensile test

I. INTRODUCTION

The development of new technologies has imported various materials into the production market, which have not been frequently used so far mostly due to their high cost or demanding production process. A typical example of such materials is modern hybrid composites [1-4], which mainly attain excellent parameters of mechanical resistance while maintaining low weight. The need for decreased prices and the necessity to implement new materials into practice owing to their unique properties indicate that the latter will be more frequently used in the future [5-8]. Thus, newly designed components and other structural elements made from them need to be tested. Since the main limiting factor is still the high cost, testing many samples is financially demanding, whereas special equipment and qualified personnel are required to carry out these tests. This constitutes a major obstacle in the case of a limited budget [9-12]. A possible solution might be to replace the practical tests of mechanical properties with a simulated test in a computer interface while using a specific calculation method [13-15].

Hybrid composite materials, combining different types of fibers and matrices, have become increasingly popular in various engineering applications due to their unique property combination. The comparison of properties obtained from real tests with those acquired through simulation programs is critical for ensuring the accuracy and reliability of computational models, therefore facilitating material design and optimization [16-19]. In [20-22], tensile and bending tests on hybrid composite specimens were conducted and the results were compared with those obtained from finite element analysis (FEA). A good agreement between experimental and simulated stress-strain curves was found, validating the predictive capabilities of the FEA model. Authors in [23-25] investigated the impact behaviour of hybrid composite laminates using both experimental testing and cohesive zone modeling. By comparing the delamination patterns observed in experiments with those predicted by simulation, they gained insights into the mechanisms governing interlaminar fracture in hybrid composites. In [26-28], a multi-scale modeling approach was employed to optimize the microstructure of hybrid composite materials for enhanced mechanical properties. Through the comparison of simulation predictions

with experimental data, an optimal fiber orientation distribution that maximized the material's strength and stiffness was identified.

In this study, different types of hybrid composites were chosen as reference materials, which had already been subjected to tensile tests in a laboratory as part of another project. The goal is to compare the real test results with the ones obtained in simulations, so that a conclusion could be drawn from this comparison.

II. USED MATERIALS FOR THE EXPERIMENT

Resin labelled L 285 MGS manufactured by Havel Composites was used as the matrix. This is a high-quality resin also employed in the aviation industry. A hardener from Havel Composites labelled 287 MGS was added to the resin in line with its intended purpose. A mixing ratio of 100:40 was utilized (with resin being the larger volume). The indicated time of possible processing after mixing these components is approximately 4 hours. The same matrix was applied for all reinforcements. All samples consist of seven layers symmetrically placed, with the outer 3 layers being the same each time: $2 \times$ carbon fabric and $1 \times$ glass fabric inward direction. The tested samples differ only in the middle layer through which the axis of symmetry passes. Carbon fabric represents the first, second, sixth and seventh layer of each tested sample. A product from the company Havel Composites was chosen with a stated weight of 160 g/m^2 .

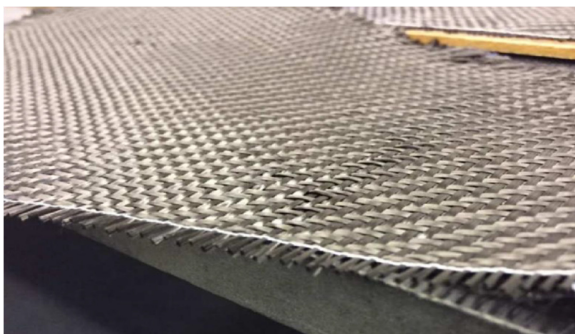


Fig. 1. Carbon fabric used for sample production.

Glass fabric titled Aeroglass 110 (type E) with a density of 102 g/m^2 was used as the third, or the fifth layer for each of the samples produced for testing. At the same time, it was utilized as a middle specific layer for type A samples, where it was placed at an angle of 45° to achieve unique mechanical properties. It is about a third lighter than the carbon layer used, but at the same time it offers significantly lower values of the modulus of elasticity in tension, and its rupture occurs at roughly half the stress. Hybrid aramid-carbon fabric was applied as a middle layer for type C samples. It combines the properties of both types of fibres. Its density is stated as 165 g/m^2 . As in the previous cases, the manufacturer is the company Havel Composites. Aramid honeycomb was chosen as a specific layer for type B samples. It is very light, has a density of only 48 g/m^2 , which is mainly due to the large ratio of "empty" space to the aramid walls of the structure. It offers

interesting values of modulus of elasticity in tension and shear due to its low weight.



Fig. 2. Glass fabric used in the production of the test samples.

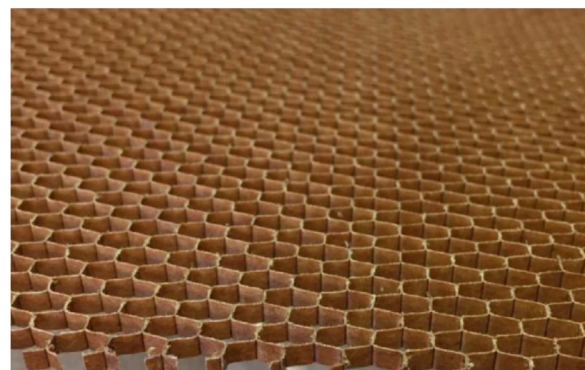


Fig. 3. Detail of the aramid honeycomb that was used as the middle layer in the type B samples.

A metal grid with a mesh size of $2 \times 2 \text{ mm}$ was implemented as a specific layer in the type D sample. From the available information, we cannot determine the exact mechanical properties of this grid. A metal grid with a mesh size of $10 \times 10 \text{ mm}$ was chosen as a specific layer for the type D sample. Similar to the steel grid with a smaller mesh, in this case as well we cannot determine the exact mechanical properties of the material from the available information. For these samples, parameters for grids of the same size that could have been made of different steel will be used for simulation purposes.



Fig. 4. Metal grid with a mesh size of $2 \times 2 \text{ mm}$ applied as a specific layer of D samples.

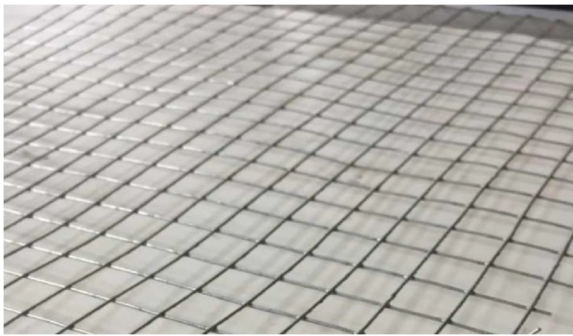


Fig. 5. 1 × 1 cm steel grid used for type E samples.

The materials needed for production were cut to a size of 26 × 33 cm. The matrix of resin and hardener were mixed according to the prescribed proportions. The individual layers, which were gradually stacked on top of each other, were subsequently saturated with the matrix. For the correct production of samples, it is necessary to observe the specified ratio of the matrix components and precisely stack the layers on top of each other. One of the last steps of the production procedure is the vacuuming, during which a layered sample is saturated with the matrix. This technology allows the removal of unwanted air from the sample, presses the individual layers together, and hardens the matrix. In practice, other technologies are additionally employed, being usually combined with increased ambient temperature during matrix hardening. After the matrix has hardened, the samples are milled from the individual composite "panels" with the assistance of special tools and they are then used in tests.



Fig. 6. Samples prepared for the tensile test.

The testing was carried out according to the TS EN ISO 527 standard. The tearing device TIRAtest 2300 was used to conduct the tensile test of the produced samples, which consists of jaws into which the tested element is firmly clamped. One of the jaws usually stays firmly in place while the other acts on

the sample with a gradually increasing force in one axis. The relative and total elongation of the sample is evaluated. To measure the elongation of a sample, the strain gauge is utilized. The maximum stress (force) that the sample can withstand without breaking is observed. Furthermore, various characteristics of flexibility, strength, and stiffness can be calculated. Elastic and plastic deformation as well as other attributes can be evaluated.



Fig. 7. Broken samples after the tensile test.

Of the fifteen samples tested, three tests failed, namely sample types A, B, and E, where the sample has come loose from the pneumatic jaws. From the tensile test results, it was found that the greatest force F_m to break the material must be exerted on the type C samples. The hybrid component of the reinforcement mostly increases the force required to break the material. Tables I-V illustrate the test outcomes on the samples, while the average value of the 3 samples tested for one type is also given. For sample types A, B, and E, only two values were calculated. Type B samples achieved the lowest values. The glass fiber samples with different orientation had the highest ultimate tensile strength. On the other hand, type E samples with 10×10 mm metal grid achieved the lowest values of ultimate tensile strength. Regarding the type E samples, each of them experienced material failure in the jaw attachment area during testing, probably due to the size of the holes in the metal grid where deformation of the composite material due to jaw pressure occurred at the sample attachment point.

III. TESTING COMPOSITE SAMPLES IN SIMULATION SOFTWARE

A key aspect before the actual modeling of composite samples and before their testing in simulation software takes place, is the knowledge of input parameters. These data should be as detailed and accurate as possible. During a later result evaluation it is useful to be aware of parameters that are not directly necessary for a specific calculation. While some data, such as tensile modulus, Poisson's number, specific density, etc., are provided by the material manufacturers themselves or can be drawn from other sources, several important input parameters must be calculated. Above all, it is necessary to know the material composition and dimensions of the sample that is to be subjected to the simulated load test. This study will be based on the values given in the previous project mentioned above as the source of samples.

The employed values of the sample dimensions could be verified by measuring as the tested samples are at this study's disposal. The problem is, however, that all samples were

subjected to tensile, possibly bending, test, during which they were irreversibly damaged. This makes it impossible for this research to accurately measure several dimensions. So, the former should rely on the data gathered from previous measurements. Recalculation of the thickness of individual layers is essential. The obtained thickness values will be utilized for modeling in the Abaqus program.

The type of matrix used, the share of its components during production, and their specific density, which is stated by the producer, are already known. The volume of the matrix that was used for individual layers is calculated based on specific density and weight. The thickness of the layer is determined by subsequently dividing the volume of the matrix by the area on which it was applied. The change in the density of the matrix after hardening must be considered. However, after rounding, it will not have a significant effect on the result. It is also necessary to consider the absorbency of the matrix reinforcement and the factor that a significant part of the applied matrix will fill the free space in the fabric to which it is applied. This phenomenon is most noticeable in sparsely woven reinforcements, such as the ones used in this study. The filling of the free space of the reinforcement in the case of metal grids or honeycombs is even more clearly observed. To calculate the thickness of the matrix, the thickness of the reinforcement is added, which significantly affects the result, especially when using metal and other grids with a relatively large diameter of the grid wire. The entire proposed formula for estimating the thickness of the layers would possibly be:

$$t_i = \frac{m_i \cdot k_p \cdot k_i}{\rho_i \cdot S_i} + t_{\text{reinforcement}, i} \text{ [mm]} \quad (1)$$

where m is the weight of the applied matrix [g], ρ is the specific density of the matrix obtained by calculation from the known specific densities of the individual components [kg/m^3], S is the area on which the matrix was applied [m^2], and k_p and k_i are the coefficients for the change in density after curing and for the saturation of the fabric. This formula will be applied when calculating layers containing fabrics. The thickness of the layers of metal grids and honeycombs is calculated from the known total thickness of the given sample by subtracting the calculated thicknesses of the other layers from the measured thickness of the sample.

For the calculations, the necessary input data are drawn from the datasheets of sellers and producers of materials, as well as the literature. After rounding, the following values were acquired:

- Thickness of the layer of saturated carbon fabric: 0.15 mm.
- Thickness of the layer of saturated glass fabric: 0.13 mm.
- Thickness of the saturated hybrid fabric layer: 0.15 mm.
- Thickness of the layer of saturated aramid honeycomb: 0.65 mm.
- Thickness of the layer of steel grid D and the matrix: 0.45 mm.
- Thickness of the layer of the steel grid E and the matrix: 1.02 mm.

All samples were produced by symmetrical layering and consist of seven layers. From both outer sides, the sample starts with two layers of carbon fabric, followed by a layer of glass fabric. This configuration is the same for all samples produced. At the core of the sample there is a specific layer that is different for each sample. Thus, the thickness of the individual tested samples varies only depending on the specific layer used in the core. The correctness of the previous considerations and calculations will be confirmed through experimental verification of the actual thickness of the tested samples.

IV. BASIC INPUT PARAMETERS OF THE MECHANICAL PROPERTIES OF THE MATERIALS USED

For correct simulation and to obtain accurate results corresponding to reality, it is crucial to know the material characteristics of flexibility, strength, and stiffness. In this study, for the simulation of the tensile test, the value of the modulus of elasticity in tension (Young's modulus), denoted as E [MPa] was utilized. Since composite materials are anisotropic, their behaviour depends on the direction in which the load is applied, the values of E for all 3 directions (in x , y , z axis) need to be known. These values are usually referred to as $E1$, $E2$, $E3$. The manufacturer of materials (Havel Composites) used to produce the examined samples without giving these parameters or used to provide only the parameters for unsaturated carbon fibers tested separately. However, in the official documentation for the materials, it is stated that the fabrics employed are produced in a license-certified manner. The original patent is owned by the company *Toray*, which lists the required values of mechanical characteristics. Materials should be produced from the same raw materials and with the same technology. So it will be assumed that the products produced under licenses do not differ qualitatively from the original ones. In case the manufacturer *Toray* indicates the parameters with a range of values, a counting with a lower value will be performed to compensate for a possible deficit in the quality of licensed fabrics.

Another important parameter is the Poisson's number μ [-] and the modulus of elasticity in shear G [MPa]. As with E there are differences for individual directions. So values $\mu1$, $\mu2$, $\mu3$, $G1$, $G2$, and $G3$ will be needed. These values will be also drawn from the manufacturer. For a complex simulation and evaluation of the tensile test, additional parameters will be needed, which describe the behavior of the materials in the area of plastic deformations, above the yield point, at stresses close to the maximum bearing capacity of the sample. However, the available sources do not offer enough reliable information on materials used specifically in this study's case. Therefore, the focus of attention will be on assessing the area of elastic deformations within the tensile test in the simulation software.

Values E_i , μ_i , G_i for the honeycomb material, were additionally drawn from previous independent experiments, since the manufacturer did not provide all the necessary parameters. Input data for steel grids of given dimensions similarly came from experimental tests carried out by independent institutions. In this case, the situation was easier, as steel exhibits isotropic behavior. However, this does not apply perfectly, and isotropy depends on the production

technology. For this study's purposes, however, it is sufficient to conceptually idealize the material used.

Parameters of the modulus of elasticity in tension, or pressure of anisotropic materials for different directions can be determined not only experimentally, but also through theoretical calculation from the available data. Computer programs designed for this purpose can facilitate the calculation procedure. Through theoretical calculation, the assumed properties of the fabric layer saturated with the matrix can also be determined, as long as the values E_i , μ_i , and G_i for the individual layer components are known. This means E_i , μ_i , and G_i reinforcements for all the necessary directions and values of the matrix, which mostly exhibits isotropic behavior.

V. MODELING IN ABAQUS

The samples were modeled according to the template in the form of the tested samples. Several approaches were offered. The given samples can be modeled as flat shells, to which the thickness of the individual layers will be assigned. Another option is to model a solid 3D body with a specified thickness that corresponds to the theoretically calculated thickness of the sample in question. In load simulations, the second variant is more effective, as it more realistically displays the modeled body. A sample sketch is initially created, to which the function *extrusion* later assigns the required thickness.

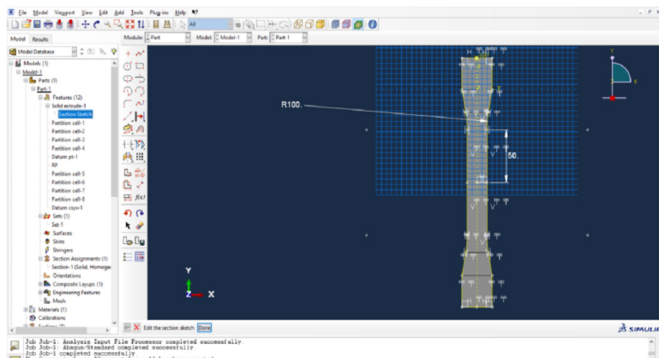


Fig. 8. Creating a sketch of a general sample (the dimensions in the image may not correspond to the final dimensions in the simulation).

Next, the model will be divided into cells, which will allow for a better simulation of a real test. Individual materials are then created. Since in most cases these are anisotropic materials, the modulus of elasticity and Poisson's number must be stated separately for each direction as the so-called engineering constants ($E1$, $E2$, $E3$, $\nu12$, $\nu13$, $\nu23$, $G1$, $G2$, $G3$). All the necessary materials that were considered in the samples will be created. Input parameters from producers' datasheets and other available sources are used, whereas in some cases, values obtained by previous calculations are employed. If the acquired input data represented an interval of values, the lowest values were chosen. The tool *Composite Layup* was utilized to stack the individual layers on top of each other, assigning them a material, and determining their relative thickness within the sample and other parameters.

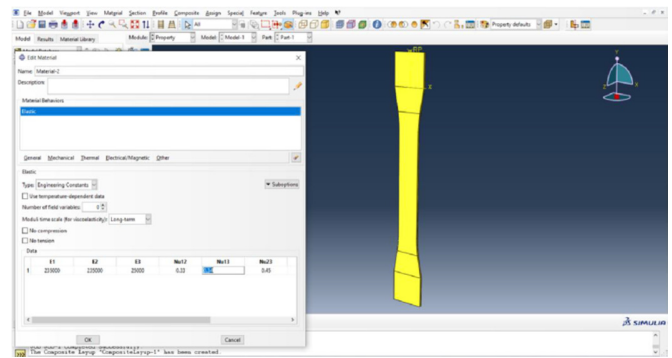


Fig. 9. Creating a material with anisotropic properties.

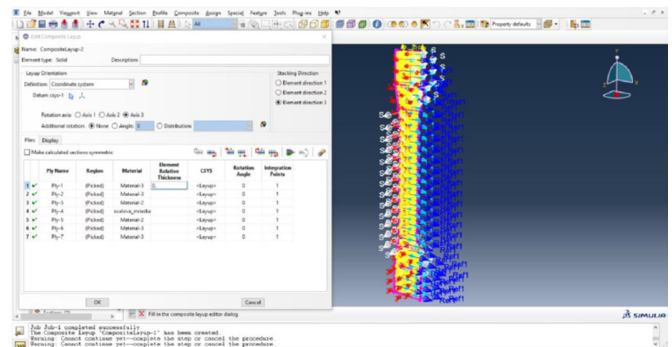


Fig. 10. Creating individual layers in Abaqus.

Further operations followed in the Abaqus interface. Their aim was to make the simulation as close as possible to the conditions to which the sample was exposed during a real test in the laboratory. A single force was then created, placed at the reference point, and was assigned the necessary magnitude and direction. It is necessary to ensure that the reference point is connected to the modeled sample. This will be realized with the assistance of a tool named *Coupling*. In the *Load* module. The weaving of the sample was also simulated into the jaws of the tearing device. The key to the finite element method is to create the sample mesh before running the calculations and simulate loading. The student version of Abaqus limits the creation to a maximum of 1000 nodes.

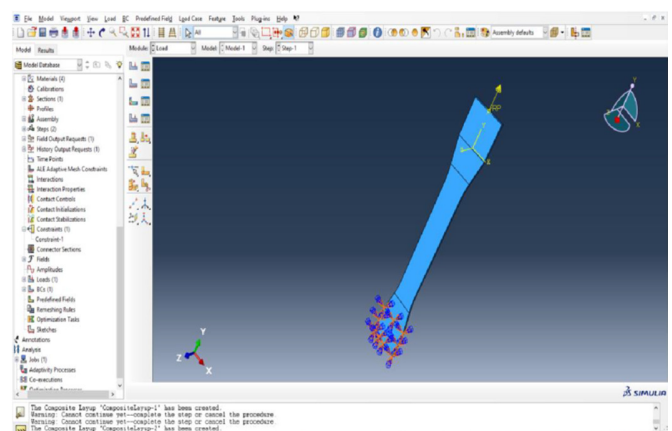


Fig. 11. Loading of the sample with the required force and weaving it.

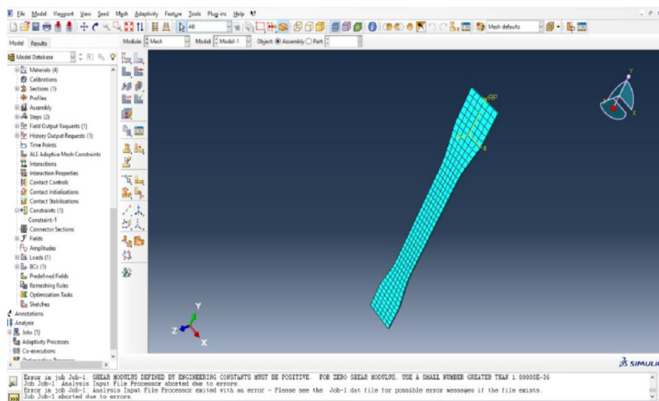


Fig. 12. Created mesh in the Mesh module in the Abaqus interface.

VI. TENSILE LOAD SIMULATION AND PROGRESS EVALUATION

A correctly modeled and loaded body will be subjected to a simulated test in the Job module. During the evaluation of the results, several inconsistencies and errors emerged. In Abaqus, it is not a problem to go back to previous incorrectly entered data entries and correct them. The same applies to changes in the sample type, when the element was not necessary to be modeled anew, but changing some previous data in the program or make other minor alterations or corrections was enough. From the simulation in Abaqus, based on calculations using the finite element method, values that describe the mechanical properties of the bodies and other important data can be acquired.

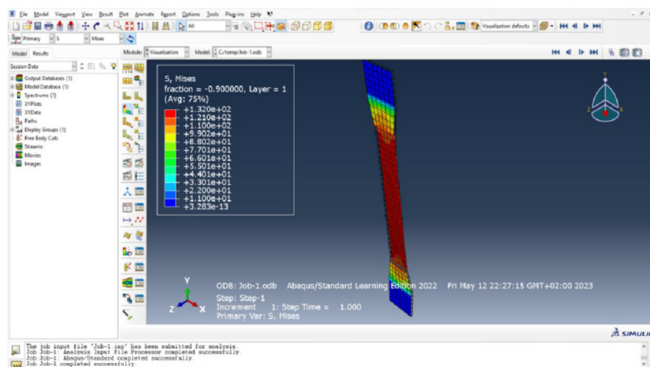


Fig. 13. Visualization of the loaded sample looks in Abaqus.

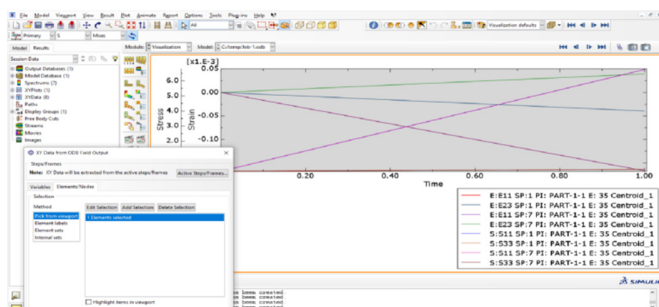


Fig. 14. Plotting the required data in Abaqus.

An approach was chosen in which a universal sample was modeled. For each specific case, a middle layer material was assigned to it. The parameters of the other layers did not change and their relative thickness was only adjusted depending on the sample. A gradual increase of the tensile force was performed manually and the values for comparison with real tests were read. The Step module can make this work easier. It allows to set steps to the required amount of force. In the resulting tables, a force of 0.5, 1.0, 1.5, 2.0, 3.0, 4.0, 5.0, and 6.0 kN was selected for each type of sample.

VII. RESULT COMPARISON

Regarding the tensile tests of the samples in question, during the tests, the tensile force was increased until the sample was torn. The elongation of the sample and the force at which the failure occurs were evaluated. Subsequently, other parameters were calculated (R_m , maximum tension, etc.). For the needs of this paper, the value of sample elongation ΔL will be compared. The elongation caused by a force ranging between 0.5 kN and 6 kN will be investigated. This investigation will be conducted due to the absence or questionable accuracy of the available information that would allow for a reliable simulation of the sample behavior just before failure and in the area of plastic deformation. The effect of the force on the sample is not continuously assessed, but for better clarity 8 values of loading force were selected and listed for the individually tested samples.

A. Type A Sample Test (Glass Fiber)

In the case of type A samples, whose middle layer was made of glass fiber fabric laid at an angle of 45° highly uniform results were generally achieved. The first two test samples differed only minimally, whereas the third showed a smaller elongation of a maximum of 0.02 mm, depending on the applied force. Compared with the tensile test simulation in the Abaqus program, the results of the real tested samples are very similar in some cases, for example, under a load of 2.0 kN, the third test sample (A-3) exhibits the same elongation value as calculated by the software.

TABLE I. COMPARISON OF ΔL WITH RESPECT TO F FOR TYPE A SAMPLE

Sample	Elongation ΔL [mm] of sample under force [kN]								Violation at force [N]
	0.5	1.0	1.5	2.0	3.0	4.0	5.0	6.0	
A-1	0.11	0.15	0.20	0.25	0.34	0.43	0.53	0.62	7182
A-2	0.12	0.17	0.22	0.26	0.35	0.44	0.53	0.62	7451
A-3	0.10	0.15	0.19	0.24	0.33	0.42	0.52	0.60	8352
A-Simul.	0.09	0.14	0.19	0.24	0.32	0.41	0.50	0.58	...

Accurate data collected from several sources concerning the matrix, the glass, and carbon fibers were used in the production of the type A sample. Thanks to this fact, it was possible to achieve a relatively accurate simulation of the tested sample. Small deviations from the theoretical model also indicate relatively precise work in the production of the sample, despite the limiting conditions.

B. Test Sample Type B (Aramid Honeycomb)

The first of the tested samples of type B displayed from the beginning the lowest elongation values compared to the other two. Nevertheless, at an applied force of 4531 N, it ruptured prematurely (unexpectedly). Due to the specific situation, this case will not be assessed and compared with the other tested samples, nor with the results achieved in the simulation software. The second and third tested samples of type B illustrate, as in the case of A samples, very similar values, approximately up to the magnitude of the applied force of 6 kN. Compared to the theoretical simulated model, the measured elongation values are higher. The difference is 0.05 mm on average and approximately increases linearly with increasing load.

TABLE II. COMPARISON OF ΔL WITH RESPECT TO F FOR TYPE B SAMPLE

Sample	Elongation ΔL [mm] of sample under force [kN]								Violation at force [N]
	0.5	1.0	1.5	2.0	3.0	4.0	5.0	6.0	
B-1	0.09	0.15	0.20	0.24	0.33	0.43	-	-	4531
B-2	0.12	0.16	0.20	0.25	0.34	0.43	0.53	0.61	6531
B-3	0.11	0.16	0.20	0.25	0.34	0.44	0.53	0.62	7047
B-Simul.	0.07	0.11	0.15	0.19	0.28	0.36	0.45	0.54	...

C. Type C Sample Test (Hybrid Fabric)

Samples marked as type C have a hybrid fabric of different types of fibers in their core. Compared to the carbon and glass fabrics used in the examined samples, this hybrid aramid layer has lower values of the quantities describing the strength and stiffness of the material. At the same time, however, the fabric is thicker (more mass per unit area), thanks to which the sample withstands greater forces. This was confirmed in the real tests, as it was necessary to develop the greatest force for rupture of all cases, but due to the larger cross-sectional area, thinner samples of type A and B were able to withstand higher tension. Very similar results were achieved for all three tested samples. Abaqus calculated lower elongation values than those found in the real tests. The difference was initially 0.04 mm and with increasing force it increased to 0.08 mm. After loading with higher forces, i.e. when approaching the area of plastic deformation, the difference in real and simulated elongation decreases, but due to missing input data, these values were no longer evaluated.

TABLE III. COMPARISON OF ΔL WITH RESPECT TO F FOR SAMPLE TYPE C

Sample	Elongation ΔL [mm] of sample under force [kN]								Violation at force [N]
	0.5	1.0	1.5	2.0	3.0	4.0	5.0	6.0	
C-1	0.13	0.17	0.21	0.25	0.32	0.40	0.48	0.55	8598
C-2	0.13	0.17	0.20	0.25	0.33	0.40	0.49	0.56	8312
C-3	0.12	0.16	0.20	0.24	0.32	0.40	0.48	0.55	8844
C-Simul.	0.08	0.12	0.15	0.19	0.27	0.34	0.42	0.51	...

Similarly to the aramid honeycomb samples, the C-type samples have tensile elongation values in the simulation software that differ from the values obtained during the real test. The results could be influenced by the production

technology. The materials vendor recommends making composites in a way that requires elevated temperature. However, the preparation of the samples for testing was carried out at room temperature, which could have resulted in worse final characteristics than expected.

D. Type D Sample Test (Metal Grid 2 x 2 mm)

D-type patterns are significantly different from the previous ones in terms of structure. They contain a steel grid in the core, with a grid size of 2 x 2 mm. Samples of type A, B, and C were completely non-metallic, which creates a more striking difference considering the material. The first and third tested samples achieved similar values of elongation under different loads, and were also broken at a similarly high applied force. The second tested (D-2) sample demonstrated a notably higher elongation even from small forces. However, it did not require less force to tear it apart than in the other two. Therefore, an error in the tensile test can be assumed, specifically in the measurement with a strain gauge. For this reason, the simulation results will not be compared with the second tested sample of type D.

TABLE IV. COMPARISON OF ΔL WITH RESPECT TO F FOR TYPE D SAMPLE

Sample	Elongation ΔL [mm] of sample under force [kN]								Violation at force [N]
	0.5	1.0	1.5	2.0	3.0	4.0	5.0	6.0	
D-1	0.10	0.14	0.19	0.24	0.33	0.41	0.50	0.59	7955
D-2	0.24	0.29	0.33	0.38	0.46	0.55	0.64	0.73	7844
D-3	0.12	0.16	0.25	0.25	0.34	0.43	0.51	0.60	8034
D-Simul.	0.10	0.14	0.18	0.21	0.28	0.35	0.42	0.46	...

At smaller loads (up to 2 kN), the difference in the elongation value obtained by simulation and real tests is minimal and comparable to the differences observed with the previous types of samples. From a loading force of 3 kN, the difference remarkably increases and at a force of 6 kN it is almost 0.2 mm. At higher forces, outside the considered range, the difference grew even more. The reason for the result variation is likely to be incorrect input data, the cause of which is a lack of information about the metal grid used. The modeling of a specific grid in the simulation software was not performed, but values from the available literature were utilized for the same grid size as the one employed in production. Their mechanical properties probably differ due to the steel used or the production technology.

E. Type E Sample Test (10 x 10 mm Metal Grid)

The last type of sample tested is similar to type D, with the difference in the size of the grid employed. In real tensile tests, uniform results were achieved without significant deviations. Equivalently to the previous case of D samples, a significant difference compared to the simulations in the Abaqus software was observed. The differences in elongation increase again with increasing force and do not correspond to realistically achieved results at high loads.

As with the load simulations of type D samples, there were no further information available about the implemented steel grid. An emphasis was placed on the experimentally determined parameters of the mechanical properties of the

metal grid with the same dimensions emerging from the available literature. To achieve more accurate results, it would be necessary to know the exact properties of the grid used in the production of the sample and to model the entire structure in the simulation software.

TABLE V. COMPARISON OF ΔL WITH RESPECT TO F FOR TYPE E SAMPLE

Sample	Elongation Δl [mm] of sample under force [kN]								Violation at force [N]
	0.5	1.0	1.5	2.0	3.0	4.0	5.0	6.0	
E-1	0.09	0.13	0.17	0.22	0.29	0.38	0.47	0.56	6538
E-2	0.10	0.14	0.18	0.23	0.31	0.38	0.48	0.57	7745
E-3	0.09	0.15	0.18	0.23	0.30	0.38	0.47	0.56	8038
E-simul.	0.07	0.10	0.13	0.17	0.23	0.29	0.35	0.42	...

VIII. DISCUSSION

At low applied forces, the sample elongation values during the real and the simulated tests are very similar. In some cases, the theoretical value even coincided with the measured one. As the force increased, the difference between the simulation and the real test increased. This could have been caused by the deviation of the materials' input parameters from reality. Type D and E samples with a metal grid constituted an exception. For those, the values obtained by simulation crucially differed from the measured ones, mainly due to the incorrect values of the modulus of elasticity or the Poisson's number, which we used due to the lack of information about the implemented grids. Table VI presents a comparison of the elongation differences between the samples. The ΔL reference value of the physical tests of each sample for comparison with the simulation results was determined as the average value calculated from the aforementioned tables of the test results.

TABLE VI. COMPARISON OF THE ΔL DIFFERENCE BETWEEN REAL TESTS AND SIMULATION

Sample		Elongation Δl [mm] of sample under force [kN]							
		0.5	1.0	1.5	2.0	3.0	4.0	5.0	6.0
A - glass fiber	Real test	0.11	0.15	0.20	0.25	0.34	0.43	0.53	0.61
	Simulation	0.09	0.14	0.19	0.24	0.32	0.41	0.50	0.58
	Difference	0.02	0.01	0.01	0.01	0.02	0.03	0.03	0.03
B - aramid honeycomb	Real test	0.11	0.16	0.20	0.25	0.34	0.43	0.53	0.61
	Simulation	0.07	0.11	0.15	0.19	0.28	0.36	0.45	0.54
	Difference	0.04	0.05	0.05	0.06	0.06	0.07	0.08	0.07
C - hybrid aramid-carbon	Real test	0.13	0.17	0.20	0.25	0.32	0.40	0.48	0.55
	Simulation	0.08	0.12	0.15	0.19	0.27	0.34	0.42	0.51
	Difference	0.05	0.05	0.05	0.06	0.05	0.06	0.06	0.04
D - metal grid 2x2 mm	Real test	0.11	0.15	0.23	0.24	0.34	0.42	0.50	0.59
	Simulation	0.10	0.14	0.18	0.21	0.28	0.35	0.42	0.46
	Difference	0.01	0.01	0.05	0.03	0.06	0.07	0.08	0.13
E - metal grid 10x10 mm	Real test	0.09	0.14	0.17	0.22	0.29	0.38	0.47	0.56
	Simulation	0.07	0.10	0.13	0.17	0.23	0.29	0.35	0.42
	Difference	0.02	0.04	0.04	0.05	0.06	0.09	0.12	0.14

The highest differences in the measured values were noticed for samples Type D and E, which were reinforced with a metal grid. Relatively more accurate findings were measured for Type A samples, where the reinforcement was glass fiber. A number of factors could have caused these differences in the measured values. In [29-30], research has been conducted in the field of integrating composite materials with other material types for selected applications in the automotive industry. The authors note that composite materials in automotive body structures showcase high potential, especially in the context of reducing vehicle weight, fuel consumption, and emissions during the vehicle's lifespan. Considering the fact that despite the significant efforts of automotive manufacturers and the implementation of regulations and standards, emissions production during vehicle operation has not significantly decreased over the past 10 years, the utilization of various composite and lightweight materials in automotive body construction is exceptionally pertinent.

When evaluating the results, it is necessary to take into account the fact that the tested samples were neither professionally produced, nor at a certified workplace. Also, the manufacturers of the components for composites usually recommend production technologies with an increased solidification temperature in the final production phase, which

can affect the resulting mechanical properties of the finished product. In this study, the production process took place at room temperature.

The importance of the difference in the behavior of the examined samples during the physical test and in the simulation software should be also highlighted. In literature, the individual parameters of the materials' mechanical properties were provided as an interval, in particular the values of Young's modulus of elasticity in tension. Each time, for the simulation purposes, the interval's lowest value is chosen. If the middle or highest value within the interval was to be estimated, there would be an even greater deviation from the real results. However, it is also necessary to consider the questionable accuracy of the declared mechanical properties of the materials in question. It is not an exception if the strength and stiffness parameters of the materials stated by the manufacturer are deliberately overestimated, due to the existing competition on the market with carbon or aramid fibres. Therefore, to obtain more accurate results, it is better to start from experimental data collected through independent tests of the given materials.

Another factor that could cause a greater difference in the measured and simulated values is the actual cross-sectional

area of the samples. Theoretically calculated thicknesses based on the existing data were used in the simulation program. The real thickness of the tested sample may be a bit larger, which was confirmed by measuring several randomly selected samples, without, though, further investigating this fact. It is also questionable why this difference with the theoretical value occurred. However, it will probably be a consequence of the limited possibilities in the sample production. If measurements were to be taken on a larger cross-sectional area and at the same time the thickness of the individual layers was to be accordingly increased, the differences between the measurement and the simulated results would be slightly augmented.

The main objective of the current study was to compare selected properties of hybrid composite materials and their simulations. These comparisons revealed that the results of the simulations and the real tests yielded different outcomes for various combinations. These outcomes were on the one hand influenced by the accuracy of the technological procedure applied to individual samples and on the other hand by the precise specification of material properties entered into the simulation program for each simulation. The research in this area builds upon previous work, which focused on the physical testing of composite material samples under tension and flexure. Further research in this direction needs to focus on refining sample preparation and accurately identifying the properties of individual materials. Additionally, simulation testing of similar samples under flexural stress is planned.

IX. CONCLUSION

Utilizing simulation software is a fast, relatively cheap, and effective way of evaluating material mechanical properties. It enables testing of several samples in a short time as well as dynamic change of material input parameters, sample dimensions, or other data. Unlike real tests, simulations do not require expensive equipment or special premises. In the same way, potential problems in the preparation of the tested samples or the actual test implementation are eliminated. An indisputable advantage in the case of computer simulations is the absence of safety risks associated with the stress tests. The disadvantage may be the questionable reliability of the results and the need for a large amount of high-quality input data to achieve satisfactory accuracy of the findings. In both cases, the human factor affects the tests' correct execution. That is, there may be errors in the production of test samples and subsequent testing, whereas numerical errors may emerge during data entry or element modeling.

The comparison of the selected properties of hybrid composite materials obtained from the real tests with the results obtained by simulations is crucial for ensuring the accuracy and reliability of the computational models. Through the validation of simulation models, cost and time-efficient analysis, enhanced understanding of material behavior, and optimization of design parameters, researchers can leverage both experimental and computational approaches to advance the development of hybrid composite materials for a wide range of engineering applications. Continued interdisciplinary research efforts are essential to further improve the accuracy and

reliability of simulation-based predictions, driving innovation in the field of composite material science and engineering.

In this paper, it was possible to achieve simulation results, which largely coincided with the physical tests in the laboratory. At the same time, however, it is evident that the calculations of the bearing capacity of materials using the finite element method are very sensitive to the accuracy of the input data. At first glance, small and negligible deviations of the boundary conditions and parameters at the input can lead to a significant deviation of the obtained results from reality. This factor can cause inaccuracies, especially in the case of anisotropic materials, such as the subject hybrid composites, since a larger number of input parameters are needed to simulate their behaviour. This gives the opportunity for more deviations to occur. Nevertheless, it is necessary to point out the prudent approach selected when choosing the input parameters. If less discrete values were to be selected from the intervals and the ideal course of physical tests was to be counted, a greater difference would arise between the results. The availability of the necessary input data and their plausibility are also key factors. As it turned out in the practical part of the work, the insufficient quality and quantity of necessary data leads to inaccuracies, which are most visible especially with larger loads.

For common practice in automobile production, or in general in mechanical engineering as well as other engineering fields, it seems to be the most advantageous to combine both approaches in a balanced way. It is necessary to assess when a real test of the mechanical properties of materials is appropriate, and when a simulated test is more suitable, which, however, carries the risk of inaccuracy of the output data. A correct assessment of this issue could lead to significant cost optimization of both the development and production processes.

ACKNOWLEDGMENT

This work has been supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic (Project KEGA 030EU- 4/2022, KEGA 020TUK-4/2023, VEGA 1/0064/23 and 1/0238/23, KEGA 019TUK-4/2022 and KEGA 003TUK-4/2024).

REFERENCES

- [1] N. Ahmad and F. Ahmad, "The Effect of Iron Content on the Thermodynamic Properties of Syndiotactic Polypropylene/Iron Composites," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11785–11788, Oct. 2023, <https://doi.org/10.48084/etasr.6241>.
- [2] L. Tesfaye Jule, K. Ramaswamy, N. Nagaprasad, V. Shanmugam, and V. Vignesh, "Design and analysis of serial drilled hole in composite material," *Materials Today: Proceedings*, vol. 45, pp. 5759–5763, Jan. 2021, <https://doi.org/10.1016/j.matpr.2021.02.587>.
- [3] A. E. Magri, K. El Mabrouk, S. Vaudreuil, and M. E. Touhami, "Mechanical properties of CF-reinforced PLA parts manufactured by fused deposition modeling," *Journal of Thermoplastic Composite Materials*, vol. 34, no. 5, pp. 581–595, May 2021, <https://doi.org/10.1177/0892705719847244>.
- [4] B. Wang, S. Zhong, T.-L. Lee, K. S. Fancey, and J. Mi, "Non-destructive testing and evaluation of composite materials/structures: A state-of-the-art review," *Advances in Mechanical Engineering*, vol. 12, no. 4, Apr.

- 2020, Art. no. 1687814020913761, <https://doi.org/10.1177/1687814020913761>.
- [5] P. Zhang, Y. Feng, T. Q. Bui, X. Hu, and W. Yao, "Modelling distinct failure mechanisms in composite materials by a combined phase field method," *Composite Structures*, vol. 232, Jan. 2020, Art. no. 111551, <https://doi.org/10.1016/j.compstruct.2019.111551>.
- [6] D. Gay, *Composite Materials: Design and Applications*, 4th edition. Boca Raton, FL, USA: CRC Press, 2022.
- [7] P. Kumar and A. Kumar, "Bending Analysis of Steel-Concrete Composite Beams with Porosity," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11230–11234, Aug. 2023, <https://doi.org/10.48084/etasr.6050>.
- [8] P. K. Penumakala, J. Santo, and A. Thomas, "A critical review on the fused deposition modeling of thermoplastic polymer composites," *Composites Part B: Engineering*, vol. 201, Nov. 2020, Art. no. 108336, <https://doi.org/10.1016/j.compositesb.2020.108336>.
- [9] F. Pierron and M. Grédiac, "Towards Material Testing 2.0. A review of test design for identification of constitutive parameters from full-field measurements," *Strain*, vol. 57, no. 1, 2021, Art. no. e12370, <https://doi.org/10.1111/str.12370>.
- [10] S. Yuan, S. Li, J. Zhu, and Y. Tang, "Additive manufacturing of polymeric composites from material processing to structural design," *Composites Part B: Engineering*, vol. 219, Aug. 2021, Art. no. 108903, <https://doi.org/10.1016/j.compositesb.2021.108903>.
- [11] G. D. Goh, Y. L. Yap, H. K. J. Tan, S. L. Sing, G. L. Goh, and W. Y. Yeong, "Process-Structure-Properties in Polymer Additive Manufacturing via Material Extrusion: A Review," *Critical Reviews in Solid State and Materials Sciences*, vol. 45, no. 2, pp. 113–133, Mar. 2020, <https://doi.org/10.1080/10408436.2018.1549977>.
- [12] T. H. Ibrahim, I. A. S. Alshaarabaf, A. A. Allawi, N. K. Oukaili, A. El-Zohairy, and A. I. Said, "Theoretical Analysis of Composite RC Beams with Pultruded GFRP Beams subjected to Impact Loading," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12097–12107, Dec. 2023, <https://doi.org/10.48084/etasr.6424>.
- [13] Y. Ivanova, "Damage Detection in Free-Free Glass Fiber Fabric Composite Beams by measuring Flexural and Longitudinal Vibrations," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10685–10690, Jun. 2023, <https://doi.org/10.48084/etasr.5788>.
- [14] C. Han *et al.*, "Microstructure and mechanical properties of (TiB+TiC)/Ti composites fabricated *in situ* via selective laser melting of Ti and B4C powders," *Additive Manufacturing*, vol. 36, Dec. 2020, Art. no. 101466, <https://doi.org/10.1016/j.addma.2020.101466>.
- [15] N. Naveed, "Investigating the Material Properties and Microstructural Changes of Fused Filament Fabricated PLA and Tough-PLA Parts," *Polymers*, vol. 13, no. 9, Jan. 2021, Art. no. 1487, <https://doi.org/10.3390/polym13091487>.
- [16] V. Mahesh, S. Joladarashi, and S. M. Kulkarni, "A comprehensive review on material selection for polymer matrix composites subjected to impact load," *Defence Technology*, vol. 17, no. 1, pp. 257–277, Feb. 2021, <https://doi.org/10.1016/j.dt.2020.04.002>.
- [17] C. Hu and Q.-H. Qin, "Advances in fused deposition modeling of discontinuous fiber/polymer composites," *Current Opinion in Solid State and Materials Science*, vol. 24, no. 5, Oct. 2020, Art. no. 100867, <https://doi.org/10.1016/j.cossms.2020.100867>.
- [18] O. Falco, R. L. Avila, B. Tijs, and C. S. Lopes, "Modelling and simulation methodology for unidirectional composite laminates in a Virtual Test Lab framework," *Composite Structures*, vol. 190, pp. 137–159, Apr. 2018, <https://doi.org/10.1016/j.compstruct.2018.02.016>.
- [19] Y. Wan and J. Takahashi, "Development of Carbon Fiber-Reinforced Thermoplastics for Mass-Produced Automotive Applications in Japan," *Journal of Composites Science*, vol. 5, no. 3, Mar. 2021, Art. no. 86, <https://doi.org/10.3390/jcs5030086>.
- [20] *The Aluminum Automotive Manual. Applications – Car body – Body structures*. European Aluminium Association, 2013.
- [21] Y. Liu, "Application of Carbon fiber-reinforced plastic composite material in automotive bumper," *IOP Conference Series: Earth and Environmental Science*, vol. 692, no. 2, Nov. 2021, Art. no. 022057, <https://doi.org/10.1088/1755-1315/692/2/022057>.
- [22] T. Kyono, "Life Cycle Assessment of Carbon Fiber-Reinforced Plastic," in *High-Performance and Specialty Fibers: Concepts, Technology and Modern Applications of Man-Made Fibers for the Future*, 1st ed., Tokyo, Japan: Springer, 2016, pp. 355–361.
- [23] L. Gornet and H. Ijaz, "A high-cyclic elastic fatigue damage model for carbon fibre epoxy matrix laminates with different mode mixtures," *Composites Part B: Engineering*, vol. 42, no. 5, pp. 1173–1180, Jul. 2011, <https://doi.org/10.1016/j.compositesb.2011.03.004>.
- [24] N. Bouhfid *et al.*, "Numerical modeling of hybrid composite materials," in *Modelling of Damage Processes in Biocomposites, Fibre-Reinforced Composites and Hybrid Composites*, M. Jawaid, M. Thariq, and N. Saba, Eds. Sawston, UK: Woodhead, 2019, pp. 57–101.
- [25] R. Thandavamoorthy and A. Palanivel, "Testing and Evaluation of Tensile and Impact Strength of Neem/Banyan Fiber-Reinforced Hybrid Composite," *Journal of Testing and Evaluation*, vol. 48, no. 1, pp. 647–655, Jan. 2020, <https://doi.org/10.1520/JTE20180640>.
- [26] R. Beigpour, H. Shokrollahi, and S. M. R. Khalili, "Experimental and numerical analysis of a biodegradable hybrid composite under tensile and three-point bending tests," *Composite Structures*, vol. 273, Oct. 2021, Art. no. 114255, <https://doi.org/10.1016/j.compstruct.2021.114255>.
- [27] T. H. Gonsalves, M. K. Garje Channabasappa, R. Motagondanahalli Rangarasaiah, and S. Joladarashi, "Dynamic characterization of hybrid composite material of rotor-bearing support system," *Mechanics of Advanced Materials and Structures*, vol. 29, no. 16, pp. 2342–2357, Jun. 2022, <https://doi.org/10.1080/15376494.2020.1861667>.
- [28] Y. L. Tee, C. Peng, P. Pille, M. Leary, and P. Tran, "PolyJet 3D Printing of Composite Materials: Experimental and Modelling Approach," *JOM*, vol. 72, no. 3, pp. 1105–1117, Mar. 2020, <https://doi.org/10.1007/s11837-020-04014-w>.
- [29] D. Sabadka and S. Kender, "Unconventional Methods of Joining Composites and Metals in the Context of Weight Reduction of Car Bodies," *Advances in Science and Technology Research Journal*, vol. 17, no. 1, pp. 230–242, 2023, <https://doi.org/10.12913/22998624>.
- [30] E. E. Evin, M. Tomas, A. Mares, D. Sabadka, and S. Kender, "The selection of materials for automotive deformation zones," *Transfer Innovations*, vol. 40, pp. 36–41, 2019.

A Research on the Influence of Flux Air Gaps on Electromagnetic Components of Shunt Reactors

Dang Chi Dung

School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, Vietnam
dung.dangchi@hust.edu.vn

Doan Thanh Bao

Faculty of Engineering and Technology, University of Quy Nhon, Vietnam
doanthanhbao@qnu.edu.vn

Phan Hoai Nam

Faculty of Electrical Engineering, University of Economic and Technical Industries, Vietnam
phnam@uneti.edu.vn

Pham Minh Tu

School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, Vietnam
tu.phamminh@hust.edu.vn

Vuong Dang Quoc

School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, Vietnam
vuong.dangquoc@hust.edu.vn (corresponding author)

Received: 31 December 2023 | Revised: 21 January 2024, 30 January 2024, and 31 January 2024 | Accepted: 3 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6842>

ABSTRACT

This paper introduces an evaluation of the flux air gaps of Shunt Reactors (SRs) to effectively mitigate fringing and leakage fluxes along the height of the iron core. The assessment of these discretely distributed flux air gaps in SRs is a rigorous and challenging process. To define their exact number, the case of one flux air gap is analyzed and investigated to observe/simulate the influence of the flux density distribution and the leakage flux along the air gaps on the reactive power and the operation conditions of the SR. Based on that, to reduce leakage flux, a large flux air gap is divided into smaller ones. Initially, an analytic model is presented to define the main parameters of the SRs. Then, a finite element method is developed to simulate electromagnetic quantities, such as the magnetic flux density, leakage flux, and electromagnetic force. The obtained results can help manufacturers define the exact number of flux air gaps along the iron core of the SR. From that, a suitable technology can be given in manufacturing high voltage SRs applied to high or super high voltage transmission lines.

Keywords-shunt reactors; leakage flux; electromagnetic force; flux air gaps, analytical method; finite element method

I. INTRODUCTION

Shunt Reactors' (SRs) purpose is to absorb voltage fluctuations resulting from variations in load and capacitive load on high-voltage transmission lines. The SRs can be continuously switched on based on the load condition of the power transmission lines. They play a crucial role in mitigating transients entering the circuit, with inrush current being a key

consideration. So far, many applications of SRs in high-voltage power systems have been studied. Authors in [1, 3] applied the Finite Element Method (FEM) to study the properties and phenomena of the SRs used in high voltage systems. In [4-9], the FEM was developed to formulate the Maxwell stress tensor for calculating the electromagnetic parameters of the SR. The relation of leakage and fringing inductances were also established. In [10], the required parameters are defined via the

analytical approach based on the theory of magnetic circuit models. In [11], the arrangement of flux air gaps along the iron core was orchestrated to alleviate the impact of fringing and leakage fluxes around these gaps. However, this research has not shown the exact number of flux air gaps that need to be split from a large flux air gap into smaller ones, something that still constitutes a big challenge for researchers and designers.

In this paper, the analytical computation process is first considered to define the main parameters of the SR. Then, a finite element model is developed to investigate and evaluate the leakage and fringing flux distribution along the flux air gaps and the electromagnetic forces acting directly on the core blocks of the SR. Thus, the exact number of flux air gaps of the SR can be obtained helping researchers and manufactures to choose a suitable technology in designing and manufacturing SRs. The proposed method is applied and validated on the practical problem of 17 MVar, 500kV.

II. BACKGROUND ON SHUNT REACTORS

A. Determination of the Main Dimensions of the SR

In this part, a single phase SR (reactive power of 17 MVar, voltage of $500/\sqrt{3}$, frequency of 50 Hz) with a single and multi-flux air gap is considered. The model of this SR is given in Figure 1.

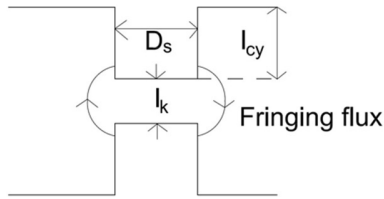


Fig. 1. Model of the flux air gap of the SR.

The electromagnetic force (EMF) and the electric current are provided in [12-14]. The volume of the air gap (V_k) is determined as [6, 12-14]:

$$V_k = A_k \cdot l_k = \frac{Q}{\mu_0 \cdot f \cdot B_m^2} \quad (1)$$

The dimension of core block (D_c) can be defined via the term A_k :

$$D_c = \sqrt{\frac{4 \cdot A_k}{\pi}} \quad (2)$$

The winding inductance is calculated by [3-5]:

$$L = N^2 \cdot \mu_0 \cdot \left(\frac{A_k}{l_k} \right) = \left(\sqrt{\frac{L}{\mu_0 \cdot \left(\frac{A_k}{l_k} \right)}} \right)^2 \cdot \Phi_{leak} \quad (3)$$

where Φ_{leak} is the leakage flux and l_k is the thickness of the air gap.

B. Analysis of the Flux Air Gap of the Shunt Reactor

In this part, the impact of the number of flux air gaps on the iron core SR is analytically considered. Figure 2 shows an SR

with a single flux air gap. From Figure 2, the flux density (B) at the flux air gap is computed by [10, 15]:

$$B = \frac{\mu_0 \cdot F}{(l_k + \pi \cdot r)} \quad (4)$$

where F is the magnetomotive force (MMF) and r is the radius of the leakage flux (Φ_{leak*}).

For a single flux air gap, the Φ_{leak*} can be defined as [6, 10]:

$$\Phi_{leak*=1} = \frac{\int B dS}{F} = \mu_0 \cdot D_s \cdot \ln \left(1 + \frac{\pi \cdot l_s}{2 \cdot l_k} \right) \quad (5)$$

where l_s is the thickness of core block.

For the multi-flux air gap, the $\Phi_{leak*>1}$ is defined as [10]:

$$\begin{aligned} \Phi_{leak*>1} &= \frac{\int B dS}{F} = \int_0^{l_{cy}} \frac{\mu_0 \pi D_s}{((g-1)l_s + g l_k + \pi \cdot r)} d_r \\ &= \frac{\mu_0 \cdot D_c}{\pi} \cdot \ln \left(1 + \frac{\pi \cdot l_s}{(g-1)l_s + g l_k} \right) \end{aligned} \quad (6)$$

where g is the air gap number.

The total magnetic flux (Φ_{total}) takes the fringing flux into account and is expressed as:

$$\Phi_{total} = \frac{\Phi_{g*} + \Phi_{leak*=1}}{g} + \Phi_{leak*>1} \quad (7)$$

Based on the analytic theory developed above, the results of the main parameters of the proposed SR of 17 MVar are displayed in Table I.

TABLE I. ANALYTICAL RESULTS

Parameter	Symbol	Value
Reactive power	Q (MVar)	17
Rated voltage	U (kV)	$500/\sqrt{3}$
Rated current	I (A)	58.89
Total inductance	L_{total} (H)	15.6
Fringing inductance	$L_{fringing}$ (H)	4.9
Gap inductance	L_{gap} (H)	9.6
Core dimension	D_s (mm)	572
Height of core	H_s (mm)	1579
Total air gap	l_k (mm)	264
Turn number	N (turn)	2761
Width of winding	Wd (mm)	218
Height of winding	Hd (mm)	1309
Gaps	g	6

III. FINITE ELEMENT METHOD

A. Maxwell's Equations

The Maxwell's equations and behavior laws are written in Euclidean space \mathbb{R}^3 as [10, 12]:

$$\text{curl } \mathbf{H} = \mathbf{J}_s \quad (11a)$$

$$\text{curl } \mathbf{E} = -j\omega \mathbf{B} \quad (11b)$$

$$\text{div } \mathbf{B} = 0 \quad (11c)$$

$$\mathbf{B} = \mu \mathbf{H} \quad (12a)$$

$$\mathbf{J} = \sigma \mathbf{E} \quad (12b)$$

where \mathbf{H} is the magnetic field (A/m), \mathbf{B} is the magnetic flux density (T), \mathbf{E} is the electric field (V/m), \mathbf{J}_s is the current density (A/m²), \mathbf{J} is the eddy current density (A/m²), and μ and σ are respectively the relative permeability and the electric conductivity (S/m).

Equations (11a) and 11(b) are solved with boundary conditions (BCs) defined on Γ , they are expressed as [11, 12]:

$$\mathbf{n} \times \mathbf{H}|_{\Gamma_h} = 0 \quad (13a)$$

$$\mathbf{n} \cdot \mathbf{B}|_{\Gamma_e} = 0 \quad (13b)$$

where \mathbf{n} is the unit normal exterior to Ω (with $\Omega = \Omega_c \cup \Omega_c^c$). Domain Ω_c is the conducting region and Ω_c^c is the non-conducting regions.

The fields \mathbf{H} , \mathbf{B} , \mathbf{E} , and \mathbf{J} belong to the function spaces $\mathbf{F}_h(\text{curl}; \Omega)$ and $\mathbf{F}_e(\text{div}; \Omega)$ that are satisfied. Tonti's diagram [6] is illustrated in Figure 2, with $\mathbf{H} \in \mathbf{F}_h(\text{curl}; \Omega)$, $\mathbf{E} \in \mathbf{F}_e(\text{curl}; \Omega)$, $\mathbf{J} \in \mathbf{H}(\text{div}; \Omega)$ and $\mathbf{B} \in \mathbf{F}_e(\text{div}; \Omega)$.

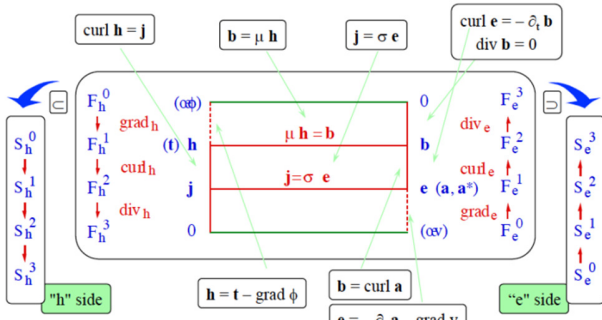


Fig. 2. Tonti's diagram.

B. Weak Formulations of Magnetic Vector Potential

The field \mathbf{B} in (11c) is derived from a vector potential \mathbf{A} :

$$\mathbf{B} = \text{curl } \mathbf{A} \quad (14)$$

From (11b) and (14), the field \mathbf{E} can be defined via the electric scalar potential v :

$$\mathbf{E} = -\partial_t \mathbf{A} - \text{grad } v \quad (15)$$

Based on the Ampere's law (11a), the weak formulation of magnetic vector potential is written as [10, 12]:

$$\begin{aligned} & \frac{1}{\mu} \oint_{\Omega} (\text{curl } \mathbf{A} \cdot \text{curl } \mathbf{t}') d\Omega - \sigma \oint_{\Omega_c} (\partial_t \mathbf{A} \cdot \text{curl } \mathbf{t}') d\Omega_c + \\ & \sigma \oint_{\Omega_c} (\text{grad } v \cdot \text{curl } \mathbf{t}') d\Omega_c + \int_{\Gamma} (\mathbf{n} \times \mathbf{H}) \cdot \mathbf{t}' d\Gamma = \\ & \oint_{\Omega_s} (\mathbf{J}_s \cdot \mathbf{t}') d\Omega_s, \forall \mathbf{t}' \in \mathbf{F}_e^0(\text{curl}, \Omega) \end{aligned} \quad (16)$$

where the function space $\mathbf{F}_e^0(\text{curl}, \Omega)$ defined in Ω contains the basis functions for \mathbf{A} and for the test function \mathbf{t}' .

The inductance value \mathbf{L} is defined via the post processing:

$$\mathbf{L} = \frac{\oint_{\Omega} \mathbf{H} \cdot \mathbf{B} d\Omega}{I^2} \quad (17)$$

IV. NUMERICAL TEST

The tested problem is a single-phase high voltage SR with a power of 17 MVar, voltage of 500/ $\sqrt{3}$ V and frequency of 50 Hz. The main parameters of the SR are already designed in Table I. The model of the single phase SR is depicted in Figure 3. The distribution of magnetic vector potential and magnetic flux density in the iron cored of the SR with a single flux air gap is presented in Figure 4.

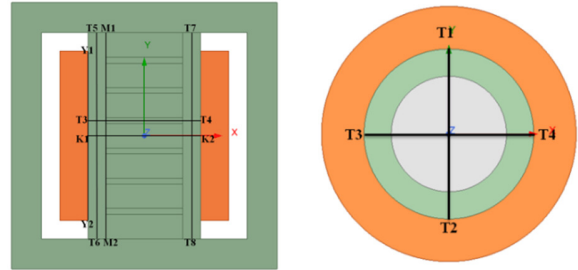


Fig. 3. Modeling of a single phase 16 MVar SR.

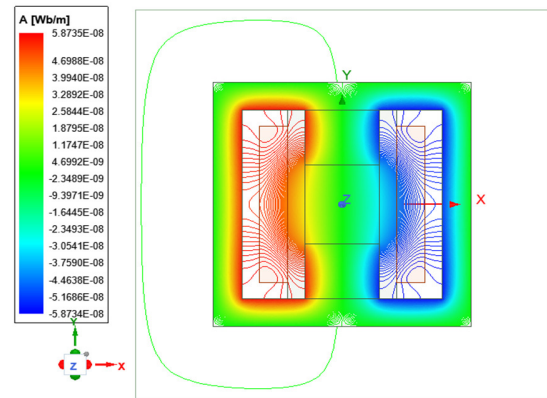


Fig. 4. Distribution of magnetic vector potential.

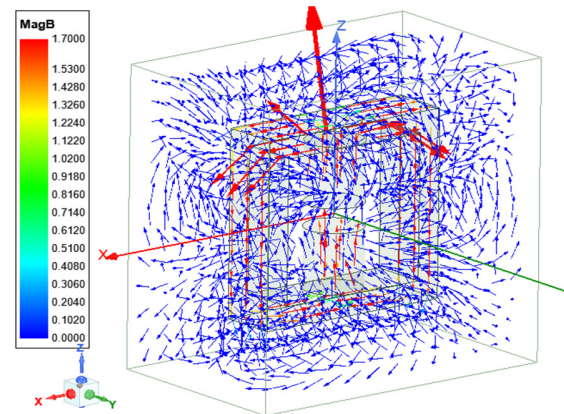


Fig. 5. Distribution of different inductance components.

It can be seen that the fringing flux appears in the air gaps along the core blocks and the height of the winding is very large, increasing the fringing/leakage inductance in the flux air

gaps as portrayed in Figure 5. The maximum value on magnetic flux density $\mathbf{B} = \text{curl} \mathbf{A}$ is 1.7 T. The obtained results indicate that when the circuit material has a high permeability and the air gap is sufficiently small, the magnetic flux is predominantly distributed within the core of the circuit without significant leakage. However, for a single flux air gap, the magnetic flux leakage will be very large, leading to a large leakage inductance. This influences directly the operational performance of the SR. In order to reduce the distribution of magnetic flux density ($\mathbf{B} = \text{curl} \mathbf{A}$) and leakage/fringing flux along the core block and the height of the winding, a divided technique of the flux air gap is performed. In this study, a single flux air gap is split into six portions. Each portion has magnetic flux density and eletromagnetic force acting directly on each core block.

The distribution of flux density along the core blocks at lines T1T2 and T3T4 with the different flux air gaps is depicted in Figure 6. It can be seen that for one and two air gaps, the flux density reaches 0.8 T in the central region of the core blocks and strongly decreases towards both edges of the core blocks. When the flux air gap is bigger than two gaps, the flux density value is reduced to 0.75 T. Of course, the minimum value at the corner of the core block for the six flux air gaps is also reduced to 0.05 T.

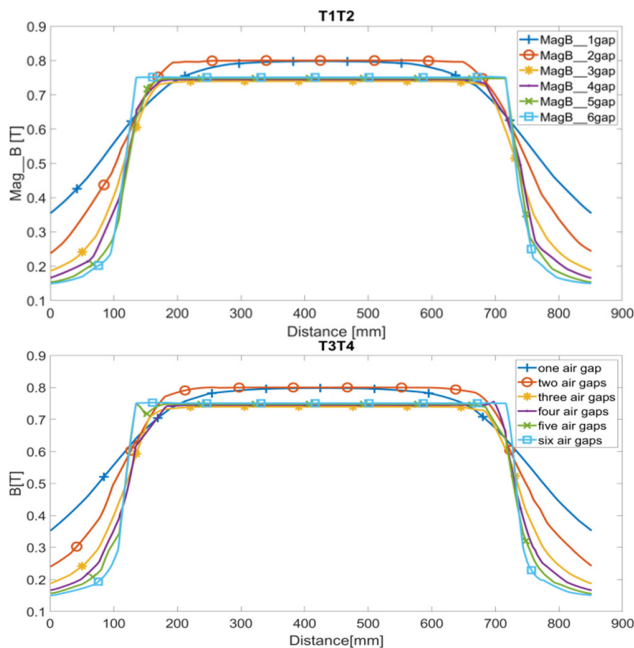


Fig. 6. Flux density in lines T1T2 (top) and T3T4 (bottom) along the core block of the SR.

The distribution of flux density along the core blocks in line T5T6 with different flux air gaps is exhibited in Figure 7. The maximum flux density is 0.55 T for a single flux air gap, and 0.18 T for six flux air gaps, decreased by 70% compared to the case of a single-flux air gap. Figure 8 shows the flux density in the line Y1Y2 along the winding of the SR. It can be observed that the maximum flux density is 0.35 T for a single flux air gap, and 0.14 T for six flux air gaps, decreased by 60%.

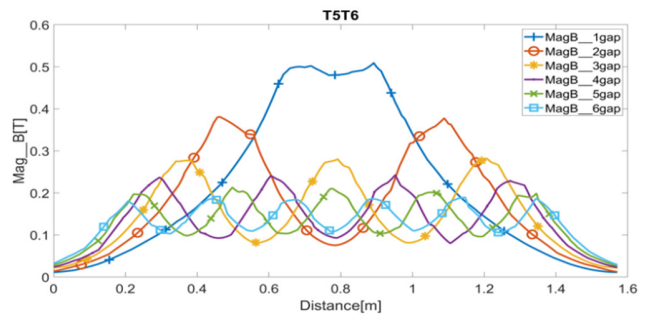


Fig. 7. Flux density in the line of T5T6 along the core block of SR with different flux air gaps.

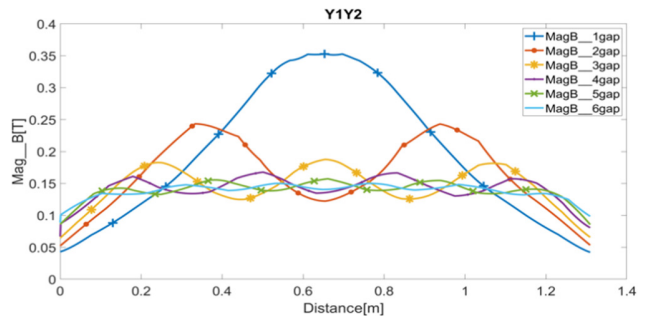


Fig. 8. Flux density in lines Y1Y2 along the core block of the SR with different flux air gaps.

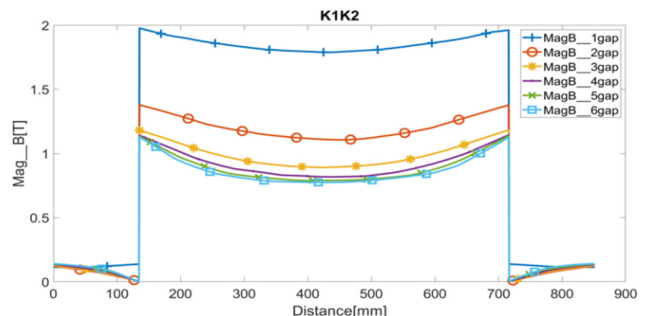


Fig. 9. Flux density in lines K1K2 across the core block of the SR with different flux air gaps.

The flux density in the lines K1K2 across the core block of the SR with different flux air gaps along the winding of the SR is presented in Figure 9. The results indicate that for six flux air gaps, the flux density is equal to 1.1 T at the corner of the core block and 0.8 T at the core middle, while for a single flux air gap, its value is 1.8 T at the corner and 1.75 T at the core middle. This means that it is reduced by 0.6 times compared to the single flux air gap case. In a similar analysis, the distribution of the flux density along the core blocks at the line M1M2 with six air gaps is pointed out in Figure 10. The EMF on the segment K1K2 in the case of six flux air gaps is depicted in Figure 11. It can be seen that for one flux air gap, the force near the corner of the core block is very large and reaches approximately 3.1×10^5 N at the corner. It should be noted that when the number of air gaps increases, the force decreases. Specifically, when the number of air gaps is 5, the value of the force change is very small in comparison with the case of a single flux air gap.

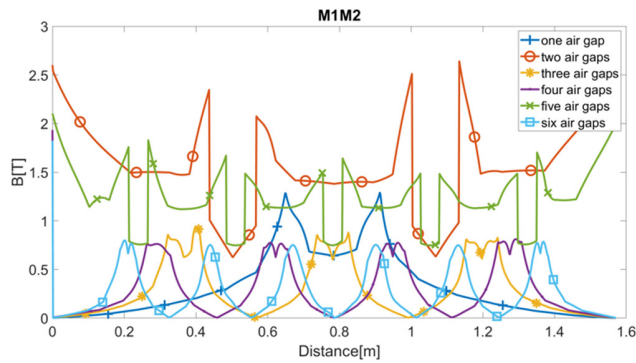


Fig. 10. Flux density in lines M1M2 along the core block of the SR with different flux air gaps.

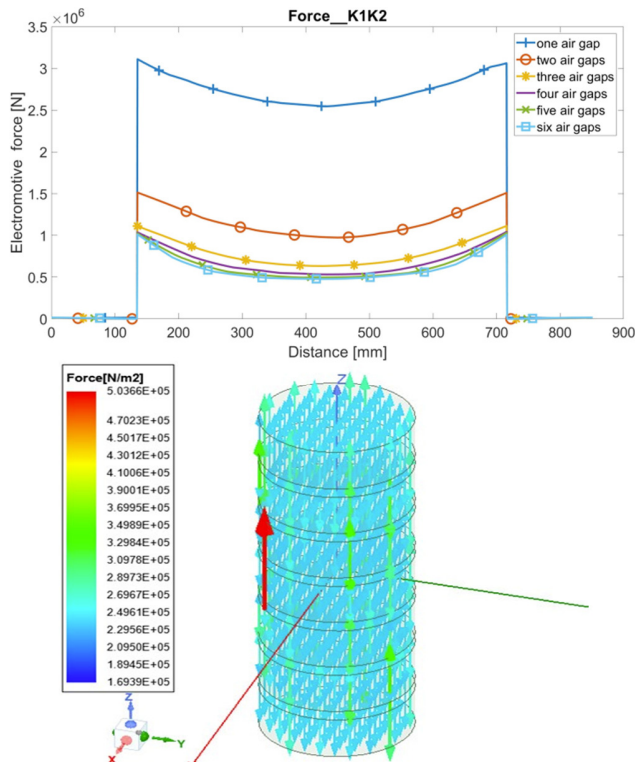


Fig. 11. Electromagnetic force at the line K1K2 acting on the core block with six flux air gaps.

TABLE II. FLUX DENSITY AND INDUCTANCE WITH DIFFERENT FLUX AIR GAPS

Air gap no	B (T)	L_{total} (H)	L_{gap} (H)	$L_{rated} = L_{total}/L_{gap}$
2	0.230	18.44	10.24	1.80
4	0.171	16.06	10.23	1.59
6	0.154	15.32	10.22	1.50
8	0.149	14.93	10.22	1.46
10	0.146	14.77	10.22	1.44
12	0.146	14.60	10.22	1.43
16	0.146	14.86	10.47	1.43
20	0.146	14.44	10.22	1.41
25	0.146	14.38	10.21	1.41
29	0.146	14.35	10.20	1.41

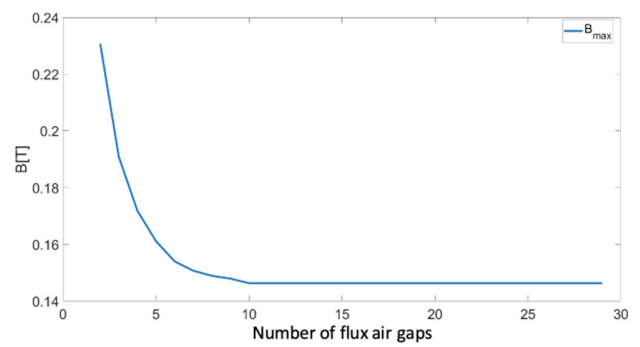


Fig. 12. Relationship between flux density and the number of flux air gaps.

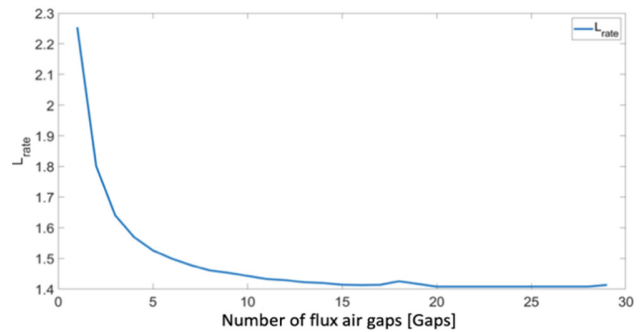


Fig. 13. Relationship between the rated inductance and the flux air gap number.

Based on these values, the suitable number of flux air gaps is proposed for researchers and manufacturers who want to define the reactive power and current during the operation of the SRs.

V. CONCLUSION

The current article has successfully combined the analytical method and finite element modeling to determine the impact of the number of air gaps on flux density, electromagnetic force, and inductance values of the SR. To achieve this, the analytical method was first developed based on the results of previous studies to determine the required parameters of the SR as shown in Table I. Subsequently, a finite element model was proposed to simulate the electromagnetic parameters. The acquired results of the paper allow to:

- Determine the exact number of air gaps based on the total inductance ratio to air gap inductance.
- Analyze the distribution of the flux density from the winding in the air gap, both radially and along the coil winding and the core blocks.
- Investigate the distribution of EMFs acting on the surface of the core blocks corresponding to different numbers of air gaps.
- Establish the values of air gap inductance and explore the relationships between the inductance values and the number of air gaps.

The obtained results may serve as reference data, assisting researchers and manufacturers in selecting suitable flux air gaps and insulation materials for SRs in practice.

ACKNOWLEDGMENT

This research is funded by the Hanoi University of Science and Technology (HUST) under project number T2023-PC-042.

REFERENCES

- [1] W. Chen, H. Chen, T. Lu, and W. Zhou, "Ultra-high voltage controllable shunt reactor," *High Voltage Engineering*, vol. 31, no. 11, pp. 26–27, Nov. 2005.
- [2] W. Xu, S.-M. Chen, and H. Jinliang, "Research on switching overvoltage caused by faults in 1000kV UHV AC transmission line," *Power System Technology-Beijing*, vol. 29, no. 21, Nov. 2005.
- [3] A. Donuk, "Modeling and design of iron-core shunt reactors with discretely distributed air-gaps," Ph.D. dissertation, Middle East Technical University, 2012.
- [4] D. I. Zaikin, S. Jonassen, and S. L. Mikkelsen, "An Air-Gap Shape Optimization for Fringing Field Eddy Current Loss Reductions in Power Magnetics," *IEEE Transactions on Power Electronics*, vol. 34, no. 5, pp. 4079–4086, Feb. 2019, <https://doi.org/10.1109/TPEL.2018.2868289>.
- [5] Q. Y. Zhou, Q. Guo, G. Q. Bu, and L. G. Ban, "Application of controllable reactors in China's power grid at extra and ultra voltage level," *Proceedings of the Chinese Society of Electrical Engineering*, vol. 27, no. 7, pp. 1–6, Mar. 2007.
- [6] H. De Gersem and K. Hameyer, "A finite element model for foil winding simulation," *IEEE Transactions on Magnetics*, vol. 37, no. 5, pp. 3427–3432, Sep. 2001, <https://doi.org/10.1109/20.952629>.
- [7] H. B. Duc, T. P. Minh, T. P. Anh, and V. D. Quoc, "A Novel Approach for the Modeling of Electromagnetic Forces in Air-Gap Shunt Reactors," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 8223–8227, Feb. 2022, <https://doi.org/10.48084/etasr.4692>.
- [8] T. P. Minh, H. B. Duc, and V. D. Quoc, "Analysis of Leakage Inductances in Shunt Reactors: Application to High Voltage Transmission Lines," *Engineering, Technology & Applied Science Research*, vol. 12, no. 3, pp. 8488–8491, Jun. 2022, <https://doi.org/10.48084/etasr.4826>.
- [9] P. M. Tú, V. D. Quoc, and B. Đ. Hùng, "Study of number and size of air-gaps in core of shunt reactors in transmission lines of high and super high voltages," *Journal of Military Science and Technology*, no. 80, pp. 23–30, Jun. 2022, <https://doi.org/10.54939/1859-1043.j.mst.80.2022.23-30>.
- [10] T. P. Minh, "Computation and Simulation of Shunt Reactors by an Analytic Method and Finite Element Method-Application to Transmission Lines of High and Super High Voltages," *Journal of Military Science and Technology*, no. 74, pp. 36–43, Aug. 2021.
- [11] T. P. Minh *et al.*, "Finite Element Modeling of Shunt Reactors Used in High Voltage Power Systems," *Engineering, Technology & Applied Science Research*, vol. 11, no. 4, pp. 7411–7416, Aug. 2021, <https://doi.org/10.48084/etasr.4271>.
- [12] K. Dawood, G. Komurgoz, and F. Isik, "Modeling of Distribution Transformer for Analysis of Core Losses of Different Core Materials Using FEM," in *2019 8th International Conference on Modeling Simulation and Applied Optimization (ICMSAO)*, Manama, Bahrain, Apr. 2019, pp. 1–5, <https://doi.org/10.1109/ICMSAO.2019.8880392>.
- [13] F. Yuan *et al.*, "Thermal Optimization for Dry Type Air Core Reactor Base on FEM," in *2018 21st International Conference on Electrical Machines and Systems (ICEMS)*, Jeju, Korea (South), Jul. 2018, pp. 1726–1730, <https://doi.org/10.23919/ICEMS.2018.8549257>.
- [14] S. Magdaleno-Adame, R. Escarela-Perez, J. C. Olivares-Galvan, E. Campero-Littlewood, and R. Ocon-Valdez, "Temperature Reduction in the Clamping Bolt Zone of Shunt Reactors: Design Enhancements," *IEEE Transactions on Power Delivery*, vol. 29, no. 6, pp. 2648–2655, Sep. 2014, <https://doi.org/10.1109/TPWRD.2014.2322994>.
- [15] A. Najafi and I. Iskender, "Comparison of core loss and magnetic flux distribution in amorphous and silicon steel core transformers," *Electrical Engineering*, vol. 100, no. 2, pp. 1125–1131, Jun. 2018, <https://doi.org/10.1007/s00202-017-0574-7>.

Impact of Safety Management Practices on Safety Performance in Workplace Environment: A Case Study in Iraqi Electricity Production Industry

Omar Munaf Tawfeeq

College of Graduate Studies (COGS), University Tenaga Nasional, Malaysia | Al Iraqia University, Iraq
omarmunaf60@gmail.com (corresponding author)

Sivadass A.L. Thiruchelvam

University Tenaga Nasional, Malaysia
sivadass@uniten.edu.my

Izham Bin Zainal Abidin

University Tenaga Nasional, Malaysia
Izham@uniten.edu.my

Received: 3 February 2024 | Revised: 24 February 2024 | Accepted: 25 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7006>

ABSTRACT

Organizations are becoming more aware of the need to ensure a safe working environment for their staff. Technological advancements and industrial growth have enhanced efficiency, however, they present new challenges and risks for employees. Accidents remain a concern despite International Labor Organization (ILO) guidelines, governmental bodies, and industry institutions promoting workplace safety. Therefore, it is crucial to evaluate the determinants of workplace safety performance, particularly in the electrical power industry. This study formulates a theoretical model to assess the predictors of safety practices of managers and staff in the Iraqi electricity sector, extending the safety climate model with four external constructs and a moderating variable. Data were collected from 374 participants using an online questionnaire and the PLS-SEM method for analysis. The factor loadings exceeded the recommended value of 0.7 and internal consistencies were greater than the threshold value of 0.8. The findings showed that the safety performance in the Iraqi electric power sector is influenced by safety communication, safety policy, safety control, prevention planning, and safety commitment. Safety commitment is affected by safety policy, prevention planning, control, and communication, while safety training and safety control were found to be insignificant. Furthermore, safety communication had the most significant effect. The results of this study provide some theoretical and practical implications for employees' safety performance toward their overall safety in the electric power industry.

Keywords- component; sustainability; safety performance; workplace safety; electric power

I. INTRODUCTION

With increasing industrialization and technological progress, efficiency and productivity have been augmented but workplaces have also seen the emergence of new risks. Organizations recognize the importance of maintaining safe work environments for the well-being of their employees and the overall success of their operations [1]. Despite advances in this area, worker safety remains a pressing concern. Although organizations that prioritize safety enjoy several benefits, many developing countries are trailing behind in terms of safety regulations and practices due to inadequate regulations,

outdated safety measures, and lack of awareness. This study examines the challenges of occupational safety and health, highlighting the pivotal role of policies, regulations, and training in shaping workplace culture and fostering secure behaviors. Safety Management Systems (SMS) represent a proactive approach to safety management to identify, mitigate, and control hazards. SMS is seamlessly integrated into the fabric of an organization, becoming an intrinsic aspect of its culture [2]. This study also explores the concept of Occupational Safety Management (OSM) and underscores the importance of behavioral-based safety management. OSM emphasizes the promotion of safety-conscious behaviors to

alleviate work-related injuries, while behavioral-based safety management underscores the crucial role of behaviors in ensuring workplace security [3]. This study assesses the influence of safety management system practices, including safety policy, safety training, prevention planning, safety control, and safety communication, as well as the impact of safety commitment on safety performance. Furthermore, the moderating role of safety knowledge on the safety performance of workers in the Iraqi electric power industry is also investigated.

II. LITERATURE REVIEW AND RESEARCH HYPOTHESES

The Ministry of Electricity (MOELC) of the Republic of Iraq, established in 2003 after transitioning from the Electricity Authority, oversees policies and electricity supply across the country, including power generation, transmission, and distribution. These responsibilities were previously managed under the Ministry of Industry and Minerals. The ministry comprises three main directorates: production, transmission, and distribution. It also ensures compliance with safety, health, and environmental standards through the electric power company. However, several challenges, such as fuel type selection and financial issues, persist. The Central Statistics Division highlighted a significant number of workplace accidents in the electric power sector between 2018 and 2022 [4]. Annually, the General Company of Production reported 5698 accidents, the General Transmission Company reported 1396, and the General Distribution Company reported 3012 [4]. To address this critical issue, a comprehensive framework, shown in Figure 1, and tools were proposed to evaluate safety performance in the electrical power industry in Iraq. Figure 1 includes safety knowledge as a moderator between safety management systems and safety performance, with safety

commitment mediating this relationship, given its recognized importance in organizational safety [5-8]. This study explores the effect of safety knowledge on the relationships between several variables and safety performance in the Iraqi electricity industries. The framework demonstrates the relationships and interdependence among seven key factors: safety policy, safety training, prevention planning, safety control, safety communication, safety commitment, and safety knowledge. Furthermore, the relationship among the seven factors is examined in the Iraqi context.

A. Safety Policy

Safety policy is an essential framework for determining a company's commitment to ensuring safety and health in the workplace. Various countries have established their own health and security guidelines, and prominent organizations have provided guidelines for safety statements [9]. The importance of these policies is demonstrated by their ability to promote a positive attitude toward safety, which subsequently improves safety performance, particularly in the electrical sector. Compliance with safety policies can increase safety productivity and performance benchmarks [10]. Additionally, the effectiveness of safety performance is not only based on policies but also requires the steadfast commitment of both employers and employees. This underscores the need for organizations to establish their safety standards following legal or regulatory requirements [11]. Given the significance of safety policies in organizational safety performance, robust safety policies are assumed to improve safety performance in the Iraqi electricity industry by setting clear expectations and guidelines. Accordingly, two hypotheses were put forward:

H1: Safety policy positively influences safety performance.

H2: Safety policy positively influences safety commitment.

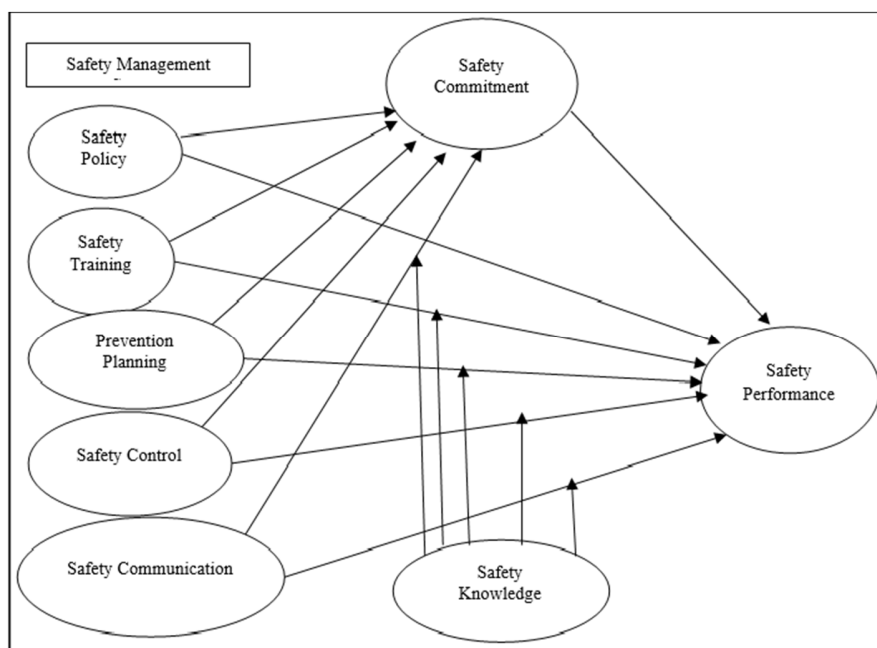


Fig. 1. Conceptual framework.

B. Safety Training

Safety training is a systematic approach to providing individuals with the knowledge, technical skills, and awareness to ensure safety at work, thus reducing the risk of injuries [12]. As many studies have consistently highlighted the relationship between comprehensive training and improved safety outcomes, training is indispensable in mitigating workplace accidents. In the construction industry, previous studies have highlighted the criticality of safety training [13]. Effective training can reduce common construction accidents, such as falls from heights and injuries from falling materials. It is crucial to note that safety training should be an ongoing effort to continuously enhance employees' security. This perspective is reinforced by management's commitment to safety and training in shaping the security climate. Hence, management commitment fosters a culture where employee safety is prioritized and guided by comprehensive training [14]. Therefore, the following two hypotheses were posed:

H3: Safety training positively influences safety performance.

H4: Safety training positively influences safety commitment.

C. Prevention Planning

Prevention planning is essential in fostering a secure work environment and ensuring the successful execution of projects. Historically, safety management was more reactive, addressing issues after they occurred. However, the modern safety approach emphasizes a proactive stance, integrating safety into broader managerial and production processes. This shift recognizes that an organization's safety culture is deeply influenced by its inherent attributes, such as its overall culture and climate. The role of upper management is pivotal in this transformation, as they are tasked with cultivating an environment that prioritizes safety. This involves the establishment of safety-centric policies, systems, and leadership styles. Organizations with a strong emphasis on prevention planning not only experience fewer accidents, but also exhibit a profound commitment to safety. Therefore, the following hypotheses were formulated:

H5: Prevention planning positively influences safety performance.

H6: Prevention planning positively influences safety commitment.

D. Safety Control

Safety control encompasses the implementation of measures, strategies, and procedures to ensure a safe and healthy work environment [5, 15]. It aims to mitigate hazards, making it the basis for a secure and efficient workplace. Beyond safeguarding employees, safety control also aims to manage and contain risks to acceptable levels. This process is multifaceted, encompassing safety equipment, managerial decisions, workforce training, and development. It operates on feedback mechanisms tailored to specific work environments, ensuring the adherence to industry safety standards [15]. Safety control not only assesses internal working conditions and incidents, but also benchmarks against the safety records of

other businesses. Employee safety evaluations are heavily based on safety control, which monitors work activities, environments, and task performance. A robust safety control system correlates with a positive safety environment, while weak controls are linked to reduced safety engagement [16-17]. Based on this, the following hypotheses were posited:

H7: Safety control positively influences safety performance.

H8: Safety control positively influences safety commitment.

E. Safety Communication

Safety communication is pivotal in disseminating health and safety information within an organization [5]. It involves an open dialogue between management, supervisors, and workers, which has been shown to enhance workplace performance and safety [5, 18-22]. Effective two-way communication is a hallmark of high safety performance, as highlighted by various studies. However, communication lapses, especially in conveying risk warnings, can negatively impact safety performance and overall productivity [18-22]. Positive reinforcement through feedback and incentives can strengthen the safety attitude of employees. Furthermore, after accidents, it is essential to focus on constructive communication about improving processes rather than placing blame [23]. Thus, communication is vital to foster a positive safety climate. Consequently, the hypotheses suggested are:

H9: Safety communication positively influences safety performance.

H10: Safety communication positively influences safety commitment.

F. Safety Commitment as a Mediator

According to [9, 24], safety commitment refers to the extent of dedication and devotion expressed by leadership and the workforce toward safety in an organization. This commitment is closely tied to safety policies, as it is reflected in the performance of both the workforce and the management. Previous studies have demonstrated that management policies, especially those responsible for promoting work motivation, boost safety commitment and enhance organizational value, effectiveness, and employee well-being [5-8]. Furthermore, a strong link has been identified between workers' dedication to health, safety, and their overall performance, with committed employees taking greater personal responsibility for their tasks and safety outcomes [25-27]. In this study, safety commitment plays a mediating role between all the variables and safety performance. Hence, the following hypotheses were posited:

H11: Safety commitment positively influences safety performance.

H12 Safety commitment mediates the relationship between safety policy and safety performance.

H13 Safety commitment mediates the relationship between safety training and safety performance.

H14 Safety commitment mediates the relationship between prevention planning and safety performance.

H15 Safety commitment mediates the relationship between safety control and safety performance.

H16 Safety commitment mediates the relationship between safety communication and safety performance.

G. Safety Knowledge as a Moderator

In [28], safety knowledge was defined as employees' understanding of the safe execution of job tasks. Safety knowledge goes beyond basic safety policies and includes aspects, such as training, prevention planning, control, and communication. Although safety policies provide foundational support, safety knowledge is crucial in effectively interpreting and implementing these policies [29-30]. Safety training is a dynamic process that integrates various elements, like skills, equipment, and social interactions [31]. However, training alone does not guarantee improved safety performance. This is because its application in real-world scenarios also requires safety knowledge for greater effectiveness [32]. According to [33-36], safety knowledge improves understanding of preventive measures and safety controls and facilitates effective communication, ultimately influencing safety performance. In [37], it was reported that effective safety knowledge can reduce accidents and construction costs, especially in developing countries. According to the aim of the current study, the following hypotheses were presented:

H17: Safety knowledge has a significant moderating role in the relationship between safety policy and safety performance.

H18: Safety knowledge has a significant moderating role in the relationship between safety training and safety performance.

H19: Safety knowledge has a significant moderating role in the relationship between prevention planning and safety performance.

H20: Safety knowledge has a significant moderating role in the relationship between safety control and safety performance.

H21: Safety knowledge has a significant moderating role in the relationship between safety communication and safety performance.

H. Materials and Methods

This study aimed to investigate the influence of safety management practices on safety performance, considering the mediating role of safety commitment and the moderating role of safety knowledge. Quantitative research was employed using an online questionnaire. The questionnaire was distributed to workers and engineers actively employed in the Ministry of Electricity in Iraq using a simple random sampling method. A sample of 374 participants was chosen from 23 power plants located in the central region of Iraq, with a response rate of 61.6%. A meticulously designed questionnaire was used to collect the required data. Not only was this questionnaire structured, but also underwent a rigorous design and validation process. It consisted of seven distinct sections and a total of 63 questions. A 5-point Likert scale was used, where a score of 1 indicated "strongly disagree" and a score of 5 represented "strongly agree". All measurements achieved acceptable reliability scores, with Cronbach's alpha scores

above 0.7. Each section of the questionnaire covered specific aspects of safety management. The sections included safety policy, safety training, prevention planning, safety control, safety communication, safety commitment, and safety knowledge. After data collection, the responses were processed and analyzed using SPSS and PLS. Specifically, the Partial Least Squares - Structural Equation Modeling (PLS-SEM) regression algorithm was employed to derive insights from the data and understand the relationships and influences between the variables in the hypothesized model.

III. RESULTS AND DISCUSSION

A. Descriptive Data

A total of 376 responses were received, comprising 355 males and 19 females. Table I presents a summary of the demographic information of the 376 respondents. The gender disparity highlights the male predominance within the electrical power industry in Iraq, aligning with global trends where the power sector is primarily male-dominated [38]. The majority of respondents fall within the 26-30 age group (39%), followed by the 18-25 age group (34%). In particular, a significant portion of respondents hold a bachelor's degree (58%), predominantly working as technicians (45%) or engineers (37%). Most of the employees had 7-10 years of experience (38%).

TABLE I. DEMOGRAPHIC PROFILES OF SAMPLES (N=374)

Demographic Profile	Frequency	Percentage (%)
Gender		
Male	355	95
Female	19	5
Age Group		
18-25	127	34
26-30	146	39
31-35	71	19
36-40	22	6
Above 40	7	2
Level of Education		
High school	86	23
Diploma	60	16
Bachelor's degree	217	58
Master's degree	11	3
Current Employment		
Technician	168	45
Engineer	138	37
Senior engineer	22	6
Head power plant assistant	26	7
Head power plant	19	5
Years of Employment		
Less than 3 years	60	16
3-6 years	101	27
7-10 years	142	38
More than 10 years	71	19

B. Statistical Analysis

This study established multiple hypotheses on the associations between the six main study variables (safety policy, safety training, prevention planning, safety control, safety communication, and safety commitment). Path analysis was performed through SmartPLS to test those hypotheses. Figure 2 illustrates the specific paths that correspond to the relationships among those six variables. The coefficient of

determination (R^2) for the endogenous variable safety performance was 0.50, indicating that it explained 50% of the variation. R^2 for the endogenous variable safety commitment was 0.48, indicating that it explained 48% of the variation.

Regarding common method bias, Harman's single-factor test displayed a ratio of extracted variables of 34.4%, below the threshold value of 50% [39]. The inner Variance Inflation Tolerance (VIF) values from the results of the measurement model were less than five. These results indicated no existence of multicollinearity in the data [40]. Table II shows that all Cronbach's Alpha values surpassed the recommended threshold of 0.7. Furthermore, the composite reliability values ranged from 0.894 to 0.947, signifying strong internal consistency [41]. The Average Variance Extracted (AVE) values were within the range of 0.742 to 0.828, confirming convergent validity. Finally, Heterotrait-Monotrait Ratio (HTMT) values were all below the threshold of 0.9, and less than the cut-off value of 0.85. Therefore, discriminant validity was established.

TABLE II. INTERNAL CONSISTENCY AND CONVERGENCE VALIDITY RESULTS

	Cronbach's Alpha	Composite Reliability (ρ_c)	Average Variance Extracted (AVE)
PP	0.902	0.932	0.773
SC	0.926	0.947	0.819
SCO	0.899	0.930	0.768
SCOM	0.896	0.935	0.828
SP	0.842	0.894	0.778
SPER	0.918	0.942	0.804
ST	0.884	0.920	0.742

C. Results from SEM

After analyzing the measurement model, the structural model was assessed. The results of the path analysis indicated that safety policy positively and significantly predicted safety performance ($\beta = 0.184$, $t = 3.463$, $p < 0.001$) and safety commitment ($\beta = -0.081$, $t = 1.725$, $p < 0.01$). Therefore, H1 and H2 were supported. The safety training did not influence safety performance or safety commitment. Hence, H3 and H4 were not supported. Prevention planning had a positive and significant influence on safety performance ($\beta = 0.153$, $t = 2.795$, $p < 0.05$) and safety commitment ($\beta = 0.327$, $t = 5.190$, $p < 0.000$). Thus, H5 and H6 were supported. Regarding safety control, it did not influence safety performance ($p > 0.05$) but had a positive and significant influence on safety commitment ($\beta = 0.084$, $t = 2.256$, $p < 0.05$). Therefore, H7 was not supported while H8 was supported.

H9 and H10 focused on the relationships between safety communication, safety performance, and safety commitment. This study hypothesized that safety communication influences safety performance and safety commitment. The results of path analyses indicated that safety communication had a positive influence on safety performance ($\beta = 0.164$, $t = 2.601$, $p < 0.05$) and safety communication ($\beta = 0.377$, $t = 5.542$, $p < 0.001$). Thus, H9 and H10 were supported. Moreover, safety commitment positively influenced safety performance ($\beta = 0.276$, $t = 4.077$, $p < 0.001$), indicating that H11 was also supported.

Regarding the moderating effect of safety knowledge, the findings revealed no moderating effect of safety knowledge on the relationship between safety policy and safety performance ($p > 0.05$). There was also no moderating effect of safety knowledge on the influence of prevention planning on safety performance ($p > 0.05$). Additionally, the relationship between safety control and safety performance was not moderated by safety knowledge ($p > 0.05$). Nonetheless, safety knowledge moderated the relationship between safety training and safety performance ($\beta = 0.100$, $t = 1.745$, $p < 0.05$). Finally, safety knowledge moderated the relationship between safety communication and safety performance ($\beta = -0.156$, $t = 2.707$, $p < 0.001$). Table III presents the results for each hypothesis' relationship. Overall, 15 out of 21 (71%) hypotheses were supported.

TABLE III. ESTIMATES AND SIGNIFICANCE OF EACH HYPOTHESIS IN THE SEM.

	Hypothesis	Result
H1	Safety Policy \rightarrow Safety Performance	Supported
H2	Safety Policy \rightarrow Safety Commitment	Supported
H3	Safety Training \rightarrow Safety Performance	Not Supported
H4	Safety Training \rightarrow Safety Commitment	Not Supported
H5	Prevention Planning \rightarrow Safety Performance	Supported
H6	Prevention Planning \rightarrow Safety Commitment	Supported
H7	Safety Control \rightarrow Safety Performance	Supported
H8	Safety Control \rightarrow Safety Commitment	Supported
H9	Safety Communication \rightarrow Safety Performance	Supported
H10	Safety Communication \rightarrow Safety Commitment	Supported
H11	Safety Commitment \rightarrow Safety Performance	Supported
H12	Safety Policy \rightarrow Safety Commitment \rightarrow Safety Performance	Supported
H13	Safety Training \rightarrow Safety Commitment \rightarrow Safety Performance	Not Supported
H14	Prevention Planning \rightarrow Safety Commitment \rightarrow Safety Performance	Supported
H15	Safety Control \rightarrow Safety Commitment \rightarrow Safety Performance	Supported
H16	Safety Communication \rightarrow Safety Commitment \rightarrow Safety Performance	Supported
H17	Safety Knowledge \times Safety Policy \rightarrow Safety Performance	Not Supported
H18	Safety Knowledge \times Safety Training \rightarrow Safety Performance	Supported
H19	Safety Knowledge \times Prevention Planning \rightarrow Safety Performance	Not Supported
H20	Safety Knowledge \times Safety Control \rightarrow Safety Performance	Not Supported
H21	Safety Knowledge \times Safety Communication \rightarrow Safety Performance	Supported

IV. DISCUSSION

This study aimed to examine the main predictors of safety performance in the Iraqi electric power sector. To this end, external variables were incorporated into the proposed model, including safety knowledge as a moderator. The empirical findings revealed that safety performance is influenced by safety communication, safety policy, safety control, prevention planning, and safety commitment. Safety commitment is affected by safety policy, prevention planning, control, and communication. Safety training and safety control were insignificant.

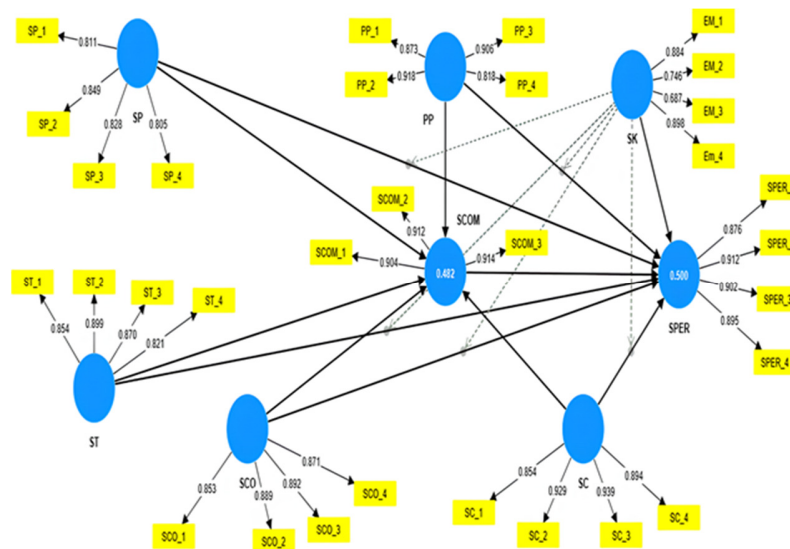


Fig. 2. Measurement model result.

Furthermore, safety communication had the most significant effect compared to other variables. As expected, safety policy had a positive impact on safety performance, aligning with previous studies suggesting that a greater emphasis on safety policy within an organization enhances overall safety performance among employees [12, 36, 42]. Consequently, effective safety policies and initiatives are likely to foster a commitment to improving safety performance among staff. However, the empirical findings show that safety training does not notably influence safety performance, in contrast to previous studies that highlighted the impact of safety training on safety performance [12, 43-44]. This unexpected result may be specific to the electric power industry context. Even though other studies showed that safety training enhances safety performance across industries [43-44], it is plausible that employees in the Iraqi power sector perceive the training they receive as inadequate or irrelevant [45-46].

As hypothesized, prevention planning positively influenced safety performance. Similarly, in [44], it was reported that effective prevention planning leads to enhanced safety performance. Thus, regarding the workplace, when employees see that their employer is dedicated to preventing incidents and ensuring their well-being, it contributes to a culture where safety is prioritized, positively influencing safety performance. Therefore, employees are likely to be more committed to enhancing safety performance [45-47]. Furthermore, the empirical findings revealed that safety control did not influence safety performance. This unexpected finding contradicts [48-49] that reported that higher safety control leads to better safety performance in the workplace. According to the literature, safety control measures are explicitly implemented to improve safety performance [5, 49-50]. The findings indicate that enhanced safety control will not improve employees' safety performance. Therefore, the employees did not perceive safety control as essential to enhance their safety performance in the electric power sector. As expected, safety control had a positive influence on safety commitment. Previous studies did not assess the impact of safety control on safety commitment. This

finding suggests that increased safety control at work will improve the safety commitment of employees. The results revealed that safety communication positively influenced safety performance, in line with [51-52] where safety communication positively affected employee safety performance. Effective communication is essential in providing health and safety-related information from management to workers and vice versa. The results also showed that safety communication positively influences safety commitment. This finding supports previous studies, which showed that effective safety communication can enhance the safety commitment of employees at the workplace [18-19]. Therefore, increased communication and awareness of safety procedures, potential hazards, and feedback will likely boost staff commitment to safety at work.

V. CONCLUSION

This study aimed to formulate a model for safety performance in the Iraqi electric power sector, integrating previous studies to develop an extended ECM that will provide some theoretical framework for future research. This study presented an extensive review of theoretical and empirical studies from the workplace safety domains to identify knowledge gaps in the safety performance of employees. The study addresses these gaps by proposing a model to forecast safety performance in the electric power sector, where security is a critical aspect of operations due to the unique and potentially hazardous nature of working with electricity. The novelty of this study is that it integrates critical safety-related variables. The model includes essential variables, such as safety policy, prevention planning, safety control, safety communication, and safety commitment. This study discloses that safety communication is critical to ensuring safety performance among employees in the electric power sector. The findings shed new light on the application of a tailored model to explain safety performance in this specific context, differentiating from previous studies, which mainly focused on past models that were not suited to this background. Although

several studies investigated the predictors of workplace safety performance [34, 53-55], limited studies have focused on the electric power sector. Safety commitment is the degree to which leadership within an organization demonstrates a strong commitment to safety [9]. The empirical findings revealed that safety performance is influenced by safety commitment, in agreement with [56].

In addition to the meaningful findings, some limitations remain to be addressed in the future. First, it would be beneficial to replicate the study in diverse cultural contexts to compare preferences and behaviors in different countries. This would improve cross-cultural generalizability and further explore the moderating effect of culture on the identified predictors of safety performance in the Iraqi electric power industry. Furthermore, future work might conduct a longitudinal study at successive time intervals to observe changes in user attitudes and intentions. This entails data collection over a longer period at different times and is particularly important as user perceptions can change over time. This method was not feasible in this study due to time constraints. Alternatively, a mixed-method approach that combines qualitative and quantitative methods can provide deeper insights and better outcomes.

REFERENCES

- [1] C. W. Su, X. Yuan, M. Umar, and O. R. Lobont, "Does technological innovation bring destruction or creation to the labor market?," *Technology in Society*, vol. 68, Feb. 2022, Art. no. 101905, <https://doi.org/10.1016/j.techsoc.2022.101905>.
- [2] E. Pereira, S. Ahn, S. Han, and S. Abourizk, "Identification and Association of High-Priority Safety Management System Factors and Accident Precursors for Proactive Safety Assessment and Control," *Journal of Management in Engineering*, vol. 34, no. 1, Jan. 2018, Art. no. 04017041, [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000562](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000562).
- [3] S. Garza Jr., "Quantitative Evaluation of Neutralization and Deterrence Theories on Security Policy User Adherence," Ph.D. dissertation, Capella University, 2021.
- [4] "Annual Reports," <https://moelc.gov.iq/?page=62>.
- [5] S. Kim, P. B. Kim, and G. Lee, "Predicting hospitality employees' safety performance behaviors in the COVID-19 pandemic," *International Journal of Hospitality Management*, vol. 93, Feb. 2021, Art. no. 102797, <https://doi.org/10.1016/j.ijhm.2020.102797>.
- [6] H. O. Kalteh, M. Salesi, and H. Mokarami, "The mediator role of safety motivation and knowledge between safety culture and safety performance: The results of a sociotechnical and macroergonomics approach," *Work*, vol. 72, no. 2, pp. 707–717, Jan. 2022, <https://doi.org/10.3233/WOR-205085>.
- [7] C. Devece, D. Palacios-Marqués, and M. Pilar Alguacil, "Organizational commitment and its effects on organizational citizenship behavior in a high-unemployment environment," *Journal of Business Research*, vol. 69, no. 5, pp. 1857–1861, May 2016, <https://doi.org/10.1016/j.jbusres.2015.10.069>.
- [8] K. Mearns, S. M. Whitaker, and R. Flin, "Safety climate, safety management practice and safety performance in offshore environments," *Safety Science*, vol. 41, no. 8, pp. 641–680, Oct. 2003, [https://doi.org/10.1016/S0925-7535\(02\)00011-5](https://doi.org/10.1016/S0925-7535(02)00011-5).
- [9] D. A. Hofmann, M. J. Burke, and D. Zohar, "100 years of occupational safety research: From basic protections and work analysis to a multilevel view of workplace safety and risk," *Journal of Applied Psychology*, vol. 102, no. 3, pp. 375–388, 2017, <https://doi.org/10.1037/apl0000114>.
- [10] J. Santos-Reyes and A. N. Beard, "A SSMS model with application to the oil and gas industry," *Journal of Loss Prevention in the Process Industries*, vol. 22, no. 6, pp. 958–970, Nov. 2009, <https://doi.org/10.1016/j.jlp.2008.07.009>.
- [11] C. Pilbeam, N. Doherty, R. Davidson, and D. Denyer, "Safety leadership practices for organizational safety compliance: Developing a research agenda from a review of the literature," *Safety Science*, vol. 86, pp. 110–121, Jul. 2016, <https://doi.org/10.1016/j.ssci.2016.02.015>.
- [12] A. Haslinda, S. Saharudin, N. H. Roslan, and R. Mohamed, "Safety Training, Company Policy and Communication for Effective Accident Management," *International Journal of Academic Research in Business and Social Sciences*, vol. 6, no. 9, pp. 141–158, Sep. 2016.
- [13] A. Tezel, E. Dobrucali, S. Demirkesen, and I. A. Kiral, "Critical Success Factors for Safety Training in the Construction Industry," *Buildings*, vol. 11, no. 4, Apr. 2021, Art. no. 139, <https://doi.org/10.3390/buildings11040139>.
- [14] S. Givehchi, E. Hemmatvaghef, and H. Hoveidi, "Association between safety leading indicators and safety climate levels," *Journal of Safety Research*, vol. 62, pp. 23–32, Sep. 2017, <https://doi.org/10.1016/j.jsr.2017.05.003>.
- [15] "Recommended Practices for Safety and Health Programs," Occupational Safety and Health Administration, 2016.
- [16] R. Eldejany, "The Relationship between Safety Management Systems, Safety Performance and Customer Satisfaction in the Australian Construction Industry: A Quantitative Research Proposal," *Journal of Research in Marketing*, vol. 10, no. 1, pp. 766–771, Feb. 2019, <https://doi.org/10.17722/jorm.v10i1.765>.
- [17] O. L. Siu and T. K. Ng, "Family-to-Work Interface and Workplace Injuries: The Mediating Roles of Burnout, Work Engagement, and Safety Violations," *International Journal of Environmental Research and Public Health*, vol. 18, no. 22, Jan. 2021, Art. no. 11760, <https://doi.org/10.3390/ijerph182211760>.
- [18] H. Sarkheil, M. Talaiean Eraghi, and S. Vatan Khah, "Hazard identification and risk modeling on runway bird strikes at Sardar-e-Jangal International Airport of Iran," *Modeling Earth Systems and Environment*, vol. 7, no. 4, pp. 2589–2598, Nov. 2021, <https://doi.org/10.1007/s40808-020-01032-0>.
- [19] M. N. Vinodkumar and M. Bhasi, "Safety management practices and safety behaviour: Assessing the mediating role of safety knowledge and motivation," *Accident Analysis & Prevention*, vol. 42, no. 6, pp. 2082–2093, Nov. 2010, <https://doi.org/10.1016/j.aap.2010.06.021>.
- [20] S. Boubaker, S. Mekni, and H. Jerbi, "Assessment of Electrical Safety Beliefs and Practices: A Case Study," *Engineering, Technology & Applied Science Research*, vol. 7, no. 6, pp. 2231–2235, Dec. 2017, <https://doi.org/10.48084/etasr.1521>.
- [21] M. Z. Hasanpour, M. R. Ahadi, A. S. Moghadam, and G. A. Behzadi, "Variable Speed Limits: Strategies to Improve Safety and Traffic Parameters for a Bottleneck," *Engineering, Technology & Applied Science Research*, vol. 7, no. 2, pp. 1535–1539, Apr. 2017, <https://doi.org/10.48084/etasr.831>.
- [22] A. Mobaraki, R. Mirzaei, and H. Ansari, "A Survey of Health, Safety and Environment (HSE) Management and Safety Climate in Construction Sites," *Engineering, Technology & Applied Science Research*, vol. 7, no. 1, pp. 1334–1337, Feb. 2017, <https://doi.org/10.48084/etasr.904>.
- [23] F. H. B. M. Taufek, Z. B. Zulkifli, and S. Z. B. A. Kadir, "Safety and Health Practices and Injury Management in Manufacturing Industry," *Procedia Economics and Finance*, vol. 35, pp. 705–712, Jan. 2016, [https://doi.org/10.1016/S2212-5671\(16\)00088-5](https://doi.org/10.1016/S2212-5671(16)00088-5).
- [24] N. Borkowski and K. A. Meese, *Organizational Behavior in Health Care*. Burlington, MA, USA: Jones & Bartlett Learning, 2020.
- [25] B. Yanar, M. Lay, and P. M. Smith, "The Interplay Between Supervisor Safety Support and Occupational Health and Safety Vulnerability on Work Injury," *Safety and Health at Work*, vol. 10, no. 2, pp. 172–179, Jun. 2019, <https://doi.org/10.1016/j.shaw.2018.11.001>.
- [26] Z. C. Tan, C. E. Tan, and Y. O. Choong, "Occupational Safety & Health Management and Corporate Sustainability: The Mediating Role of Affective Commitment," *Safety and Health at Work*, vol. 14, no. 4, pp. 415–424, Dec. 2023, <https://doi.org/10.1016/j.shaw.2023.10.006>.
- [27] R. Kaynak, A. T. Toklu, M. Elci, and I. T. Toklu, "Effects of Occupational Health and Safety Practices on Organizational Commitment, Work Alienation, and Job Performance: Using the PLS-

- SEM Approach," *International Journal of Business and Management*, vol. 11, no. 5, 2016.
- [28] K. Y. Kao, C. Spitzmuller, K. P. Cigularov, and C. L. Thomas, "A Moderated Mediation Model of Safety Knowledge, Safety Attitude, and Safety Performance," *Academy of Management Proceedings*, vol. 2016, no. 1, Jan. 2016, Art. no. 16128, <https://doi.org/10.5465/ambpp.2016.219>.
- [29] K. Smith-Crowe, M. J. Burke, and R. S. Landis, "Organizational climate as a moderator of safety knowledge–safety performance relationships," *Journal of Organizational Behavior*, vol. 24, no. 7, pp. 861–876, 2003, <https://doi.org/10.1002/job.217>.
- [30] K. H. Sharif and S. Y. Ameen, "A Review of Security Awareness Approaches With Special Emphasis on Gamification," in *2020 International Conference on Advanced Science and Engineering (ICOASE)*, Duhok, Iraq, Dec. 2020, pp. 151–156, <https://doi.org/10.1109/ICOASE51841.2020.9436595>.
- [31] P. U. Okoye, J. U. Ezeokkonkwo, and F. O. Ezeokoli, "Building Construction Workers' Health and Safety Knowledge and Compliance on Site," *Journal of Safety Engineering*, vol. 5, no. 1, pp. 17–26, 2016.
- [32] J. P. Hatala and P. R. Fleming, "Making Transfer Climate Visible: Utilizing Social Network Analysis to Facilitate the Transfer of Training," *Human Resource Development Review*, vol. 6, no. 1, pp. 33–63, Mar. 2007, <https://doi.org/10.1177/1534484306297116>.
- [33] D. Ramos, T. Cotrim, P. Arezes, J. Baptista, M. Rodrigues, and J. Leitão, "Frontiers in Occupational Health and Safety Management," *International Journal of Environmental Research and Public Health*, vol. 19, no. 17, Jan. 2022, Art. no. 10759, <https://doi.org/10.3390/ijerph191710759>.
- [34] J. K. Wachter and P. L. Yorio, "A system of safety management practices and worker engagement for reducing and preventing accidents: An empirical and theoretical investigation," *Accident Analysis & Prevention*, vol. 68, pp. 117–130, Jul. 2014, <https://doi.org/10.1016/j.aap.2013.07.029>.
- [35] H. Liu, H. Chen, R. Hong, H. Liu, and W. You, "Mapping knowledge structure and research trends of emergency evacuation studies," *Safety Science*, vol. 121, pp. 348–361, Jan. 2020, <https://doi.org/10.1016/j.ssci.2019.09.020>.
- [36] G. M. A. Naji, A. S. N. Isha, A. Alazzani, M. S. Saleem, and M. Alzoraiki, "Assessing the Mediating Role of Safety Communication Between Safety Culture and Employees Safety Performance," *Frontiers in Public Health*, vol. 10, 2022.
- [37] S. Kumar and V. K. Bansal, "Construction safety knowledge for practitioners in the construction industry," *Journal of Frontiers in Construction Engineering*, vol. 2, no. 2, pp. 34–42, 2013.
- [38] H. Beides and E. Maier, "Getting more women into the energy sector: A RENEW'ed approach for MENA," Jul. 21, 2022, <https://blogs.worldbank.org/arabvoices/getting-more-women-energy-sector-renewed-approach-mena>.
- [39] J. Hair and A. Alamer, "Partial Least Squares Structural Equation Modeling (PLS-SEM) in second language and education research: Guidelines using an applied example," *Research Methods in Applied Linguistics*, vol. 1, no. 3, Dec. 2022, Art. no. 100027, <https://doi.org/10.1016/j.rmal.2022.100027>.
- [40] J. F. Pallant and A. Tennant, "An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS)," *British Journal of Clinical Psychology*, vol. 46, no. 1, pp. 1–18, 2007, <https://doi.org/10.1348/014466506X96931>.
- [41] J. F. Hair Jr., G. T. M. Hult, C. M. Ringle, M. Sarstedt, N. P. Danks, and S. Ray, *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook*. Springer Nature, 2021.
- [42] B. Fernández-Muñiz, J. M. Montes-Peón, and C. J. Vázquez-Ordás, "The role of safety leadership and working conditions in safety performance in process industries," *Journal of Loss Prevention in the Process Industries*, vol. 50, pp. 403–415, Nov. 2017, <https://doi.org/10.1016/j.jlp.2017.11.001>.
- [43] X. Ye, S. Ren, X. Li, and Z. Wang, "The mediating role of psychological capital between perceived management commitment and safety behavior," *Journal of Safety Research*, vol. 72, pp. 29–40, Feb. 2020, <https://doi.org/10.1016/j.jsr.2019.12.004>.
- [44] D. Schlesinger, "Organizational Culture," presented at the 2017 Joint Rail Conference, Jul. 2017, <https://doi.org/10.1115/JRC2017-2247>.
- [45] Z. Noureddine, L. Madi, S. Ullah, H. Alrawashdeh, and L. Naseralallah, "A prospective observational study to evaluate the safety of COVID-19 mRNA vaccines administered to Qatar Rehabilitation Institute patients," *Qatar Medical Journal*, vol. 2023, no. 1, May 2023, Art. no. 10, <https://doi.org/10.5339/qmj.2023.10>.
- [46] S. Ahn, T. Kim, Y.-J. Park, and J.-M. Kim, "Improving Effectiveness of Safety Training at Construction Worksite Using 3D BIM Simulation," *Advances in Civil Engineering*, vol. 2020, Feb. 2020, Art. no. e2473138, <https://doi.org/10.1155/2020/2473138>.
- [47] K. Czarnocki, E. Czarnocka, J. Szer, B. Hoła, M. Rebelo, and K. Czarnocka, "Scaffold use risk assessment model for construction process safety," presented at the Joint CIB W099 & TG59 International Safety, Health, and People in Construction Conference, 2017.
- [48] S. Supardi and G. Chandrarin, "The Role of Safety Leadership and Safety Culture in improving Safety Performance," *East African Scholars Journal of Economics, Business and Management*, vol. 4, no. 4, pp. 55–62, May 2021.
- [49] F. Saleem and M. I. Malik, "Safety Management and Safety Performance Nexus: Role of Safety Consciousness, Safety Climate, and Responsible Leadership," *International Journal of Environmental Research and Public Health*, vol. 19, no. 20, Jan. 2022, Art. no. 13686, <https://doi.org/10.3390/ijerph192013686>.
- [50] L. A. Snyder, A. D. Krauss, P. Y. Chen, S. Finlinson, and Y.-H. Huang, "Safety performance: The mediating role of safety control," *Work*, vol. 40, no. 1, pp. 99–111, Jan. 2011, <https://doi.org/10.3233/WOR-2011-1210>.
- [51] A. Al-Refaie, "Factors affect companies' safety performance in Jordan using structural equation modeling," *Safety Science*, vol. 57, pp. 169–178, Aug. 2013, <https://doi.org/10.1016/j.ssci.2013.02.010>.
- [52] A. Makin and C. Winder, "Managing hazards in the workplace using organisational safety management systems: a safe place, safe person, safe systems approach," *Journal of Risk Research*, vol. 12, no. 3–4, pp. 329–343, Jun. 2009, <https://doi.org/10.1080/13669870802658998>.
- [53] E. N. K. Nkrumah, S. Liu, D. Doe Fiergbor, and L. S. Akoto, "Improving the Safety–Performance Nexus: A Study on the Moderating and Mediating Influence of Work Motivation in the Causal Link between Occupational Health and Safety Management (OHSM) Practices and Work Performance in the Oil and Gas Sector," *International Journal of Environmental Research and Public Health*, vol. 18, no. 10, Jan. 2021, Art. no. 5064, <https://doi.org/10.3390/ijerph18105064>.
- [54] H. P. Pei, L. K. Hock, and F. S. A. Aziz, "Understanding The Relationship between Safety Climate and Safety Performance: A Narrative Review," *Journal of Safety, Health & Ergonomics*, vol. 3, no. 1, Aug. 2021.
- [55] N. Ansori, A. Widyanti, Yassierli, N. Ansori, A. Widyanti, and Yassierli, "The Influence of Safety Climate, Motivation, and Knowledge on Worker Compliance and Participation: An Empirical Study of Indonesian SMEs," *Ingeniería e Investigación*, vol. 41, no. 3, Dec. 2021, <https://doi.org/10.15446/ing.investig.v41n3.83763>.
- [56] J. Zara, S. M. Nordin, and A. S. N. Isha, "Influence of communication determinants on safety commitment in a high-risk workplace: a systematic literature review of four communication dimensions," *Frontiers in Public Health*, vol. 11, 2023.

Model-based Design of a High-Throughput Canny Edge Detection Accelerator on Zynq-7000 FPGA

Ahmed Alhomoud

Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
aalhomoud@nbu.edu.sa

Refka Ghodhbani

Department of Computer Sciences, Faculty of Computing and information Technology, Northern Border University, Saudi Arabia | Electronics and Micro-Electronics Laboratory, Faculty of Sciences, Monastir University, Tunisia
refka.ghodhbani@nbu.edu.sa

Taoufik Saidani

Department of Computer Sciences, Faculty of Computing and information Technology, Northern Border University, Saudi Arabia | Electronics and Micro-Electronics Laboratory, Faculty of Sciences, Monastir University, Tunisia
taoufik.saidani@nbu.edu.sa (corresponding author)

Hafedh Mahmoud Zayani

Department of Electrical Engineering, College of Engineering, Northern Border University, Saudi Arabia
hafedh.zayani@nbu.edu.sa

Yahia Said

Department of Electrical Engineering, College of Engineering, Northern Border University, Saudi Arabia
yahia.said@nbu.edu.sa

Mohamed Ben Ammar

Department of Information Systems, Faculty of Computing and Information Technology, Northern Border, Saudi Arabia
mohammed.ammar@nbu.edu.sa

Jihane Ben Slimane

Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
jehan.saleh@nbu.edu.sa

Received: 16 February 2024 | Revised: 24 February 2024 and 25 February 2024 | Accepted: 27 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7081>

ABSTRACT

This paper presents a novel approach for fast FPGA prototyping of the Canny edge detection algorithm using High-Level Synthesis (HLS) based on the HDL Coder. Traditional RTL-based design methodologies for implementing image processing algorithms on FPGAs can be time-consuming and error-prone. HLS offers a higher level of abstraction, enabling designers to focus on algorithmic functionality while the tool automatically generates efficient hardware descriptions. This advantage was exploited by implementing

the Canny edge detection algorithm in MATLAB/Simulink and utilizing the HDL Coder to automatically convert it into synthesizable VHDL code. This design flow significantly reduces development time and complexity compared to the traditional RTL approach. The experimental results showed that the HLS-based Canny edge detector achieved real-time performance on a Xilinx FPGA platform, showcasing the effectiveness of the proposed approach for fast FPGA prototyping in image processing applications.

Keywords-FPGA; high-level synthesis; HDL coder; Canny edge detection; image processing; fast prototyping

I. INTRODUCTION

Canny edge detection is a fundamental image processing technique widely used in computer vision applications. While software implementations offer flexibility, FPGAs are preferred for real-time and resource-constrained applications due to their parallel processing capabilities. However, traditional RTL-based design methodologies for FPGAs can be time-consuming and require expertise in hardware description languages and manual optimization techniques. To address this challenge, High-Level Synthesis (HLS) has emerged as a promising approach to accelerate FPGA design by enabling algorithm implementation in familiar C/C++ or MATLAB/Simulink environments [1-2].

In the realm of computer vision, where raw pixels transform into meaningful insights, edge detection reigns as a fundamental image processing technique [3-4]. It acts as a cartographer, meticulously tracing the boundaries between objects, shapes, and textures, revealing the very essence of a scene. Within this domain, Canny edge detection stands as a beacon of precision and efficiency, with its algorithm meticulously crafting intricate edge maps with unparalleled clarity [3-5]. However, harnessing the power of Canny edge detection for real-time applications often presents a conundrum. Although software implementations offer flexibility, their computational demands can lag behind the lightning-fast pace of real-time processing. FPGAs boast parallel processing prowess, as their ability to crunch numbers at blistering speeds, which makes them ideal candidates for image processing tasks, particularly when latency matters [6]. Unlocking the full potential of FPGAs for Canny edge detection can be a laborious endeavor. Traditional RTL-based design methodologies, while offering fine-grained control, demand expertise in hardware description languages and meticulous hand-coding, often extending development timelines and introducing the risk of human error [4].

High-Level Synthesis (HLS) is a revolutionary approach. Bypassing the intricacies of RTL, HLS can express the Canny edge detection algorithm in familiar environments, such as MATLAB/Simulink, leveraging intuitive visual blocks and pre-built image processing libraries. HDL Coder, the interpreter within HLS, transforms this high-level description into synthesizable VHDL or Verilog code, effortlessly bridging the gap between algorithm and hardware [7]. This study showcases the transformative power of HLS for fast FPGA prototyping of Canny edge detection, by proposing a streamlined design flow, charting a course from algorithm modeling in MATLAB/Simulink to real-time edge detection on an FPGA platform. This research aims to evaluate the advantages of real-time performance, resource efficiency, and development time of the proposed approach, contrasting it with the traditional

RTL paradigm [7]. More than just entailing technical advancements, this investigation unlocks exciting possibilities for the future of real-time image processing applications. From autonomous vehicles to medical imaging systems that reveal previously unseen details, the swift and precise edges identified by the proposed implementation of HLS-powered Canny edge detection hold the promise of transforming diverse fields with their clarity and insight [4].

II. EMBEDDED IMAGE AND VIDEO SYSTEM DESIGN BASED ON ZYNQ7000

The system used Zedboard, a Zynq-7000 development board, an Avnet FMC-HDMI-CAM module, and a Python 1300-C camera [8]. FMC stands for FPGA Mezzanine Card, a standardized interface that facilitates high-speed connections between FPGAs and peripheral devices [9]. This function enabled seamless integration of all cards equipped with the particular interface. UART communication between the system and a computer allowed for monitoring via the serial port, but the system functioned autonomously, independently of a computer connection. An integrated system consisting of a camera and a display constitutes a comprehensive vision system, facilitating the capture of video data. The system can process the data in real-time, depending on the specific application it is designed for. This study proposes an embedded architecture for the acquisition of video and processing modules in the design of a vision system, using a platform-based design method. This enables the seamless transmission of live video data from the sensor to the monitor using DDR3 memory and FPGA-based hardware processing. Figure 1 shows the design of the proposed system. The system consists of a VITA-2000 image sensor, an FMC module [10], the Xilinx ZedBoard platform, and an HDMI monitor to display the video output. These boards facilitate the software and hardware development of applications that are built on the Xilinx circuit from the Zynq family. The circuit utilizes an All-Programmable System-on-Chip (AP SoC) design, with a dual-core ARM™ Cortex-A9™ processor coupled with a Xilinx FPGA. This platform provides all the necessary resources to design a multitude of high-efficiency applications in the domains of video processing, motor control, software acceleration, and even system creation using embedded Linux [2]. The Zedboard has buttons and switches that allow the developer to select the desired filter and other parameters. Figure 1 displays the overall structure of the system. The heart of the system used in this study is the Zedboard development board, featuring a Zynq-7020 series SoC and various peripherals [10]. High-speed Advanced Expandable Interface (AXI) buses seamlessly bridge the gap between the CPU and FPGA, enabling their synergistic cooperation for dedicated tasks and defying their usual segregation.

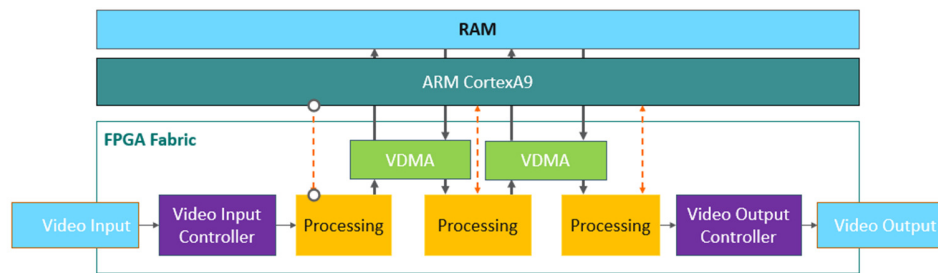


Fig. 1. General schematic of the designed system.

A. Model Design Based on HDL Coder Methodology

Model-Based Design (MBD) has the power to virtually build and test systems before constructing them in the real world, a concept explored in this study, focusing on the versatile HDL Coder from MathWorks. Although LabVIEW from National Instruments offers another MBD option, it lacks dedicated support for vendor-specific HLTs [11-12]. In contrast to these specialized HLTs optimized for specific hardware, such as filter chains through hardwired blocks, HDL Coder takes a holistic approach by analyzing the entire design and prioritizing functionality over platform limitations. From MATLAB functions, Simulink models, and Stateflow charts, it effortlessly generates portable and synthesizable Verilog and VHDL code, unlocking a world of possibilities from FPGA programming and ASIC prototyping to hardware/software co-design on platforms such as Xilinx Zynq and Intel SoC (complete with generated HDL and C code for embedded CPUs) [1]. HDL Coder goes beyond mere code generation, as its workflow advisor streamlines FPGA programming for platforms such as Xilinx, Microsemi, and Intel. The designer can gain granular control over the HDL architecture and implementation, identifying critical paths and estimating resource utilization. Furthermore, seamless traceability between the Simulink model and the generated code enables rigorous verification, even meeting the strict standards of DO-254 and beyond [7].

Unlike vendor-specific tools that are burdened with a plethora of detailed configurations, often requiring deep hardware knowledge, HDL Coder takes a simpler approach. Its streamlined interface drastically reduces design time and makes it the go-to solution for rapid FPGA prototyping. Whether designing intricate filter chains or complex SoC systems, HDL Coder provides the ability to explore, refine, and deploy designs with unprecedented efficiency and flexibility [2]. In essence, HDL Coder is not just a tool but a design philosophy. Mathworks offers two workflows for Zynq programming: Embedded Coder Hardware Support Package and SoC Blockset. Although both allow communication between Zynq's FPGA and ARM processors, their model construction and capabilities differ [7].

SoC Blockset uses separate Simulink models for each component (FPGA/processor), qualifying fine-grained control. A third model is required for communication, explicitly defining interfaces. This approach allows advanced features, such as direct memory access and interrupts, but can be unreliable for tasks within tight timeframes (<100 μ s).

Mathworks support confirms this limitation, making it less ideal for power electronics prototyping [13-17]. The Embedded Coder Hardware Support Package implements all code (FPGA and processor) in a single Simulink model, simplifying model management and communication. While basic communication channels are available, the trade-off is less flexibility and control.

Before building the physical components of the system, it is recommended to create a prototype design that can capture camera images [13, 15]. Following the correct acquisition of the camera picture, the filters are created and simulated within the Matlab/Simulink environment. The filters are then enclosed as regular IP blocks with the HDL Coder utility [16]. To incorporate IP blocks into the camera reference design, they need to be integrated into the Vivado environment. The Xilinx SDK environment was eventually used to write the system's software. Subsequently, the required drivers were incorporated into the system, culminating in its final configuration [11]. Building a robust system requires careful planning and execution. Here is how the referenced work tackled it:

1. Reference design: Before stepping into hardware, a blueprint was established, which is a reference design that encompasses camera image recording capabilities [10]. This ensured that all components and functionalities were clearly defined before physical construction.
2. Filter design and simulation: Capturing quality images is crucial. Therefore, the next step involved designing and simulating filters within the Matlab/Simulink environment. These filters, which act as image-processing algorithms, were fine-tuned to enhance the captured footage [5].
3. IP block encapsulation: To seamlessly integrate the designed filters with the hardware, they were packaged as reusable IP blocks using HDL Coder. This standardized format facilitated easy incorporation into the camera reference design within the Vivado environment [7].
4. Software development and integration: The system's software was then developed using the Xilinx SDK. Finally, essential drivers were added, finalizing the system's configuration and making it ready for action [10].
5. Visualization: Although diagrams and figures were not provided in the original text, incorporating visuals related to specific stages of the workflow could further enhance the reader's understanding. For example, an image depicting the reference design or a block diagram illustrating the filter

development and integration process could be helpful additions.

The overwhelming majority of contemporary chip systems use integrated processors. When designing a chip with numerous ICs, it is necessary to install additional software or firmware to handle the concurrent operations of the microprocessors, DSPs, memory, and personalized logic. With the use of a common functional specification, automated HLS allows architects and designers to explore different algorithmic and implementation choices, and evaluate and enhance the tradeoffs associated with space, power, and performance. The implementation of commercial HLS technology is now a more plausible prospect due to advancements in RTL synthesis techniques. Prominent semiconductor design companies, including IBM, Motorola, Philips, and Siemens, have presented exclusive software tools. Leading Electronic Design Automation (EDA) companies support a range of HLS products. An instance of developing RTL implementations using behavioral HDL code and linking to subsequent tools is demonstrated by Synopsys's "Behavioral Compiler" tool, which was introduced in 1995 [16]. Other comparable tools include Mentor Graphics' Catapult HLS [4] and Cadence's Stratus High Level Synthesis [18].

B. Development Workflow Setup

Development requires the appropriate configuration of many tools and components. Figure 2 presents the general configuration procedure. The following software is required to transfer control algorithms built in Matlab/Simulink to Zynq-based platforms:

- Simulink/HDL coder and embedded coder tools in Matlab/Simulink.
- Xilinx Vivado, using the free Webpack if the selected board is supported, else System Edition is advised.

This study employed Matlab R2018b and Vivado 2017.4 [7]. The System Generator tool is unable to officially support the necessary version of Matlab for HDL Coder compatibility when utilizing Vivado System Edition. Hence, it is imperative to manually adjust the list of compatible versions to ensure its compatibility with Matlab. As it will immediately establish a connection to Xilinx tools, it is recommended that Matlab be started using the System Generator tool when Vivado System Edition is being utilized. On the other hand, the HDL generation tool necessitates manual configuration every time Matlab is initiated by using the command `hdlsetuptoolpath('ToolName', 'XilinxVivado', 'ToolPath', 'C:\Xilinx\Vivado\2017.4\bin\vivado.bat')`. The following Matlab Add-Ons must be installed from the Add-On Explorer:

- The HDL Coder Extension packages for the FPGA Zynq-7000 Platform.
- Xilinx Zynq-7000 Platform Embedded Coder Support package.
- A compatible compiler (MinGW-w64 is a recommended option, available for direct installation from the Toolbox Add-On Explorer).

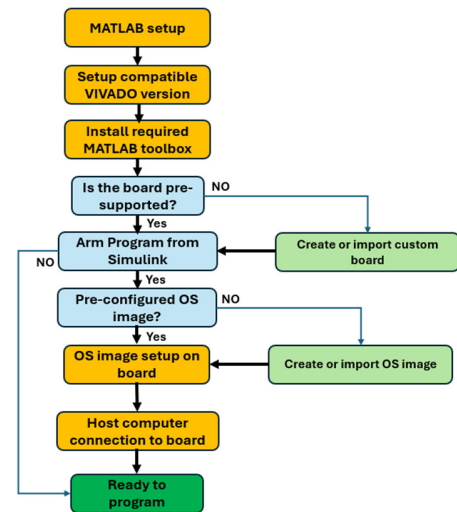


Fig. 1. System design workflow setup.

III. DESIGN AND SIMULATION OF CANNY EDGE DETECTION IN SIMULINK HDL CODER

A. Conventional Canny Edge Detection Algorithm

The Canny edge detection algorithm is highly popular for its robust performance in detecting edges, particularly in the presence of visual noise. Its demanding computing requirements result in the need for high clock frequencies and substantial power consumption on typical microprocessor architectures, especially when real-time constraints must be satisfied. Figure 3 illustrates the stages of the Canny edge detector to identify the texture of edges. The Canny edge detector uses a Gaussian blur mask to effectively smooth and filter out noise. The next step involves convolving the picture utilizing partial derivatives of a 2D Gaussian function, G_x and G_y , to obtain the edge direction edge *Angle* matrix and the strength matrix G . The G and edge *Angle* matrices are employed in the nonmaxima suppression operation. The determination of the local maximum involves a comparison of the pixel with its neighboring pixels in the direction of the gradient. The pixel is removed if it does not have a local maximum gradient magnitude. Along the gradient's direction, a comparison is performed between the real pixel and its neighbors to eliminate any extraneous pixels that are not part of the edge. Hysteresis thresholding is the final stage, which employs a high and a low threshold. Edge pixels with intensities exceeding the high threshold are classified as strong, while edge pixels with intensities below the low threshold are suppressed. Edge pixels falling between the two thresholds are classified as weak. These actions will result in a subtle boundary in the resulting image.

Canny devised a method to obtain an ideal edge detector for handling step edges that are distorted by white Gaussian noise. Figure 3 provides a concise overview of the initial Canny edge detection algorithm [5] for an input image with dimensions $M \times M$ and each pixel represented by n bits. The process consists of the following stages:

1. Using gradient masks, convolve each pixel position to determine the horizontal and vertical gradients G_x and G_y .

2. Calculate the magnitude G and the direction θG of the gradient at every pixel position.
3. NMS: Restore crisp edges in an image with blurred edges and eliminate minima while keeping local maxima in a gradient image.
4. Threshold calculation involves determining potential edges using both high and low thresholds. These thresholds are derived based on the histogram of the gradient magnitude of the entire image.

Hysteresis thresholding generates a smooth edge map by comparing the gradient magnitude values of each pixel with predefined low- and high-threshold values. This process eliminates border pixels resulting from noise and fluctuations in lighting conditions.

B. Simulink HDL Coder Design

This study applied filters within the HDL Coder environment to design the video system. Each HDL Toolbox block was designed individually. Initially, MATLAB was transformed into a subsystem. After performing simulations, the HDL work IP blocks that were compatible with the flux advisor were transformed. Four unique filter designs, all of

which can be simultaneously activated in the system, were consolidated into two IP blocks to achieve a size reduction. By doing so, it is possible to simultaneously activate all filters without adding any complication to the design [7].

C. Design Synthesis and Results

Once finished, each filter system was individually transformed into HDL code using the Matlab/Simulink HDL Coder and HDL Workflow Advisor add-on. The Matlab and Simulink systems were transformed into HDL code and the required settings were configured throughout the conversion procedure. The system was transformed into an IP block within the Vivado Suite Camera reference. This IP block was connected to the design. Consequently, filter systems were built sequentially on the Zedboard. In the future, the system will have reduced on-chip components to optimize space utilization and streamline software complexity. The IP blocks that were first built were redesigned and integrated into Simulink [8]. The edge detection algorithm was integrated into a unified IP block. The second IP block embedded sharpening and median filters. The system was reassembled and all necessary connections were made. Figure 4 portrays the schematic representation of the fully implemented system, indicating the created IP blocks.

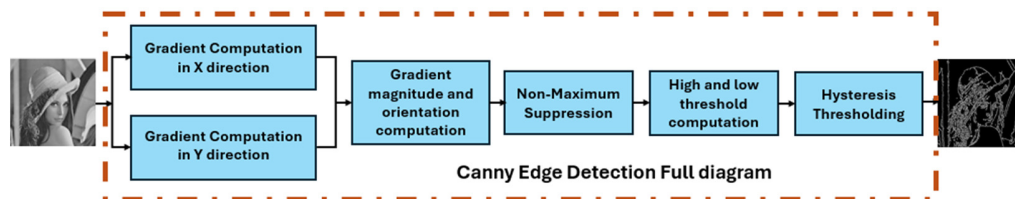


Fig. 2. Block diagram of the Canny edge detection algorithm.

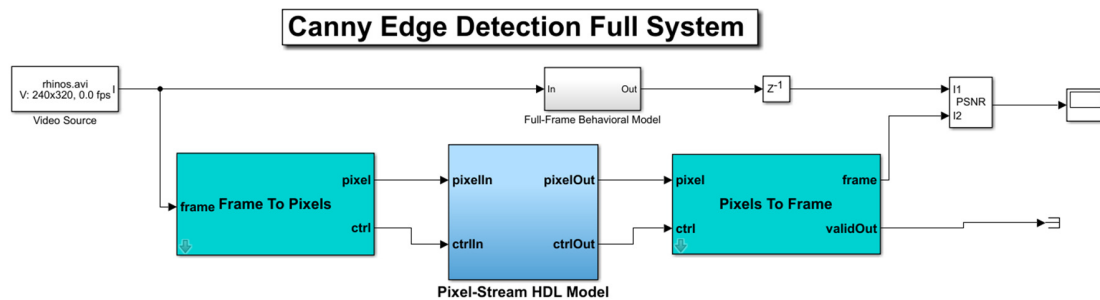


Fig. 3. HDL code for Canny edge detector.

D. Simulation Results

The Vivado xSim tool was utilized to simulate the non-synthesizable testbench employing the generated VHDL RTL code. The reference and recommended approaches yield identical pixel values. Additionally, it was discovered that they were indistinguishable from the high-level simulation results achieved using MATLAB on the optimal bit-width model. This was demonstrated by performing a comparison of the resulting photographs generated by both paths utilizing the identical input image. Additionally, the quantization error that resulted from selecting the narrowed "optimal" signal widths was contrasted with that of the MATLAB-based double precision

model. This was achieved using the "FPGA in the Loop" co-simulation feature of the MathWorks' HDL Verifier [7].

E. Synthesis Results

Table I presents the allocation of system resources. A maximum of 12% of the available resources were utilized. Furthermore, it is important to mention that no optimization efforts were undertaken in this context. The system's primary clock speed was set at 150 MHz, while the camera and image processing blocks operated at a pixel clock speed of 110 MHz. These findings demonstrate that the proposed image processing system is capable of functioning at an image resolution of 1280×1024 with a refresh rate of 60 Hz, allowing it to process

60 images per second [11]. This shows that the system is capable of satisfying the operational requirements in real-time.

1) Implementation of the System Software

Once the hardware design was completed, the required software was developed in parallel with the hardware. The design was executed in the C programming language using the Vivado Suite SDK tool. The software prepares and runs the IP blocks and modifies the filter and input-output unit parameters.

This technology allows for real-time adjustment of filters, and numerous filters can be simultaneously active. The Zedboard's system functions can be accessed via its built-in buttons and switches. Moreover, the LEDs on the Zedboard can be engaged to monitor the functioning status of the filters. Multiplexer logic was put into service to configure the system to select the appropriate filter depending on the Zedboard switches' states. The system was programmed to automatically activate the appropriate filter according to the condition of the keys.

TABLE I. PROPOSED CANNY EDGE DETECTION SYSTEM'S IMPLEMENTATION RESULTS

	Bit Depth	LUT-FF Pairs	LUTs as Logic	LUTs as Memory	Slice Registers	fmax	FPS at 1920×1080
Number	3x8	4740	2380	1930	1920	190MHz	89

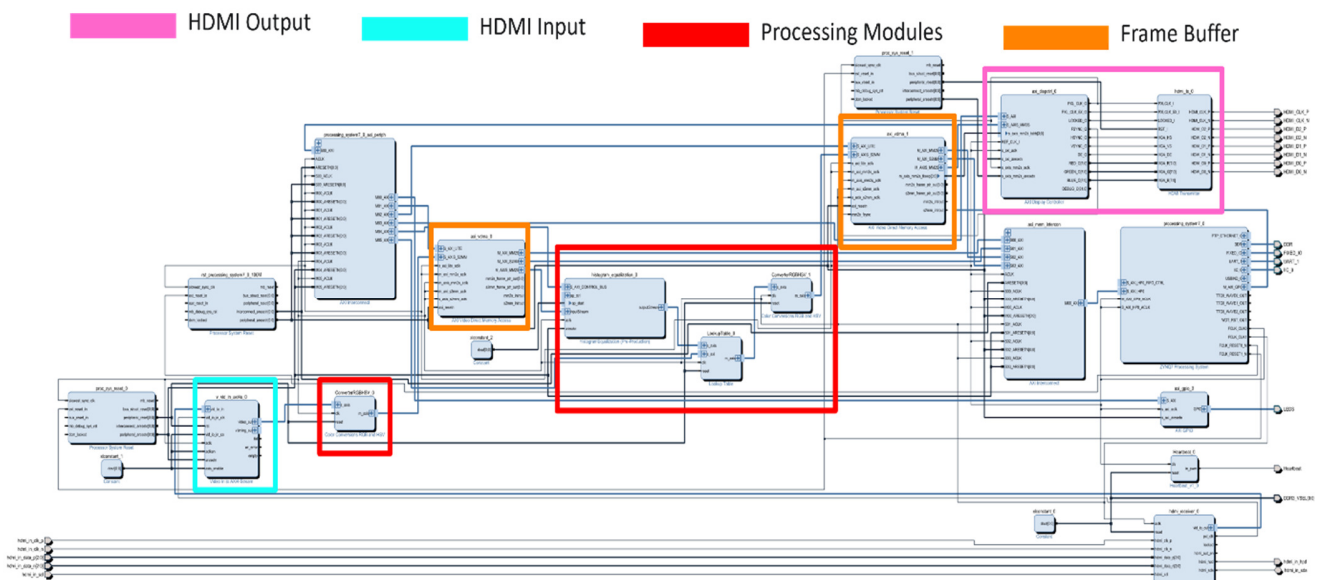


Fig. 4. RTL design of the proposed system based on Zynq 7000.

IV. CONCLUSION

This study uses a model-based real-time image processing system running on the Zynq 7000 development board. This study's hardware and software were developed applying the Zynq 7000 architecture due to the high processing power requirements. The Vivado Suite and related tools were employed for the system's hardware and software design, while Matlab/Simulink and Mathworks' HDL Coder and Vision HDL Toolbox were utilized for the design of the image processing systems. By adopting this approach, there is no requirement to develop HDL code manually, resulting in a reduction in design time. A Zedboard development board was used to design a system capable of operating at a frequency of 60 Hz and a resolution of 1280×1024 pixels. The system permits controlling the Canny edge detection algorithm for video processing through the input and output units on the card. This results in a versatile system that can easily adjust the settings. The designed system was capable of satisfying the requirements and operating in real-time. Nevertheless, resource utilization is believed to be rather low. The usage might be even more reduced, though, if the necessary optimizations are implemented in the HDL Coder tool. The proposed concept

was successfully implemented with full reusability. By adopting this approach, it will be feasible to reuse IP pieces or the entire system developed in subsequent research and across diverse systems. This technology can be utilized as a preprocessing unit to achieve more sophisticated real-time object detection and tracking applications in future designs. The implementation of hardware acceleration for corner edge detection was carried out on the Xilinx Zynq-7000 SoC hardware platform, specifically for image resolutions of 1920×1080. The simulation and synthesis were acquired employing Vivado 2017.4.

ACKNOWLEDGMENT

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through project number "NBU-FFR-2024-2225-01".

REFERENCES

- [1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 779–788, <https://doi.org/10.1109/CVPR.2016.91>.

- [2] R. Ghodhbani, T. Saidani, A. Alhomoud, A. Alshammari, and R. Ahmed, "Real Time FPGA Implementation of an Efficient High Speed Harris Corner Detection Algorithm Based on High-Level Synthesis," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12169–12174, Dec. 2023, <https://doi.org/10.48084/etasr.6406>.
- [3] T. Saidani and R. Ghodhbani, "Hardware Acceleration of Video Edge Detection with Hight Level Synthesis on the Xilinx Zynq Platform," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 8007–8012, Feb. 2022, <https://doi.org/10.48084/etasr.4615>.
- [4] T. Saidani, R. Ghodhbani, A. Alhomoud, A. Alshammari, H. Zayani, and M. B. Ammar, "Hardware Acceleration for Object Detection using YOLOv5 Deep Learning Algorithm on Xilinx Zynq FPGA Platform," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 13066–13071, Feb. 2024, <https://doi.org/10.48084/etasr.6761>.
- [5] CongJason *et al.*, "FPGA HLS Today: Successes, Challenges, and Opportunities," *ACM Transactions on Reconfigurable Technology and Systems (TRETS)*, Aug. 2022, <https://doi.org/10.1145/3530775>.
- [6] Z. Tan and J. S. Smith, "Real-time Canny Edge Detection on FPGAs using High-level Synthesis," in *2020 7th International Conference on Information Science and Control Engineering (ICISCE)*, Sep. 2020, pp. 1068–1071, <https://doi.org/10.1109/ICISCE50968.2020.00217>.
- [7] A. Fuentes-Alventosa, J. Gómez-Luna, and R. Medina-Carnicer, "GUD-Canny: a real-time GPU-based unsupervised and distributed Canny edge detector," *Journal of Real-Time Image Processing*, vol. 19, no. 3, pp. 591–605, Jun. 2022, <https://doi.org/10.1007/s11554-022-01208-0>.
- [8] F. Siddiqui *et al.*, "FPGA-Based Processor Acceleration for Image Processing Applications," *Journal of Imaging*, vol. 5, no. 1, Jan. 2019, Art. no. 16, <https://doi.org/10.3390/jimaging5010016>.
- [9] P. Babu and E. Parthasarathy, "Hardware acceleration for object detection using YOLOv4 algorithm on Xilinx Zynq platform," *Journal of Real-Time Image Processing*, vol. 19, no. 5, pp. 931–940, Oct. 2022, <https://doi.org/10.1007/s11554-022-01234-y>.
- [10] F. N. Taher, M. Kishani, and B. C. Schafer, "Design and Optimization of Reliable Hardware Accelerators: Leveraging the Advantages of High-Level Synthesis," in *2018 IEEE 24th International Symposium on On-Line Testing And Robust System Design (IOLTS)*, Platja d'Aro, Spain, Jul. 2018, pp. 232–235, <https://doi.org/10.1109/IOLTS.2018.8474222>.
- [11] "AXI4-Stream Video IP and System Design Guide," Xilinx, UG934, Oct. 2019.
- [12] Stephen Neuendorffer, Thomas Li, and Devin Wang, "Accelerating OpenCV Applications with Zynq-7000 All Programmable SoC using Vivado HLS Video Libraries," Xilinx, Aug. 2013.
- [13] K. Kintali and Y. Gu, "Model-Based Design with Simulink, HDL Coder, and Xilinx System Generator for DSP," Mathworks, 2015.
- [14] S. Liu *et al.*, "Real-time implementation of harris corner detection system based on FPGA," in *2017 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, Jul. 2017, pp. 339–343, <https://doi.org/10.1109/RCAR.2017.8311884>.
- [15] S. Chumpol, P. Solod, K. Thongnoo, and N. Jindapetch, "Model-Based Design Optimization using CDFG for Image Processing on FPGA," *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, vol. 17, no. 4, pp. 479–487, Oct. 2023, <https://doi.org/10.37936/ecti-cit.2023174.252417>.
- [16] S. Titri, C. Larbes, and K. Y. Toumi, "Rapid prototyping of PVS into FPGA: From model based design to FPGA/ASICs implementation," in *2014 9th International Design and Test Symposium (IDT)*, Algeria, Algeria, Dec. 2014, pp. 162–167, <https://doi.org/10.1109/IDT.2014.7038606>.
- [17] I. El Hajjouji, S. Mars, Z. Asrih, and A. El Mourabit, "A novel FPGA implementation of Hough Transform for straight lane detection," *Engineering Science and Technology, an International Journal*, vol. 23, no. 2, pp. 274–280, Apr. 2020, <https://doi.org/10.1016/j.jestch.2019.05.008>.

A Data Acquisition System with sEMG Signal and Camera Images for Finger Classification with Machine Learning Algorithms

Ismail Mersinkaya

Department of Electronics and Automation, Soke Vocational School, Aydin Adnan Menderes University, Türkiye | Biomedical Engineering Department, Faculty of Engineering, Karabuk University, Türkiye
imersinkaya@adu.edu.tr (corresponding author)

Ahmet Resit Kavsaoğlu

Medical Engineering Department, Faculty of Engineering, Karabuk University, Türkiye
kavsaoğlu@karabuk.edu.tr

Received: 8 February 2024 | Revised: 20 February 2024 | Accepted: 23 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7040>

ABSTRACT

Advances in robotics and biomedical engineering have expanded the possibilities of Human-Computer Interaction (HCI) in the last few years. The identification of hand movements is the accurate and real-time signal acquisition of hand movements through the use of image-based systems and surface electromyography sensors. This study uses multithreading to record motion signals from the forearm muscles in conjunction with a surface electromyography (sEMG) sensor and a camera image. The finger movement information labels were tabulated and analyzed along with the simultaneous acquisition of surface electromyography signals and these gestures through the camera. After the acquisition, signal processing techniques were applied to the sEMG signal marked from the camera. Therefore, once the interface is established, data sets suitable for machine learning can be generated.

Keywords-sEMG; image processing; real-time data acquisition; signal processing

I. INTRODUCTION

Surface electromyography (sEMG) signals are being explored for robotic control, but their multidimensional nature poses challenges in motion translation. Integrating sEMG signals with camera-based image signals is a promising approach. Robotic arms play a critical role in the integration of artificial intelligence. They are in high demand and have drawn numerous research investigations from both academia and business [1]. sEMG is a non-invasively technique, recording data from the skin's surface, and is thought to be a useful source for Human-Computer Interaction (HCI), which offers a natural manner of control. It represents a person's neuromuscular activity and neurological information [2]. Authors in [3] developed a sEMG signal acquisition instrument for body rehabilitation training. They used the ADS1294 chip and STM32 master hardware circuit to achieve the sEMG signal acquisition; then the acquisition software was to read and send the sEMG signal data. Authors in [4] reached a consensus on the acceptance of using a universal user interface model. The users' attitude towards the suggested application was reported as effective, pleasant, and enjoyable. They measured ease of use, consistency, operability, perceived usefulness, minimal memory load, system usability scale, understandability, and learnability through a survey

questionnaire. The results illustrate that their proposed solution is more robust, easy to use, and adaptable than other solutions operated through accessibility services [4]. Regarding real-time data acquisition, authors in [5] presented an automated weather station for real-time and local measurements based on an embedded system that continuously measures several weather factors.

II. MATERIALS AND METHODS

Nine healthy participants, with a mean age of 41.3 ± 12.0 years (two females and seven males), volunteered to take part in this study. Using the right hand, the subjects' sEMG signals and the camera-derived finger movement states of six positions (Thumb, Index, Middle, Ring, Pinky, and Open, which is the relaxed position of the fingers without movement) were simultaneously recorded in real time. None of the participants had a prior history of neuromuscular disease, and all were thoroughly informed about the experimental protocol prior to their participation.

A. Integrated Method: Combining Camera and sEMG Data

The current study proposes a novel way of HCI using forearm sEMG) biosignal data and camera image processing data. There are several steps in the process, which are depicted in the block diagram of Figure 1.

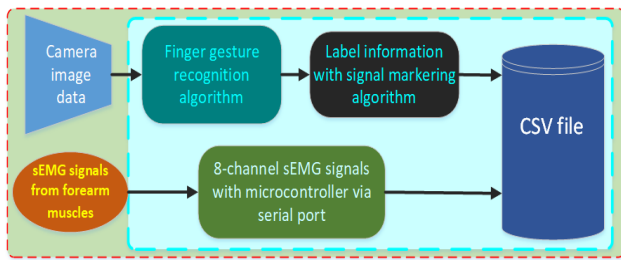


Fig. 1. Real-time hybrid data acquisition HCI for sEMG signals and camera images.

A finger identification algorithm records hand movement data acquired from the camera in order to identify hand movements. This was made possible by the Python-programmed HCI implementation shown in Figure 2. Thirty five seconds were recorded during the real-time sEMG and camera data collection. To precisely capture muscle action using the sEMG signal, an image processing package was used. In order for their movements to be classified, volunteers had to follow the instructions displayed on the screen for each movement (3 seconds on, 3 seconds off, 5 repeats). Multichannel wearable technologies, such the Myo Armband, greatly improve the recognition of numerous hand gestures from EMG data obtained using multichannel devices, providing several advantages over single-channel devices in biomedical applications [6].

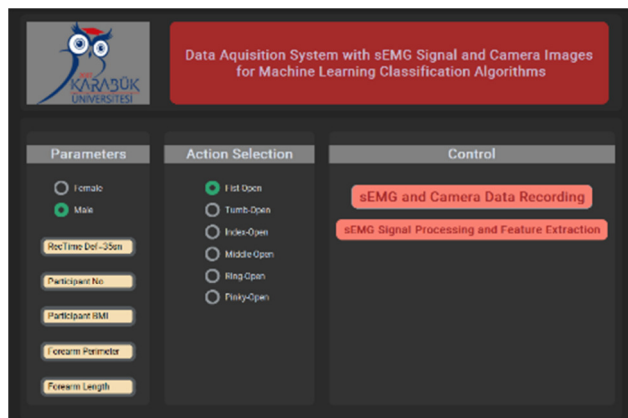


Fig. 2. sEMG-HCI data recording interface and finger identification algorithm.

Biosignal data were collected from every channel at a sampling frequency of 240 Hz. As shown in Figure 3, a microcontroller with a 16-bit ADC input was then used to transport the data to the computer environment via a serial port. A forearm band featuring dry surface electrodes was created in order to gather sEMG data from the forearm muscles. Eight sEMG amplifier circuits were linked to the electrodes by a 3.5" TTS cable, and a microcontroller with eight 16-bit ADC inputs was attached to the amplifier circuit's signal outputs. The USB cable of the microcontroller, which was connected to the data recording interface made in the computer environment for 8-channel sEMG recording, had the DC power line and the data line separated by a USB isolator in order to guarantee the accuracy of the data.

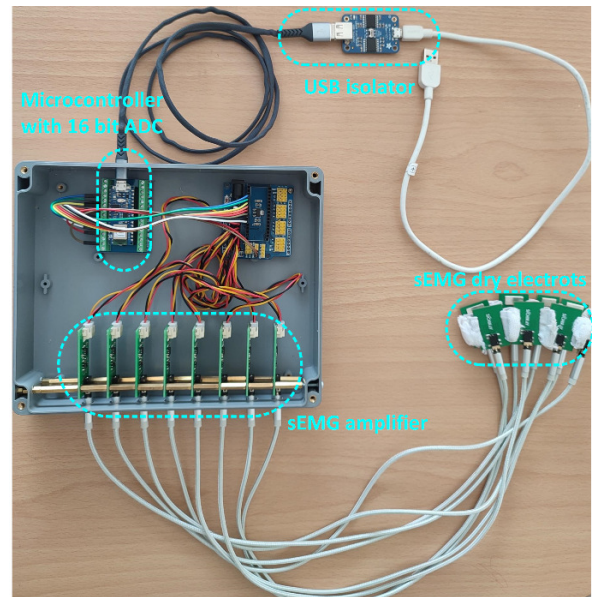


Fig. 3. Real-time hybrid data acquisition HCI for sEMG signals and camera images prototype.

The 8 sEMG sensor electrodes were placed on the cross-section of the forearm as sensors 5 and 4 [7] were replaced with sEMG-0 and sEMG-1, respectively, and the others were replaced in this order. The remaining sensors were put together into a ring and were assigned numbers. The study employed dry electrodes, and the (+), (-), and reference points were positioned side by side to form one electrode, obviating the need for a separate reference electrode.

B. Image Processing and Camera Image Data for Finger Detection

The camera picture data were generated by the MediaPipe (Hand Tracking) library, which uses digital image processing to identify finger joint points. Within the identified hand regions, the library is able to locate the crucial point of the 21 hand-joint coordinates [8]. With the MediaPipe library, the finger identification algorithm was developed as shown in Figure 4.

The hand was in an open position and was marked as "Open" if the number of fingers with function was equal to five. In order to label the sEMG signal for no movement in real-time, the marking subroutine has been invoked using the algorithm in the red-lined area. If the counted finger number was equal to 0 instead of 5, it meant that all fingers were closed and this gesture was marked as "Fist." To mark the sEMG signal for hand movement, the marking subroutine has been called. One finger had to be closed in order to have the smallest distance possible from the distance measurement if the counted fingers were equal to 4. To calculate the distance between each fingertip and the wrist point (0) in Figure 5, we used the Euclidean distance formula:

$$dst = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (1)$$

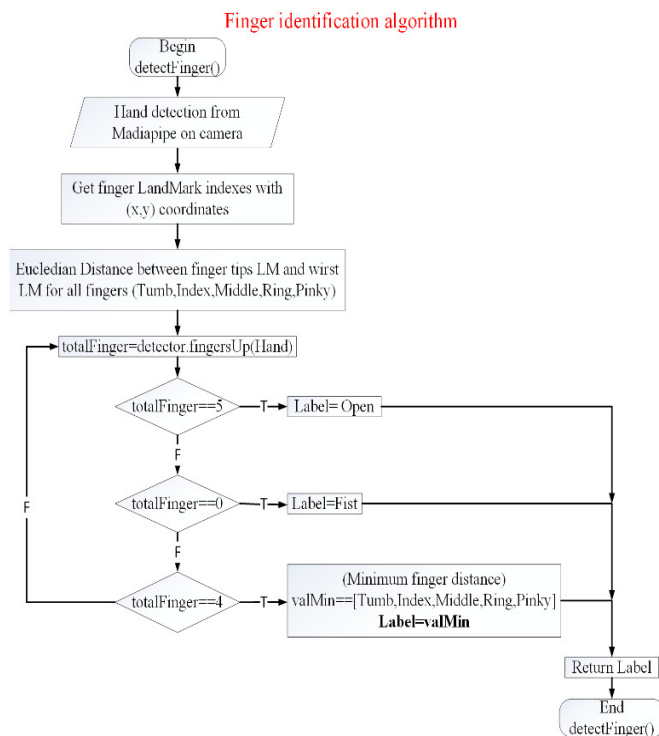


Fig. 4. Finger identification algorithm.

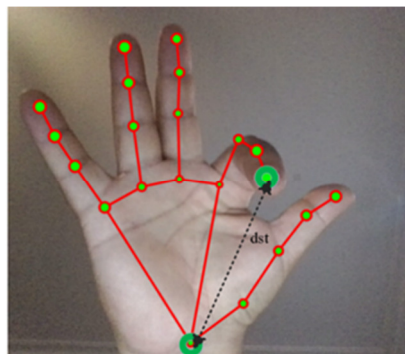


Fig. 5. The changeable distance between the terminal point (8) of the index finger and the wrist point (0).

To provide a regulated recording process for the participants, finger identifications were shown on the screen during the recording of sEMG signals (Figure 6). The interface screen featured finger motion data validated by the created algorithm in the left corner, a time counter in the center, and on-off motion commands in the right corner. This made it possible to record the same moment every time and to capture each movement more than once. The stages of picture processing were carried out at a speed of 15 fps.

As can be observed in Figure 7, the sEMG signal was registered on finger movement timings when the label changed in the signal marking subroutine. It indicated if volunteers closed all or just one finger. Their hand was considered open if they did not close any fingers. Because of the camera-based image processing, the start and end points of the sEMG signal were marked, and a valid dataset was constructed.



Fig. 6. Finger identification algorithm on recording.

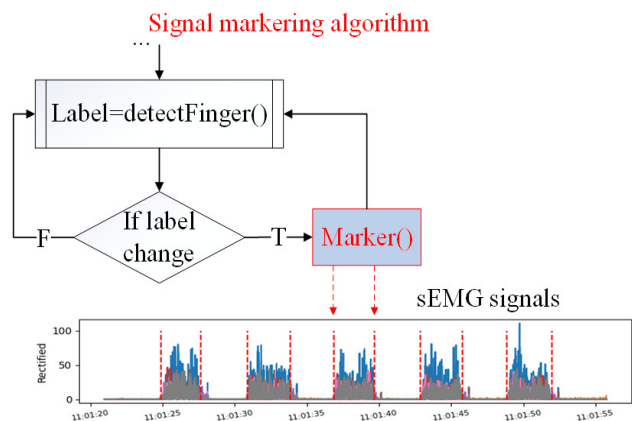


Fig. 7. Signal marking subroutine.

C. sEMG Signal Processing and Feature Extraction

In the study, the sEMG signal was first subjected to a band-pass filter to extract features. The filtered signal was then rectified and subjected to smoothing and overlapping windowing techniques [9]. A feature extraction function was developed by selecting a period of 150 ms and a processing time of 25 ms [10]. It can be difficult to distinguish between signals from different muscles and noise when interpreting EMG signals for limb motions. Using popular feature types like RMS, VAR, MAV, IEMG, WL, MNP, TP, and MNF, a high-performing EMG classification system was created to solve this problem [11]. These characteristics were chosen because they are common in earlier research.

Root Mean Square (RMS): RMS is a feature used in EMG analysis. Equation (2) illustrates how it relates to non-fatiguing contraction and constant force [12].

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (2)$$

Variance (VAR) is generally defined as the average of square values of the variable's deviation. However, the EMG signal's mean value is very near to zero ($\sim 10^{-10}$) [13].

$$\text{VAR} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (3)$$

In clinical practice, integrated electromyography (iEMG) is a measurement that is frequently used to identify patterns of muscle activity without depending on EMG data using (4). In patient assessments, it is frequently utilized as an initial detection index [14].

$$\text{iEMG} = \sum_{i=1}^N |x_i| \quad (4)$$

In MAV, $|x_i|$ is the magnitude of the i^{th} sample of x and N represents the total number of samples in a segment [15]:

$$\text{MAV} = \frac{1}{N} \sum_{i=1}^N |x_i| \quad (5)$$

WL is defined as the wavelength of an EMG signal over a segment [16]:

$$\text{WL} = \sum_{i=2}^N |x_i - x_{i-1}| \quad (6)$$

Furthermore, using (7)-(9), the frequency domain characteristics MNP, TP, and MNF are computed [17-19]:

$$\text{MNP} = \frac{1}{N} \sum_{i=1}^N x_i^2 \quad (7)$$

$$\text{TP} = \sum_{i=1}^N X(f_i)^2 \quad (8)$$

$$\text{MNF} = \frac{\sum_{i=1}^N f_i X(f_i)^2}{\sum_{i=1}^N X(f_i)^2} \quad (9)$$

As a result of applying signal processing techniques to the sEMG signal, the signals shown in Figure 8 were created.

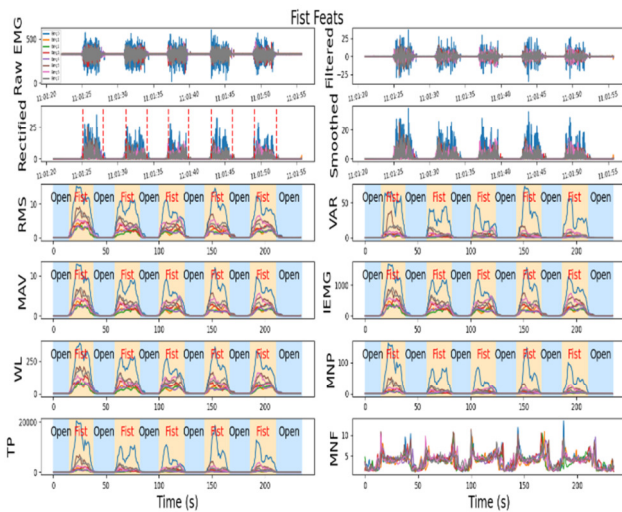


Fig.8. The results of applying signal processing techniques to the 8-channel sEMG signal.

With this study, 8-channel sEMG signal data can be recorded in real-time with the help of the finger identification algorithm on the created interface. With this approach, the process of recording muscle data with the sEMG sensor while users make finger movements can be done with visual guidance. With the recording data being more controlled, a functional interface has been developed to create the necessary data set for machine learning as a result of signal processing techniques.

III. CONCLUSION

The finger movement detection method in this system was created using image processing and multithreading. It comes with an armband that can record 8-channel sEMG data. In contrast to the sEMG hybrid data collection and classification research, the purpose of the current study was to assess the efficacy of a real-time hybrid data acquisition system. Table I presents a comparison of the developed system with related studies. Image data source, sEMG data source, and features are detailed in Table I. In this study, the sEMG signals and image data sources were used in real-time, which facilitated the comparison of data.

TABLE I. COMPARISON OF THE DEVELOPED SYSTEM WITH RELEATED STUDIES

Work	Image source	sEMG source	Features
[20]	LeapMotion	4-channel sEMG data aquisition system	MAV, ZC, Slope sign change, DAMV, VAR, joint angle, finger wrist distance
[21]	Dynamic Vision Sensor (DVS)	MyoArmband (8-channel sEMG band) 200 Hz	MAV, RMS
[22]	LeapMotion	MyoArmband (8-channel sEMG band) 200 Hz	RMS + LeapMotion features + feature fusion
[23]	Motion capture system with 12 cameras	Delsys's TrignoTM Wireless EMG System	RMS
[24]	Azure Kinect + Google Mediapipe	-	Movement area average, Coefficient of Variation
Proposed	PC Camera + Google Mediapipe	8-channel sEMG armband design 240 Hz	RMS, VAR, MAV, IEMG, WL, MNP, TP, MNF
	Real-Time data aquisition		

Promising outcomes have been found in the study of the 8-channel sEMG signal and its real-time motion estimation by image processing methods. The integration of sEMG signals with the MediaPipe library and finger gesture identification algorithm has demonstrated greater effectiveness in real-time acquisition, particularly when visual input sources are utilized. Thus, the real-time data collection interface design required to classify finger movements with machine learning algorithms could be realized with a new method.

ACKNOWLEDGMENT

The present research involved the acquisition of sEMG sensor and camera image processing data from volunteers, with the approval of Karabuk University Rectorate Non-Interventional Clinical Research Ethics Committee under the reference number 2022/846.

This study was supported by the Karabuk University Scientific Research Projects Coordination Unit. Project Number: KBÜBAP-22-DR-127.

The authors have no conflicts of interest to disclose.

REFERENCES

- [1] S. Husnain and R. Abdulkader, "Fractional Order Modeling and Control of an Articulated Robotic Arm," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12026–12032, Dec. 2023, <https://doi.org/10.48084/etasr.6270>.
- [2] Md. R. Ahsan, M. Ibrahimy, and O. Khalifa, "EMG Signal Classification for Human Computer Interaction A Review," *European Journal of Scientific Research*, vol. 33, pp. 480–501, Jan. 2009.
- [3] R. Zhou, K. Wang, and M. Li, "Design of a sEMG Signal Acquisition Instrument for Physical Rehabilitation Training," in *2017 10th International Symposium on Computational Intelligence and Design (ISCID)*, Hangzhou, China, Sep. 2017, vol. 2, pp. 136–139, <https://doi.org/10.1109/ISCID.2017.106>.
- [4] A. Khan, S. Khusro, and I. Alam, "BlindSense: An Accessibility-inclusive Universal User Interface for Blind People," *Engineering, Technology & Applied Science Research*, vol. 8, no. 2, pp. 2775–2784, Apr. 2018, <https://doi.org/10.48084/etasr.1895>.
- [5] M. Abdul-Niby, M. Farhat, M. Abdullah, and A. Nazzal, "A Low Cost Automated Weather Station for Real Time Local Measurements," *Engineering, Technology & Applied Science Research*, vol. 7, no. 3, pp. 1615–1618, Jun. 2017, <https://doi.org/10.48084/etasr.1187>.
- [6] K. Englehart, B. Hudgin, and P. A. Parker, "A wavelet-based continuous classification scheme for multifunction myoelectric control," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 3, pp. 302–311, Mar. 2001, <https://doi.org/10.1109/10.914793>.
- [7] J. Amezcua-García, M. Bravo-Zanoguera, F. F. Gonzalez-Navarro, R. Lopez-Avitia, and M. A. Reyna, "Applying Machine Learning to Finger Movements Using Electromyography and Visualization in Opensim," *Sensors*, vol. 22, no. 10, Jan. 2022, Art. no. 3737, <https://doi.org/10.3390/s22103737>.
- [8] "Gesture Recognition," *MediaPipe Studio*. https://mediapipe-studio.webapps.google.com/demo/gesture_recognizer.
- [9] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 7, pp. 848–854, Jul. 2003, <https://doi.org/10.1109/TBME.2003.813539>.
- [10] L. H. Smith, L. J. Hargrove, B. A. Lock, and T. A. Kuiken, "Determining the Optimal Window Length for Pattern Recognition-Based Myoelectric Control: Balancing the Competing Effects of Classification Error and Controller Delay," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 19, no. 2, pp. 186–192, Apr. 2011, <https://doi.org/10.1109/TNSRE.2010.2100828>.
- [11] K. Xing, P. Yang, J. Huang, Y. Wang, and Q. Zhu, "A real-time EMG pattern recognition method for virtual myoelectric hand control," *Neurocomputing*, vol. 136, pp. 345–355, Jul. 2014, <https://doi.org/10.1016/j.neucom.2013.12.010>.
- [12] M. Ariyanto *et al.*, "Finger movement pattern recognition method using artificial neural network based on electromyography (EMG) sensor," in *2015 International Conference on Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology (ICACOMIT)*, Bandung, Indonesia, Jul. 2015, pp. 12–17, <https://doi.org/10.1109/ICACOMIT.2015.7440146>.
- [13] K. H. Lee, J. Y. Min, and S. Byun, "Electromyogram-Based Classification of Hand and Finger Gestures Using Artificial Neural Networks," *Sensors*, vol. 22, no. 1, Jan. 2022, Art. no. 225, <https://doi.org/10.3390/s22010225>.
- [14] S. Bhagwat and P. Mukherji, "Electromyogram (EMG) based fingers movement recognition using sparse filtering of wavelet packet coefficients," *Sādhanā*, vol. 45, no. 1, Dec. 2019, Art. no. 3, <https://doi.org/10.1007/s12046-019-1231-9>.
- [15] S. Kim *et al.*, "Development of an Armband EMG Module and a Pattern Recognition Algorithm for the 5-Finger Myoelectric Hand Prosthesis," *International Journal of Precision Engineering and Manufacturing*, vol. 20, no. 11, pp. 1997–2006, Nov. 2019, <https://doi.org/10.1007/s12541-019-00195-w>.
- [16] M. A. Oskoei and H. Hu, "GA-based Feature Subset Selection for Myoelectric Classification," in *2006 IEEE International Conference on Robotics and Biomimetics*, Sep. 2006, pp. 1465–1470, <https://doi.org/10.1109/ROBIO.2006.340145>.
- [17] A. Saikia *et al.*, "Combination of EMG Features and Stability Index for Finger Movements Recognition," *Procedia Computer Science*, vol. 133, pp. 92–98, Jan. 2018, <https://doi.org/10.1016/j.procs.2018.07.012>.
- [18] P. Phukpattaranont, S. Thongpanja, K. Anam, A. Al-Jumaily, and C. Limsakul, "Evaluation of feature extraction techniques and classifiers for finger movement recognition using surface electromyography signal," *Medical & Biological Engineering & Computing*, vol. 56, no. 12, pp. 2259–2271, Dec. 2018, <https://doi.org/10.1007/s11517-018-1857-5>.
- [19] H. Su, S. E. Ovr, X. Zhou, W. Qi, G. Ferrigno, and E. De Momi, "Depth vision guided hand gesture recognition using electromyographic signals," *Advanced Robotics*, vol. 34, no. 15, pp. 985–997, Aug. 2020, <https://doi.org/10.1080/01691864.2020.1713886>.
- [20] H. Zhou *et al.*, "Toward Hand Pattern Recognition in Assistive and Rehabilitation Robotics Using EMG and Kinematics," *Frontiers in Neurobotics*, vol. 15, 2021, Art. no. 659876, <https://doi.org/10.3389/fnbot.2021.659876>.
- [21] E. Ceolini *et al.*, "Hand-Gesture Recognition Based on EMG and Event-Based Camera Sensor Fusion: A Benchmark in Neuromorphic Computing," *Frontiers in Neuroscience*, vol. 14, 2020, Art. no. 00637, <https://doi.org/10.3389/fnins.2020.00637>.
- [22] J. Li, J. Zhong, and N. Wang, "A multimodal human-robot sign language interaction framework applied in social robots," *Frontiers in Neuroscience*, vol. 17, 2023, Art. no. 1168888, <https://doi.org/10.3389/fnins.2023.1168888>.
- [23] Q. Song, X. Ma, and Y. Liu, "Continuous online prediction of lower limb joints angles based on sEMG signals by deep learning approach," *Computers in Biology and Medicine*, vol. 163, Sep. 2023, Art. no. 107124, <https://doi.org/10.1016/j.compbiomed.2023.107124>.
- [24] G. Amprimo, C. Ferraris, G. Masi, G. Pettiti, and L. Priano, "GMH-D: Combining Google MediaPipe and RGB-Depth Cameras for Hand Motor Skills Remote Assessment," in *2022 IEEE International Conference on Digital Health (ICDH)*, Barcelona, Spain, Jul. 2022, pp. 132–141, <https://doi.org/10.1109/ICDH55609.2022.00029>.

Advancing IoT Cybersecurity: Adaptive Threat Identification with Deep Learning in Cyber-Physical Systems

C. Atheeq

GITAM University, India

atheeq.prof@gmail.com (corresponding author)

Ruhiat Sultana

Lords Institute of Engineering and Technology, India

ruhiatsultana@lords.ac.in

Syeda Asfiya Sabahath

King Khalid University, Saudi Arabia

assyed@kku.edu.sa

Murtuza Ahmed Khan Mohammed

Universiti Teknologi Malaysia

ahmedkhan@graduate.utm.my

Received: 27 January 2024 | Revised: 16 February 2024 | Accepted: 18 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6969>

ABSTRACT

Securing Internet of Things (IoT)-enabled Cyber-Physical Systems (CPSs) can be challenging because security solutions intended for typical IT/OT systems may not be as effective in a CPS setting. The goal of this study is to create a mechanism for identifying and attributing two-level ensemble attacks that are specifically designed for use against Industrial Control Systems (ICSs). An original ensemble deep representation learning model is combined with decision tree algorithm to identify assaults on unbalanced ICS environments at the first level. An attack attribution network, which constitutes a collection of deep neural networks, is formed at the second level. The proposed model is tested using real-world datasets, notably those pertaining to water purification and gas pipelines. The results demonstrate that the proposed strategy outperforms other strategies with comparable computing complexity and that the recommended model outperforms the existing mechanisms.

Keywords-cyber-attacks; deep learning; threat detection; industrial control system; industrial IoT; cyber-physical systems

I. INTRODUCTION

Cyber-Physical Systems (CPSs) are increasingly fusing IoT-connected devices, even in areas of critical infrastructure like dams and power plants [1-7]. The Industrial Control System (ICS), which is in charge of ensuring that the infrastructure is operating efficiently, commonly includes connected devices to the Internet of Things (IoT), sometimes referred to as Industrial IoT or IIoT in certain contexts. ICSs may also refer to Distributed Control Systems (DCSs), Programmable Logic Controllers (PLCs), and Modbus protocols in addition to Supervisory Control And Data Acquisition (SCADA) systems. On the other side, linking IoT or ICS to public networks expands their attack surfaces and the likelihood that hackers will target them. One famous

example is the 2010 Stuxnet attack, which reportedly caused serious damage to Iranian centrifuges used for nuclear enrichment. An assault against Iranian centrifuges happened in 2010. Another illustration of this is the Illinois pump attack that occurred in 2011 and led to the shutdown of the state's water treatment facility.

Physical behavior analysis and ensuring that system operations are always available require system-level security solutions. Unlike most IT and OT systems, which normally emphasize availability, integrity, and confidentiality, ICS security objectives are prioritized in the reverse order. ICS attacks have the potential to have detrimental effects. This highlights the value of having exceedingly strong security and safety processes in place to identify and prevent attacks

that target ICSs [8–10]. Attack identification and attribution methods based on anomalies and signatures are widely utilized. Hybrid-based detection and attribution systems have been created to solve the recognized limitations of signature-based and anomaly-based detection and attribution techniques. Hybrid-based approaches are effective at identifying anomalous behavior, but they are unreliable since they cause distinctive Intrusion Detection System (IDS) typologies due to frequent network upgrades. Hybrid-based approaches can identify strange conducts well, but are unreliable. Traditional methods for the detection and attribution of attacks heavily rely on the study of network metadata, including things like transmission ports, traffic volume, packet intervals, and IP addresses. Recently, there has been a revived interest in attack detection and attribution techniques based on Deep Neural Networks (DNNs) or Machine Learning (ML). Numerous cyber-physical systems have been created as a result of the IoT proliferation, allowing industrial equipment and IT processes to transmit and receive data over the internet [11-13]. These systems rely on sensors to detect the condition of equipment and report their findings to a centralized server using an internet connection. However, these sensors can be vulnerable to attacks from malicious actors, leading to false reports being sent to the centralized server and incorrect actions being taken. In response to this threat, researchers have developed algorithms to detect attacks on IoT-enabled CPSs. However, the data are usually unbalanced (e.g. normal records are usually significantly more than attack records). This imbalance can lead to inaccurate predictions and detection failures [14-17]. To address this issue, researchers have proposed a novel technique that engages an autoencoder and DNNs to detect and attribute assaults from the IoT on CPSs [18]. The technique consists of two modules. The first one involves training an autoencoder on an imbalanced dataset and utilizing the extracted features to train a decision tree algorithm to predict attack labels. In the second module, a DNN is trained on known and unknown attacks to identify attack labels or classes [19, 20]. To test the effectiveness of this technique, the Secure Water Treatment (SWaT) dataset was utilized, which contains IoT request and response signatures associated with unique attack labels. The dataset includes various types of cyber-attacks, such as command injection, water level manipulation, and flow rate tampering [21-23].

CPSs are made possible by the IoT. However, protecting IoT-enabled CPS might be challenging since security protocols designed for standard IT/OT systems might not work as well in a CPS setting [24-26]. This study aims to provide a framework for categorizing and assigning responsibility for two-level ensemble attacks created specifically for usage in an ICS and designed for CPS. In an original ensemble deep learning method to detect attacks on unbalanced ICS settings at the first level, a decision tree algorithm and a representation learning model are used together. At the second level, a group of DNNs called attack attribution network is created. Real-world datasets are applied to test the suggested model, notably those about

water purification and gas pipelines. The results demonstrate that the proposed model outperforms rivals' tactics at a comparable degree of computer complexity.

According to the mechanism put out in [1], the increasing frequency of cyber-physical system attacks in recent years raises questions about the ICS cyber security. The majority of the current ICS cyber security efforts are based on firewalls, data diodes, and other intrusion prevention techniques. However, it is possible that these strategies will not be enough to guard against the rising number of cyber threats posed by determined attackers. A defense-in-depth-based cyber-attack detection system is currently being developed to raise the level of cyber security provided by ICS. Data from the host system, network traffic, and process parameter measurements will be used by this system. The latter also provides multiple levels of defense in order to buy the defenders some crucial extra time before the physical system is irreparably harmed.

The model developed in [2] is based on the Stealthy Attack against Redundant Controller Architecture. The controller is a crucial component that must be present in an Industrial Cyber-Physical System (ICPS) to guarantee dependability and stability. Several businesses employ the redundant controller architecture technique, including those that use DCS, SCADA, and other common ICPSs. Power production, chemical industry, water treatment, and other crucial industrial processes are under their strict observation and control. Given that some mechanical failures are unpredictable, redundant controller architecture has been created and, to a large part, implemented. This structure, which was first recommended for ensuring dependability and safety, nevertheless, has the potential to increase the surface area vulnerable to cyberattacks. As a result, there is a chance that a hostile entity could use its design as a cover for stealth strikes. The vulnerability caused by the redundant controller design is analyzed, and a combined attack methodology that may be utilized covertly against systems that employ the redundant controller architecture is provided.

Electric grids, water networks, and transport systems are all important metropolitan infrastructure that is frequently the target of cyber-attacks. A method for guaranteeing the cyber security of such critical urban infrastructure was devised in [4]. The network of linked objects that makes up these systems is referred to as the IIoT. An assault on key urban infrastructure IIoT would have a profoundly detrimental effect on society. SCADA systems are frequently used in IIoT control for critical urban infrastructure. Although it is vital to comprehend the cyber threat to the latter, there is currently no data-driven method for assessing the risk that SCADA software poses to IIoT devices. Using cosine similarity tests, this paper involves a comparison of SCADA systems to other types of control systems and it is found that SCADA, as a software subclass, carries particular risk characteristics for the IIoT. The widely held belief that the SCADA subclass of software is not vulnerable to attack is refuted based on the standard vulnerability score system risk criteria of exploitability and impact. In order to identify

SCADA risk metrics, a variety of statistical models were created. These models can be used to calculate the likelihood that a SCADA vulnerability will be exploited. Authors in [5] describe a method for the identification of data abnormalities. A developing trend in conventional industrial systems is the integration of the physical and cyber realms. The goal of this integration is to increase the adaptability and effectiveness of management, control, and supervision. The risk of security breaches has increased due to the deep integration of ICPSSs. An essential component of the overall security protection offered is the initial protective barrier that attack detection generates. Contrarily, the majority of conventional systems focuses primarily on digital information and ignores any potential limitations that might result from the characteristics of the physical world. This study develops a zone partition-based strategy for the detection of anomalies in ICPSSs. To ensure that important system states can be observed in several zones, an automatic zone division mechanism is created in the initial step of this process. Authors in [7] suggested an intrusion detection system similar to that in [6]. ICS relied on "air-gap" security measures until recently. As a result, every node of the ICS network was physically cut off from the Internet and all other networks. Businesses and professionals that use ICS networks benefit from connecting them to the internet. The protocols employed by ICSs, however, entail extremely few or no security measures because these systems were created for use in an environment with a high level of air-gapped security, which makes them vulnerable to a number of attacks. The described approach for detecting intrusions into network-attached ICSs employs network telemetry. With the aid of simulated PLC units, the newly developed IDS was able to distinguish between the machines used by an engineer and an attacker on the same network with an accuracy rate of 94.3% and between an engineer and an attacker on the Internet with a rate of 99.5%.

The literature review agrees that cyber-physical systems are becoming more risky, emphasizing the need for ICS cyber security. Most research focuses on redundant controller topologies and covert attacks or on SCADA systems' IIoT concerns in important urban infrastructures. Further research demonstrates that ICPSSs' integration of cyber and physical realms requires increased anomaly detection and that the transition from air-gapped to network-attached systems demands unique intrusion detection to protect against sophisticated cyber-attacks.

II. THE PROPOSED SYSTEM

The increasing reliance on IoT-enabled CPSs, including operational IT and industrial machinery, has led to a growing concern about cyber-attacks targeting these systems. Attack detection and attribution in these systems can be challenging due to the imbalanced nature of the training data, where one class may contain significantly more samples than the other. Traditional machine learning algorithms often struggle to accurately detect attacks in such imbalanced datasets. This research study presents an original deep learning-based assault detection and attribution approach using a two-stage

ensemble methodology that has been specifically created for imbalanced ICS data.

Traditional sampling methods, such as over- and under-sampling, might be ineffective when dealing with imbalanced datasets. In order to enable the DNN to handle unbalanced datasets, this research aims to introduce a novel deep representation learning approach that does not alter, produce, or remove any samples. Although the suggested framework has a complex design, the amount of training samples required (n) is comparable to that of other known DNN-based techniques. To evaluate the proposed framework's effectiveness, we used the SWaT dataset, which contains IoT request and response signatures associated with unique attack labels. The proposed framework demonstrated better recall and f-measure in detecting and attributing attacks than those exhibited by previous works.

The proliferation of data volume has increased as a result of IoT technologies in CPSs, including industrial machinery and operational ITs. These systems use sensors to monitor the condition of equipment, which sends data to a centralized server using internet connections. However, there is a risk of malicious users hacking into these sensors and altering the transmitted data, which could result in false actions being taken. To address this problem, attack detection algorithms have been developed. However, these algorithms frequently are faced with imbalanced datasets that may lead to inaccurate predictions.

The first part of the proposed technique involves implementing an autoencoder deep learning algorithm to extract features from an imbalanced dataset. The autoencoder is trained on the dataset, and then the extracted features are utilized to train a decision tree algorithm to predict whether an attack is labeled as known or unknown. Principal Component Analysis (PCA)'s reduced set of features is applied to train the decision tree. By using an autoencoder, we are able to extract meaningful features from the imbalanced dataset, which can then be employed to train a more accurate decision tree algorithm. The second component of the proposed technique puts into service a DNN algorithm to train on both known and unidentified threats. The DNN will recognize the attack label or class and attribute it if a record includes an attack signature. The ability to precisely identify both known and unidentified attacks using a DNN is crucial for spotting and avoiding cyber-attacks on CPS with IoT support. For the goal of locating and determining who is responsible for cyberattacks in IoT-enabled CPSs, the proposed technique has several advantages over the existing techniques. First, it avoids under- or over-sampling methods, which might produce incorrect predictions. Second, the proposed method enhances the decision tree algorithm's accuracy by employing an autoencoder to extract useful features from unbalanced datasets. Also, the proposed method utilizes a DNN to precisely identify both known and unidentified attacks, which is essential for spotting and avoiding cyberattacks in IoT-enabled CPSs. For imbalanced ICS data, assault detection and attack attribution is conducted adopting a

unique two-stage ensemble deep learning approach. An autoencoder and a DNN are utilized to accurately detect known and unidentified assaults and extract features, respectively. The SWaT dataset was used to test the suggested method, which performed better than other known methods that employed under- or over-sampling strategies. The proposed method could be utilized to identify and stop cyber-attacks in IoT-enabled CPSs due to its advantages over the current techniques.

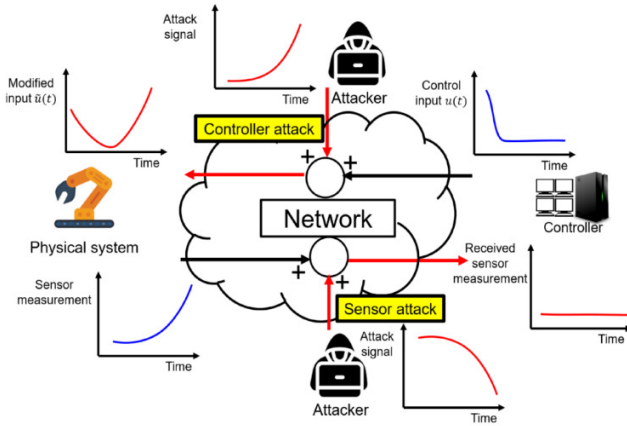


Fig. 1. The proposed architecture.

A. Algorithm: Two-Stage Ensemble Deep Learning for IoT Cyber-Attack Detection and Attribution (2SDL-CA)

1) Step 1: Data Collection

Collect data from sensors and define the dataset D as a set of records:

$$D = \{(x_i, y_i) | x_i \in \mathbb{R}^n, y_i \in \{0, 1\}\} \quad (1)$$

where x_i represents a data vector of dimension n and y_i represents labels (0 for normal, 1 for attack).

2) Step 2: Data Preprocessing

Split the dataset D into D_{normal} (normal records) and D_{attack} (attack records):

$$D_{normal} = \{(x_i, y_i) | (x_i, y_i) \in D, y_i = 0\} \quad (2)$$

$$D_{attack} = \{(x_i, y_i) | (x_i, y_i) \in D, y_i = 1\} \quad (3)$$

3) Step 3: Autoencoder Feature Extraction

Utilize an autoencoder deep learning algorithm to extract features from the imbalanced dataset D . The autoencoder is trained on the dataset, and the extracted features are used as D_{PCA} to train a decision tree: $D_{PCA} = \text{Autoencoder}(D)$.

4) Step 4: Decision Tree Training

Train a decision tree algorithm (DT) employing the reduced features obtained from PCA:

$$DT(D_{PCA}) = \operatorname{argmind} \left(\sum_{i=1}^n \sum_{j=1}^m (x_{ij} - \mu_j)^2 \right) \quad (4)$$

where D_{PCA} contains the reduced features z_i obtained through PCA, n is the number of features, m is the number of data points, x_{ij} is the value of feature i for data point j , and μ_j is the mean value of feature i for all data points.

5) Step 5: Decision Tree Prediction

Use the trained DT to predict whether an attack is known or unknown for new data points.

6) Step 6: Deep Neural Network (DNN) Training

Train a DNN model (DNN) on both known and unknown threats:

$$DNN(D) = \operatorname{argmin}_{\theta} \left(\sum_{i=1}^m L(y_i, f_{\theta}(x_i)) \right) \quad (5)$$

where D represents the entire dataset, m is the number of data points, θ represents the model parameters, L is the loss function, y_i is the true label, and $f_{\theta}(x_i)$ is the predicted label for data point i .

7) Step 7: DNN Prediction

Use the trained DNN to identify the attack label or class and attribute it when a record includes an attack signature.

8) Step 8: Evaluation and Testing

Evaluate the proposed method's performance utilizing datasets. Performance metrics include precision (P), recall (R), F1-score ($F1$), and accuracy (ACC), which are computed based on the true labels y_i . F1 is the harmonic mean of P and R , providing a balance between the two metrics.

$$P = \frac{TP}{TP+FP} \quad (6)$$

$$R = \frac{TP}{TP+FN} \quad (7)$$

$$F1 = \frac{2 \cdot P \cdot R}{P+R} \quad (8)$$

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

where TP (True Positives) represents the number of correctly identified attacks, FP (False Positives) represents the number of normal instances incorrectly identified as attacks, FN (False Negatives) represents the number of actual attacks incorrectly classified as normal, and TN (True Negative) represents the correctly identified normal instances. The DNN Prediction Probability (P_{pred}) is:

$$P_{pred}(y_i = 1) = \operatorname{sigmoid}(\sum_{j=1}^n w_j x_{ij} + b) \quad (10)$$

where n is the number of features, w_j symbolizes the weights associated with each feature, x_{ij} represents the j -th feature of data point i , b is the bias term, and $\operatorname{sigmoid}(\cdot)$ is the sigmoid activation function.

DNN Model Loss (L) is defined by:

$$L(y_i, f_{\theta}(x_i)) = - \left(y_i \log(P_{pred}(y_i)) + (1 - y_i) \log(1 - P_{pred}(y_i)) \right) \quad (11)$$

The loss function L measures the dissimilarity between true labels y_i and predicted probabilities $P_{pred}(y_i)$ for each data point. It is commonly used in binary classification problems. These complex formulas help assess the performance of the model through various metrics and describe the probability prediction generated by the DNN during testing. Partially connected components, inspired by dropout techniques, introduce random dropout of neurons during training. The fully connected layers integrate the high-level features extracted by the autoencoder. The output y_i of a neuron in the fully connected layer is calculated adopting the parametric rectified linear unit (PReLU) activation function. The mathematical formulation for dropout during training is given by:

$$x_{out} = x_{in} \odot \left(\text{Mask} \cdot \frac{1}{1 - \text{dropout}_{rate}} \right) \quad (12)$$

$$\text{Mask} \sim \text{Bernoulli}(1 - \text{dropout}_{rate})$$

$$y_i = \text{PReLU}(\sum_j w_{ij} \cdot x_j + b_i) \quad (13)$$

$$\text{PReLU}(x) = \max(0, x) + a_i \cdot \min(0, x) \quad (14)$$

$$a_i \sim \text{Uniform}(0.01, 0.1)$$

where \odot represents element wise multiplication, w_{ij} represents the weight between the i -th neuron and the j -th input, x_j is the j -th input, b_i is the bias term for the i -th neuron, and a_i is a learnable parameter.

B. Attack Attribution Mechanism:

The DNN is trained to recognize and attribute cyber-attacks based on learned features. Let X represent the input features, Y denote the true labels, and Y' signify the predicted labels. The training process involves minimizing a novel attack attribution loss function:

$$L_{attr}(\theta) = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C Y_{ic} \cdot \log(Y'_{ic}) \cdot (1 - \text{Attribution}_{ic}) \quad (15)$$

where Attribution_{ic} is a binary indicator (0 or 1) specifying whether an instance is attributed to a known attack class. The ensemble DNN excels in handling both known and unknown threats. By employing a one-vs-all approach, individual classifiers within the ensemble are specialized for different attack attributes. The final prediction is determined through a weighted combination of these individual predictions, promoting a comprehensive and nuanced attribution process.

$$\text{Final Prediction} = \text{softmax}(\sum_i w_i \cdot \text{Classifier}_i(X)) \quad (16)$$

where w_i represents the weight associated with the i -th classifier Classifier_i .

C. Model Robustness

Ensuring the robustness of the proposed model is crucial for practical deployment in ICS environments. To enhance model robustness, adversarial training techniques, such as the Fast Gradient Sign Method (FGSM), were incorporated during the training phase. Additionally, to improve model robustness, integrate methods like Layer-wise Relevance

Propagation (LRP) were utilized to highlight important features contributing to the final decision. To strengthen the model's defenses against possible attacks, hostile instances are added to the training dataset through the process of adversarial training. Generating adversarial instances is a common application of the Fast Gradient Sign Method (FGSM). The definition of the adversarial loss function (L_{adv}) is:

$$L_{adv}(\theta) = \frac{1}{N} \sum_{i=0}^N \max(0, \|\theta - \epsilon \cdot \text{sign}(\nabla_{\theta} J(\theta, x_i, y_i))\|_2 - \|\theta\|_2) \quad (17)$$

where θ represents the model parameters, N is the number of training samples, ϵ controls the magnitude of the perturbation, $J(\theta, x_i, y_i)$ is the model's loss function for the input x_i and true label y_i , $\nabla_{\theta} J(\theta, x_i, y_i)$ is the gradient of the loss with respect to the model parameters.

LRP is a technique used to attribute the model's decision to input features, providing insight into feature importance. The relevance (R_i) for each input feature i is calculated applying the LRP formula. This recursive formula propagates relevance from the output layer to the input layer, highlighting the contribution of each feature to the final decision. To optimize model robustness, an integrated objective ($L_{integrated}$) can be defined as a combination of the adversarial loss and the LRP-based loss:

$$R_i = \sum_j \frac{a_{ij} R_j}{\sum_k a_{ik}} \quad (18)$$

$$L_{integrated}(\theta) = \alpha \cdot L_{adv}(\theta) - \beta \cdot L_{expl}(\theta) \quad (19)$$

where α and β are hyperparameters controlling the trade-off for adversarial training, $L_{expl}(\theta)$ is the explainability loss derived from LRP, a_{ij} represents the connection weight between the i -th input and the j -th neuron in the preceding layer, and R_j is the relevance of the j -th neuron in the preceding layer.

D. Computational Efficiency Considerations

Efficiency is a critical factor for the practical deployment of the proposed ensemble DNN in large-scale ICS environments. Focus is placed on two aspects: model size reduction through weight pruning and knowledge distillation, and inference time reduction through model quantization. To reduce the model size by eliminating less significant weights, the pruning mask M_{ij} is applied to the weight w_{ij} . To transfer knowledge from the ensemble DNN ($DNN_{ensemble}$) to a smaller, distilled DNN ($DNN_{distill}$), the distillation loss is minimized during training:

$$w_{ij}^{pruned} = w_{ij} \cdot M_{ij} \quad (20)$$

$$L_{distill} = \frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C T_c \cdot \log \left(\sigma \left(\frac{y_{ic}}{T} \right) \right) \quad (21)$$

where M_{ij} is determined based on a predefined threshold, T_c is the softened target distribution from the ensemble model, σ is the softmax function, and T is the temperature parameter.

Quantization reduces the precision of weights and activations, decreasing the memory footprint and speeding

up inference. The quantized weight $w_{ij}^{quantized}$ is given in (22). The predicted inference time t_{pred} can be estimated using a linear regression model:

$$w_{ij}^{quantized} = Round\left(\frac{w_{ij}}{\Delta}\right) \cdot \Delta \quad (22)$$

$$t_{pred} = \alpha \cdot Size_{model} + \beta \cdot Size_{input} + \gamma \quad (23)$$

where Δ is the quantization step size, α , β , and γ are the regression coefficients, and $Size_{model}$ and $Size_{input}$ represent the sizes of the model and input data, respectively.

III. RESULTS AND DISCUSSION

To implement the proposed technique, the SWaT dataset, which contains IoT request and response signatures, associated with unique attack labels was utilized. The dataset includes various types of cyber-attack labels, such as replay, denial-of-service, and false data injection attacks. The proposed technique was evaluated based on its ability to use the SWaT dataset to identify and attribute cyber-attacks. The findings demonstrate that the proposed method can reliably identify and attribute cyber-attacks even in the presence of

imbalanced data. The proposed technique outperformed existing algorithms that used under- or over-sampling techniques, indicating its effectiveness.

In Table I, Test Case ID is a unique identifier for each test case conducted in this study. Test Case Name provides a brief description of what is being tested each time. Test Case Description provides further details on the test case and its purpose. Test Step field outlines the steps taken to complete each test case. Test Case Status describes whether the test case was successful or not. Test Priority indicates the level of importance assigned to each test case. The first test case, Test Case ID 01, aimed to test whether the Personality Dataset was successfully uploaded into the system. Further operations could not be conducted until the dataset was successfully uploaded. This test case was assigned a high priority due to its importance in the study. The second test case, Test Case ID 02, aimed to test whether the preprocessed dataset was successfully uploaded into the system, which was the expected result. In this case, further operations could not be conducted until the dataset's successful upload was achieved. This test case was also assigned a high priority due to its importance in the study.

TABLE I. ACTIONS ON THE DATASET FOR THE PROPOSED MODEL

Test Case Id	Test Case Name	Test Case Description	Test Steps			Test Case Status	Test Priority
			Step	Negative	Positive		
01	Upload SWaT dataset	Check if the dataset has been uploaded to the system.	Perhaps the dataset was not uploaded.	Any further operations cannot be performed.	Dataset uploaded. Further operations will be done.	High	High
02	Upload preprocessed dataset	Check if the preprocessed dataset has been uploaded to the system.	Perhaps the dataset was not uploaded.	Any further operations cannot be performed.	Dataset uploaded. Further operations will be done.	High	High
03	Run auto encoder algorithm	Test whether auto encoder algorithm was run successfully or not.	Is the encoder sent.	The algorithm cannot be run.	Further operations can be conducted.	High	High
04	Run decision tree with PCA	Test whether decision tree was run successfully or not.	PCA.	Extended PCA cannot be performed.	Further operations can be conducted.	High	High
05	Run DNN algorithm	Verify run rule biased attack detector.	Detection.	DNN cannot be run.	Further operations can be conducted.	High	High
06	Detection and attack attribute type	Verify if the attack detection graph was successful or not.	Attribute detection.	Attribute detection cannot be done.	Further operations can be conducted.	High	high
07	Comparison graph	Verify run rule biased attack detector.	Without comparison.	The graph cannot be done.	Further operations can be conducted.	High	High
08	Comparison table	Verify if the attack detection graph was successful or not.	Without comparison table.	The table cannot be done.	Further operations can be conducted.	High	high

Test Case ID 03 tested the auto encoder algorithm success. The method was anticipated to run, but if the encoder was not supplied, it may not have achieved this target. Without fixing the algorithm, no more operations could be done. High priority was given to this test case also. All the test cases were completed and the findings were reported. The dataset is shown in Figure 2 as a graph with the name of the attack on the x-axis and the number of times that it was detected in the dataset on the y-axis. As it can be observed, the attacks have fewer records than the Normal class, which results in data imbalance. This issue can be corrected using DNN, Autoencoder, or Decision Tree. The dataset was then preprocessed to fill in any missing values, and then the option "Normalize Values Using

MIN-MAX" was chosen to normalize the numbers. The result section of the current research presents the outcomes of the proposed methodology. The dataset was first normalized to convert all values between 0 and 1, and the total number of records in the dataset, along with the train and test split counts, were displayed on the screen. The Autoencoder algorithm was trained on the dataset, resulting in 90% accuracy. To further enhance accuracy, the Decision Tree algorithm was implemented with PCA, and the accuracy and precision values were improved.

The graph in Figure 3 depicts the performance over time for the proposed and the existing method. The proposed method's performance exhibits an overall upward trend with some

fluctuations throughout the observation period, suggesting an increase in performance or output over time. On the other hand, the existing method also shows fluctuations but with a general trend that seems to plateau or increase less steeply compared to the proposed method. This representation implies that while both methods improve over time, the proposed method might offer a more pronounced enhancement in performance.

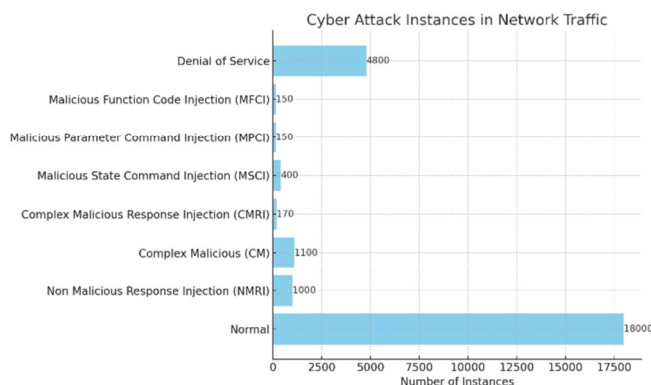


Fig. 2. Attack instance count.



Fig. 3. Efficiency comparison.

In Figure 4, the efficiency of both methods is assessed over the same time span. This graph uses a percentage scale on the y-axis to measure efficiency, which might correspond to the effectiveness or resource utilization of the methods. The proposed method displays a dynamic pattern, with its efficiency percentage rising and falling sharply, indicating significant variability in its efficiency. The existing method, follows a similar trend with less pronounced peaks and troughs, which may suggest a more stable but less efficient performance. The comparison reflects that the proposed method has moments of high efficacy, possibly outperforming the existing method, but also periods where its effectiveness crucially drops. The DNN algorithm was then applied to the dataset, which resulted in 99% accuracy. The detection and attribute of attack types were performed on the uploaded test data. Various types of attacks were spotted, and the results were displayed in a text area. Overall, the proposed methodology proved to be effective in identifying and

attributing attack types on the dataset. The normalization of the data, followed by the application of Autoencoder, Decision Tree with PCA, and DNN algorithms, resulted in high accuracy and precision values. The Comparison graph provides a visual representation of the encountered attacks, which could be useful in analyzing and understanding the data. Figure 5 and Table II display the precision, recall, accuracy, and F1 score for each considered algorithm.

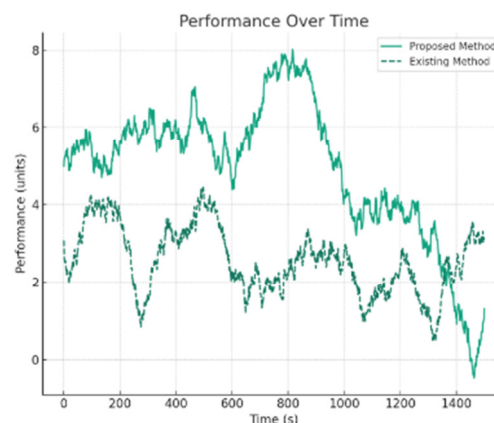


Fig. 4. Performance comparison.

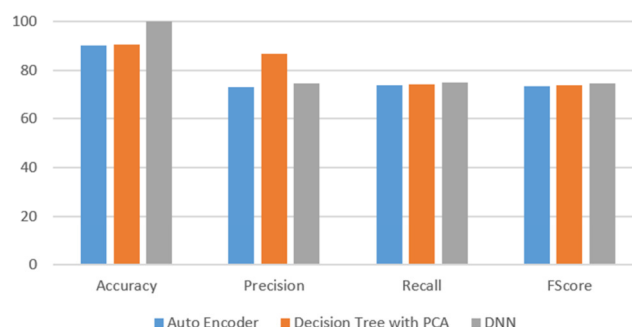


Fig 5. Algorithm performance comparison.

IV. CONCLUSION

Two-stage ensemble deep learning was used in this paper to offer a unique assault detection and attribution methodology for imbalanced ICS data. Deep representation learning converts the data to a higher-dimensional space for attack detection and Decision Tree was utilized to find plausible assault samples. This level can identify new assaults and survive skewed datasets. Multiple one-vs-all classifiers trained on different assault features are merged during attack attribution. The model is a complicated DNN with partially linked and completely connected components that can appropriately attribute cyber-attacks. As with earlier DNN-based methodologies, the proposed framework's training and testing phases are computationally intensive. This applies despite the framework's complex architecture. The proposed framework has greater recall and f-measure values than previous efforts and can swiftly recognize and attribute samples.

TABLE II. PERFORMANCE COMPARISON OF THE PROPOSED MODEL WITH EXISTING METHODS

Algorithm	Accuracy	Precision	Recall	F1 score
Auto encoder	90.091911764	73.168251465	74.120354616	73.598642216
DT withPCA	90.459558826	86.654346153	74.235468468	73.895321646
DNN	99.963512465	74.562458621	75.246864126	74.774652165

REFERENCES

- [1] F. Zhang, H. A. D. E. Kodituwakku, J. W. Hines, and J. Coble, "Multilayer Data-Driven Cyber-Attack Detection System for Industrial Control Systems Based on Network, System, and Process Data," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4362–4369, Jul. 2019, <https://doi.org/10.1109/TII.2019.2891261>.
- [2] R. Ma, P. Cheng, Z. Zhang, W. Liu, Q. Wang, and Q. Wei, "Stealthy Attack Against Redundant Controller Architecture of Industrial Cyber-Physical System," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9783–9793, Sep. 2019, <https://doi.org/10.1109/JIOT.2019.2931349>.
- [3] E. Nakashima, "Foreign hackers targeted U.S. water plant in apparent malicious cyber attack, expert says," *Washington Post*, Jun. 30, 2023, https://www.washingtonpost.com/blogs/checkpoint-washington/post/foreign-hackers-broke-into-illinois-water-plant-control-system-industry-expert-says/2011/11/18/gIQAgmTZYN_blog.html.
- [4] G. Falco, C. Caldera, and H. Shrobe, "IIoT Cybersecurity Risk Modeling for SCADA Systems," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4486–4495, Dec. 2018, <https://doi.org/10.1109/JIOT.2018.2822842>.
- [5] J. Yang, C. Zhou, S. Yang, H. Xu, and B. Hu, "Anomaly Detection Based on Zone Partition for Security Protection of Industrial Cyber-Physical Systems," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4257–4267, May 2018, <https://doi.org/10.1109/TIE.2017.2772190>.
- [6] S. A. Alshaya, "IoT Device Identification and Cybersecurity: Advancements, Challenges, and an LSTM-MLP Solution," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 11992–12000, Dec. 2023, <https://doi.org/10.48084/etasr.6295>.
- [7] S. Ponomarev and T. Atkison, "Industrial Control System Network Intrusion Detection by Telemetry Analysis," *IEEE Transactions on Dependable and Secure Computing*, vol. 13, no. 2, pp. 252–260, Mar. 2016, <https://doi.org/10.1109/TDSC.2015.2443793>.
- [8] J. F. Clemente, "No cyber security for critical energy infrastructure," Ph.D. dissertation, Naval Postgraduate School, Monterey, CA, USA, 2018.
- [9] C. Bellinger, S. Sharma, and N. Japkowicz, "One-Class versus Binary Classification: Which and When?," in *11th International Conference on Machine Learning and Applications*, Boca Raton, FL, USA, Dec. 2012, vol. 2, pp. 102–106, <https://doi.org/10.1109/ICMLA.2012.212>.
- [10] M. A. Lateef, C. Atheeq, M. A. Rahman, and M. A. Faizan, "Data Aegis Using Chebyshev Chaotic Map-Based Key Authentication Protocol," in *Intelligent Manufacturing and Energy Sustainability*, A. R. Manchuri, D. Marla, and V. V. Rao, Eds. New York, NY, USA: Springer, 2023, pp. 187–195.
- [11] M. M. N. Aboelwafa, K. G. Seddik, M. H. Eldefrawy, Y. Gadallah, and M. Gidlund, "A Machine-Learning-Based Technique for False Data Injection Attacks Detection in Industrial IoT," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8462–8471, Sep. 2020, <https://doi.org/10.1109/JIOT.2020.2991693>.
- [12] W. Yan, L. K. Mestha, and M. Abbaszadeh, "Attack Detection for Securing Cyber Physical Systems," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8471–8481, Oct. 2019, <https://doi.org/10.1109/JIOT.2019.2919635>.
- [13] M. A. Alqarni and S. H. Chaudhary, "A Security Scheme for Statistical Anomaly Detection and the Mitigation of Rank Attacks in RPL Networks (IoT Environment)," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12409–12414, Dec. 2023, <https://doi.org/10.48084/etasr.6433>.
- [14] T. K. Das, S. Adepu, and J. Zhou, "Anomaly detection in Industrial Control Systems using Logical Analysis of Data," *Computers & Security*, vol. 96, Sep. 2020, Art. no. 101935, <https://doi.org/10.1016/j.cose.2020.101935>.
- [15] Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013, <https://doi.org/10.1109/TPAMI.2013.50>.
- [16] M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan, and R. Jain, "Machine Learning-Based Network Vulnerability Analysis of Industrial Internet of Things," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6822–6834, Aug. 2019, <https://doi.org/10.1109/JIOT.2019.2912022>.
- [17] I. A. Khan, D. Pi, Z. U. Khan, Y. Hussain, and A. Nawaz, "HML-IDS: A Hybrid-Multilevel Anomaly Prediction Approach for Intrusion Detection in SCADA Systems," *IEEE Access*, vol. 7, pp. 89507–89521, 2019, <https://doi.org/10.1109/ACCESS.2019.2925838>.
- [18] C. Atheeq and M. M. A. Rabbani, "Mutually authenticated key agreement protocol based on chaos theory in integration of internet and MANET," *International Journal of Computer Applications in Technology*, vol. 56, no. 4, pp. 309–318, Jan. 2017, <https://doi.org/10.1504/IJCAT.2017.089088>.
- [19] R. Alsulami, B. Alqarni, R. Alshomrani, F. Mashat, and T. Gazdar, "IoT Protocol-Enabled IDS based on Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12373–12380, Dec. 2023, <https://doi.org/10.48084/etasr.6421>.
- [20] J. J. Q. Yu, Y. Hou, and V. O. K. Li, "Online False Data Injection Attack Detection With Wavelet Transform and Deep Neural Networks," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 3271–3280, Jul. 2018, <https://doi.org/10.1109/TII.2018.2825243>.
- [21] A. Cook, A. Nicholson, H. Janicke, L. Maglaras, and R. Smith, "Attribution of Cyber Attacks on Industrial Control Systems," *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, vol. 3, no. 7, Apr. 2016, Art. no. e3, <https://doi.org/10.4108/eai.21-4-2016.151158>.
- [22] N. A. Alsharif, S. Mishra, and M. Alshehri, "IDS in IoT using Machine Learning and Blockchain," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11197–11203, Aug. 2023, <https://doi.org/10.48084/etasr.5992>.
- [23] L. Maglaras, M. A. Ferrag, A. Derhab, M. Mukherjee, H. Janicke, and S. Rallis, "Threats, Countermeasures and Attribution of Cyber Attacks on Critical Infrastructures," *EAI Endorsed Transactions on Security and Safety*, vol. 5, no. 16, Oct. 2018, Art. no. e1, <https://doi.org/10.4108/eai.15-10-2018.155856>.
- [24] C. Atheeq and M. M. A. Rabbani, "CACK—A Counter Based Authenticated ACK to Mitigate Misbehaving Nodes from MANETs," *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, vol. 14, no. 3, pp. 837–847, Apr. 2021, <https://doi.org/10.2174/2213275912666190809104054>.
- [25] M. Alaeiyan, A. Dehghantanha, T. Dargahi, M. Conti, and S. Parsa, "A Multilabel Fuzzy Relevance Clustering System for Malware Attack Attribution in the Edge Layer of Cyber-Physical Networks," *ACM Transactions on Cyber-Physical Systems*, vol. 4, no. 3, Nov. 2020, Art. no. 31, <https://doi.org/10.1145/3351881>.
- [26] U. Noor, Z. Anwar, T. Amjad, and K.-K. R. Choo, "A machine learning-based FinTech cyber threat attribution framework using high-level indicators of compromise," *Future Generation Computer Systems*, vol. 96, pp. 227–242, Jul. 2019, <https://doi.org/10.1016/j.future.2019.02.013>.

Optimizing the Supercritical Carbon Dioxide Extraction of Hibiscus Flower Essential Oil using Response Surface Analysis

Tahani Y. A. Alanazi

Department of Chemistry, College of Science, University of Ha'il, Saudi Arabia
ty.alanazi@yahoo.com (corresponding author)

Received: 15 February 2024 | Revised: 26 February 2024 | Accepted: 28 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7076>

ABSTRACT

The development of Supercritical Fluid Extraction (SFE) has opened the door to the harvesting of plants for a wide range of chemical compounds. This study distilled essential oil from hibiscus flowers utilizing the supercritical CO₂ method. Different extraction parameters, including pressure (100-300 bar) and temperature (300-350 K), were studied to visualize how they affected oil recovery. Response surface analysis was used to fine-tune the extraction process. The chemical composition of the recovered oil was analyzed by Gas Chromatography-Mass Spectrometry (GC-MS). According to the findings, 13.11% per 80 g of dry flowers is the ideal oil extracted from Hibiscus flowers, using SFE at 200 bar pressure and 325 K extraction temperature. Six compounds were provisionally identified in the extracted oil from hibiscus flowers under optimum SFE conditions.

Keywords-essential oil; response surface analysis; hibiscus flower; Supercritical Fluid Extraction (SFE); Gas Chromatography-Mass Spectrometry (GC-MS)

I. INTRODUCTION

Essential oils, also called volatile or ethereal oils, are fluid mixtures obtained by distilling aromatic plant materials [1]. In manufacturing, volatile oils with aromatic properties are known as "fundamental oils". Phenols, esters, alcohols, ketones, aldehydes, and hydrocarbons are all part of the complex chemical makeup of these oils [2]. Essential oils extracted from numerous plants are commercially available [3]. Several discussions have been held about the biological activity of essential oils [1-4]. An example is olive leaf oil, which has been found to have a strong and high antioxidant capacity [5].

The Malvaceae family encompasses many well-known evergreen and grassy plants, including the popular Hibiscus rosa-sinensis. This plant can grow to a width of 1.5-3.0 m (5-10 ft) and a height of 2.5-5.0 m (8.0-16 ft). Most hibiscus trees originate in tropical regions, such as Southeast Asia, Australia, and Africa. Nowadays, this plant can be found almost anywhere in the world. Hibiscus flowers are used predominantly for decorative purposes. For example, flowers are often cut in half and used to create laurel wreaths. The medical, cosmetic, decorative, and industrial applications of hibiscus have increased its market value [6-7]. Its leaves, roots, and flowers have medical properties, serving as a laxative, an aphrodisiac, and a contraceptive. The oil extracted from the hibiscus flower acts as a stimulant, cough suppressant, germicide, and muscle relaxant [7]. It contains hydrocarbons and glycerides of various types. The chemical compounds found in this oil include eugenol, cis-jasmone, linalool, benzyl

acetate, benzyl alcohol, farnesol, isophytol, benzyl benzoate, acid derivatives, phytol, and geraniol [8]. The oil from this flower is beneficial for depression, migraines, sensitive skin, and fatigue, among other conditions [9]. Different types of the same plant, grown for its flowers or essential oil, include Hibiscus schizopetalus, Hibiscus rosa-sinensis, Hibiscus mutabilis, Hibiscus sabdariffa, and Hibiscus tiliaceus [6]. Hydro-distillation, solvent extraction, and maceration are adopted to extract oil from the hibiscus plant. However, these methods entail certain drawbacks, including decreased oil yield [10-11].

There are several methods to extract hibiscus flower oil, involving the liquid-liquid extraction method [12], where the flowers are soaked for a very long time and require special solvents, but it is a more restrictive and expensive method and its results may be unsatisfactory. This study used the cutting-edge method of supercritical liquid extraction. Supercritical Fluid Extraction (SFE) offers a blend of advantages, including selectivity and efficiency due to the adjustable density of supercritical fluids, and environmental friendliness through the use of non-toxic and recyclable CO₂. Furthermore, low-temperature operations benefit the extraction of thermally labile compounds, and the high purity of extracts with minimal residual solvents makes them suitable for products intended for human consumption. However, SFE faces limitations, such as high equipment costs, restrictions on volatile and semivolatile compounds, operational complexity that requires specialized knowledge, and scaling challenges for industrial applications, which affect its broader adoption. The state of a supercritical

liquid can be altered by changing the temperature and/or pressure. In the health and medical industries, this is crucial. Due to its specific physical properties, such as low temperature and moderate pressure, CO₂ is used in supercritical liquid extraction [13]. It has also been claimed that this technique produces a higher yield than either hydro-distillation or steam. However, the high price of the necessary hardware limits the applications of SFE to only the most fundamental mechanical parameters [14]. Essential oils of Apricot [15], Myrtle [16], Palm [17], Juniperus [18], Soybeans [19], Rosemary [20], Sunflower [21], Jojoba [22], Sesame [23], Celery [24], Parsley [2], Almond [26], and Pistachio [27], have been extracted more efficiently utilizing SFE.

This study investigated the employment of supercritical CO₂ to recover oil from hibiscus flowers. Furthermore, Gas Chromatography-Mass Spectrometry (GC-MS) was implemented to analyze the chemical composition of the extracted oil under ideal conditions. Response surface analysis was also used to determine how SFE operating conditions affected output quality [28].

II. MATERIALS AND METHODS

A. Materials

The hibiscus flowers were acquired from a natural flower shop supplied by Nature Flower Enterprise. The flowers were baked at 75°C for 1 hour to dry. The crushing of dried flowers increased the available contact area. 80 g of flowers were washed in water to remove contaminants.

B. Instrumentation

This study used a Retsch Ultra Centrifugal Mill ZM 200 grinder, a hot air-drying oven (Thermo Scientific Heracus), an SFE system by Separex (Champigneulle, France), an experimental plan, and a GC-MS (Agilent 5975C inert, USA) analyzer.

C. Supercritical CO₂ Extraction

The Separex 4219 extraction unit (Separex, France) was utilized to perform the steps required for SFE. This tool is compatible with 5.0, 10.5, and 20.0 cm³ autoclave breaking points (Axes). 5.0 cm of autoclave breaking points were put into service in this investigation. Hibiscus bloom powder of a known mass was weighed and measured, the reactor's pressure and temperature were adjusted, and the autoclave was stacked. The reactor was allowed to cool down after the oil was extracted.

D. Experimental Design

The yield (*Y*) and properties of the essential oil extracted from Hibiscus flowers with supercritical CO₂ depend on the pressure and temperature used to extract the oil. Employing a 32-central composite design from response surface analysis, the best SFE factors for getting oil from hibiscus flowers were found. Table I displays the results of an analysis of the extraction yield as a function of two independent variables: pressure (100-300 bar, *A*) and temperature (300-350 K, *B*). The experimental matrix (Table I), ANOVA analysis, regression coefficient calculation, and data visualization were all carried out in Design Expert 7.1.6 with a response surface plot and a

contour plot. Equation (1), a second-order response surface model, was applied to figure out the results of the experiments. Comparing the predicted and obtained values shows how well the model worked.

$$Y = a_0 + \sum_{i=1}^2 a_i A_i + \sum_{j=1}^2 a_j B_j + \sum_{i=1}^2 \sum_{j=1}^2 a_{ij} A_i B_j + \sum_{i=1}^2 a_{ii} A_i^2 + \sum_{j=1}^2 a_{jj} B_j^2 + \sum_{i=1}^2 \sum_{j=1}^2 a_{ij} A_i^2 B_j + \sum_{i=1}^2 \sum_{j=1}^2 a_{ij} A_i B_j^2 \quad (1)$$

where *Y* refers to the response for variables, and *a*₀, *a*_{*i*}, *a*_{*ii*}, *a*_{*ij*}, and *a*_{*ijj*} are consistent coefficients of the object, direct, quadratic, and intelligent terms, respectively. *A_i* and *B_j* are the autonomous factors of pressure and temperature, accordingly.

TABLE I. FACTORS AND STANDARDS EXAMINED FOR THE EXPLORATORY OUTLINE (RSM)

Symbol	Independent variables	Coded levels		
		Low (-1)	Middle (0)	High (+1)
<i>A</i>	Pressure (bar)	100	200	300
<i>B</i>	Temperature (K)	300	325	350

E. GC-MS Analysis

The Agilent 5975C series GC-MS was implemented to analyze the oil's chemical composition, which is made of 100% dimethylpolysiloxane and DB-WAX (30 m, 0.25 mm, ID 2.5 m) under ideal SFE conditions. Starting at 60°C, the working temperature increased at a rate of 20°C/min to a peak of 250°C, where it was maintained for 10 minutes. A speed of 30 cm/s for the helium gas was utilized. The oils extracted at 200 bar and 325 K, as well as at 200 bar and 300 K, were compared. To determine what chemicals they were, the National Institute of Standards and Technology (NIST) library was consulted. Mass spectral data obtained from oil and pure standards injected under identical conditions allowed differentiation of the compositions.

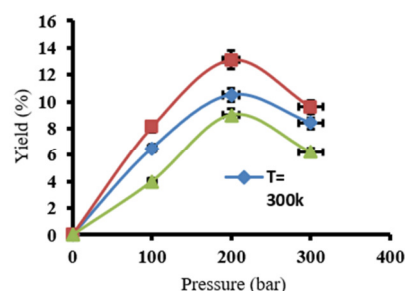


Fig. 1. The yield for oil at various pressures and temperatures.

III. RESULTS AND DISCUSSION

A. Supercritical CO₂ Extraction

The influences of pressure and temperature on the supercritical CO₂ extraction yield were analyzed and Table II summarizes the results of the various experimental iterations. Both pressure and temperature affected oil production. Oil production changed significantly when the pressure increased from 100 to 200 bar. At 200 bar, oil production was at its highest level. However, after 200 bar, the oil production began to decrease. As pressure increases, the supercritical liquid's

thickness increases, making it difficult to dissolve [29]. One possible explanation for the decrease in yield above 200 bar is that higher pressure can sometimes make the solute less soluble, allowing liquid particles to move more quickly into the solid pores. The solute is thus dissociated. As more pressure is put on the substantial grid-free space, the extraction yield may also change. In this example, the product increased by increasing the temperature from 300 to 325 K. At temperatures higher than 325 K, oil production decreased [30]. However, the oil yield was almost temperature-free at 100 bars. The wide range of possible temperature changes affects the oil yield, as a rise in temperature can dissolve solids more easily in supercritical liquids [30]. Solubility in oil is controlled by the equilibrium between the vapor pressure of the solute and the dissolvable consistency [31].

TABLE II. THE YIELD OF HIBISCUS OIL EXTRACTION (% OIL EXTRACTED/80 G DRY FLOWER)

Run	Pressure (bar)	Temperature (K)	Observed yield (%)
1	100	300	6.5
2	100	350	4.1
3	100	325	8.1
4	200	325	13.11
5	200	300	10.5
6	200	350	9.1
7	300	300	8.4
8	300	325	9.6
9	300	350	6.2

B. Hibiscus Flower Oil Extracted Characterization

Table III lists several chemicals, including methyl palmitate, decanoic acid ethyl ester, 7-formylbicyclo[4.1.0]heptanes, (Z)-9-pentadecadien-1-ol, 2-methyl-propanamide, and 1-(2-adamantylidene) semicarbazide. Hibiscus blossom oil also contained a large amount of aliphatic aldehyde and vegetable hydrocarbons. This made the 40% oil content and the total mass at 200 bar of pressure abundantly clear. Potential causes include the impact of factors, such as altitude, precipitation, and temperature on oil production.

TABLE III. MAJOR COMPOSITIONS IN THE HIBISCUS FLOWER OIL EXTRACTED

RT (min)	Compounds	Composition at 200 bar, 300 K (%)	Composition at 200 bar, 325 K (%)	Composition at 300 bar, 350 K (%)
39.672	Methyl palmitate	9.52	12.70	10.74
41.843	Decanoic acid, ethyl ester	55.32	57.31	53.43
44.544	7-Formylbicyclo[4.1.0]heptanes	8.87	6.79	4.68
44.635	(Z)-9-pentadecadien-1-ol	22.61	25.91	27.29
44.849	Propanamide, 2-methyl-	11.54	9.99	10.11
55.349	1-(2-Adamantylidene) semicarbazide	4.69	5.21	7.55

RT: Retention Time

C. Response Surface Analysis

The results disclosed that hibiscus blossom oil yield ranged from 4.11 to 13.11%. This demonstrates the significance of pressure and temperature in determining the oil yield. ANOVA was used to earn a second-degree equation:

$$Y = +12.09 + 1.16 * A - 1.52 * B + 0.35 * A * B - 3.52 * A^2 - 1.01 * B^2 - 0.238 * A^2 * B + 0.034 * A * B^2 \quad (2)$$

The fit of the model was quite satisfactory, as indicated by the multiple correlation coefficient R^2 value of 0.9986. Table IV portrays the nine experiments and their observed and predicted results. Most points are clustered around the first bisector line, confirming that the experimental and predicted data agree. The quality of the model was assessed in terms of Fisher's Exact Test via the ANOVA F-test. To do this, the F-value, the ratio between the square of the model's mean, and the sum of the residual error were employed. There is statistical significance in the model, as the F-value is 503.68. Table IV exhibits that the model's p-value is less than 0.05, indicating that it is statistically significant. The effects of pressure (A), heat (B), two-tier relationships between the two (AB), pressure (A^2), and heat (B^2) were all relatively significant. The fact that many other terms had values higher than 0.1000 indicates that the model terms could be more helpful. In terms of oil production, the hibiscus flower was most sensitive to pressure's immediate effect (A). Since pressure is polar, it can dynamically interact with temperature. This result agrees with the results of [32]. The positioning of the crucial components was determined by the predicted F-value. Therefore, in this study, the following is the order of importance for the features: When comparing similar sets of numbers, $A^2 > B > B^2 > A > AB > A^2B > AB^2$. Furthermore, the p-value for the lack of fit is only 3.49%. R^2 is used to determine whether a relapse is the right move. The determined R^2 -value was 0.9986, which is close to 1 and is therefore acceptable. There was some ambiguity between the reported R^2 of 0.9349 and the adjusted R^2 of 0.9966. A ratio greater than 4 is preferable. In this case, the estimation was significantly higher than 4, with a balance of 60.510, revealing a superb indication.

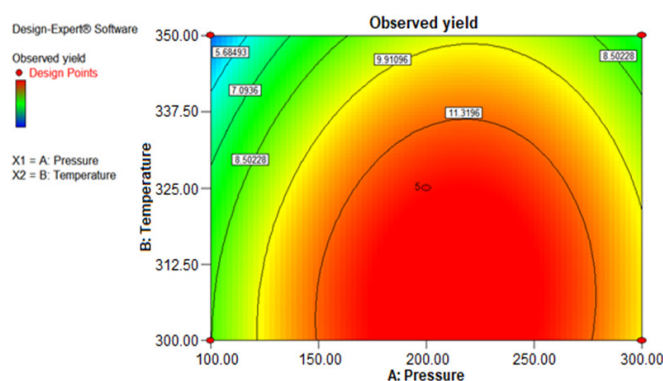


Fig. 2. Optimum yield for oil extraction.

D. Optimum Conditions

Response surface analysis was implemented to determine how changes in extraction parameters affect the amount of oil

that can be extracted from hibiscus blooms. The highest oil yield, 13.11%, was achieved at a pressure (A) of 200 bar and a temperature (B) of 325 K. Figure 2 depicts the three-dimensional response surface discovered by analyzing the optimal pressure and temperature for oil production. At first, the oil yield increased but then decreased as the temperature rose at all three pressures (100, 200, and 300 bar). The experiments showed that the amount of oil extracted from Hibiscus flowers depends on how much pressure will be applied and how hot it is.

IV. CONCLUSION

This study investigated the use of supercritical CO₂ to extract oil from hibiscus blooms. This innovative approach highlights the way pressure directly affects the quantity of oil extracted, while the temperature of the extraction environment impacts the oil's quality. Through the application of response surface analysis, this study successfully identified the optimal values for pressure and temperature during the SFE process. Under these optimal conditions, the extraction of oil from hibiscus flowers revealed the presence of six distinct compounds, highlighting the technique's efficiency and specificity. At 325 K and 200 bar, the oil yield was maximized at 13.11% per 80 g of dry flower, indicating that the hibiscus flower, when processed with SFE, produces a substantial amount of oil. The novelty of this study lies in its approach to optimizing the SFE conditions, and its contribution extends to demonstrating the potential of hibiscus flowers as a source of valuable oil, utilizing supercritical CO₂ for efficient and environmentally friendly extraction.

ACKNOWLEDGEMENTS

The author would like to thank the Department of Chemistry, College of Science, University of Ha'il for assisting with this study.

REFERENCES

- [1] S. Burt, "Essential oils: their antibacterial properties and potential applications in foods—a review," *International Journal of Food Microbiology*, vol. 94, no. 3, pp. 223–253, Aug. 2004, <https://doi.org/10.1016/j.ijfoodmicro.2004.03.022>.
- [2] A. Younis, A. Riaz, M. A. Khan, A. A. Khan, and M. A. Pervez, "Extraction and identification of chemical constituents of the essential oil of Rosa species," in *XXVII International Horticultural Congress-IHC2006: International Symposium on Ornamentals, Now!* 766, 2006, pp. 485–492, <https://doi.org/10.17660/ActaHortic.2008.766.65>.
- [3] K. H. Kubeczka and V. Formacek, *Essential Oils Analysis by Capillary Gas Chromatography and Carbon-13 NMR Spectroscopy*, 2nd ed. John Wiley & Sons, 2022.
- [4] M. Oussalah, S. Caillet, L. Saucier, and M. Lacroix, "Antimicrobial effects of selected plant essential oils on the growth of a *Pseudomonas putida* strain isolated from meat," *Meat Science*, vol. 73, no. 2, pp. 236–244, Jun. 2006, <https://doi.org/10.1016/j.meatsci.2005.11.019>.
- [5] N. Elboughdiri, "Effect of Time, Solvent-Solid Ratio, Ethanol Concentration and Temperature on Extraction Yield of Phenolic Compounds From Olive Leaves," *Engineering, Technology & Applied Science Research*, vol. 8, no. 2, pp. 2805–2808, Apr. 2018, <https://doi.org/10.48084/etasr.1983>.
- [6] P. Green and D. Miller, *The genus Jasminum in cultivation*. Richmond, UK: Royal Botanic Gardens, 2009.
- [7] J. Lawless, *The encyclopedia of essential oils*. Newburyport, MA, USA: Conari Press, 2013.
- [8] J. Lawless, *The illustrated encyclopedia of essential oils: the complete illustrated guide to the use of oils in aromatherapy and herbalism*. Shaftesbury, UK: Element, 1995.
- [9] G. Mojay, *Aromatherapy for healing the spirit: Restoring emotional and mental balance with essential oils*. Inner Traditions/Bear & Co, 2000.
- [10] E. Reverchon, "Supercritical fluid extraction and fractionation of essential oils and related products," *The Journal of Supercritical Fluids*, vol. 10, no. 1, pp. 1–37, Apr. 1997, [https://doi.org/10.1016/S0896-8446\(97\)00014-4](https://doi.org/10.1016/S0896-8446(97)00014-4).
- [11] I. Zizovic, M. Stamenić, A. Orlović, and D. Skala, "Supercritical carbon dioxide extraction of essential oils from plants with secretory ducts: Mathematical modelling on the micro-scale," *The Journal of Supercritical Fluids*, vol. 39, no. 3, pp. 338–346, Jan. 2007, <https://doi.org/10.1016/j.supflu.2006.03.009>.
- [12] F. E. Al-Damluji and A. A. Mohammed, "Performance Evaluation of Emulsion Liquid Membrane on Chlorpyrifos Pesticide Removal: Stability, Mass Transfer Coefficient, and Extraction Efficiency Studies," *Engineering, Technology & Applied Science Research*, vol. 13, no. 1, pp. 9872–9878, Feb. 2023, <https://doi.org/10.48084/etasr.5401>.
- [13] R. N. Carvalho, L. S. Moura, P. T. V. Rosa, and M. A. A. Meireles, "Supercritical fluid extraction from rosemary (*Rosmarinus officinalis*): Kinetic data, extract's global yield, composition, and antioxidant activity," *The Journal of Supercritical Fluids*, vol. 35, no. 3, pp. 197–204, Oct. 2005, <https://doi.org/10.1016/j.supflu.2005.01.009>.
- [14] M. A. A. Meireles, "Supercritical extraction from solid: process design data (2001–2003)," *Current Opinion in Solid State and Materials Science*, vol. 7, no. 4, pp. 321–330, Aug. 2003, <https://doi.org/10.1016/j.cossms.2003.10.008>.
- [15] S. G. Özkal, M. E. Yener, and L. Bayındır, "Response surfaces of apricot kernel oil yield in supercritical carbon dioxide," *LWT - Food Science and Technology*, vol. 38, no. 6, pp. 611–616, Sep. 2005, <https://doi.org/10.1016/j.lwt.2004.08.003>.
- [16] A. Zermane, O. Larkeche, A.-H. Meniai, C. Crampon, and E. Badens, "Optimization of essential oil supercritical extraction from Algerian *Myrtus communis* L. leaves using response surface methodology," *The Journal of Supercritical Fluids*, vol. 85, pp. 89–94, Jan. 2014, <https://doi.org/10.1016/j.supflu.2013.11.002>.
- [17] I. Md Zaidul, N. Nik Norulaini, and A. Mohd Omar, "Separation/fractionation of triglycerides in terms of fatty acid constituents in palm kernel oil using supercritical CO₂," *Journal of the Science of Food and Agriculture*, vol. 86, no. 7, pp. 1138–1145, 2006, <https://doi.org/10.1002/jsfa.2479>.
- [18] O. Larkeche, A. Zermane, A.-H. Meniai, C. Crampon, and E. Badens, "Supercritical extraction of essential oil from *Juniperus communis* L. needles: Application of response surface methodology," *The Journal of Supercritical Fluids*, vol. 99, pp. 8–14, Apr. 2015, <https://doi.org/10.1016/j.supflu.2015.01.026>.
- [19] E. Reverchon and L. S. Osséo, "Comparison of processes for the supercritical carbon dioxide extraction of oil from soybean seeds," *Journal of the American Oil Chemists' Society*, vol. 71, no. 9, pp. 1007–1012, Sep. 1994, <https://doi.org/10.1007/BF02542270>.
- [20] A. Zermane, A.-H. Meniai, and D. Barth, "Supercritical CO₂ Extraction of Essential Oil from Algerian Rosemary (*Rosmarinus officinalis* L.)," *Chemical Engineering & Technology*, vol. 33, no. 3, pp. 489–498, 2010, <https://doi.org/10.1002/ceat.200900381>.
- [21] L. Calvo, M. J. Cocero, and J. M. Díez, "Oxidative stability of sunflower oil extracted with supercritical carbon dioxide," *Journal of the American Oil Chemists' Society*, vol. 71, no. 11, pp. 1251–1254, Nov. 1994, <https://doi.org/10.1007/BF02540546>.
- [22] U. Salgın, "Extraction of jojoba seed oil using supercritical CO₂+ethanol mixture in green and high-tech separation process," *The Journal of Supercritical Fluids*, vol. 39, no. 3, pp. 330–337, Jan. 2007, <https://doi.org/10.1016/j.supflu.2006.03.013>.
- [23] J. Xu, S. Chen, and Q. Hu, "Antioxidant activity of brown pigment and extracts from black sesame seed (*Sesamum indicum* L.)," *Food Chemistry*, vol. 91, no. 1, pp. 79–83, Jun. 2005, <https://doi.org/10.1016/j.foodchem.2004.05.051>.

- [24] I. Papamichail, V. Louli, and K. Magoulas, "Supercritical fluid extraction of celery seed oil," *The Journal of Supercritical Fluids*, vol. 18, no. 3, pp. 213–226, Oct. 2000, [https://doi.org/10.1016/S0896-8446\(00\)00066-8](https://doi.org/10.1016/S0896-8446(00)00066-8).
- [25] V. Louli, G. Folas, E. Voutsas, and K. Magoulas, "Extraction of parsley seed oil by supercritical CO₂," *The Journal of Supercritical Fluids*, vol. 30, no. 2, pp. 163–174, Jul. 2004, <https://doi.org/10.1016/j.supflu.2003.07.003>.
- [26] C. Marrone, M. Poletto, E. Reverchon, and A. Stassi, "Almond oil extraction by supercritical CO₂: experiments and modelling," *Chemical Engineering Science*, vol. 53, no. 21, pp. 3711–3718, Nov. 1998, [https://doi.org/10.1016/S0009-2509\(98\)00150-X](https://doi.org/10.1016/S0009-2509(98)00150-X).
- [27] T. K. Palazogilu and M. O. Balaban, "Supercritical CO₂ Extraction of Lipids from Roasted Pistachio Nuts," *Transactions of the ASAE*, vol. 41, no. 3, pp. 679–684, 1998, <https://doi.org/10.13031/2013.17193>.
- [28] H. E. Kolli and M. E. Kolli, "Preparation and Characterization of Gelatin-Based Films Cross-Linked by Two Essential Oils at Different Concentrations and Plasticized with Glycerol," *Engineering, Technology & Applied Science Research*, vol. 11, no. 4, pp. 7489–7494, Aug. 2021, <https://doi.org/10.48084/etasr.4283>.
- [29] I. Haloui and A.-H. Meniai, "Supercritical CO₂ extraction of essential oil from Algerian Argan (*Argania spinosa* L.) seeds and yield optimization," *International Journal of Hydrogen Energy*, vol. 42, no. 17, pp. 12912–12919, Apr. 2017, <https://doi.org/10.1016/j.ijhydene.2016.12.012>.
- [30] S. Pereda, S. B. Bottini, and E. A. Brignole, *Fundamentals of supercritical fluid technology*. CRC Press - Taylor & Francis Group, 2007.
- [31] B. C. Roy, M. Goto, and T. Hirose, "Extraction of Ginger Oil with Supercritical Carbon Dioxide: Experiments and Modeling," *Industrial & Engineering Chemistry Research*, vol. 35, no. 2, pp. 607–612, Jan. 1996, <https://doi.org/10.1021/ie950357p>.
- [32] T. Zhu, H. J. Heo, and K. H. Row, "Central composite design for optimized extraction of polysaccharides from *Undaria pinnatifida*," *Chemical Research in Chinese Universities*, vol. 28, no. 4, pp. 620–623, 2012.

An MCDM Approach for Evaluating Construction-Related Risks using a Combined Fuzzy Grey DEMATEL Method

Rana Jabbar Kasid Jalhoom

Civil Engineering Department, College of Engineering, University of Technology, Iraq
rana.j.kasid@uotechnology.edu.iq (corresponding author)

Ahmed Mohammed Raoof Mahjoob

Civil Engineering Department, College of Engineering, University of Baghdad, Iraq
ahmed.mahjoob@coeng.uobaghdad.edu.iq

Received: 24 January 2024 | Revised: 9 February 2024 | Accepted: 14 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6959>

ABSTRACT

There is a need for more research into prioritizing project risks based on a sound technique due to the complicated and disorganized character of this stage. The project risk management process typically begins with the identification of critical hazards. This study presents a Grey Fuzzy Decision-Making Trial and Evaluation Laboratory (FGDEMATEL) approach to prioritize potential causes of project risks within Multi-Criteria Decision-Making (MCDM). This framework organizes the numerous risks using the Risk Breakdown Structure (RBS) of the Project Management Institute (PMI). The risk information used in this analysis comes mostly from the views and choices of project experts. Grey theory, which takes language phrases for preference collections and translates them into numerical intervals, is responsible for controlling uncertainty and variance in experts' preferences. As each expert has unique skills and experiences, it evaluates the significance of their opinions using a fuzzy number system that incorporates three dimensions. In the end, the FGDEMATEL approach devised a method to rank various project risks.

Keywords-grey system theory; DEMATEL; project risk management; fuzzy set theory

I. INTRODUCTION

Project risk considers the possibility that something may go wrong and hinder progress toward the project's goals. The risk of a project is the product of two variables: the likelihood that an event will take place and the severity of its potential consequences [1-4]. The total effect of all separate risks in a project is referred to as the aggregate project risk. Risk management is an essential procedure that must be carried out to complete projects successfully and on schedule. A detailed process of systematically determining particular risk sources at a consistent degree of specificity can be accomplished using risk categories as a framework. A Risk Breakdown Structure (RBS) is a document that provides a list of several categories and subcategories within which a typical project may experience risks [5-6]. A risk manager uses it to better manage risks and familiarize himself with elements common to a typical project [7-8]. Figure 1 shows a graphic description of the RBS that appears regularly in the PMI PMBOK Guide.

Many tools and techniques have emerged to help project managers predict and control structural risks during the life of a project. In [9], the possibility of using fuzzy DEMATEL, based on the PMBOK standard, was explored to rate the project risks from most to least significant. In [10], a risk assessment

method for construction projects was presented, based on MCDM methodologies such as Grey TOPSIS and COPRAS-G. The ELECTRE method was used in [11] to classify the risks incurred throughout the tunneling process for the Tehran metro project. In [12], the Analytic Network Process (ANP) was presented to manage the interrelationships that exist between risk-related components in multinational construction projects. The ANP method's ranking was used as an input to a decision support tool and depended on it through the bidding phase. In [13], the significance of risk rating in megaprojects was investigated using fuzzy compromise software methods, such as TOPSIS, VIKOR, and LINMAP, and three MADM algorithms, to analyze data in a fuzzy setting. In addition, a new fuzzy VIKOR approach was proposed to help managers better deal with the risks associated with large projects. In [14], high risks in Iran's onshore gas refinery plants were prioritized using fuzzy TOPSIS and fuzzy LINMAP algorithms, with fuzzy LINMAP being more effective. In [15], a formal method was developed for qualitative risk assessment based on a hierarchical risk breakdown structure. An applicable fuzzy MCDM was presented in [16] as a means to identify and simultaneously prioritize project hazards within an EPC project. In [17], a fuzzy MCDM method was used to perform an in-depth risk assessment for a building project in a

metropolitan area. Consistent Fuzzy Preference Relations (CFPR) were used to quantify the relative influence of 20 specified risk variables on the success of the project. Additionally, an approach known as Fuzzy Multiple Attributes Direct Rating (FMADR) was used to evaluate the possibility of the occurrence of various risk factors. In [18], the most significant risk factors in a building project were arranged in a hierarchical structure, using a modified logical MCDM with fuzzy logic to choose an efficient risk factor.

This study combined fuzzy logic with the Grey theory. In most cases, the risks associated with a project depend on each other, and there is the possibility of some degree of interaction between them. As a result, the DEMATEL approach was used to investigate the complex interconnections between the various risk categories. Another reason for adopting the Grey system theory is that it makes it easier to assess human opinions. This study proposes the Fuzzy Grey DEMATEL (FGDEMATEL) approach to help risk managers in the Iraqi construction industry prioritize various types of project risks.

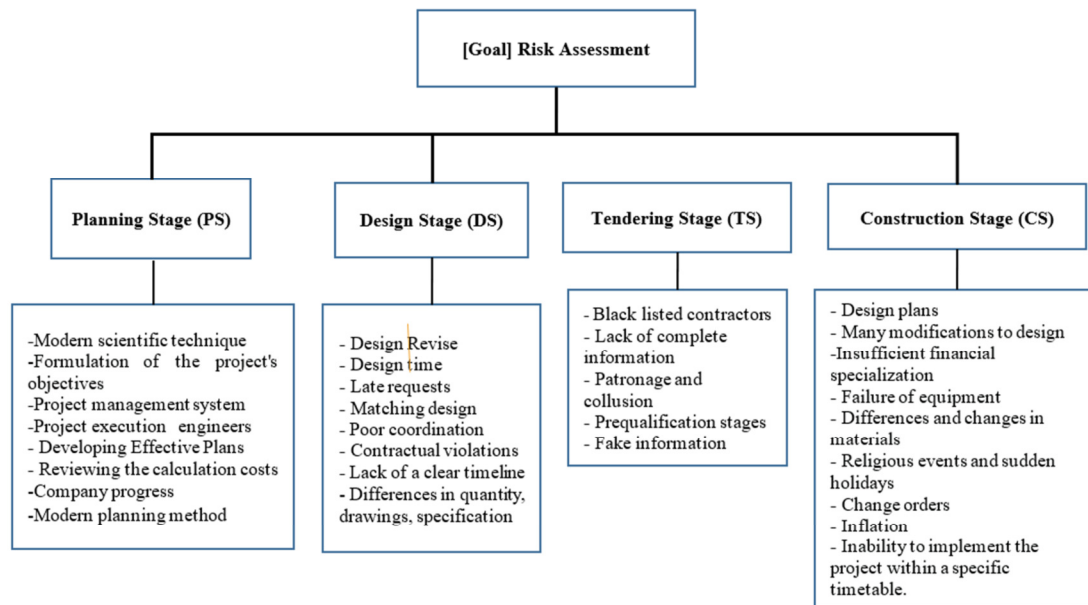


Fig. 1. Risk Breakdown Structure (RBS)

II. RESEARCH METHODOLOGY

The steps of the proposed method are summarized as:

- Identify structural risks that occurred in specific finished construction projects across all stages of their lifecycle by assessing periodic updates on projects from 2019-2022, in addition to consulting with industry experts.
- Distribute a questionnaire on risks occurring at each stage with a five-point linguistic rating scale to identify the influence of each risk on the others.
- Determine the fuzzy weights of the expert's importance based on his self-evaluation of his level of expertise in terms of his knowledge and experience.
- Apply the FGDEMATEL method to these risks.

III. RISK RATING

A. Establishing an Expert Group

This team consists of specialists and experts with specific expertise and experience in project management to identify the primary kinds of project risks, considering the RBS, as described in the fourth edition of the PMBOK Guide.

B. Determine the Fuzzy Weights of Experts

Each expert p has a fuzzy established weight of importance \tilde{W}_p (experts' relative significance weight) based on his self-evaluation of his level of expertise in terms of knowledge and experience. A 5-point scale questionnaire, shown in Table I, was used to determine the relative significance weights (\tilde{W}_p) of the experts using a self-assessing method [16], based on their previous experience and understanding of working on projects.

TABLE I. EXPERTS' RELATIVE SIGNIFICANCE WEIGHT (\tilde{W}_p) [16]

Lingual variables	Fuzzy number
Very Small (VS)	(0, 0.1, 0.3)
Small (L)	(0.1, 0.3, 0.5)
Moderate (M)	(0.3, 0.5, 0.7)
Large (L)	(0.5, 0.7, 0.9)
Very Large (VL)	(0.7, 0.9, 1)

C. Representation of Grey Numbers

The evaluation criteria were established, and a Grey linguistic scale was constructed to represent more accurately the uncertainties of human assessments. The use of a 5-point scale for the questionnaire allows for calculating both the linguistic scale and the related Grey values. To measure the

interrelations between risks using a 5-point linguistic classification system that reflects the influence of any risk on the others, a questionnaire was distributed to the experts, in addition to a matrix containing linguistic scores. In place of linguistic information, a Grey-scale linguistic representation employs Grey values to change the effect scores of lingual data in the initial relation matrix to produce a matrix with Grey numbers. The influence scores of linguistic information in the initial relation matrix are replaced by the Grey values, as shown in Table II [16, 19].

TABLE II. REPRESENTATION OF GRAY NUMBERS AND LINGUISTIC SCORES [16, 19]

lingual expression	Impact score	Grey number
No influence	0	[0, 0]
lower influence	1	[0,0.25]
Moderate influence	2	[0.25, 0.5]
large influence	3	[0.5, 0.75]
Very large influence	4	[0.75,1]

D. Convert Gray Numbers to Crisp Values

Using (1)-(5) a matrix with crisp values was created. Grey aggregation methods are required to arrive at a clear number. This study employed a modified version of the de-fuzzification approach known as Conversion of Fuzzy data into Crisp Scores (CFCS) to eliminate fuzziness [20-21]. The Grey number for an expert p , who will assess the effect that a risk i has on a risk j , is denoted by the sign $\otimes X_{ij}^p$. The higher and lesser Grey values of the Grey number $\otimes X_{ij}^p$ are identified by the notations $\otimes \overline{X}_{ij}^p$ and $\otimes \underline{X}_{ij}^p$ as: $\otimes X_{ij}^p = [\otimes \underline{X}_{ij}^p, \otimes \overline{X}_{ij}^p]$. The modified CFCS method consists of three stages:

- Normalization:

$$\underline{X}_{ij}^p = \frac{(X_{ij}^p - \min_j X_{ij}^p)}{\Delta_{\min}^{\max}} \quad (1)$$

$$\overline{X}_{ij}^p = \frac{(\overline{X}_{ij}^p - \min_j \overline{X}_{ij}^p)}{\Delta_{\min}^{\max}} \quad (2)$$

$$\Delta_{\min}^{\max} = \max_j \overline{X}_{ij}^p - \min_j \underline{X}_{ij}^p \quad (3)$$

- The total normalized crisp value is calculated by:

$$Y_{ij}^p = \frac{(X_{ij}^p (1 - \underline{X}_{ij}^p) + (\overline{X}_{ij}^p - \underline{X}_{ij}^p))}{(1 - \underline{X}_{ij}^p + \overline{X}_{ij}^p)} \quad (4)$$

- Determine crisp value:

$$Z_{ij}^p = \min_j \underline{X}_{ij}^p + Y_{ij}^p \Delta_{\min}^{\max} \quad (5)$$

E. Determine the Aggregated Opinion of the Experts

Equation (6) is used to acquire the aggregated opinion of the n experts to evaluate the influence that risk i has on risk j . This is because each expert possesses his own unique fuzzy importance weight. \tilde{a}_{ij} is the aggregated fuzzy value of the influence that risk i has on risk j , and \tilde{W}_p is the fuzzy important weight that expert p assigns to the influence. The calculation of

(6) needs the implementation of the basic laws of fuzzy functioning in the ending. Equation (7) is used to obtain an explicit evaluation of \tilde{a}_{ij} , which is depicted by $\tilde{a}_{ij}(a_1, a_2, a_3)$, and Z_{ij}^p describes the crisp value of risk i on risk j measured by expert p .

$$\tilde{a}_{ij} = \frac{\sum_{p=1}^n \tilde{W}_p Z_{ij}^p}{\sum_{p=1}^n \tilde{W}_p} \quad (6)$$

$$a_{ij} = \frac{(a_1 + a_2 + a_3)}{3} \quad (7)$$

F. Applying the DEMATEL Approach

The method's steps are outlined in [22-23].

- The direct relationship matrix has to be created. A pairwise analysis of criteria performed by an expert team generates the $A(n \times n)$ matrix, where each entry in the matrix a_{ij} reflects the influence value of factor i on factor j . The influence of i on j is reflected in how a change in one criterion (factor) i can affect another factor j .

- Normalization of the direct-relation matrix using:

$$X = k \times A \quad (8)$$

$$k = \frac{1}{\max \sum_{j=1}^n a_{ij}}, \quad 1 \leq i \leq n \quad (9)$$

- The following equation should provide the whole relation matrix, where I is the identity matrix,

$$T = (I - X) - 1 \quad (10)$$

- The system produces a causal diagram. When applied to matrix T , the following equations determine the sum of rows (D) and the sum of columns (R). A factor's D value is the extent to which it affects other variables. The value of the factor's R indicates the effect it has on other variables [24].

$$T = [t_{ij}]_{n \times n}, \quad i, j = 1, 2, \dots, n \quad (11)$$

$$R = [\sum_{i=1}^n t_{ij}]_{1 \times n} = [t_{.j}]_{1 \times n} \quad (12)$$

$$D = [\sum_{j=1}^n t_{ij}]_{n \times 1} = [t_{i.}]_{n \times 1} \quad (13)$$

- On the horizontal axis, D+R denotes the weight given to each criterion. D-R stands for relation in mathematics. The condition belongs to the group of causes if and only if $(D-R) > 0$. If $(D-R)$ is less than zero, then the parameter belongs to the effect group. The success or failure of the entire system depends on the actions of the cause factors. In addition, more focus needs to be given to cause group criteria. Factors in the effect group are unsuitable and likely to be easily influenced by others, making them a significant success element [25].

IV. APPLICATION

The proposed method was applied in an Iraqi construction project in Baghdad at the Shanashil residential complex during the project's implementation stages, holder of Investment Permit No. 79A of 2019. The project's location is south of Daura, on the new Daura-Youssafiyah highway, which is 17 km away from the two-floor bridge within the city of Jawhara,

Baghdad. It contains 777 residential horizontal units and is divided into 5 residential areas ranging from 160 to 480 m. The four basic stages were adopted, namely Planning, Tender, Design, and Construction phases, during which the construction risks were assessed and diagnosed by seven specialists and experts. The following steps were applied:

- Classifying major risks to the project: PMBOK, fourth edition RBS, as shown in Figure 1.
- Assembled the professional team of experts: The seven members of the team had extensive experience in managing construction projects for Iraqi enterprises that operate on a project basis, and they were asked to fill out the surveys. Table III shows the experts' relative importance weights.
- Establish causal connections: The experts were asked to fill out questionnaires that used the five-point linguistic rating system to investigate the correlation between risks. This scale shows how one risk affects another. Table IV shows the thoughts of the sixth expert.

TABLE III. RELATED SIGNIFICANCE WEIGHT OF EXPERT (\bar{W}_P)

Lingual variables	Fuzzy number
Expert 1,2,3	Small (L)
Expert 4	large (H)
Experts 5,7	Moderate (M)
Expert 6	Very large (VH)
Expert 1,2,3	Small (L)

TABLE IV. SIXTH PROFESSIONAL'S ASSESSMENTS OF RELATED RISKS

Phase	Planning	Tendering	Design	Construction
Planning	0	3	2	3
Tendering	2	0	2	2
Design	3	2	0	3
Construction	3	2	3	0

- The language data are switched out for a Grey-scale version. Grey numbers were substituted for the influence scores of language data in the initial relation matrix Table IV, and (1)-(5) were used to obtain the crisp values. Tables V and VI are shown below. Since each expert has his own unique fuzzy importance weight, (6) can be applied with $n = 7$, as in Table VII, to get an averaged view of how the experts feel about the impact of risk i on risk j .

TABLE V. GRAY EQUIVALENTS OF THE 6TH EXPERT'S OPINIONS OF THE INTERRELATIONSHIP BETWEEN RISKS

Phase	Planning	Tendering	Design	Construction
Planning	[0, 0]	[0.5, 0.75]	[0.25, 0.5]	[0.5, 0.75]
Tendering	[0.25, 0.5]	[0, 0]	[0.25, 0.5]	[0.25, 0.5]
Design	[0.5, 0.75]	[0.25, 0.5]	[0, 0]	[0.5, 0.75]
Construction	[0.5, 0.75]	[0.25, 0.5]	[0.5, 0.75]	[0, 0]

TABLE VI. THE 6TH EXPERT'S CRISP ASSESSMENTS OF THE RISKS' INTERDEPENDENCE

Phase	Planning	Tendering	Design	Construction
Planning	0	0.6875	0.3750	0.6875
Tendering	0.4167	0	0.4167	0.4167
Design	0.6875	0.3750	0	0.6875
Construction	0.6875	0.3750	0.6875	0

TABLE VII. THE MEAN FUZZY WEIGHTS OF ALL EXPERTS' OPINIONS IN A MATRIX

Phase	Planning	Tendering	Design	Construction
Planning	0	0.6959	0.6625	0.9063
Tendering	0.7958	0	0.7405	0.7764
Design	0.5492	0.5572	0	0.8465
Construction	0.7586	0.5136	0.8718	0

- Compile a cause-and-effect diagram: For the purposes of DEMATEL, Table VII was used as the direct relation matrix. Tables VIII and IX display the computed total relation matrix alongside the values for D+R and D-R.

TABLE VIII. TOTAL RELATION MATRIX

Phase	Planning	Tendering	Design	Construction
Planning	3.3653	3.1148	3.8861	4.1787
Tendering	3.6807	2.9447	3.9599	4.2161
Design	3.1864	2.7664	3.2520	3.7337
Construction	3.4504	2.9406	3.7569	3.7145

TABLE IX. VALUES OF D+R AND D-R

Phase	D-R rank	D-R	D+R rank	D+R
Planning	3	-1.9165	4	26.5679
Tendering	4	-1.9806	2	28.2277
Design	2	0.8621	3	27.7934
Construction	1	3.0350	1	29.7054

Figure 2 shows the acquired causal diagram after mapping the dataset of (D+R, D-R), which displays the ranking and relation between stages as shown above.

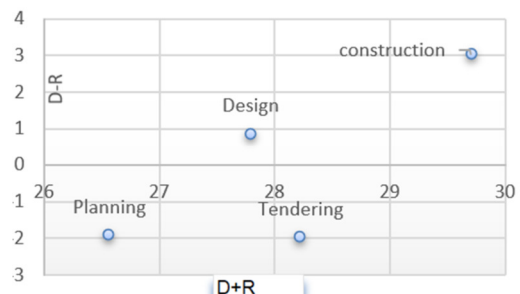


Fig. 2. The causal diagram.

V. FINDINGS AND DISCUSSION

Planning and Tendering risks are classified as part of the effect group in both the causal diagram shown in Figure 2, and the total relation matrices shown in Tables VIII and IX. This is because the D-R scores for these risks are negative and they tend to be easily influenced by other risks. On the contrary, risks associated with Design and Construction are classified as part of the cause group. This is because both categories have positive scores on the D-R value scale, indicating that they are significant risks that have the potential to affect the overall success of the project. Construction, Planning, Tendering, and Design risk values were also examined in the D+R ranking. Construction ranks first in terms of relative importance compared to other hazards. It is necessary to focus on both the D+R and D-R rankings to arrive at a DEMATEL ranking that makes sense. First, the D-R ranking and risks that are part of

the cause group are considered. Specifically their D-R values must be positive. Next, the D+R ranking is examined. In this instance, the Construction and Design risks are members of the cause group and rank first and second in importance. However, in the D+R ranking, they are in ranks 1 and 3, respectively. As a cause group is of the utmost significance, and the disparity between the risk values associated with Construction and Design poses a significant challenge. The risks associated with Construction come at the top, followed by the Design concerns. The D-R ranking does not help with the third and fourth positions, because both are effectively in the same group with negative D-R values, and the difference between their respective values within this group is insignificant. Therefore, the Tendering risk is the third important risk category, while the Planning risk is the fourth, according to the D+R ranking. The following is the final order in risk classification: 1. Construction risks, 2. Design risks, 3. Tendering risks, and 4. Planning risks.

VI. CONCLUSION

This study provides useful insights, especially for risk managers. Construction and Design risks, which fall under the umbrella of cause group risks, are extremely important to consider. According to the findings, the technology used in the project should be seen as one of the main sources of risk and should be carefully chosen to reduce the effects of Construction risks. It is also shown that Design risks are substantial and can bring about many drawbacks for effective project accomplishment, such as a lack of a clear timeline and poor coordination, even though these components are outside the purview of the authorization of a project manager but cause a wide scope of issues, and administrators should be attentive to lessen their adverse effects. Due to the numerous complexities of modern projects, companies must be able to prioritize project risks to complete more productive projects.

This study established FGDEMATEL to account for the language ambiguity and uncertainty inherent in human judgment, as well as the interrelationships between the various project risk categories based on the RBS of the PMBOK Guide. Based on the results, managers in Iraqi project-based construction companies need to pay more attention to the Construction risk category than they already do. This is monitored from the Design and Tendering risk categories, as well as the Planning risk category. To obtain more precise and unique results that would be more relevant for specific project types, a future study should be carried out in other industry sectors, such as industrial, IT, and research. The proposed framework allows the user to examine the project at various stages, identify uncertainties, and assess risks associated with specific parties. It also allows the user to add an unlimited number of unpredictable risks to the project and see how they affect time or cost. To explore the most significant stages and provide information on the highest-order risks, this study investigated structural risk assessment while investigating uncertainties and ambiguities utilizing fuzzy logic, combining it with the Grey system and DEMATEL methodology.

The results of this study demonstrate that the proposed framework can be used to predict most risk factors and that, with careful planning, the project can also determine the effect

of a risk factor on time and cost. There is still a great problem with risks in construction projects. Knowing the cause-and-effect groups allows us to clearly indicate how controlling one factor in one phase can affect another factor in another phase. This, in turn, helps reduce delay or cost by speculating and knowing the predetermined relationships through the DEMATEL approach, which is how the proposed framework can assess risks and create interrelationships between them.

ACKNOWLEDGMENT

This article would not have been possible without the help of the people and engineers who participated, the experts who reviewed the potential risks at each level, and the faculty and staff of the University of Baghdad's Civil Engineering Department.

REFERENCES

- [1] M. Askari and H. R. Shokrizade, "An Integrated Method for Ranking of Risk in BOT Projects," *Procedia - Social and Behavioral Sciences*, vol. 109, pp. 1390–1394, Jan. 2014, <https://doi.org/10.1016/j.sbspro.2013.12.642>.
- [2] R. M. Wideman, *Project and Program Risk Management: A Guide to Managing Project Risks and Opportunities*. Drexel Hill, PA, USA: Project Management Institute, 1992.
- [3] B. Carter, T. Hancock, J.-M. Morin, and N. Robins, *Introducing Riskman: The European Project Risk Management Methodology*. Oxford, UK: Blackwell Pub, 1994.
- [4] R. J. Chapman, "The effectiveness of working group risk identification and assessment techniques," *International Journal of Project Management*, vol. 16, no. 6, pp. 333–343, Dec. 1998, [https://doi.org/10.1016/S0263-7863\(98\)00015-5](https://doi.org/10.1016/S0263-7863(98)00015-5).
- [5] *A Guide to the Project Management Body of Knowledge (PMBOK guide)*, Fifth edition. Newtown Square, PA, USA: Project Management Institute, 2013.
- [6] A. A. Alfahad and A. M. Burhan, "BIM-Supporting System by Integrating Risk Management and Value Management," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12130–12137, Dec. 2023, <https://doi.org/10.48084/etasr.6427>.
- [7] S. H. N. Alani and A. M. R. Mahjoob, "Corruption Risk Analysis at the Project Planning Stage in the Iraqi Construction Sector using the Bowtie Methodology," *Engineering, Technology & Applied Science Research*, vol. 11, no. 3, pp. 7223–7227, Jun. 2021, <https://doi.org/10.48084/etasr.4060>.
- [8] A. Alameri, A. S. A. M. Alhammadi, A. H. Memon, I. A. Rahman, and N. A. N. Nasaruddin, "Assessing the Risk Level of the Challenges Faced In Construction Projects," *Engineering, Technology & Applied Science Research*, vol. 11, no. 3, pp. 7152–7157, Jun. 2021, <https://doi.org/10.48084/etasr.4020>.
- [9] S. M. A. K. Firouzabadi and A. V. Nikjoo, "Ranking Of Project Risks Based on the PMBOK Standard by Fuzzy Dematel," in *Proceedings of the 10th International Conference on Manufacturing Research ICMR 2012*, 2012.
- [10] E. K. Zavadskas, Z. Turskis, and J. Tamošaitiene, "Risk assessment of construction projects," *Journal of Civil Engineering and Management*, vol. 16, no. 1, pp. 33–46, Jan. 2010, <https://doi.org/10.3846/jcem.2010.03>.
- [11] A. Yazdani-Chamzini, S. H. Yakhchali, and M. Mahmoodian, "Risk ranking of tunnel construction projects by using the ELECTRE technique under a fuzzy environment," *International Journal of Management Science and Engineering Management*, vol. 8, no. 1, pp. 1–14, Feb. 2013, <https://doi.org/10.1080/17509653.2013.783185>.
- [12] A. S. Bu-Qammaz, I. Dikmen, and M. T. Birgonul, "Risk assessment of international construction projects using the analytic network process," *Canadian Journal of Civil Engineering*, vol. 36, no. 7, pp. 1170–1181, Jul. 2009, <https://doi.org/10.1139/L09-061>.

- [13] S. Ebrahimnejad, S. M. Mousavi, R. Tavakkoli-Moghaddam, and M. Heydar, "Risk ranking in mega projects by fuzzy compromise approach: A comparative analysis," *Journal of Intelligent & Fuzzy Systems*, vol. 26, no. 2, pp. 949–959, Jan. 2014, <https://doi.org/10.3233/IFS-130785>.
- [14] S. Ebrahimnejad, S. M. Mousavi, and S. M. H. Mojtahedi, "A fuzzy decision-making model for risk ranking with an application to an onshore gas refinery," *International Journal of Business Continuity and Risk Management*, vol. 1, no. 1, pp. 38–66, Jan. 2009, <https://doi.org/10.1504/IJBCRM.2009.02895>.
- [15] N. V. Thuyet, S. O. Ogunlana, and P. K. Dey, *Risk management in oil and gas construction projects in Vietnam*. Routledge, 2019.
- [16] S. Q. Wang, M. F. Dulaimi, and M. Y. Aguria, "Risk management framework for construction projects in developing countries," *Construction Management and Economics*, vol. 22, no. 3, pp. 237–252, Mar. 2004, <https://doi.org/10.1080/0144619032000124689>.
- [17] J. H. M. Tah and V. Carr, "A proposal for construction project risk assessment using fuzzy logic," *Construction Management and Economics*, vol. 18, no. 4, pp. 491–500, Jun. 2000, <https://doi.org/10.1080/01446190050024905>.
- [18] R. Tavakkoli-Moghaddam, S. M. Mousavi, and H. Hashemi, "A fuzzy comprehensive approach for risk identification and prioritization simultaneously in EPC projects," in *Risk management in environment, production and economy*, vol. 12, Rijeka, Croatia: InTech, 2011, pp. 123–46.
- [19] B. Oztaysi, "A decision model for information technology selection using AHP integrated TOPSIS-Grey: The case of content management systems," *Knowledge-Based Systems*, vol. 70, pp. 44–54, Nov. 2014, <https://doi.org/10.1016/j.knosys.2014.02.010>.
- [20] Y. Dou, Q. Zhu, and J. Sarkis, "Evaluating green supplier development programs with a grey-analytical network process-based methodology," *European Journal of Operational Research*, vol. 233, no. 2, pp. 420–431, Mar. 2014, <https://doi.org/10.1016/j.ejor.2013.03.004>.
- [21] A. Gabus and E. Fontela, "World problems, an invitation to further thought within the framework of DEMATEL," Battelle Geneva Research Center, 1972.
- [22] A. Gabus and E. Fontela, "Perceptions of the world problematique: Communication procedure, communicating with those bearing collective responsibility," Battelle Geneva Research Center, 1973.
- [23] W. W. Wu and Y. T. Lee, "Developing global managers' competencies using the fuzzy DEMATEL method," *Expert Systems with Applications*, vol. 32, no. 2, pp. 499–507, Feb. 2007, <https://doi.org/10.1016/j.eswa.2005.12.005>.
- [24] D. Julong, "Introduction to Grey System Theory," *The Journal of Grey System*, vol. 1, pp. 1–24, 1989.
- [25] S. Opricovic and G.-H. Tzeng, "Defuzzification within a multicriteria decision model," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 11, no. 05, pp. 635–652, Oct. 2003, <https://doi.org/10.1142/S0218488503002387>.

Predictive Modeling of Groundwater Recharge under Climate Change Scenarios in the Northern Area of Saudi Arabia

Rabie A. Ramadan

Computer Engineering Department, College of Computer Science and Engineering, Hail University, Saudi Arabia | Computer Engineering Department, Faculty of Engineering, Cairo University, Egypt
rabieramadan@gmail.com (corresponding author)

Sahbi Boubaker

Department of Computer and Network Engineering, College of Computer Science and Engineering, University of Jeddah, Saudi Arabia
sboubaker@uj.edu.sa

Received: 6 February 2024 | Revised: 22 February 2024 | Accepted: 1 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7020>

ABSTRACT

Water scarcity is considered a major problem in dry regions, such as the northern areas of Saudi Arabia and especially the city of Hail. Water resources in this region come mainly from groundwater aquifers, which are currently suffering from high demand and severe climatic conditions. Forecasting water consumption as accurately as possible may contribute to a high level of sustainability of water resources. This study investigated different Machine Learning (ML) algorithms, namely Support Vector Machine (SVM), Random Forest (RF), Linear Regression (LR), and Gradient Boosting (GB), to efficiently predict water consumption in such areas. These models were evaluated using a set of performance measures, including Mean Squared Error (MSE), R-squared (R^2), Mean Absolute Error (MAE), Explained Variance Score (EVS), Mean Absolute Percentage Error (MAPE), and Median Absolute Error (MedAE). Two datasets, water consumption and weather data, were collected from different sources to examine the performance of the ML algorithms. The novelty of this study lies in the integration of both weather and water consumption data. After examining the most effective features, the two datasets were merged and the proposed algorithms were applied. The RF algorithm outperformed the other models, indicating its robustness in capturing water usage behavior in dry areas such as Hail City. The results of this study can be used by local authorities in decision-making, water consumption analysis, new project construction, and consumer behavior regarding water usage habits in the region.

Keywords-groundwater; recharge; Hail city; machine learning; prediction

I. INTRODUCTION

Today, many concerns are raised about environmental conditions and their effects on the global population. This raises the issue of water management as a natural resource. Water in countries such as Saudi Arabia, especially in northern areas such as Hail City, is a precious resource [1] and management and prediction of water consumption are challenging problems. This study uses Machine Learning (ML) capabilities in water management in the northern areas of Saudi Arabia, focusing on predictive analytics and using available datasets of water consumption and weather data [2-3]. Instead of traditional forecasting methods, which depend on historical data, this study aims to evaluate and compare four ML models: Support Vector Machine (SVM), Random Forest (RF), Linear Regression (LR), and Gradient Boosting (GB) in the context of water consumption prediction. These models were selected

because of their proven capabilities in similar forecasting frameworks when applied to complex multidimensional data.

Data collection is the first step of a successful prediction. In this regard, the two datasets used were collected from different sources [2-3]. This study uses a water consumption dataset, which contains records of monthly water consumption, and comprehensive hourly weather observations. Modeling the relationship between weather data and water consumption is crucial for managing water resources under climate change conditions, as water demand may increase following an increase in temperature [4]. Additionally, other weather data, such as wind speed and visibility, greatly affect water consumption.

This study aims to make a significant contribution to the prediction of water consumption in the northern area of Saudi Arabia, which has a special climate with a scarcity of rain and

groundwater. Four methods were evaluated with the Mean Square Error (MSE), R square (R^2), Mean Absolute Error (MAE), Explained Variance Score (EVS), Mean Absolute Percentage Error (MAPE), and Median Absolute Error (MedAE). These measures were selected to reflect the accuracy and reliability of the proposed algorithms in forecasting groundwater consumption in northern Saudi Arabia. For example, MSE reflects the sensitivity of large errors, which critically assesses the accuracy of the prediction, whereas R^2 and EVS measure the proportion of variance in water consumption that the models could explain, indicating the fitting of the models used to the datasets and their ability to capture the underlying patterns in water usage. Therefore, they perfectly reflect the study's objective of developing a model that accurately reflects consumption dynamics. MAE and MedAE indicate the average magnitude and central tendency of these errors, respectively. These measures are less sensitive to outliers than MSE. Therefore, a more robust view of model performance in different scenarios is very important in water resource management. Finally, MAPE is a relative error metric that was chosen to reflect the perspective on the error size relative to actual consumption values and is particularly important to policymakers because it gives an intuitive understanding of model accuracy in percentage terms.

The results could benefit the local authorities of Saudi Arabia in managing water consumption in these areas and extend the results to many similar locations. However, in leveraging predictive analytics for water management, ethical integrity requires transparency, privacy protection, and equity, ensuring that the models do not perpetuate biases. Crucially, engaging communities in decision-making processes fosters trust and ensures that water management practices are inclusive and fair, respect diverse needs, and promote equitable access to water resources.

II. RELATED WORKS

The prediction of water consumption and Groundwater Level (GWL) plays a crucial role in water resource management. Therefore, researchers and water systems practitioners have been very interested in this topic for the last few decades. Traditional forecasting techniques have shown several limitations, although the potential they exhibited when applied to water time series. Some of these methods have linear relationships between variables. However, water consumption and GWLs often exhibit non-linear behavior patterns. Additionally, many techniques frequently require the time series data to be stationary, meaning the statistical properties of the series (mean, variance, autocorrelation, etc.) are constant over time. Additionally, they may focus on historical data patterns and not effectively incorporate external factors or variables. Moreover, they face difficulty in handling large and complex datasets. On the other hand, machine learning techniques consider the variability in the dataset and can handle large datasets.

Water consumption greatly affects water resources, including GWLs, since climatic, topographic, and hydrological factors interact, making simulation and prediction challenging. In [5], a comprehensive overview of soft computing techniques, including ML, was provided to predict GWLs

while considering the techniques used, the study location, the dataset, and the main performance metrics. This study showed that the prediction of GWLs is case-sensitive. In [6], a real-time framework was presented to predict short-term water consumption, based on a back-propagation ML algorithm and Long-Short-Term Memory (LSTM). The results obtained were promising, as they were found to have an accuracy of a few liters. Combining human mobility, historical water consumption, and applying ML algorithms such as SVM and tree-based ensemble methods yielded accurate forecasts for a case study in Wroclaw, Poland [7]. Compared to classical techniques such as ARIMA, ML techniques have improved accuracy. In [8], the water irrigation demand was estimated based on an RF algorithm. The results provided a projected spatial resolution of 6 km, considering evapotranspiration and meteorological data.

GWL prediction, as among the drivers of water consumption, has been extensively considered using ML algorithms [9-17]. GWL is in mutual interaction with water demand/consumption, mainly in dry regions with extensive agricultural activities such as Hail City, Saudi Arabia. GWL is affected by several factors, including climatic conditions [9] and regional characteristics [10]. The groundwater depth usually exhibits fluctuations due to climate, water extraction/demand, surface water flows, and rainfall. The relationship between GWL and these factors is known to be highly nonlinear. The classic techniques usually fail to capture those complex relationships. In [11], ML methods were used to model these relationships, and the tested Deep Learning (DL) models provided good results, although they suffered from relatively limited datasets. Authors in [12] highlighted that the studied aquifer was in a critical situation and recommended urgent management actions. A new trending method to predict GWL consists of using ensemble DL methods [13], as the results have shown their superiority over single methods. Combined methods have also been investigated while using different techniques, sometimes from different categories, such as ML with wavelet transforms [14]. ML and DL have also been investigated in the domain of water prediction [15-17]. However, case studies from arid regions with scarce water resources and difficult climatic conditions, such as in the north of Saudi Arabia, were not adequately investigated.

This study fills this research gap by investigating water consumption in the northern area of Saudi Arabia using different algorithms, including RF, GB, SVM, and LR. Similar frameworks to these methods can be found in [18-24]. The novel contributions of this study are that it is carried out for a special area with special characteristics, it also merges two datasets for efficient water consumption prediction, including water and weather data, and finally applies different ML algorithms to the merged datasets.

III. PROPOSED SOLUTION APPROACH

Figure 1 shows the proposed approach, which is broken down into five key steps:

- Data collection: At this stage, data are collected from different sources [2-3]. The collected water data include historical observations of water consumption and historical

weather data. The quality of the data collected certainly affects the accuracy of the prediction process.

- **Data Processing:** It is a critical step that cleans the data and transforms them into a suitable format for analysis. Preprocessing operations include finding missing values, normalization, and selecting relative features. Preprocessing also involves recognizing seasonal variations and timestamp conversion.
- **Data Split:** In this phase, the data are divided into training and testing sets. The training data are used to train the

investigated models, and the test data are used to test the accuracy of the prediction.

- **Apply Models:** ML models are applied to the data. Each model requires specific data handling. This step is critical to evaluate the performance of the proposed models.
- **Evaluation of the models:** In the last phase, the results are evaluated. Different evaluation methods were chosen, such as MSE, R^2 , MAE, EVS, MAPE, and MedAE.



Fig. 1. Proposed approach.

This study used the following ML models.

A. Random Forest (RF) Regressor

RF is an ensemble learning technique mostly used for classification and regression tasks [18]. It constructs multiple trees during the training phase, based on the type of data, and gives a chance to avoid overfitting, which is common for individual decision trees. It does this by averaging over a multitude of decision trees. The algorithm is also suitable for complex and nonlinear data. It also provides valuable information about the feature's importance, leading to the selection of important and effective features. In addition, it can handle various data types, which makes it suitable for ML applications.

B. Gradient Boosting (GB) Regressor

GB is known for its computational efficiency and performance, proven in large- and complex datasets. It involves regularization (L1 and L2) that helps avoid overfitting. An important feature of GB is its efficient handling of missing data, which makes it robust for modeling without extensive preprocessing. Moreover, it has tree pruning, which limits tree growth during the learning process [19].

C. Linear Regression (LR)

LR is an ML technique that is used to model the relationship between two variables by fitting a linear equation to the observed data. Some variables are considered explanatory, and another variable is considered the dependent variable. In this study, the water demand is considered the dependent variable, and the explanatory variables are the climatic conditions [25].

D. Support Vector Machines (SVMs)

SVMs consist of a group of supervised learning methods that are usually used in classification and regression problems. SVMs are extended to nonlinear models of the generalized algorithm developed for linear regressions [20].

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The RF was designed with 100 estimators and a fixed random state for reproducibility. The RF model, as shown in

Figure 2, showed reasonable results, with an R^2 of approximately 0.694. This indicates that the model was able to explain 69.4% of the variance in water consumption. At the same time, MSE was 0.549 and MAE was 0.529, which are relatively low. In addition, a substantial portion of the variance in water consumption was shown through EVS, which was 0.694. The MAPE of 3.05% indicates a lower prediction error, while the MSLE of 0.0016 and MedAE of 0.371 again confirm its accuracy.

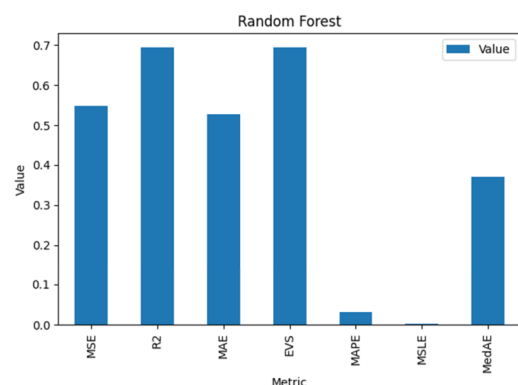


Fig. 2. RF evaluation metrics.

Figure 3 shows the results of the LR, indicating that it is a moderate fit to predict water consumption. LR had R^2 , MSE, MAE, EVS, MAPE, and MAE of 0.470, 0.951, 0.846, 0.470, 4.89%, and 0.808, respectively. These results indicate that the model was able to explain nearly 47% of the variance in water consumption with a moderate level of accuracy. The results assume that the model accounts for a substantial portion of the variance in water consumption and offers a robust assessment of central prediction errors. The SVM results, shown in Figure 4, indicate weak performance, as R^2 was approximately 0.0039, indicating that the model captured only 0.39% of the variance in water consumption. This evaluation was also confirmed by MAE, EVS, MAPE, MSLE, and MedAE. The GB was configured similarly to the RF with 100 estimators.

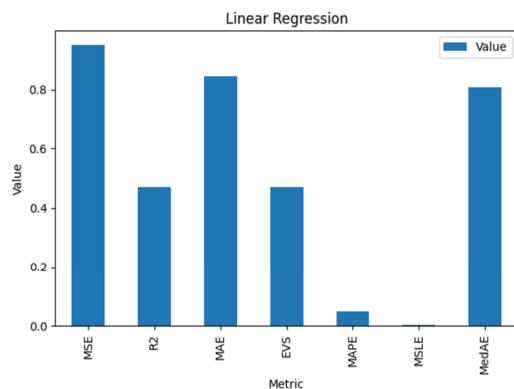


Fig. 3. LR evaluation metrics.

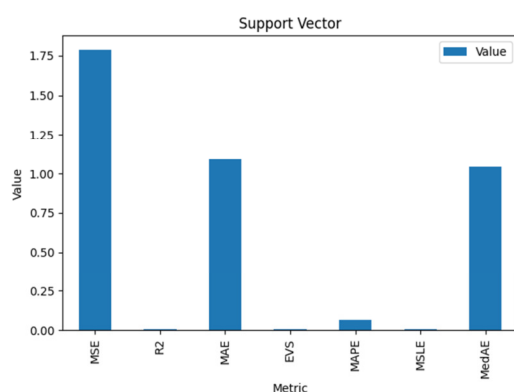


Fig. 4. SVM evaluation metrics.

Figure 5 shows the GB loss curve, plotting MSE against the number of estimators for training and validation data. The figure suggests effective learning with increasing estimators. As the number of estimators continues to increase, the curve plateaus indicating diminishing returns on error reduction, and the model is generalizing well and not overfitting. Figure 6 presents the comparison between the actual and the predicted values from a GB model. The red (predicted) and blue (actual) points show the overall accuracy of the model. Figure 7 shows the performance of the GB model, showing an R^2 score of 0.596, indicating that it can account for approximately 59.63% of the variance in water consumption. Similarly, MSE equals 0.724, implying that the model's predictions have a relatively moderate level of accuracy. Also, MAE of 0.711 and EVS of 0.596 indicate that, on average, the model's predictions deviate by this amount from the actual values. Other measures confirm the given results.

Table I summarizes the evaluation results of the considered algorithms. RF outperformed the other models in almost all metrics. GB follows closely behind, showing robust performance. LR and SVM lagged, with LR performing moderately, and SVM showed the worst results. Although RF shows the best results among the other models, its accuracy is still not up to the expected results. These results reflect the complexity of the prediction based on two different datasets that were not collected for water consumption. To the best of our knowledge, no previous study has combined weather and

water data, especially in the northern areas of Saudi Arabia. However, this study is a step forward in the relationship between weather data and water use and consumption in different areas.

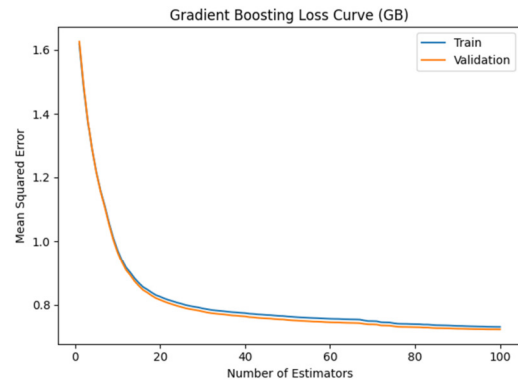


Fig. 5. GB loss curve.

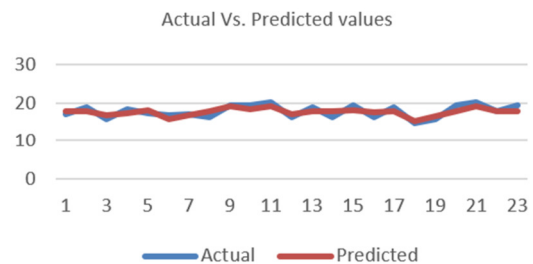


Fig. 6. GB actual vs predicted values.

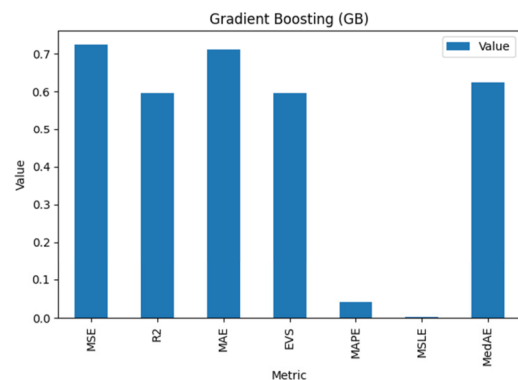


Fig. 7. GB performance metrics.

TABLE I. PERFORMANCE COMPARISON

Algorithm	MSE (lt ²)	R ² (%)	MAE (lt)	EVS (unitless)	MAPE(%)	MedAE (lt)
SVM	1.7864	0.0039	1.0929	0.0053	0.0634	1.0411
RF	0.5489	0.6939	0.5288	0.6939	0.0305	0.3705
LR	0.9506	0.4699	0.8461	0.4700	0.0489	0.8081
GB	0.7240	0.5962	0.7111	0.5963	0.0411	0.6239

V. CHALLENGES AND LIMITATIONS

Studying water consumption in areas such as northern regions of Saudi Arabia faces challenges such as data scarcity and quality, reliance on non-renewable groundwater, climate change, socioeconomic growth, technological and infrastructural limitations, and environmental concerns. This area has limited historical quality datasets, which hinder accurate modeling and forecasting of water usage. This study had to collect data from various sources to obtain enough for the investigation. Also, the combination of underground water and weather data demands careful resource management. Additionally, government policies and the need to balance ecological sustainability with human water use present complex challenges. The complexity of predicting water consumption is increased by the interplay of various external factors, making accurate forecasting challenging.

VI. IMPLICATIONS OF THE FINDINGS FOR WATER RESOURCE MANAGEMENT POLICIES

The results of this study can have a great impact on water resource management policies in northern Saudi Arabia, highlighting the critical role of predictive analytics in the formation of sustainable strategies. Using predictive analytics leads to a better understanding of patterns of water consumption and prediction of future demands. Therefore, policymakers could have results to make informed decisions. The results of this research identify the key factors driving water consumption in the region. Policies can target these factors to reduce water consumption and usage. Adaptive analytics could optimize the allocation of water resources. Therefore, policymakers can ensure that water distribution is managed more efficiently, decreasing wastage and ensuring water is available where and when it is most needed. Intelligent irrigation systems for agriculture could be developed, ensuring that water is used precisely and efficiently to meet crop needs without excess. Finally, the results of this study highlight the importance of integrating predictive analytics into long-term water resource planning and infrastructure development. Therefore, policymakers can better plan for the infrastructure of water storage and distribution.

VII. CONCLUSION

The results of this study underscore the potential of ML techniques, especially RF, in predicting water consumption in areas such as northern Saudi Arabia and Hail City. RF showed superior performance with the lowest MSE and the highest R^2 and EVS, indicating its ability to handle non-linear and interesting water usage patterns. However, SVM was unable to integrate and perform well in the given water and weather datasets. This study is one of the first that integrates weather data into water usage data in such areas, where weather and water consumption usage and sources are exceptional cases. This study could be extended to incorporate the interaction between climate conditions and agricultural activities or between underground water consumption and consumer behavior. Future research will focus on the use of different algorithms, including DL algorithms, to refine the prediction. Also, real-time analytics is expected to have more value than offline prediction.

ACKNOWLEDGMENT

This research was funded by the Ali Aljomaiya Scientific Research Chair for Sustainable Development at the University of Hail, Saudi Arabia, through Project Number SCR-22 045.

REFERENCES

- [1] A. I. Almulhim and I. R. Abubakar, "Developing a sustainable water conservation strategy for Saudi Arabian cities," *Groundwater for Sustainable Development*, vol. 23, Nov. 2023, Art. no. 101040, <https://doi.org/10.1016/j.gsd.2023.101040>.
- [2] "Saudi Arabia Hourly Climate Integrated Surface Data." <https://datasource.kapsarc.org/explore/dataset/saudi-hourly-weather-data/information>.
- [3] "Per Capita Water Consumption In Saudi Regions." https://datasource.kapsarc.org/explore/dataset/per_capita_average_water_use_in_saudi_regions/information.
- [4] D. Fiorillo, Z. Kapelan, M. Xenochristou, F. De Paola, and M. Giugni, "Assessing the Impact of Climate Change on Future Water Demand using Weather Data," *Water Resources Management*, vol. 35, no. 5, pp. 1449–1462, Mar. 2021, <https://doi.org/10.1007/s11269-021-02789-4>.
- [5] H. Tao *et al.*, "Groundwater level prediction using machine learning models: A comprehensive review," *Neurocomputing*, vol. 489, pp. 271–308, Jun. 2022, <https://doi.org/10.1016/j.neucom.2022.03.014>.
- [6] A. Boudhaouia and P. Wira, "A Real-Time Data Analysis Platform for Short-Term Water Consumption Forecasting with Machine Learning," *Forecasting*, vol. 3, no. 4, pp. 682–694, Dec. 2021, <https://doi.org/10.3390/forecast3040042>.
- [7] K. Smolak, B. Kasieczka, W. Fialkiewicz, W. Rohm, K. Siła-Nowicka, and K. Kopańczyk, "Applying human mobility and water consumption data for short-term water demand forecasting using classical and machine learning models," *Urban Water Journal*, vol. 17, no. 1, pp. 32–42, Jan. 2020, <https://doi.org/10.1080/1573062X.2020.1734947>.
- [8] S. Wei, T. Xu, G. Y. Niu, and R. Zeng, "Estimating Irrigation Water Consumption Using Machine Learning and Remote Sensing Data in Kansas High Plains," *Remote Sensing*, vol. 14, no. 13, Jan. 2022, Art. no. 3004, <https://doi.org/10.3390/rs14133004>.
- [9] F. B. Banadkooki *et al.*, "Enhancement of Groundwater-Level Prediction Using an Integrated Machine Learning Model Optimized by Whale Algorithm," *Natural Resources Research*, vol. 29, no. 5, pp. 3233–3252, Oct. 2020, <https://doi.org/10.1007/s11053-020-09634-2>.
- [10] H. Cai, H. Shi, S. Liu, and V. Babovic, "Impacts of regional characteristics on improving the accuracy of groundwater level prediction using machine learning: The case of central eastern continental United States," *Journal of Hydrology: Regional Studies*, vol. 37, Oct. 2021, Art. no. 100930, <https://doi.org/10.1016/j.ejrh.2021.100930>.
- [11] D. Kumar, T. Roshni, A. Singh, M. K. Jha, and P. Samui, "Predicting groundwater depth fluctuations using deep learning, extreme learning machine and Gaussian process: a comparative study," *Earth Science Informatics*, vol. 13, no. 4, pp. 1237–1250, Dec. 2020, <https://doi.org/10.1007/s12145-020-00508-y>.
- [12] H. Kardan Moghaddam, S. Ghordoyee Milan, Z. Kayhomayoon, Z. Rahimzadeh kivi, and N. Arya Azar, "The prediction of aquifer groundwater level based on spatial clustering approach using machine learning," *Environmental Monitoring and Assessment*, vol. 193, no. 4, Mar. 2021, Art. no. 173, <https://doi.org/10.1007/s10661-021-08961-y>.
- [13] H. A. Afan *et al.*, "Modeling the fluctuations of groundwater level by employing ensemble deep learning techniques," *Engineering Applications of Computational Fluid Mechanics*, vol. 15, no. 1, pp. 1420–1439, Jan. 2021, <https://doi.org/10.1080/19942060.2021.1974093>.
- [14] A. T. M. S. Rahman, T. Hosono, J. M. Quilty, J. Das, and A. Basak, "Multiscale groundwater level forecasting: Coupling new machine learning approaches with wavelet transforms," *Advances in Water Resources*, vol. 141, Jul. 2020, Art. no. 103595, <https://doi.org/10.1016/j.advwatres.2020.103595>.
- [15] M. Sapitang, W. M. Ridwan, A. N. Ahmed, C. M. Fai, and A. El-Shafie, "Groundwater level as an input to monthly predicting of water level

- using various machine learning algorithms," *Earth Science Informatics*, vol. 14, no. 3, pp. 1269–1283, Sep. 2021, <https://doi.org/10.1007/s12145-021-00654-x>.
- [16] W. Liu, H. Yu, L. Yang, Z. Yin, M. Zhu, and X. Wen, "Deep Learning-Based Predictive Framework for Groundwater Level Forecast in Arid Irrigated Areas," *Water*, vol. 13, no. 18, 2021, <https://doi.org/10.3390/w13182558>.
- [17] W. Li, M. M. Finsa, K. B. Laskey, P. Houser, and R. Douglas-Bate, "Groundwater Level Prediction with Machine Learning to Support Sustainable Irrigation in Water Scarcity Regions," *Water*, vol. 15, no. 19, Jan. 2023, Art. no. 3473, <https://doi.org/10.3390/w15193473>.
- [18] F. Mlaw, E. Mkoba, and N. Mduma, "A Machine Learning Model for detecting Covid-19 Misinformation in Swahili Language," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10856–10860, Jun. 2023, <https://doi.org/10.48084/etasr.5636>.
- [19] M. Sipper and J. H. Moore, "AddGBoost: A gradient boosting-style algorithm based on strong learners," *Machine Learning with Applications*, vol. 7, Mar. 2022, Art. no. 100243, <https://doi.org/10.1016/j.mlwa.2021.100243>.
- [20] K. Theofilatos, S. Likothanassis, and A. Karathanasopoulos, "Modeling and Trading the EUR/USD Exchange Rate Using Machine Learning Techniques," *Engineering, Technology & Applied Science Research*, vol. 2, no. 5, pp. 269–272, Oct. 2012, <https://doi.org/10.48084/etasr.200>.
- [21] S. Benítez-Peña, R. Blanquero, E. Carrizosa, and P. Ramírez-Cobo, "Cost-sensitive probabilistic predictions for support vector machines," *European Journal of Operational Research*, vol. 314, no. 1, pp. 268–279, Apr. 2024, <https://doi.org/10.1016/j.ejor.2023.09.027>.
- [22] R. G. Siqueira *et al.*, "Modelling and prediction of major soil chemical properties with Random Forest: Machine learning as tool to understand soil-environment relationships in Antarctica," *CATENA*, vol. 235, Feb. 2024, Art. no. 107677, <https://doi.org/10.1016/j.catena.2023.107677>.
- [23] X. C. Nguyen *et al.*, "Estimating ammonium changes in pilot and full-scale constructed wetlands using kinetic model, linear regression, and machine learning," *Science of The Total Environment*, vol. 907, Jan. 2024, Art. no. 168142, <https://doi.org/10.1016/j.scitotenv.2023.168142>.
- [24] S. Mondal, S. Ghosh, and A. Nag, "Brain stroke prediction model based on boosting and stacking ensemble approach," *International Journal of Information Technology*, vol. 16, no. 1, pp. 437–446, Jan. 2024, <https://doi.org/10.1007/s41870-023-01418-0>.

Deep Learning for Tomato Disease Detection with YOLOv8

Hafedh Mahmoud Zayani

Department of Electrical Engineering, College of Engineering, Northern Border University, Saudi Arabia
hafedh.zayani@nbu.edu.sa (corresponding author)

Ikhlass Ammar

Computer Science Department, Faculty of Sciences of Tunis (FST), University of Tunis El Manar, Tunisia | OASIS Laboratory, National Engineering School of Tunis, University of Tunis El Manar, Tunisia
ikhlass_ammam@yahoo.fr

Refka Ghodhbani

Department of Computer Sciences Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
refka.ghodhbani@nbu.edu.sa

Albia Maqbool

Department of Computer Sciences Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
albia.alam@nbu.edu.sa

Taoufik Saidani

Department of Computer Sciences Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
taoufik.saidan@nbu.edu.sa

Jihane Ben Slimane

Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
jehan.saleh@nbu.edu.sa

Amani Kachoukh

Department of Information Systems, Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
amani.khasookh@nbu.edu.sa

Marouan Kouki

Department of Information Systems, Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
marouan.kouki@nbu.edu.sa

Mohamed Kallel

Department of Physics, Faculty of Sciences and Arts, Northern Border University, Saudi Arabia
mohamed.kallel@nbu.edu.sa

Amjad A. Alsuwaylimi

Department of Information Technology, College of Computing and Information Technology, Northern Border University, Saudi Arabia
amjad.alsuwaylimi@nbu.edu.sa

Sami Mohammed Alenezi

Department of Information Technology, College of Computing and Information Technology, Northern Border University, Saudi Arabia
sami.m.alenezi@nbu.edu.sa

Received: 13 February 2024 | Revised: 28 February 2024 | Accepted: 29 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7064>

ABSTRACT

Tomato production plays a crucial role in Saudi Arabia, with significant yield variations due to factors such as diseases. While automation offers promising solutions, accurate disease detection remains a challenge. This study proposes a deep learning approach based on the YOLOv8 algorithm for automated tomato disease detection. Augmenting an existing Roboflow dataset, the model achieved an overall accuracy of 66.67%. However, class-specific performance varies, highlighting challenges in differentiating certain diseases. Further research is suggested, focusing on data balancing, exploring alternative architectures, and adopting disease-specific metrics. This work lays the foundation for a robust disease detection system to improve crop yields, quality, and sustainable agriculture in Saudi Arabia.

Keywords-disease detection tomato; deep learning; YOLOv8

I. INTRODUCTION

Tomatoes, grown worldwide and yielding approximately 170 million tons annually, are essential in vegetable crop production [1]. Despite the harsh environmental conditions in Saudi Arabia, characterized by arid and hot climates, the Greenhouse Horticulture Business Unit of Wageningen University & Research (WUR) has achieved remarkable success, boasting an annual tomato yield of 80 kg/m² [2]. This achievement is attributed to advanced greenhouse technology and precise irrigation methods. In Saudi Arabia, tomato cultivation spans 1260 hectares of greenhouse area, exhibiting varied productivity ranging from 8-10 kg/m² under uncontrolled conditions to 35-45 kg/m² in managed greenhouses, notably lower than in countries such as the Netherlands where yields can exceed 80-90 kg/m². The disparity in crop yields is influenced by agricultural practices, climate, water quality, and tomato cultivar selection, factors that still lack precise understanding despite the introduction of several tomato varieties annually.

In addition, tomato harvesting faces challenges with damage rates as high as 10% [3-4]. However, advances in automation technology offer promising solutions [5], focusing on image recognition, precise positioning, and efficient picking mechanisms. Traditionally, early fruit recognition relied on complex machine learning algorithms, but the advent of Convolutional Neural Networks (CNNs) has significantly improved efficiency and precision [6-9]. Object detection algorithms, crucial for automating tomato disease detection, include two-stage algorithms such as R-CNN [10] and one-stage algorithms such as SSD [11-12] and YOLO [13], with various iterations of YOLO (v3 [14-15], v4 [16], v5 [17-20], v7 [21], and v8 [22]) that improve object detection capabilities.

This study proposes a deep learning-based approach for tomato disease detection, leveraging a YOLO-based algorithm trained on a custom dataset to improve tomato quality and increase crop yields in remote sensing technologies. The key contributions of this study are:

- Prepare a custom dataset and apply the YOLOv8 algorithm to improve disease detection in tomatoes and potentially increase crop yields.
- Present a carefully trained model that uses a wide array of images, including those collected from databases. The effectiveness of the model is thoroughly evaluated on validation and testing sets, as well as on external photos.

II. RELATED WORKS

Since the advent of DL algorithms, significant advances have been made in plant disease detection, particularly in leaf disease identification and diagnosis, contributing to the progress of precision agriculture. These developments, driven by computer vision techniques and neural networks, have led to the creation of machine learning and deep learning models tailored to accurately diagnose leaf diseases such as early blight, late blight, bacterial spot, and powdery mildew in various plant species, including tomatoes. Using image processing techniques and CNNs, these models analyze leaf images and classify diseases according to visual symptoms. In addition, spectral and hyperspectral imaging techniques have been explored for early disease detection, bolstering plant health monitoring. Although disease detection efforts have shown high accuracy rates, there is a lag in tomato fruit disease detection. Although many models have been developed to identify fruit diseases such as blossom end rot, fruit rot, and sunscald, the focus remains primarily on leaf diseases. This discrepancy underscores the need to expand research efforts

toward fruit disease detection to advance tomato disease management and overall crop productivity. Currently, datasets for tomato fruit disease detection are limited, typically containing fewer than five classes, with less emphasis compared to leaf disease datasets, reflecting the research imbalance between the two domains.

In [23], a 15-layer CNN served as the backbone of the Single-Shot Detector (SSD) to improve the detection of healthy tomato fruits and three classes of tomato fruit diseases. The proposed CNN-SSD model outperformed state-of-the-art models, achieving a significantly higher detection accuracy. In [24], the detection and classification of tomato plant diseases was automated using a Raspberry Pi. This study used image processing techniques and CNN-based classification to successfully detect diseases such as late blight, gray spot, and bacterial canker. In [25], YOLOv5m was used to classify tomato fruits into ripe, immature, and damaged categories. These models demonstrated high prediction accuracy, particularly in ripe and immature tomatoes, offering potential applications in automated tomato fruit harvesting. In [26], an improved YOLOv5 tomato detection algorithm, called CAM-YOLO, was introduced that incorporated a Convolutional Block Attention Module (CBAM) to improve accuracy in detecting small and overlapped tomatoes, achieving an average precision of 88.1%. These studies show ongoing efforts to improve the detection of tomato fruit diseases, paving the way for advances in agricultural practices and crop management.

III. MATERIALS AND METHODS

A. Data Improvement

This study used a dataset called "balanceddata Dataset" [27] in Roboflow, published in October 2023, and contained three types of tomato diseases. Figure 1 depicts an illustrative image of a class. This dataset was augmented to improve the model performance. Although the original dataset contains 906 images, the augmented dataset consists of 2152 images. The augmented dataset was divided into three sets for training, validation, and testing tasks. Specifically, 87% or 1872 images were allocated for the training set, 9% or 188 images for the validation set, and 4% or 92 images for the testing set. The dataset was partitioned into these subsets with an average ratio of 87:9:4. Figure 2 shows the number of labels corresponding to each class and the size of the labels used in the dataset. YOLOv8 was used as the foundational architecture for the proposed model, as it offers enhanced efficiency and flexibility and addresses three key computer vision tasks:

- **Classification:** This task involves the identification of a predominant class within an input image, with the model outputting a class index along with a confidence score. Classification proves to be beneficial for discerning the presence of specific classes in an image. YOLOv8 offers five sizes, namely yolov8n (nano), yolov8s (small), yolov8m (medium), yolov8l (large), and yolov8x (extra large), each tailored for specific purposes [28].
- **Detection:** Building on classification, detection involves identifying various classes present in an image and precisely locating them using bounding boxes.

- **Segmentation:** Positioned beyond object detection, segmentation delves into identifying individual pixels belonging to objects, offering a more precise understanding compared to object detection alone. This technique finds diverse applications and provides detailed insight into objects within an image.

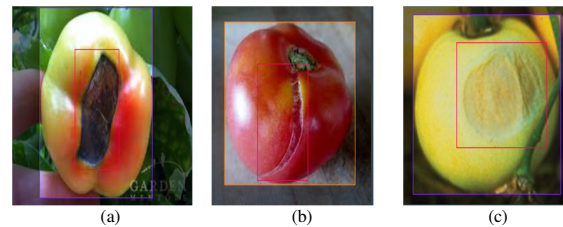


Fig. 1. Dataset classes: (a) Blossom end rot, (b) splitting, (c) sun scaled.

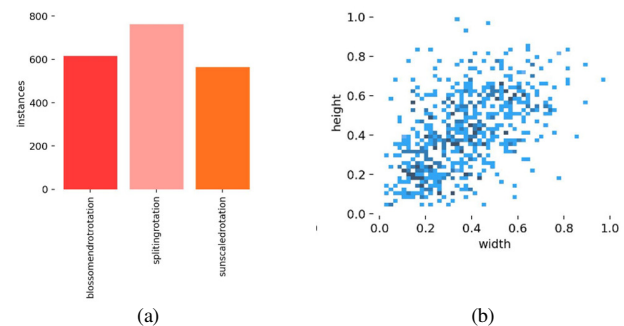


Fig. 2. (a) Number of labels in a class, (b) and different sizes of the labels.

B. Architecture of YOLOv8l

The YOLOv8 design consists of 53 convolutional layers and is implemented as a CNN. In particular, it integrates cross-stage partial connections, enabling improved information transmission between layers for enhanced performance. YOLOv8 stands out for its ability to efficiently detect objects at multiple scales because it employs a feature pyramid network. This network uses multiple layers specifically engineered to detect objects of different sizes and scales within an image, guaranteeing extensive coverage across various object dimensions. The structure of YOLOv8 can be dissected into three main elements [28].

1) Backbone

This component consists of a pre-trained CNN that is responsible for extracting low-, medium-, and high-level feature maps from the input images. Figure 3 shows the dimensions of the input images in four different sizes. Functionally, this component of YOLOv8 largely approximates that of YOLOv5, with the main divergence being the mosaic data augmentation technique. Mosaic data augmentation includes the simultaneous presentation of many resized images to the model, improving its ability to recognize objects from different viewpoints and in situations where they are partially obscured by merging four images. However, it has been observed that the implementation of mosaic data augmentation in YOLOv8 tends to reduce performance, terminating its discontinuation during the last 10 training epochs. To boost

model efficiency, YOLOv5's C3 module was replaced with the C2f module, inspired by YOLOv7's ELAN concept. C2f, shown in Figure 4, unlike C3, which uses only the last output, uses all the outputs of the choke unit. The C2f structure consists of two convolution blocks (this process reduces the number of channels in the feature map usually by half) and three bottleneck convolution blocks. The structural bottleneck blocks, shown in Figure 5, are composed of two convolutional blocks and a contact. This block reduces the calculations and improves the training speed. C2f optimizes the network structure by regulating gradient routes, improving trainability. Additionally, it effectively learns multiscale features and expands receptive fields through feature diversion and multilevel convolution, all without altering the network's shortest and longest gradient paths. This leads to a lighter model with preserved feature richness. Figure 6 shows a backbone block, which comprises a sequence of convolutional layers responsible for extracting significant features from the input image at different sizes. Toward the end of the backbone, the Spatial Pyramid Pooling Fusion (SPPF) layer, followed by

convolutional layers, processes features across various scales, while upsample layers improve the resolution of feature maps.



Fig. 3. The different sizes of an image in backbone.

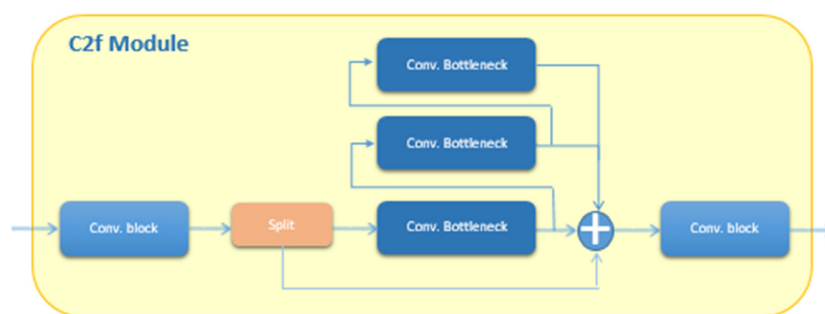


Fig. 4. The structure of the C2f convolutional block.

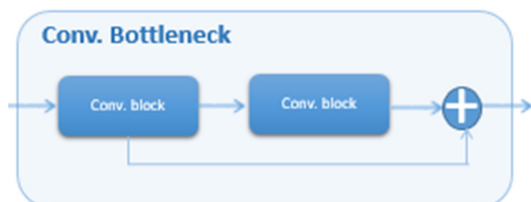


Fig. 5. The structure of the convolutional bottleneck block.

2) Neck

The neck is the bridge between the backbone's raw feature maps and the head's final predictions. It refines and merges information from the backbone's raw features before feeding it to the head for final predictions. It employs path aggregation blocks, such as FPNs, to optimize multiscale feature integration [29], further enhanced by upsampling and numerous C2f modules. Figure 7 shows three components: the upsampling block that aims to increase the spatial resolution of the feature maps extracted from the backbone, the c2f module that manipulates the feature maps to extract more relevant information for object detection, and the conventional block. These components can extract higher-level features, reduce noise, and improve the model's ability to differentiate between different objects. YOLOv8 also decouples confidence and

regression boxes in the later stages of the neck, increasing accuracy. Through these mechanisms, the neck prepares high-quality data for the head, ensuring accurate object detection.

3) Head

The head of YOLOv8 acts as the decision-maker, taking the refined information from the neck and turning it into actionable results. Unlike previous iterations, the head no longer combines classification and regression tasks. Instead, it performs these tasks independently, resulting in enhanced model performance. The feature maps are transmitted to the head, where objects are classified, and predictions for bounding boxes, objectness scores, and class probabilities are made for detected objects within an image. The head comprises multiple convolutional layers followed by a sequence of fully connected layers, as shown in Figure 8. YOLOv8 takes object detection up a notch with its innovative "anchor-free" approach. Ditching the limitations of predefined shapes, it predicts bounding boxes directly, seamlessly adapting to diverse object sizes and shapes. However, YOLOv8 incorporates a self-attention mechanism that mimics this focus, allowing it to prioritize relevant image features for even more precise detection. This powerful combination of independent analysis and keen attention to detail makes the YOLOv8 head a critical player in delivering accurate results.

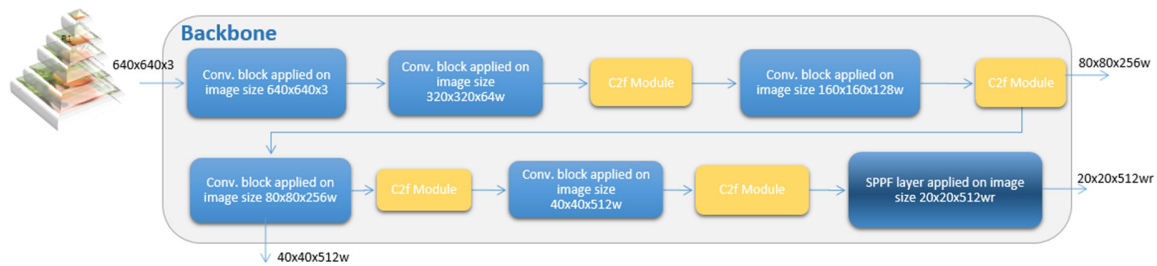


Fig. 6. The structure of the backbone block.

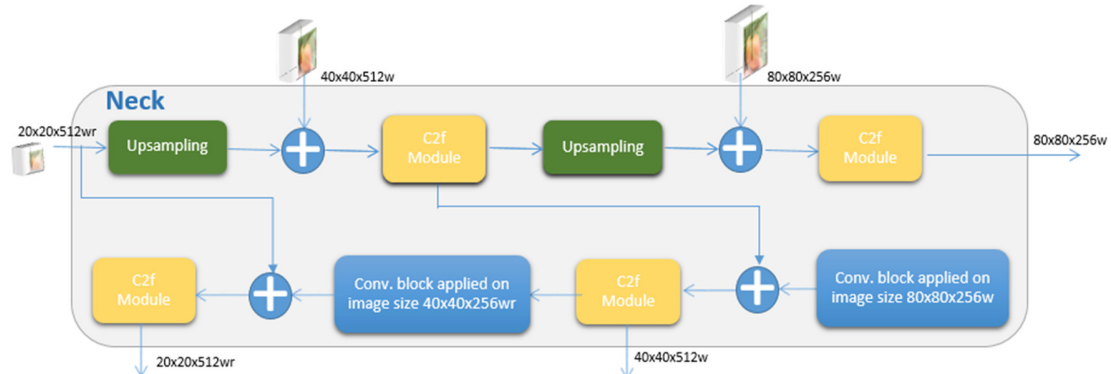


Fig. 7. The structure of the neck.



Fig. 8. The structure of the head block.

IV. EXPERIMENTAL RESULTS ANALYSIS

A. Hardware Environment

The experiments were conducted using Google Colab, a free cloud-based platform that offers a convenient notebook environment for research because of its seamless integration with popular machine learning libraries such as TensorFlow and PyTorch. Google Colab provides the flexibility to choose between different hardware options: CPUs, GPUs, and even TPUs, each offering a 12-hour continuous execution window. To maximize performance, the 12GB NVIDIA Tesla T4 GPU was used, achieving significant computational power at no cost. The models were trained for 50 epochs, processing small batches of 3 images at a time. Each image was resized to 640 pixels to balance model performance with training efficiency.

B. Results

Figure 10 shows the effectiveness of the proposed YOLO-based tomato disease detection algorithm evaluated using precision, recall, and mean Average Precision (mAP). Precision ensures minimal false alarms by measuring the proportion of true positives in detections, reflected by a consistently high precision curve. Recall measures the efficiency in identifying actual diseased plants, depicted by a high recall curve. Finally, mAP combines both, providing a holistic view of performance across various diseases. These metrics were used to objectively evaluate the model's ability to accurately and efficiently detect diseases, minimizing both false positives and missed infections.



Fig. 9. Some experimental results.

Figure 11(a) shows the confidence-precision curves. Each curve shows how accurate the model is for different object types as its confidence increases. In general, higher confidence leads to higher precision, indicating that the model effectively distinguishes confident detections from less confident ones. Figure 11(b) shows the recall vs. confidence curves, indicating how many true positives the model finds at different confidence levels. Initially, it detects many diseases even with low confidence, suggesting a good sensitivity. However, as the confidence threshold increases, the model becomes stricter. Figure 11(c) shows the precision-recall curve, which visualizes the trade-off between precision and completeness (recall). The model starts by selectively prioritizing accuracy. As it tries to capture more diseases, it includes some false positives. The peak shows the optimal balance, where it identifies many diseases while maintaining good accuracy. The chosen

threshold determines the specific precision-recall values. Figure 11(d) shows the F1-confidence curve, which combines precision and recall and is a measure of balanced performance at different confidence levels. The peak indicates the confidence level where the model achieves the best balance between precision and recall for this specific dataset.

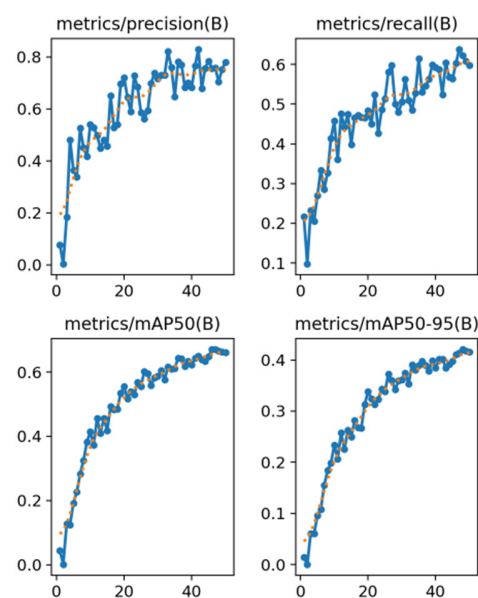


Fig. 10. Performance results.

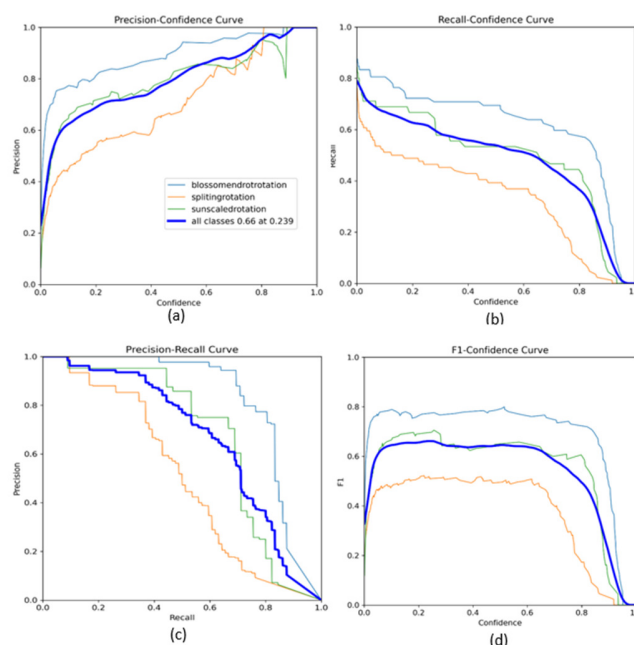


Fig. 11. (a) Precision-confidence, (b) recall-confidence, (c) precision-recall, and (d) F1-confidence curves.

The confusion matrix in Figure 12 shows the model's performance in different types of tomato diseases. Each cell shows how often the model confused one disease for another,

or correctly identified it. For example, 51 "blossom end rot rotation" instances were correctly identified, while 8 were misclassified. In general, the model achieved 66.67% accuracy across all diseases. However, this masks differences in performance for each disease. Table I illustrates the precision, recall, and accuracy of each class in the chosen model.

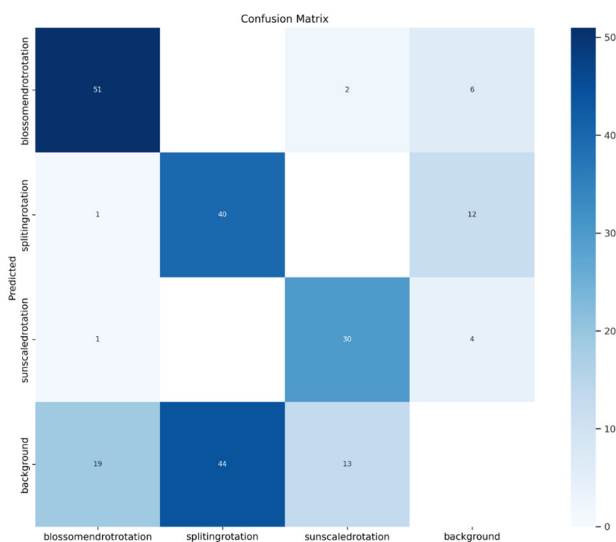


Fig. 12. Confusion matrix.

TABLE I. PRECISION AND RECALL FOR EACH CLASS IN THE MODEL

	Blossom end rot rotation	Splitting rotation	Sun scaled rotation
Precision	86.44%	74.07%	85.71%
Recall	70.84%	90.90%	66.67%
Accuracy	82.50%	50.00%	67.10%

V. CONCLUSION

This study trained the YOLOv8 model to automatically detect three tomato fruit diseases. Although the overall accuracy was 66.67%, discriminatory disease detection occurred. Although the splitting rotation was accurately identified, the visual similarity between blossom end rot and sun-scaled rotation posed difficulties. Although the precision for blossom end rot rotation remained satisfactory at 86.44%, the recall of 70.84% indicates the possibility of overlooking genuine instances of the disease. The recall and precision for sun-scaled rotation were 66.67% and 85.71%, respectively, suggesting that there is potential to improve the ability to classify this category with greater accuracy. The model performed exceptionally well in identifying dividing rotation with a high recall of 90.90% and a precision rate of 74.07%. The limitations mentioned can be ascribed to intrinsic visual similarities among diseases, which can complicate the precision of the identification process.

Future work should focus on addressing data imbalance by adding healthy class, data collection or augmentation, exploring alternative architectures or incorporating domain

knowledge, and evaluating disease-specific metrics for a more practical perspective. These improvements can pave the way for a robust and reliable automated tomato disease detection system, enhancing crop monitoring and contributing to sustainable agriculture.

ACKNOWLEDGMENT

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA, for funding this research work through project number NBU-FFR-2024-1563-02.

REFERENCES

- [1] J. M. Costa and E. Heuvelink, "The global tomato industry," in *Tomatoes*, CABI, 2018, pp. 1–26.
- [2] "Huge increase in tomato production in Saudi Arabia due to targeted watering," Jan. 20, 2020. <https://www.wur.nl/en/research-results/research-institutes/plant-research/greenhouse-horticulture/show-greenhouse/huge-increase-in-tomato-production-in-saudi-arabia-due-to-targeted-watering.htm>.
- [3] I. Tiwari *et al.*, "Post-harvest practices and loss assessment in tomato (*Solanum lycopersicum* L.) in Kathmandu, Nepal," *Journal of Agriculture and Natural Resources*, vol. 3, no. 2, pp. 335–352, Oct. 2020, <https://doi.org/10.3126/janr.v3i2.32545>.
- [4] M. Benavides, M. Cantón-Garbín, J. A. Sánchez-Molina, and F. Rodríguez, "Automatic Tomato and Peduncle Location System Based on Computer Vision for Use in Robotized Harvesting," *Applied Sciences*, vol. 10, no. 17, 2020, <https://doi.org/10.3390/app10175887>.
- [5] Y. Bai, S. Mao, J. Zhou, and B. Zhang, "Clustered tomato detection and picking point location using machine learning-aided image analysis for automatic robotic harvesting," *Precision Agriculture*, vol. 24, no. 2, pp. 727–743, Apr. 2023, <https://doi.org/10.1007/s11119-022-09972-6>.
- [6] Y. Gulzar, "Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique," *Sustainability*, vol. 15, no. 3, 2023, <https://doi.org/10.3390/su15031906>.
- [7] M. Afonso *et al.*, "Tomato Fruit Detection and Counting in Greenhouses Using Deep Learning," *Frontiers in Plant Science*, vol. 11, 2020, <https://doi.org/10.3389/fpls.2020.571299>.
- [8] G. Moreira, S. A. Magalhães, T. Pinho, F. N. dos Santos, and M. Cunha, "Benchmark of Deep Learning and a Proposed HSV Colour Space Models for the Detection and Classification of Greenhouse Tomato," *Agronomy*, vol. 12, no. 2, Feb. 2022, Art. no. 356, <https://doi.org/10.3390/agronomy12020356>.
- [9] Y. Mu, T. S. Chen, S. Ninomiya, and W. Guo, "Intact Detection of Highly Occluded Immature Tomatoes on Plants Using Deep Learning Techniques," *Sensors*, vol. 20, no. 10, 2020, <https://doi.org/10.3390/s20102984>.
- [10] J. Wu, Z. Kuang, L. Wang, W. Zhang, and G. Wu, "Context-Aware RCNN: A Baseline for Action Detection in Videos," in *Computer Vision – ECCV 2020*, Glasgow, UK, Aug. 2020, pp. 440–456, https://doi.org/10.1007/978-3-030-58595-2_27.
- [11] F. Zeng, Y. Liu, Y. Ye, J. Zhou, and X. Liu, "A detection method of Edge Coherent Mode based on improved SSD," *Fusion Engineering and Design*, vol. 179, Jun. 2022, Art. no. 113141, <https://doi.org/10.1016/j.fusengdes.2022.113141>.
- [12] H. Peng *et al.*, "General improved SSD model for picking object recognition of multiple fruits in natural environment," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 34, no. 16, pp. 155–162, 2018.
- [13] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A Review of Yolo Algorithm Developments," *Procedia Computer Science*, vol. 199, pp. 1066–1073, Jan. 2022, <https://doi.org/10.1016/j.procs.2022.01.135>.
- [14] G. Liu, J. C. Nouaze, P. L. Touko Mbouembe, and J. H. Kim, "YOLO-Tomato: A Robust Algorithm for Tomato Detection Based on YOLOv3," *Sensors*, vol. 20, no. 7, 2020, <https://doi.org/10.3390/s20072145>.

- [15] Y. Yang, J. Li, J. Nie, S. Yang, and J. Tang, "Cotton Stubble Detection Based on Improved YOLOv3," *Agronomy*, vol. 13, no. 5, 2023, <https://doi.org/10.3390/agronomy13051271>.
- [16] R. Gai, N. Chen, and H. Yuan, "A detection algorithm for cherry fruits based on the improved YOLO-v4 model," *Neural Computing and Applications*, vol. 35, no. 19, pp. 13895–13906, Jul. 2023, <https://doi.org/10.1007/s00521-021-06029-z>.
- [17] R. Li, Z. Ji, S. Hu, X. Huang, J. Yang, and W. Li, "Tomato Maturity Recognition Model Based on Improved YOLOv5 in Greenhouse," *Agronomy*, vol. 13, no. 2, 2023, <https://doi.org/10.3390/agronomy13020603>.
- [18] R. Rajamohanan and B. C. Latha, "An Optimized YOLO v5 Model for Tomato Leaf Disease Classification with Field Dataset," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12033–12038, Dec. 2023, <https://doi.org/10.48084/etasr.6377>.
- [19] T. Saidani, "Deep Learning Approach: YOLOv5-based Custom Object Detection," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12158–12163, Dec. 2023, <https://doi.org/10.48084/etasr.6397>.
- [20] T. Saidani, R. Ghodhbani, A. Alhomoud, A. Alshammari, H. Zayani, and M. B. Ammar, "Hardware Acceleration for Object Detection using YOLOv5 Deep Learning Algorithm on Xilinx Zynq FPGA Platform," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 13066–13071, Feb. 2024, <https://doi.org/10.48084/etasr.6761>.
- [21] J. Zhou, Y. Zhang, and J. Wang, "RDE-YOLOv7: An Improved Model Based on YOLOv7 for Better Performance in Detecting Dragon Fruits," *Agronomy*, vol. 13, no. 4, 2023, <https://doi.org/10.3390/agronomy13041042>.
- [22] G. Yang, J. Wang, Z. Nie, H. Yang, and S. Yu, "A Lightweight YOLOv8 Tomato Detection Algorithm Combining Feature Enhancement and Attention," *Agronomy*, vol. 13, no. 7, 2023, <https://doi.org/10.3390/agronomy13071824>.
- [23] B. N. E. Nyarko, W. Bin, Z. Jinzhi, and J. Odoom, "Tomato fruit disease detection based on improved single shot detection algorithm," *Journal of Plant Protection Research*, vol. 63, no. 4, pp. 405–417, Oct. 2023, <https://doi.org/10.24425/jppr.2023.146877>.
- [24] A. Santosh, S. Bikesh, B. Bibek, and K. K. C. Saban, "Tomato Plant Diseases Detection System Using Image Processing," in *1st KEC Conference Proceedings*, Sep. 2018, pp. 81–86.
- [25] Q. H. Phan, V. T. Nguyen, C. H. Lien, T. P. Duong, M. T. K. Hou, and N. B. Le, "Classification of Tomato Fruit Using Yolov5 and Convolutional Neural Network Models," *Plants*, vol. 12, no. 4, 2023, <https://doi.org/10.3390/plants12040790>.
- [26] S. N. Appe, A. G, and B. Gn, "CAM-YOLO: tomato detection and classification based on improved YOLOv5 using combining attention mechanism," *PeerJ Computer Science*, vol. 9, Jul. 2023, Art. no. e1463, <https://doi.org/10.7717/peerj-cs.1463>.
- [27] "balanceddata Computer Vision Project." Roboflow Universe, Oct. 2023, [Online]. Available: <https://universe.roboflow.com/research-proj-mgap9/balanceddata-y4ox0>.
- [28] S. Mukherjee, "YOLOv8: Pioneering Breakthroughs in Object Detection Technology," *Paperspace Blog*, Nov. 03, 2023. <https://blog.paperspace.com/yolov8-a-revolutionary-advancement-in-object-detection-2/>.
- [29] Z. Zhang, "Drone-YOLO: An Efficient Neural Network Method for Target Detection in Drone Images," *Drones*, vol. 7, no. 8, 2023, <https://doi.org/10.3390/drones7080526>.

A Study on the Influence of FDM Parameters on the Compressive Behavior of PET-G Parts

Dragos Gabriel Zisopol

Mechanical Engineering Department, Petroleum-Gas University Ploiesti, Romania
zisopold@upg-ploiesti.ro

Mihail Minescu

Mechanical Engineering Department, Petroleum-Gas University Ploiesti, Romania
mminescu@upg-ploiesti.ro

Dragos Valentin Iacob

Production Department, Marelli Ploiesti Romania
dragoshicb@gmail.com (corresponding author)

Received: 13 February 2024 | Revised: 24 February 2024 | Accepted: 3 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7063>

ABSTRACT

This article presents the results of a study on the influence of Fused Deposition Modeling (FDM) 3D printing parameters on the compressive behavior of test specimens made of PET-G. In this context, 45 test specimens, made by FDM on the Anycubic 4 Max Pro 2.0 printer, were compressive tested on a universal testing machine Barrus White 20 kN, with the height of the layer applied in one pass being $L_h = 0.10/0.15/0.20$ mm and filling percentage $I_d = 50/75/100\%$. The two considered variable parameters, L_h and I_d influence the compression resistance of the PET-G parts, with I_d having a more significant influence. The scope and novelty of this work is to find the optimal parameters for maximum compressive strength (C_s) of PET-G samples made of FDM.

Keywords-FDM parameters; PET-G; compressive stress; experimental tests

I. INTRODUCTION

Additive manufacturing is the process of producing parts by adding material in overlapping layers according to the 3D model profiles of the part [1-2]. Additive manufacturing technologies have continuously evolved since their emergence, this evolution being remarkable for new methods and materials used, improvement of mechanical and material quality characteristics of 3D manufactured parts, etc. [3-10]. One of the most widely used additive manufacturing technologies is Fused Deposition Modeling (FDM), which is characterised by the wide range of materials used for printing, such as PLA (polylactic acid), PET-G (polyethylene terephthalate glycol), ASA (acrylonitrile styrene acrylate), ABS (acrylonitrile butadiene styrene), TPU (thermoplastic polyurethane), and nylon, its ease of use, and the advantageous price of materials and equipment [11, 23-29]. However, in order to maximize the mechanical and qualitative characteristics of the parts produced by FDM, it is necessary to perform an optimization of the process parameters, i.e. the height of the layer applied in one pass L_h , the filling percentage I_d , the filling pattern I_p , the part orientation, the printing speed P_s , the extruder temperature E_t , and the table temperature B_t , [7-20]. In [2], a comparative study of the impact of FDM parameters, (part orientation, L_h , I_d , and P_s) on the mechanical properties of PLA, ABS, PEEK, and

PET-G is presented. Out of the 4 parameters studied, I_d has the most decisive influence on the mechanical properties of the studied FDM parts. In [7], the compressive behavior was studied by varying I_d and L_h of FDM parts made of PLA, ABS, and heat-treated PLA. Heat treatment of PLA parts was performed by holding the parts in an oven at a temperature of 75°C for a 3-hour period. The results show that among the two varied parameters, I_d had higher impact on the compressive behavior, with the compressive strength increasing with increasing filling percentage. At the same time, it is observed that PLA parts which were heat treated, had 8.20% higher compressive strength on average than the not heated PLA parts. In [8], a study on the effects of printing parameters on the mechanical characteristics and mathematical modeling of FDM printed PET-G parts is presented. The best compressive results were obtained for parts made of PET-G with triangle filling pattern and the highest filling percentage. The obtained results for triangle are higher by 6.45 – 35.16% than other filling patterns (grid, rectilinear, honeycomb, concentric).

The aim and novelty of this work are the determination of the influence of the FDM parameters (height of the layer applied in one pass L_h and filling percentage I_d) on the compressive behavior of parts made from PET-G in order to find the optimum parameters for the best mechanical

properties. Specimens were manufactured and tested in compression in the laboratories of the Faculty of Mechanical and Electrical Engineering of the Petroleum – Gas University of Ploiești.

II. DETERMINATION OF THE INFLUENCE OF FDM PARAMETERS ON THE TENSILE BEHAVIOR OF PARTS MADE OF PET-G

A. Methodology

Figure 1 presents the steps of the work methodology in studying the influence of FDM parameters on the compressive behavior of PET-G parts.

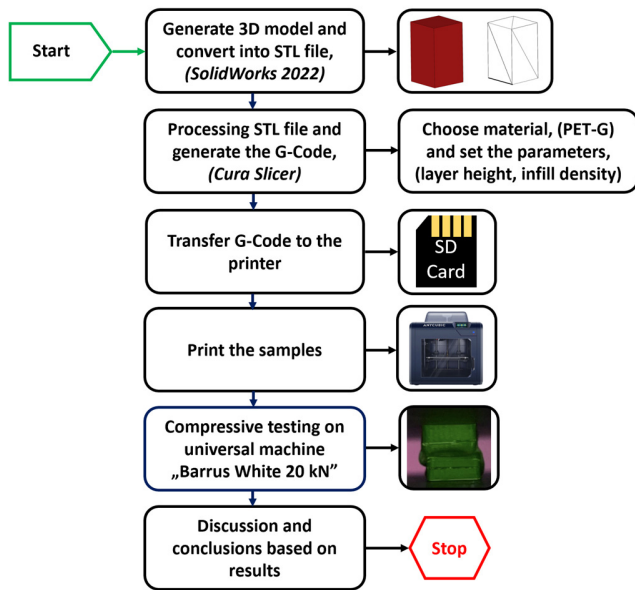


Fig. 1. Steps of the work methodology.

Using Solidworks 2022 software [31], we produced the 2D sketch and 3D model of the specimen for compression, converting it from SLD to STL format (Figure 2).

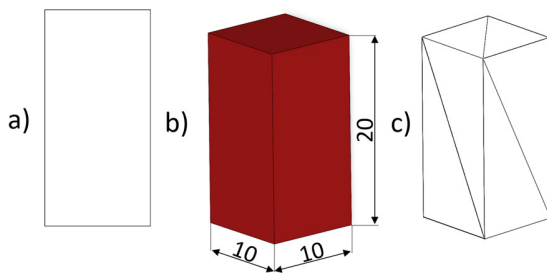


Fig. 2. Specimen for compressive testing in Solidworks 2022 software: (a) 2D model, (b) 3D model, (c) STL model.

The file in STL format, corresponding to the specimen shown in Figure 2, was processed in the Cura Slicer program, in which we set the PET-G material, entered the parameters shown in Table I and generated the G-Code file.

TABLE I. FDM PRINTING PARAMETERS FOR PET-G COMPRESSION SPECIMENS

Constant parameters		Variable parameters		Material
		Layer height in one pass	Filling percentage	PET-G
		L _h	I _d	-
Part orientation	X-Y	(mm)	(%)	(parts)
		0.10	100	5
75				
50				
Temperature of the extruder E _t	250 °C	0.15	100	
			75	
			50	
Temperature of the table B _t	70 °C	0.20	100	
			75	
			50	
Printing speed P _s	30 mm/s		100	
			75	
		50		
Filling pattern I _p	Grid	75		
		50		

Figure 3 depicts the rendering of the PET-G specimen for compression, made in Cura Slicer [32] with constant parameters from Table I and $L_h = 0.10$ mm, $I_d = 100\%$. The G-Code file of the compression specimen illustrated in Figure 3 contains 40187 command lines, its structure being as shown in Figure 4 [32].

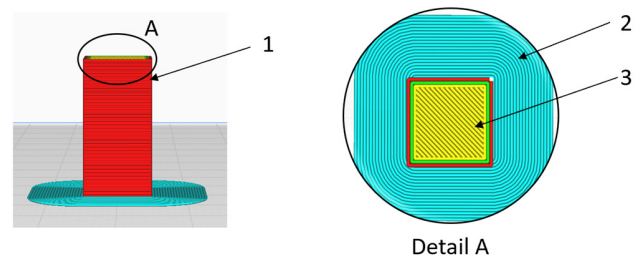


Fig. 3. PET-G specimen for compressive testing in Cura Slicer software: 1 - specimen made layer by layer, 2 - specimen holder, 3 - filling pattern.

```

;LAYER_COUNT:198
;LAYER:0
M107
M204 S300
M205 X8.18 Y8.18
G0 F2400 X99.372 Y99.567 Z0.3
M205 X3 Y3
;TYPE:SKIRT
G1 F1500 E0
G1 F1200 X100.749 Y98.458 E0.08821
G1 X102.725 Y97.474 E0.19834
G1 X104.358 Y97.244 E0.28061
G1 X115.619 Y97.2 E0.84243
G1 X117.676 Y97.623 E0.9472
G1 X119.55 Y98.681 E1.05457
G1 X120.433 Y99.372 E1.11051
G1 X121.542 Y100.749 E1.19872

```

Fig. 4. G-Code file structure of PET-G compression specimen.

The G-Code file shown in Figure 4 was transferred to the Anycubic 4 Max Pro 2.0 3D printer (Figure 5), on which 45 compression specimens shown in Figure 6 were manufactured using Everfill 1.75 mm PET-G filament. The 45 specimens presented in Figure 6, manufactured using the parameters in Table I, were tested for compression on the Barrus White 20 kN universal testing machine (Figure 7), according to the ISO 604:2002, with a speed of 10 mm/min [30].



Fig. 5. Anycubic 4 Max Pro 2.0 printer, used for FDM manufacturing of PET-G specimens for compression.

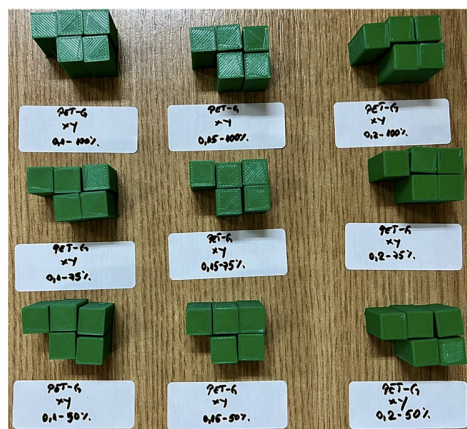


Fig. 6. Compression specimens made of PET-G by FDM.

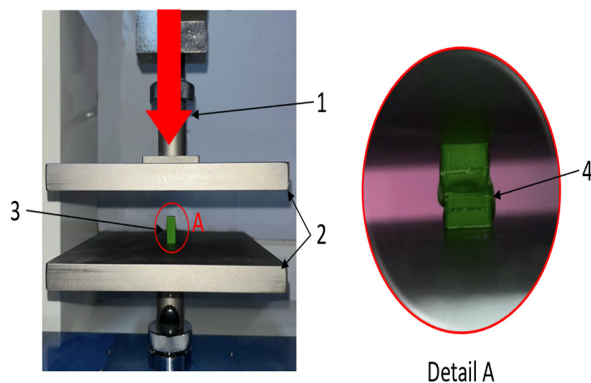


Fig. 7. Compression test on Barrus White 20 kN machine: 1 - loading direction, 2 - machine plates, 3 - specimen before compression test, 4 - specimen after compression test.

B. Results and Discussion

The results of the compression tests on the 45 specimens are summarized in Tables II-IV and plotted in Figures 8-10. An analysis of Figure 8 shows that the compressive strength (C_s) increases with increasing I_d . The best results, 29.79 MPa - 30.84 MPa, were obtained for specimens having $I_d = 100\%$. By increasing the I_d from 50% to 75%, the compressive strength values increased by 38.78% - 46.94%, and by increasing the I_d from 75% to 100% they increased by 47.66% - 56.30%.

TABLE II. RESULTS OF COMPRESSION TESTS FOR PET-G WALLS WITH $L_h = 0.10$ mm

Infill I_d (%)	Compressive stress C_s (mpa)					
	Specimen					
	1	2	3	4	5	Average
50	14.20	13.73	14.00	14.16	14.21	14.06
75	19.17	19.06	19.76	20.29	20.88	19.83
100	29.79	30.84	30.02	30.47	30.54	30.33

TABLE III. RESULTS OF COMPRESSION TESTS FOR PET-G WALLS WITH $L_h = 0.15$ mm

Infill I_d (%)	Compressive stress C_s (MPa)					
	Specimen					
	1	2	3	4	5	Average
50	11.82	12.25	12.47	12.18	12.24	12.20
75	20.05	18.58	20.61	21.12	20.74	20.22
100	30.61	30.41	30.73	30.64	30.47	30.57

TABLE IV. RESULTS OF COMPRESSION TESTS FOR PET-G RIVERS WITH $L_h = 0.20$ mm

Infill I_d (%)	Compressive stress C_s (MPa)					
	Specimen					
	1	2	3	4	5	Average
50	11.10	10.65	10.65	11.86	12.07	11.27
75	20.39	19.74	19.73	19.10	20.12	19.82
100	28.67	29.55	29.06	29.30	29.41	29.20

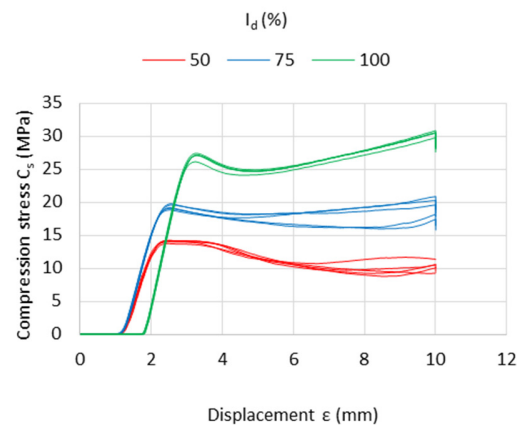


Fig. 8. Average compressive strength values for PET-G specimens with $L_h = 0.10$ mm and $I_d = 50/75/100$ %.

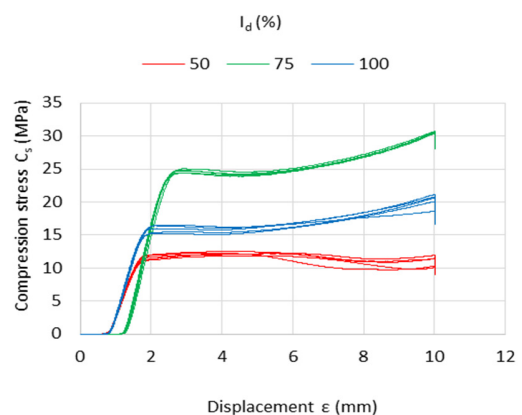


Fig. 9. Average compressive strength values for PET-G specimens with $L_h = 0.15$ mm and $I_d = 50/75/100$ %.

It is observed in Figure 9 that the compressive strength is decisively influenced by the filling percentage. The best results, 30.41 MPa - 30.73 MPa, were obtained for specimens having $I_d = 100\%$. By increasing the filling percentage from 50% to 75%, the compressive strength values increased by 57.13% - 69.28%, whereas by increasing filling percentage from 75% to 100%, they increased by 45.52% - 63.70%. Looking at Figure 10, it can be observed that C_s is strongly influenced by I_d . The best results, 28.67 MPa - 29.55 MPa, were obtained for specimens with $I_d = 100\%$. By increasing the filling percentage from 50% to 75%, the compressive strength values increased by 68.89% - 79.32%, while by increasing the filling percentage from 75% to 100%, they increased by 44.89% - 50.11%.

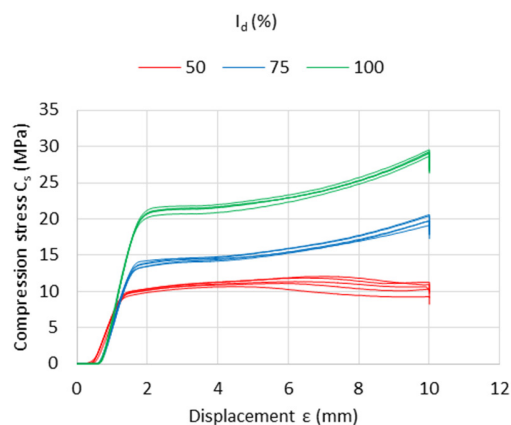


Fig. 10. Average compressive strength values for PET-G specimens with $L_h = 0.20$ mm and $I_d = 50/75/100\%$.

Based on the average compressive strength results (see Tables II-IV), the graph in Figure 11 was plotted with Minitab software [33]. The Figure illustrates how the variable parameters L_h and I_d of the FDM influence the C_s of PET-G specimens [33].

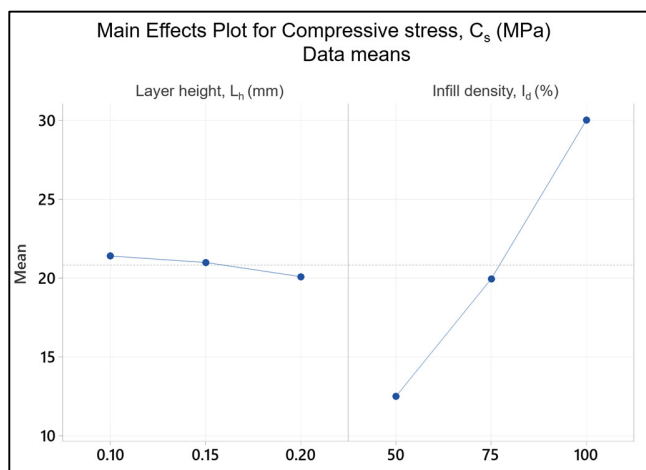


Fig. 11. Influence of variable FDM parameters L_h and I_d on the C_s of PET-G specimens.

According to Figure 11, out of the two variable parameters, I_d has the highest impact on the C_s of PET-G specimens, whereas the second variable parameter L_h , has an insignificant influence. The same conclusions can be drawn from an analysis of the Pareto chart in Figure 12. The Pareto chart shows that the filling percentage ($B = I_d$) has a strong influence on the compressive strength values obtained from compression tests of PET-G specimens on the Barrus White 20 kN universal testing machine (Figure 7). Using Minitab [33] and the FDM parameters in Table I, $L_h = 0.10/0.15/0.20$ mm and $I_d = 50/75/100\%$, the contour graph of compressive strengths for PET-G specimens was plotted, as shown in Figure 13.

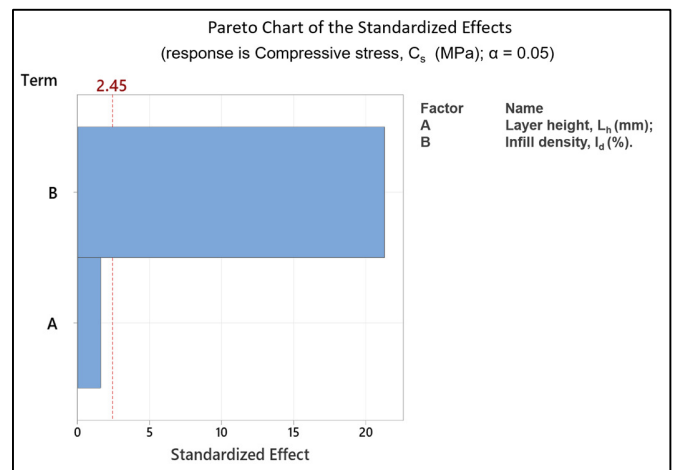


Fig. 12. Pareto plot of the influence of the variable parameters $A = L_h$ and $B = I_d$ of FDM on the compressive strength of PET-G specimens.

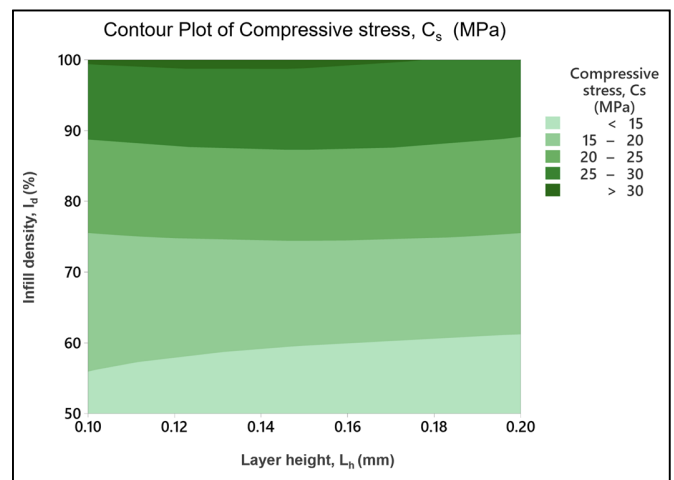


Fig. 13. Contour plot of the compressive strength of PET-G specimens manufactured by FDM.

Analyzing the contour plot in Figure 13, the way the variable parameters L_h and I_d of the FDM influence the compressive strength of the specimens can be concluded. Increasing I_d has a significant influence on the increase of the compressive strength of the specimens.

III. CONCLUSIONS

This paper presents the results of the research on the influence of FDM 3D printing parameters (see Table I) on the compression behavior of Everfill PET-G parts.

In this context, 45 specimens were produced on the Anycubic 4 Max Pro 2.0 printer, with layer height having values $L_h = 0.10/0.15/0.20$ mm and filling percentage having values $I_d = 50/75/100$ %. The specimens were compression tested on a Barrus White 20 kN universal testing machine.

Increasing the filling percentage (I_d) increases the compressive strength of the specimens made by FDM from PET-G, in accordance with the findings in [2, 7, 8]. The minimum value of the compressive strength of the specimens, i.e. 11.27 MPa, was recorded for parameter values $L_h = 0.20$ mm and $I_d = 50\%$, and the maximum, i.e. 30.57 MPa, was recorded for $L_h = 0.15$ mm and $I_d = 100\%$. The obtained results for PLA samples in [7] are bigger by 131.56 – 187.24% than the ones of this study for PET-G samples with the same parameters. For the same parameters for ABS samples, the obtained results are bigger by 99.12 – 104.65% than our own. When comparing with the findings in [8], our results for the same parameters, but for different filament brand are smaller by 7.09%.

Our study can be applied in the manufacturing process of parts by FDM from PET-G using optimal parameters to acquire the best mechanical proprieties. In addition, with this work, we filled knowledge gaps by investigating the effects of two crucial parameters (L_h and I_d) of 3D printing by FDM and opportunities for future research were found. It is recommended to extrapolate the study to other materials such as ASA, recycled PET-G, and ASA with varying percentages of the recycled material.

REFERENCES

- [1] S. Valvez, A. P. Silva, and P. N. B. Reis, "Compressive Behaviour of 3D-Printed PETG Composites," *Aerospace*, vol. 9, no. 3, Mar. 2022, Art. no. 124, <https://doi.org/10.3390/aerospace9030124>.
- [2] M. Algarni and S. Ghazali, "Comparative Study of the Sensitivity of PLA, ABS, PEEK, and PETG's Mechanical Properties to FDM Printing Process Parameters," *Crystals*, vol. 11, no. 8, Aug. 2021, Art. no. 995, <https://doi.org/10.3390/cryst11080995>.
- [3] I. M. Alarifi, "Mechanical properties and numerical simulation of FDM 3D printed PETG/carbon composite unit structures," *Journal of Materials Research and Technology*, vol. 23, pp. 656–669, Mar. 2023, <https://doi.org/10.1016/j.jmrt.2023.01.043>.
- [4] A. Yankin *et al.*, "Optimization of Printing Parameters to Enhance Tensile Properties of ABS and Nylon Produced by Fused Filament Fabrication," *Polymers*, vol. 15, no. 14, Jan. 2023, Art. no. 3043, <https://doi.org/10.3390/polym15143043>.
- [5] A. Kholil, E. Asyaefudin, N. Pinto, and S. Syaripuddin, "Compression Strength Characteristics of ABS and PLA Materials Affected by Layer Thickness on FDM," *Journal of Physics: Conference Series*, vol. 2377, no. 1, Aug. 2022, Art. no. 012008, <https://doi.org/10.1088/1742-6596/2377/1/012008>.
- [6] S. Skere, A. Zvirioniene, K. Juzenas, and S. Petraitiene, "Optimization Experiment of Production Processes Using a Dynamic Decision Support Method: A Solution to Complex Problems in Industrial Manufacturing for Small and Medium-Sized Enterprises," *Sensors*, vol. 23, no. 9, Jan. 2023, Art. no. 4498, <https://doi.org/10.3390/s23094498>.
- [7] D. G. Zisopol, I. Nae, and A. I. Portoaca, "Compression Behavior of FFF Printed Parts Obtained by Varying Layer Height and Infill Percentage," *Engineering, Technology & Applied Science Research*, vol. 12, no. 6, pp. 9747–9751, Dec. 2022, <https://doi.org/10.48084/etasr.5488>.
- [8] R. Kumaresan, M. Samykano, K. Kadirgama, A. K. Pandey, and Md. M. Rahman, "Effects of printing parameters on the mechanical characteristics and mathematical modeling of FDM-printed PETG," *The International Journal of Advanced Manufacturing Technology*, vol. 128, no. 7, pp. 3471–3489, Oct. 2023, <https://doi.org/10.1007/s00170-023-12155-w>.
- [9] D. G. Zisopol, M. Minescu, and D. V. Iacob, "A Study on the Evaluation of the Compression Behavior of PLA Lattice Structures Manufactured by FDM," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11801–11806, Oct. 2023, <https://doi.org/10.48084/etasr.6262>.
- [10] D. G. Zisopol, D. V. Iacob, and A. I. Portoaca, "A Theoretical-Experimental Study of the Influence of FDM Parameters on PLA Spur Gear Stiffness," *Engineering, Technology & Applied Science Research*, vol. 12, no. 5, pp. 9329–9335, Oct. 2022, <https://doi.org/10.48084/etasr.5183>.
- [11] D. G. Zisopol, M. Minescu, and D. V. Iacob, "A Theoretical-Experimental Study on the Influence of FDM Parameters on the Dimensions of Cylindrical Spur Gears Made of PLA," *Engineering, Technology & Applied Science Research*, vol. 13, no. 2, pp. 10471–10477, Apr. 2023, <https://doi.org/10.48084/etasr.5733>.
- [12] D. G. Zisopol, I. Nae, A. I. Portoaca, and I. Ramadan, "A Theoretical and Experimental Research on the Influence of FDM Parameters on Tensile Strength and Hardness of Parts Made of Polylactic Acid," *Engineering, Technology & Applied Science Research*, vol. 11, no. 4, pp. 7458–7463, Aug. 2021, <https://doi.org/10.48084/etasr.4311>.
- [13] D. G. Zisopol, M. Tanase, and A. I. Portoaca, "Innovative Strategies for Technical-Economical Optimization of FDM Production," *Polymers*, vol. 15, no. 18, Jan. 2023, Art. no. 3787, <https://doi.org/10.3390/polym15183787>.
- [14] D. G. Zisopol, A. I. Portoaca, and M. Tanase, "Dimensional Accuracy of 3D Printed Dog-bone Tensile Samples: A Case Study," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11400–11405, Aug. 2023, <https://doi.org/10.48084/etasr.6060>.
- [15] D. G. Zisopol, A. I. Portoaca, and M. Tanase, "Improving the Impact Resistance through Annealing in PLA 3D Printed Parts," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11768–11772, Oct. 2023, <https://doi.org/10.48084/etasr.6281>.
- [16] D. G. Zisopol, A. I. Portoaca, I. Nae, and I. Ramadan, "A Comparative Analysis of the Mechanical Properties of Annealed PLA," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 8978–8981, Aug. 2022, <https://doi.org/10.48084/etasr.5123>.
- [17] D. G. Zisopol, I. Nae, A. I. Portoaca, and I. Ramadan, "A Statistical Approach of the Flexural Strength of PLA and ABS 3D Printed Parts," *Engineering, Technology & Applied Science Research*, vol. 12, no. 2, pp. 8248–8252, Apr. 2022, <https://doi.org/10.48084/etasr.4739>.
- [18] M. Tanase, D. G. Zisopol, and A. I. Portoaca, "A Study regarding the Technical-Economical Optimization of Structural Components for enhancing the Buckling Resistance in Stiffened Cylindrical Shells," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11511–11516, Oct. 2023, <https://doi.org/10.48084/etasr.6135>.
- [19] D. G. Zisopol, N. Ion, and A. I. Portoaca, "Comparison of the Charpy Resilience of Two 3D Printed Materials: A Study on the Impact Resistance of Plastic Parts," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10781–10784, Jun. 2023, <https://doi.org/10.48084/etasr.5876>.
- [20] A. Dinita, A. Neacsu, A. I. Portoaca, M. Tanase, C. N. Ilinca, and I. N. Ramadan, "Additive Manufacturing Post-Processing Treatments, a Review with Emphasis on Mechanical Characteristics," *Materials*, vol. 16, no. 13, Jan. 2023, Art. no. 4610, <https://doi.org/10.3390/ma16134610>.
- [21] D. G. Zisopol, M. Minescu, and D. V. Iacob, "A Study on the Influence of aging of the Butt-welded PE100 SDR11 on Shore A Hardness and Tensile Strength," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12722–12727, Feb. 2024, <https://doi.org/10.48084/etasr.6635>.

- [22] D. Zisopol, M. Minescu, and D. Iacob, "A Study on the Influence of FDM Parameters on the Tensile Behavior of Samples made of PET-G," *Engineering, Technology and Applied Science Research*, vol. 14, pp. 13487–13492, Apr. 2024, <https://doi.org/10.48084/etasr.6949>.
- [23] D. G. Zisopol and A. Dumitrescu, *Materiale și tehnologii primare. Aplicații practice și studii de caz*. Ploiesti, Romania: Editura Universității Petrol-Gaze din Ploiești, 2005.
- [24] D. G. Zisopol and A. Dumitrescu, *Ecotehnologie. Studii de caz*. Ploiesti, Romania: Editura Universității Petrol-Gaze din Ploiești, 2021.
- [25] D. G. Zisopol, A. Dumitrescu, and C. N. Trifan, *Ecotehnologie: Noțiuni teoretice, aplicații și studii de caz*. Ploiesti, Romania: Editura Universității Petrol-Gaze din Ploiești, 2010.
- [26] D. G. Zisopol and M. J. Săvulescu, *Bazele tehnologiei*. Ploiesti, Romania: Editura Universității Petrol-Gaze din Ploiești, 2003.
- [27] M. J. Săvulescu and D. G. Zisopol, *Tehnologii industriale și de construcții*. Ploiesti, Romania: Editura Universității Petrol-Gaze din Ploiești, 2002.
- [28] D. G. Zisopol, *Tehnologii industriale și de construcții, Aplicații practice și studii de caz*. Ploiesti, Romania: Editura Universității Petrol-Gaze din Ploiești, 2003.
- [29] M. J. Săvulescu, D. G. Zisopol, and I. Nae, *Bazele tehnologiei materialelor. Îndrumar de lucrări practice*. Ploiesti, Romania: Editura Premier Ploiești, 1997.
- [30] *ISO 604:2002 Plastics: Determination of compressive properties*. ISO, 2002.
- [31] "3D CAD Design Software," SOLIDWORKS. <https://www.solidworks.com/home-page-2021>.
- [32] "UltiMaker Cura," *UltiMaker*. <https://ultimaker.com/software/ultimaker-cura/>.
- [33] "Data Analysis, Statistical & Process Improvement Tools," Minitab. <https://www.minitab.com/en-us/>.

Using a Chaotic Digital System to Generate Random Numbers for Secure Communication on 5G Networks

Haider Th. Salim Alrikabi

Department of Electrical Engineering, College of Engineering, University of Wasit, Iraq
hdhiyab@uowasit.edu.iq (corresponding author)

Ibtisam A. Aljzaery

Department of Electrical Engineering, College of Engineering, University of Babylon, Iraq
sci.ibtisam.abdulwahid@uobabylon.edu.iq

Abdul Hadi Mohammed Alaidi

Department of Programming, Computer Science and Information Technology College, University of Wasit, Iraq
alaidi@uowasit.edu.iq

Received: 21 January 2024 | Revised: 23 February 2024 | Accepted: 24 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6938>

ABSTRACT

There are several encryption system applications in 5G networks where rapid response is needed, particularly in the military, health sector, traffic, and vehicular movement. This article presents a proposed data security system for 5G networks that fortifies the security of the network through the use of synchronized chaotic systems to produce pseudo-random numbers. The technique by which random numbers are generated during the encryption procedures is closely associated with 5G network security. Many synchronized chaotic systems are used to produce chaotic random models which are used as encryption bases for a wide variety of data. In this study, the encryption was carried out using a variety of data, including two and three-dimensional color images and audio signals of varying lengths, in addition to the use of Fast Fourier Transform (FFT) for encryption of the ingredient energy wave. The results revealed that the algorithm deployed in the process of encryption performed well. Simulations were performed in MATLAB.

Keywords-synchronized chaotic systems; 5G networks; Pseudo-Random Number Generator (PRNG); data security; FFT

I. INTRODUCTION

The ability of the 5G networks to deliver data is almost 40 times faster than other networks, and when they are used fewer delays occur in the transfer and reception of files. Through the 5G networks, the advancement of other new technologies such as the Internet of Things (IoT), virtual reality, etc. can be driven [7]. Cryptography is a requirement for the secure transmission of data and information, [1-4]. Using a robust technique for encryption and decryption of data becomes imperative in the transfer of data, and a 5G network enables the security of the data being transferred. There are two kinds of random number generators, the True Random Number Generator (TRNG) and the Pseudo Random Number Generator (PRNG) [5, 6]. In the PRNG, a random number is intentionally generated, but it is not completely random and it is generated by software. The TRNG, on the other hand, utilizes hardware

to generate the random numbers. Nevertheless, PRNG is a deterministic system, and due to this, the safety and security of the number that has been generated is guaranteed when the initial value has high entropy. In this work, high entropy was obtained as a result of the use of the statistical algorithm for random chaotic systems, which in turn enabled the realization of high security for the data that were employed for encryption.

Chaotic signals are stable and constant. More so, they are characterized by random time evolution and a broadband spectrum that is generated through a deterministic nonlinear dynamical system whose behavior is irregular. Chaos communications is referred to as the application of chaos theory to the security of information that is transmitted using telecommunication technologies. Secured communication refers to communication that is carried out in a way potential eavesdroppers cannot access the content of the transmitted

message. The security of communication in chaotic communications is determined by the complex and dynamic behaviors these systems demonstrate. In this type of communication, data are encoded using some features of the chaotic dynamics like spread spectrum, pseudorandom noise which is also referred to as noise-like dynamics, and complex behavior [8]. The decoding of chaos can be done using its deterministic characteristic given that it is a deterministic phenomenon. Practically, when the devices of chaos communications are being implemented, either chaos synchronization or chaos control is considered [9, 10]. The use of such properties of chaos for the implementation of chaos communications requires using two chaotic oscillators as a transmitter (or master) and receiver (or slave). At the transmitter, the message is hidden in the chaotic signal. Since the chaotic signal plays the role of conveying the information, it is also referred to as the chaotic carrier. The synchronization of the oscillators can be compared to the synchronization of random neural nets in neural cryptography. When chaos synchronization is used, a basic scheme of a communications device [11] is made by two identical chaotic oscillators, i.e., transmitter and receiver. A connection between the two oscillators is established in a configuration where the transmitter drives the receiver toward the realization of identical synchronization of chaos between them. Information is transmitted from the transmitter by adding a message in the form of a small perturbation to the chaotic signal, which acts as a driver of the chaotic signal. Through this, the chaotic signal embeds the transmitted message. Upon the synchronization completion between the receiver and the transmitter, the decoding of the message occurs through a subtraction between the signal that has been transmitted by the transmitter, with a copy produced at the receiver through a mechanism known as chaos synchronization. This is possible because the chaotic carrier and the message are contained in the transmitter's output, while the output of the receiver is produced by just one copy of the chaotic carrier without the message.

II. RELATED WORK

Authors in [4] introduced versatile steering, which is based on 5G Quality of Service (QoS) and Mobile Ad Hoc Networks (MANETs). To achieve versatile 5G steering, the geography of a virtualized climate was planned and CHAOS was also customized for network traffic flooding. Authors in [2] constructed a novel two-parameter chaotic system that is characterized by three dimensions. The evaluation of the system was first conducted by computing its diagrams of bifurcation and diagrams of Lyapunov exponents. Afterward, application of the system was done to two problems associated with encryption. In [12], a new voice encryption wireless system was presented. The proposed system was designed based on the characteristics of a massive Multiple Input Multiple Output (MIMO) wireless channel. The permutation of the channel fading values is carried out using the Minimum Mean Square Error (MMSE) precoding method, and afterward, the use of different chaotic generators was employed in substituting the permuted channel fading values. A combination of the voice samples with the chaotic sequence and channel values was carried out before transmission. The authors deployed their new system on a 5G network that is

inclusive of Massive MIMO, Parallel Spatial Modulation (PSM), and Generalized Frequency Division Multiplexing (GFDM).

Authors in [13] proposed the Boltzmann machine (BMKG)-based encryption algorithm that could provide security for IoT-based 5G network device connectivity and coverage and expand the Encryption and Authentication Scheme (EAS) framework. For the exchange of keys, the authors did a comparison of different asymmetric algorithms. An efficient and highly secure encryption-decryption method was proposed in [14]. Moreover, other algorithms, e.g. those in [15, 16], can be used to optimize the chaotic shuffling process and the dynamic generation of substitution boxes. Authors in [17, 18] proposed a system that enable secured end-to-end communication through user authentication and key agreement. They aimed to address the security challenges that occur when multiple devices network for data collection and analysis. To achieve this aim, the authors presented a Subtree-Based Online/Offline Signature Procedure (SBOOSP) alongside its aggregation (AggSBOOSP) for massive devices in 5G WSNs using conformable chaotic maps. The authors reported that the proposed system is efficient, lightweight, and secure.

A TRNG algorithm was proposed in [19-21]. The source of noise used in this work was the visible spectrum. The authors assumed that if the data used by the cryptography system belong to the visible spectrum, high entropy will be obtained for the random numbers generated by the proposed TRNG. With the use of the random number generator, keys and seeds are generated randomly in numerous cryptographic systems. To this end, the use of keys for the encryption and decryption of the information being transferred is suggested.

III. METHODOLOGY

This section presents the mathematical model that was deployed in the generation of random numbers. The model involved devising a set of statistical equations, denoting the synchronized chaotic system for the formation of random statistical matrices with different dimensions as given below:

$$F1 = (1/5) x1 + (1/4) x2 - (1/3) x3 \quad (1)$$

$$F2 = (2/5) x1 + (2/4) x2 - (1/3) x3 \quad (2)$$

$$F3 = (1/5) x1 - (2/4) x2 + (2/3) x3 \quad (3)$$

where, $x1$, $x2$, and $x3$ denote the chaotic system's initial values. Chaotic systems are known to be overly sensitive to the initial values, and as such, it was ensured that the equations of the synchronized chaotic system were handled in a manner that guaranteed high accuracy. The aim of this was to realize the most appropriate values required for the construction of the random number generator.

The use of the synchronous chaotic system was employed in obtaining random numerical matrices characterized by different shapes and dimensions as in Figure 1.

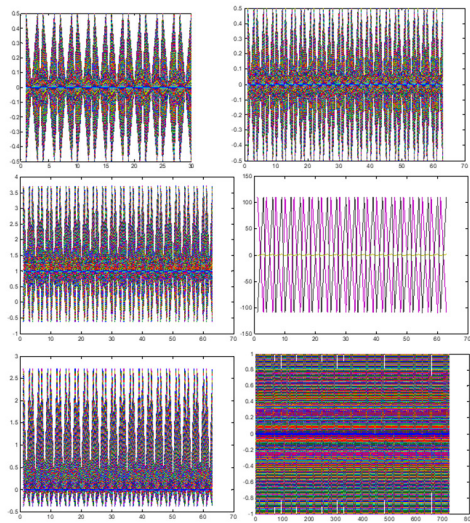


Fig. 1. Examples of the Chaotic Random Number Generator (CRNG).

IV. PROPOSED SYSTEM AND ALGORITHMS

In this section, the algorithms and schemes used for the processes of data decryption and encryption are presented. Figure 2 shows the algorithm which is proposed for the encryption operation.

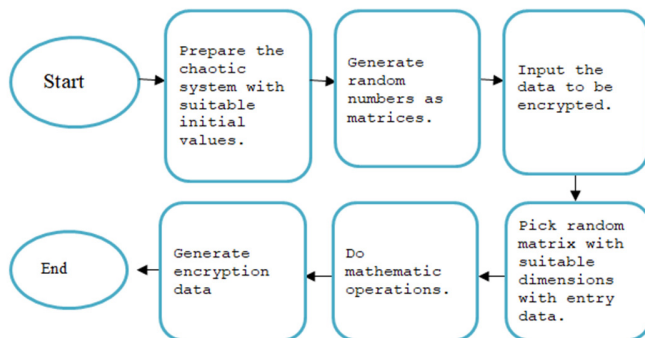


Fig. 2. The encryption algorithm.

The proposed algorithm for encryption operation follows:

Begin:

Step 1. Input initial values of synchronized chaotic equations.

Step 2. Input synchronized chaotic equations as a group of system equations.

Step 3. $F1 = (1/5) x1 + (1/4) x2 - (1/3) x3$, $F2 = (2/5) x1 + (2/4) x2 - (1/3) x3$, $F3 = (1/5) x1 - (2/4) x2 + (2/3) x3$.

Step 4. for $i=1:t$
 $f1r(:, i+1) = f1r(:, 1) .* f1r(:, i);$
 $f2r(:, i+1) = f2r(:, 1) .* f2r(:, i);$
 $f3r(:, i+1) = f3r(:, 1) .* f3r(:, i);$
 end

Step 5. Generate a new random matrix, $x = [f1r \ f2r \ f3r]$.

Step 6. Input data to encrypt.

Step 7. Choose suitable x-array dimensions

Step 8. Make math. Operations (data, x-matrix).

Step 9. Plot encrypted data (stego data).
End.

The proposed algorithm for decryption operation follows:

Begin:

Step 1. Input encrypted data.

Step 2. Perform inverse math operations.

Step 3. Distinguish the result from the previous step.

Step 4. Obtain the original data.

End.

The block diagram in Figure 3 shows the steps required for the decryption.

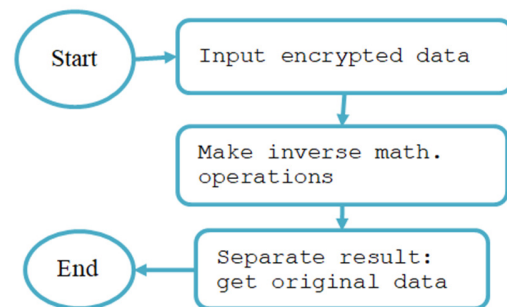


Fig. 3. The decryption algorithm.

V. EXPERIMENTAL RESULTS AND DISCUSSION

Based on the experiments performed in this study, the proposed algorithm demonstrated a high level of performance in terms of data protection and security. The experiments involved the use of a wide range of data with PRNG alongside synchronized chaotic parameters. Tables I-III show the experimental results. It was observed that total concealing of the original data's features was achieved by the high entropy values of the CRNG models, and these in turn resulted in results that were encrypted with high entropy values. This result shows that a random generator is efficient in concealing the features of the original data. The PSNR values also reveal that higher values were obtained for the original data, as compared with the encoded values, as a result of the loss of clarity features. Consequently, this results in distortion which is regarded as a kind of high noise to the outputs obtained from the proposed technique. The results obtained in this study show the output of the correlation equation. A significantly low correlation coefficient was found, showing that there was a great difference in the data after the security procedure was applied. It was also observed that more time was required in the process of encryption than in decryption.

Figure 4 shows that the high entropy of the CRNG models completely hid the features of the original 3D image.

The various kinds of data used in testing the proposed algorithms, along with their encryption stages are shown below.

- 3D image is shown in Figure 4.
- 2D image is shown in Figure 5.
- Audio signal 1 is shown in Figure 6.
- Audio signal 2 is shown in Figure 7.
- The FFT power file is shown in Figure 8.

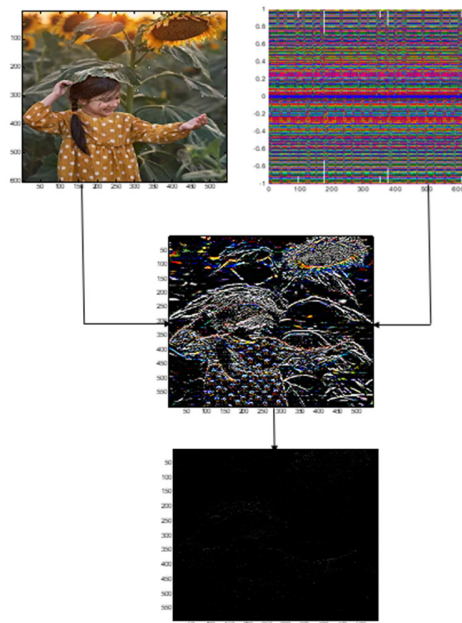


Fig. 4. Encryption stages of a 3D image.

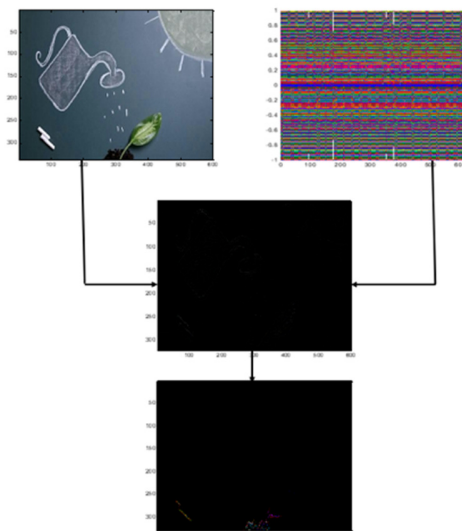


Fig. 5. Encryption stages of a 2D image.

Figure 5 shows a completely encrypted 2D image with small entropy. Figure 7 shows the hiding features of the

original audio signal, while Figure 8 shows that a great difference was obtained in the FFT signal after the security processes.

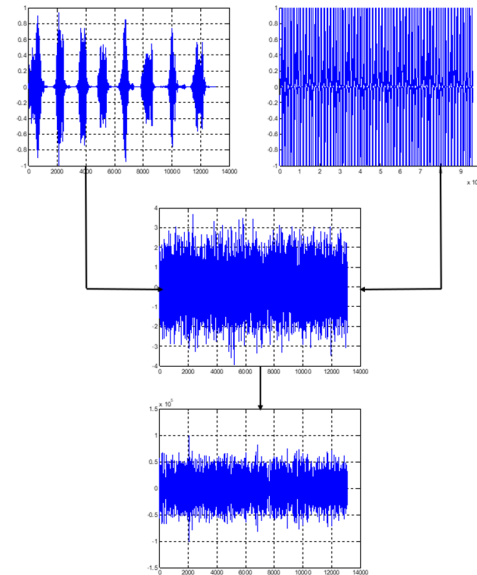


Fig. 6. Encryption stages of audio signal1.

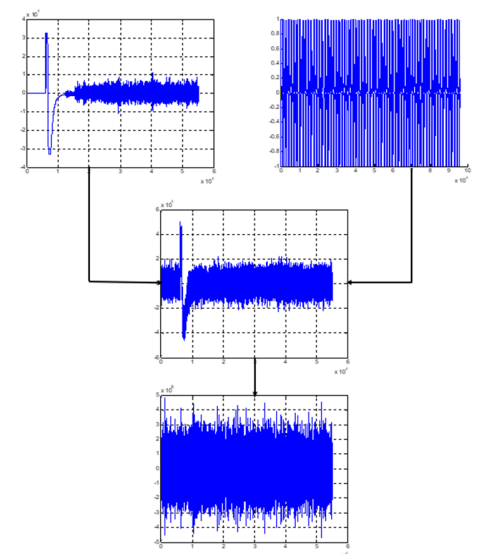


Fig. 7. Encryption stages of audio signal 2.

VI. QUALIFYING MEASUREMENTS

A. Mean Square Error (MSE)

This evaluation parameter measures the total squared error in all kinds of data. MSE is obtained by:

$$E = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n [c(i, j) - c(i, j)']^2 \quad (4)$$

where m denotes the dimensions of the origin matrices and $c(i, j)$ the dimensions of the stego data.

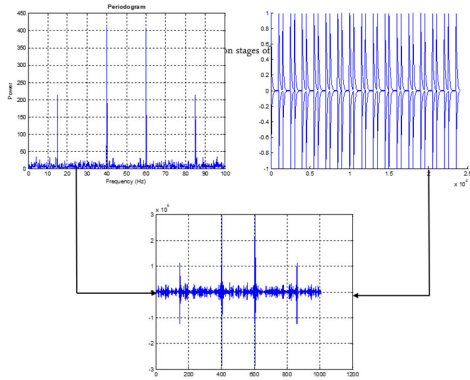


Fig. 8. Encryption stages of an FFT power signal.

B. Peak Signal to Noise Ratio (PSNR)

The PSNR parameter is a measure through which the peak error is determined. Thus, it is determined by the quality of the image, thus higher PSNR can be achieved when the quality of the image is high. The following equation can be used to calculate the PSNR:

$$PSNR = 10 \log_{10} \frac{R^2}{MSE} \quad (5)$$

where R represents the higher potential value of the pixel's density.

Performance can also be evaluated using the time frame used to encode and decode the data:

$$Corr = \frac{\sum_{i=1}^N \sum_{j=1}^M (I_1(i,j) - \bar{I}_1)(I_2(i,j) - \bar{I}_2)}{\sqrt{[\sum_{i=1}^N \sum_{j=1}^M (I_1(i,j) - \bar{I}_1)^2][\sum_{i=1}^N \sum_{j=1}^M (I_2(i,j) - \bar{I}_2)^2]}} \quad (6)$$

where $I_1(i,j)$ is the value of the (i,j) pixel of the original data, \bar{I}_1 is the mean of the original data, calculated by (7), $I_2(i,j)$ presents the value of the (i,j) pixel of the reconstructed data, \bar{I}_2 is the mean of the reconstructed data calculated by (8), M is the height of the image, N is its width, and i and j represent the row and column numbers. Entropy is defined by (9) where S is the entropy and P_i is the probability.

$$\bar{I}_1 = \frac{1}{M \times N} \sum_{i=1}^N \sum_{j=1}^M I_1(i,j) \quad (7)$$

$$\bar{I}_2 = \frac{1}{M \times N} \sum_{i=1}^N \sum_{j=1}^M I_2(i,j) \quad (8)$$

$$S = -\sum_i P_i \log P_i \quad (9)$$

TABLE I. QUALITY MEASUREMENTS FOR ONE-DIMENSIONAL DATA

Audio signal 1	Signals	PSNR (db.)	Entropy	Corr.
Audio signal 1	Origin signal (1)	48	2.891	0.0003168
	CRNG model		5.857	
	Encrypted signal (1)	31	0.6527	
Audio signal 2	Origin signal (2)	45	3.034	0.000581
	CRNG model		5.857	
	Encrypted signal (2)	32	0.7531	
FFT power signal	Origin signal	41	1.352	0.000201
	CRNG model		2.855	
	Encrypted signal	30	0.301	

TABLE II. QUALITY MEASUREMENTS FOR TWO-DIMENSIONAL DATA

	Images	PSNR (db.)	Entropy	Corr.
3-D Image	Origin Image	43.78	4.583	0.0000384
	CRNG model		7.304	
	Encrypted Image	33.21	0.8980	
2-D Image	Origin Image	47	4.169	0.000022
	CRNG model		7.304	
	Encrypted Image	35.83	0.7998	

TABLE III. TIME OF ENCRYPTION AND DECRYPTION PROCESSES

Data type	Time of encryption (sec)	Time of decryption (sec)
Audio signal 1	1.005	0.2035
Audio signal 2	1.845	0.375
FFT power signal	1.013	0.210
3-D Image	3.590	1.001
2-D Image	2.720	0.990

VII. CONCLUSIONS

In this study, highly encrypted data were produced as a result of the construction of a random number generator from a synchronized chaotic system. The high rate of encrypted data is attributed to the high entropy values achieved by the models proposed in this study. This is why the features of the original data were totally hidden. With regard to the urgency with which the data are needed and considering the speedy nature of the 5G network, the focus of this work is to shorten the process of data recovery. The decryption algorithm has high efficiency in terms of returning data to its original values despite its straightforwardness and short steps. It was observed that the proposed algorithm consumed more time during data encryption and lesser time for decryption. The central idea of the current research is to create a mathematical model that is highly chaotic during the process of encoding.

REFERENCES

- [1] A. A. A. El-Latif, B. Abd-El-Atty, W. Mazurczyk, C. Fung, and S. E. Venegas-Andraca, "Secure Data Encryption Based on Quantum Walks for 5G Internet of Things Scenario," *IEEE Transactions on Network and Service Management*, vol. 17, no. 1, pp. 118–131, Mar. 2020, <https://doi.org/10.1109/TNSM.2020.2969863>.
- [2] M. K. Abdul-Hussein and H. T. S. AlRikabi, "Secured Transfer and Storage Image Data for Cloud Communications," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 19, no. 06, pp. 4–17, May 2023, <https://doi.org/10.3991/ijoe.v19i06.37587>.
- [3] A. H. M. Alaidi, R. M. Al-airaji, H. T. S. Alrikabi, I. A. Aljazeera, and S. H. Abbood, "Dark Web Illegal Activities Crawling and Classifying Using Data Mining Techniques," *International Journal of Interactive Mobile Technologies (IJIM)*, vol. 16, no. 10, pp. 122–139, May 2022, <https://doi.org/10.3991/ijim.v16i10.30209>.
- [4] H. T. S. Alrikabi and H. T. Hazim, "Secure Chaos of 5G Wireless Communication System Based on IOT Applications," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 18, no. 12, pp. 89–105, Sep. 2022, <https://doi.org/10.3991/ijoe.v18i12.33817>.
- [5] M. Gafsi, N. Abbassi, R. Amdouni, M. A. Hajjaji, and A. Mtibaa, "Hardware implementation of a strong pseudo-random numbers generator with an application to image encryption," in *2022 IEEE 9th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)*, Hammamet, Tunisia, Feb. 2022, pp. 510–515, <https://doi.org/10.1109/SETIT54465.2022.9875453>.

- [6] M. Tuna, "A novel secure chaos-based pseudo random number generator based on ANN-based chaotic and ring oscillator: design and its FPGA implementation," *Analog Integrated Circuits and Signal Processing*, vol. 105, no. 2, pp. 167–181, Nov. 2020, <https://doi.org/10.1007/s10470-020-01703-z>.
- [7] M. Eisenbarth *et al.*, "Toward Smart Vehicle-to-Everything-Connected Powertrains: Driving Real Component Test Benches in a Fully Interactive Virtual Smart City," *IEEE Vehicular Technology Magazine*, vol. 16, no. 1, pp. 75–82, Mar. 2021, <https://doi.org/10.1109/MVT.2020.3008018>.
- [8] M. P. Kennedy and G. Kolumbán, "Digital communications using chaos," *Signal Processing*, vol. 80, no. 7, pp. 1307–1320, Jul. 2000, [https://doi.org/10.1016/S0165-1684\(00\)00038-4](https://doi.org/10.1016/S0165-1684(00)00038-4).
- [9] S. Banerjee, Ed., *Chaos Synchronization and Cryptography for Secure Communications: Applications for Encryption*, 1st ed. Hershey, PA, USA: Information Science Reference, 2010.
- [10] M. Roy *et al.*, "Data Security Techniques Based on DNA Encryption," in *Proceedings of International Ethical Hacking Conference 2019*, Singapore, 2020, pp. 239–249, https://doi.org/10.1007/978-981-15-0361-0_19.
- [11] K. M. Cuomo and A. V. Oppenheim, "Chaotic signals and systems for communications," in *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Minneapolis, MN, USA, Apr. 1993, vol. 3, pp. 137–140 vol.3, <https://doi.org/10.1109/ICASSP.1993.319454>.
- [12] M. J. M. Ameen and S. S. Hreshee, "Securing Physical Layer of 5G Wireless Network System over GFDM Using Linear Precoding Algorithm for Massive MIMO and Hyperchaotic QRDecomposition," *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 5, pp. 579–591, Oct. 2022, <https://doi.org/10.22266/ijies2022.1031.50>.
- [13] B. Bordel, R. Alcarria, T. Robles, and M. S. Iglesias, "Data Authentication and Anonymization in IoT Scenarios and Future 5G Networks Using Chaotic Digital Watermarking," *IEEE Access*, vol. 9, pp. 22378–22398, 2021, <https://doi.org/10.1109/ACCESS.2021.3055771>.
- [14] M. O. Al-Dwairi, A. Y. Hendi, and Z. A. AlQadi, "An Efficient and Highly Secure Technique to Encrypt and Decrypt Color Images," *Engineering, Technology & Applied Science Research*, vol. 9, no. 3, pp. 4165–4168, Jun. 2019, <https://doi.org/10.48084/etasr.2525>.
- [15] A. H. Alaidi, C. Soong Der, and Y. Weng Leong, "Increased Efficiency of the Artificial Bee Colony Algorithm Using the Pheromone Technique," *Engineering, Technology & Applied Science Research*, vol. 12, no. 6, pp. 9732–9736, Dec. 2022, <https://doi.org/10.48084/etasr.5305>.
- [16] A. H. Alaidi, S. D. Chen, and Y. Weng Leong, "Artificial Bee Colony with Crossover Operations for Discrete Problems," *Engineering, Technology & Applied Science Research*, vol. 12, no. 6, pp. 9510–9514, Dec. 2022, <https://doi.org/10.48084/etasr.5250>.
- [17] W. Jinqiu, Q. Gang, and K. Pengbin, "Emerging 5G Multicarrier Chaotic Sequence Spread Spectrum Technology for Underwater Acoustic Communication," *Complexity*, vol. 2018, Oct. 2018, Art. no. e3790529, <https://doi.org/10.1155/2018/3790529>.
- [18] B. Jovic, "Chaotic Signals and Their Use in Secure Communications," in *Synchronization Techniques for Chaotic Communication Systems*, B. Jovic, Ed. Berlin, Heidelberg, Germany: Springer, 2011, pp. 31–47.
- [19] H. B. Meitei and M. Kumar, "FPGA Implementation of a Wireless Communication System for Secure IR Sensor Data Transmission using TRNG," *International Journal of Engineering Trends and Technology*, vol. 70, no. 7, pp. 220–237, 2022, <https://doi.org/10.14445/22315381/IJETT-V70I7P223>.
- [20] K. Lee, S.-Y. Lee, C. Seo, and K. Yim, "TRNG (True Random Number Generator) Method Using Visible Spectrum for Secure Communication on 5G Network," *IEEE Access*, vol. 6, pp. 12838–12847, 2018, <https://doi.org/10.1109/ACCESS.2018.2799682>.
- [21] J. S. Teh, A. Samsudin, M. Al-Mazrooe, and A. Akhavan, "GPUs and chaos: a new true random number generator," *Nonlinear Dynamics*, vol. 82, no. 4, pp. 1913–1922, Dec. 2015, <https://doi.org/10.1007/s11071-015-2287-7>.

A Spectrotemporal EEG Mapping Approach for Decoding Imagined Marathi Language Phonemes

Umesh Mhapankar

Agnel Charities' Fr. C. Rodrigues Institute of Technology, India
mhapankar.umesh@fcrit.ac.in (corresponding author)

Milind Shah

Agnel Charities' Fr. C. Rodrigues Institute of Technology, India
milind.shah@fcrit.ac.in

Received: 22 January 2024 | Revised: 6 February 2024 and 9 February 2024 | Accepted: 11 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6954>

ABSTRACT

Individuals facing verbal communication impairments resulting from brain disorders like paralysis or autism encounter significant challenges when unable to articulate speech. This research proposes the design and development of a wearable system capable of decoding imagined speech using electroencephalogram (EEG) signals obtained during the mental process of speech generation. The system's main objective is to offer an alternative communication method for individuals who can hear and think but face challenges in articulating their thoughts verbally. The design suggested includes user-friendliness, wearability, and comfort for seamless integration into daily life. A minimal number of electrodes are strategically placed on the scalp to minimize invasiveness. Achieving precise localization of the cortical areas responsible for generating the EEG patterns during imagined speech is vital for accurate decoding. Literature studies are utilized to determine the cortical positions associated with speech processing. Due to the inherent limitations in EEG spatial resolution, meticulous experiments are conducted to map the scalp positions onto their corresponding cortical counterparts. Specifically, we focus on identifying the scalp location over the superior temporal gyrus (T3) using the internationally recognized 10-20 electrode placement system by employing a circular periphery movement with a 2 cm distance increment. Our research involves nine subjects spanning various age groups, with the youngest being 23 and the oldest 65. Each participant undergoes ten iterations, during which they imagine six Marathi syllables. Our work contributes to the development of wearable assistive technology, enabling mute individuals to communicate effectively by translating their imagined speech into actionable commands. This innovation ultimately enhances their social participation and overall well-being.

Keywords-EEG; imagined speech; location mapping; Wernicke area

I. INTRODUCTION

Speech plays a vital role in human communication; however, a portion of the population may be mute due to brain disorders. Despite possessing the ability to listen, hear, and internally imagine speech, these individuals are unable to overtly reproduce it. About 10% of the global population is affected by such brain disorders [1], emphasizing the pressing need for a non-invasive system capable of decoding internally generated speech. Several methods exist for measuring the brain's activity, including invasive techniques such as Electrocorticography (ECoG) and non-invasive approaches like electroencephalography (EEG). While ECoG offers superior spatial and temporal resolution, it requires surgical intervention and lacks portability. In contrast, with its good temporal resolution, non-invasiveness, and portability, EEG has gained significant traction in applications, such as Brain-Computer

Interfaces (BCIs) and clinical settings [2]. Other non-invasive methods to test brain activity are fMRI (functional Magnetic Resonance Imaging) and PET (Positron Emission Tomography) scans, but these techniques have poor temporal resolution and are not portable.

This research focuses on the non-invasive decoding of imagined speech, also known as covert speech, using an EEG system. However, the EEG signals acquired from the scalp are weak, on the order of micro-volts. They are susceptible to contamination from various body artifacts like electrocardiogram (ECG) and electromyogram (EMG) signals, as well as from electromagnetic interference (EMI) signals. Clinical methods employ a minimum of 24 to 64 electrodes using the standard international 10-20 system with wet electrodes and many systems are within the 0-40 Hz bandwidth.

II. RELATED WORK

A. Wernicke area

The existing literature highlights the significance of specific brain regions in language processing. Wernicke area, located in the Superior Temporal Gyrus (STG), as shown in Figure 1, is known to play a crucial role in comprehension. Additionally, the angular gyrus and Broca area, supported by motor neurons, contribute significantly to language articulation prior to real-time speech. These regions collectively assist the pre-processing stages of speech within the brain.

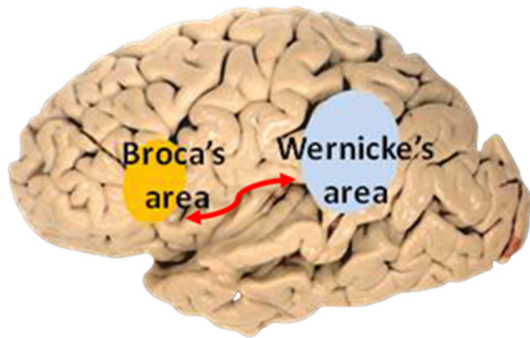


Fig. 1. Broca and Wernicke areas.

The dorsal stream is surrounded by the lateral sulcus, Broca area, and Wernicke area. The superior longitudinal fasciculus/arcuate fasciculus is the major fiber tract of the dorsal stream. It comprises four subcomponents (superior longitudinal fasciculus I–III), with the arcuate fasciculus constituting the fourth subcomponent. The superior longitudinal fasciculus III and arcuate fasciculus play a significant function in the dorsal language stream.

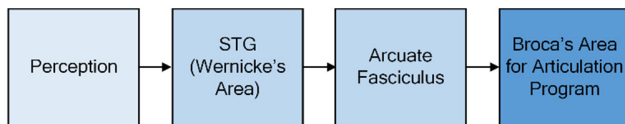


Fig. 2. Wernicke–Geschwind model.

B. Electrical Dipole at the Superior Temporal Gyrus (STG) due to Speech and other Activities

Many researchers agree that electrical dipoles arise from the dynamic activity in the brain. These dipoles are likely created by a moment that correlates with both location and strength of the activity [3]. However, this potential is reduced when reflected on the scalp due to its layers and the skull's high resistance. For cognitive activity in the brain, the dual-stream model, the dorsal and the ventral is conceptualized in [4]. Understanding the current density and potential development in specific brain areas is crucial for analyzing various activities related to synapses, membranes, fluids, and tissues. In particular, the STG, also known as Wernicke area, plays a vital role in speech comprehension [5].

Figure 3 highlights the STG area, which is responsible for speech comprehension. The electrical dipole influences in this

area arise from various activities and can be analyzed using the superposition principle, similar to network theory. The location was selected using the theory mentioned in anatomy books [3]. The potential developed at the STG is influenced by current vectors E_1, E_2, E_3 , up to E_n and the potential due to local speech activity. These vectors possess different amplitudes, directions, and phases. The vectors can be separated and analyzed using Independent Component Analysis (ICA). Apart from speech signals, the remaining signals are considered Event-Related Potentials (ERP). These signals are characterized by low frequency and delayed responses hence they do not affect the speech signals. The envelopes for speech signals also exhibit low-frequency characteristics, allowing separating the speech information using spectral filtering, particularly in the gamma and theta frequency ranges. To map the electrical dipole at the STG onto the scalp, a three-sphere model (scalp, skull, and CSF) or a four-sphere model (including the inner brain or cortex) is considered to account for volume conduction as displayed in Figure 4. The placement of electrodes within this boundary is determined by calculations based on Poisson's theorem.

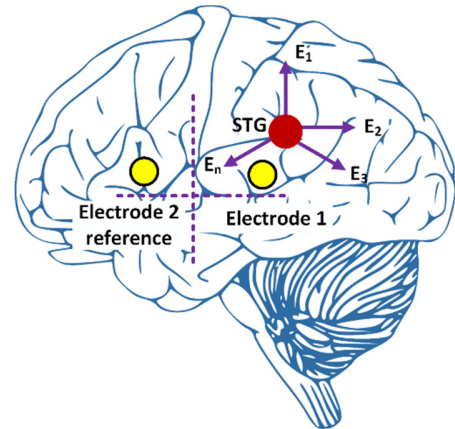


Fig. 3. Electrode positioning as fMRI, standard 10-20 method on the scalp.

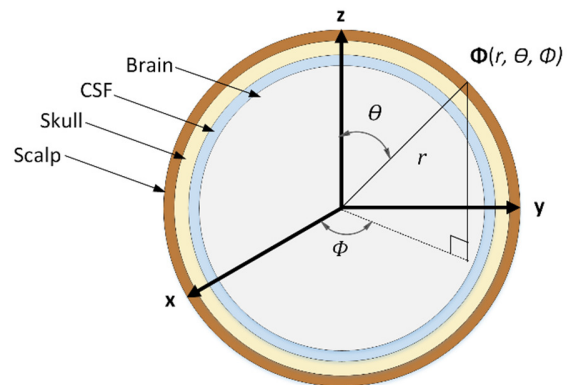


Fig. 4. A schematic representation of the brain, depicting four layers: the brown-coloured scalp layer, the yellow-coloured skull layer, the light blue-coloured cerebrospinal fluid (CSF), and the light gray-coloured inner brain or cortex. Spherical coordinates (r, θ, ϕ) are used to represent the location of the STG.

$$\Phi = \sum_1^n E_n e^{-i\omega_n \theta_n} \quad (1)$$

In (1), E_n is the amplitude of potential due to various inputs, ω_n is the frequency, θ_n is the phase angle of the vector, and n is the number of sources that influence the location at the cortex, involved in speech comprehension. Equation (1) suggests that multiple inputs contribute to the potential with different frequencies and phase components. After acquiring EEG at the proposed location, Fast Fourier Transform (FFT) determines the frequency spectrum, which is then filtered using various bandpass filters through LabVIEW. The resulting drop in potential will be dependent on the equivalent circuit portrayed in Figure 5. Impedance $Z_c + Z_{sk} + Z_{sc}$ will attenuate the voltage at the scalp. The EEG voltage V_{ee} in (2) is given by:

$$V_{ee} = I_s R_t - Z_t I \quad (2)$$

where I_s is the source current, R_t is the source internal resistance, Z_t is the total impedance including electrode and contact resistance ($Z_c + Z_{sk} + Z_{sc}$), and I is the total current. Before analysis in the LabVIEW, the signal is amplified using a front-end amplifier.

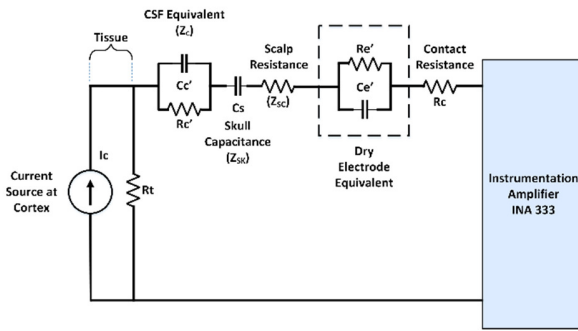


Fig. 5. Equivalent circuit from cortex to electrode.

C. Electrodes

The choice of electrode type is another critical consideration. Wet Ag/AgCl electrodes are considered the gold standard due to their superior conductivity. However, their application necessitates the use of conductive gel to minimize the contact resistance between skin and electrodes. There are two issues with the gel. First, it creates discomfort for the user, and it dries every five hours, so it needs to be replaced to maintain conductivity [6]. This reduces the system's portability. Additionally, the bandwidth of existing EEG systems poses limitations, as decoding imagined speech requires access to theta waves (0-4 Hz) and high gamma signals (70 Hz-200 Hz and above). To address bandwidth and wet electrode issues, a graphene-coated dry electrode has been developed [7]. Graphene is a highly conducting and sensitive engineering material. These electrodes introduce capacitive reactance into the circuit, creating a high-pass filter to address the low bandwidth nature of the dry electrodes. Additionally, it offers low skin contact resistance, exhibiting characteristics closely resembling the gold standard.

During the mapping process, EEG signals are influenced by low-pass filtering from the cortex to the scalp. Therefore, an expanded bandwidth is essential, posing a challenge for the

current EEG systems. ECoG, an invasive technique involving electrode placement on the cortex via surgery, can accurately decode imagined speech, specifically targeting the STG. However, the ECoG method is unsuitable for general wearable applications due to its invasive nature [3, 6]. To address these challenges, this study aims to map the specific scalp region associated with decoding imagined speech using EEG. By accurately projecting the cortex data onto the scalp, we aim to reduce the number of electrodes required. Minimizing the electrode count is a crucial step towards achieving a practical and user-friendly decoding technique. While some researchers have utilized multiple electrodes across the head periphery, others have employed a minimum of 14 channels (electrodes) for EEG-based decoding. Improving the placement of these electrodes based on precise scalp mapping derived from cortex data [8] will contribute to more efficient decoding techniques.

D. Localization

Brain activity signals exhibit three-dimensional characteristics with respect to the normal vector of the activity source. One of the locations where these signals are reflected is the scalp. However, the reflected signal captured by the EEG on the scalp suffers from poor spatial resolution. Therefore, it becomes imperative to accurately map the precise location of the brain activity [9, 10]. The exact localization of primary and secondary speech areas is still being under research [5]. The design should be user-friendly for precise electrode placement on the scalp with minimal electrodes. The literature on localization presents varying perspectives on mapping brain activity and the abstraction of neurons. A widely accepted perspective is based on the modular theory, suggesting that distinct brain regions handle specific tasks and activities, exerting control over corresponding areas of the body [3, 11]. Conversely, some researchers propose a modified perspective, indicating that brain organization may vary across individuals [12]. Furthermore, the literature also delves into the auditory processing pathway within the brain, which elicits differing opinions regarding its course. Two distinct viewpoints emerge: the dorsal and ventral pathways [12-14]. Understanding the precise localization of the electrical potential resulting from speech within the brain's cortex is of utmost importance. This process entails source localization, involving the identification of the cortical region responsible for generating the electrical signal, followed by mapping the same signal onto the scalp for further analysis and interpretation.

Researchers have noticed similarities between the signal properties observed at the cortex and the scalp, apart from differences in bandwidth [15, 16]. While these properties have been primarily investigated in clinical contexts, their confirmation of cognitive activity remains uncertain. Consequently, decoding speech signals necessitates further investigation into the scalp's location. Mapping from the cortex to the scalp is essential as it addresses the issue of volume conduction, as numerous simultaneous activities can influence speech-related activity, making it challenging to isolate clear speech signals. The inter-individual variations in head geometry need also to be considered. It is crucial to study EEG signals simultaneously in the vicinity of the speech location, as indicated in the literature. Such investigations, conducted with

an appropriate experimental setup can provide further evidence to support the correlation between cortical activity and EEG signals during speech-related activities. However, a comprehensive study involving simultaneous analysis of both cortical and EEG activity falls beyond the scope of the study. Instead, this study focuses on conducting multiple iterations of EEG measurements on a diverse group of subjects spanning different age groups and possessing varying head structures.

III. MATERIALS AND METHODS

A. Participants

Nine male participants within the age group of 23 to 65 were selected for this study. All participants had bald heads and no known medical history. The experiments were conducted with the guidance of medical practitioners at Niramay Clinic in Panvel (MS India). Prior to participating in the study, all participants provided an informed consent.

B. Data Acquisition

The experiments utilize multiple electrode placements at different locations to accurately capture speech-related activity. Various speech prompts involving vowels and consonants were utilized, and statistical features as well as spectrotemporal traits were extracted from the EEG data to facilitate analysis. EEG data were recorded using a 32-channel Medicaid, Chandigarh EEG machine with a bandwidth of 512 Hz. Six electrodes were placed on the scalp according to the standard 10-20 electrode placement system, as depicted in Figure 1. The chosen electrode positions were between the T3 (Temporal) and F5 (Frontal) areas. During the recording session, the subjects were instructed to perform three modalities: no speech, overt speech, and imagined speech (covert speech), each lasting for 10 seconds. A total of six Marathi phonemes, including three vowels and three consonants, were used to test the results. They were asked to pronounce the Marathi phoneme [kə] [17] ten times within a duration of 10 seconds, following a one-second starting cue. Subsequently, they were instructed to speak the phonemes [tsə] and [jə], and three vowels [ə], [i], and [u] for a total duration of 90 seconds, with a 5-second break between each set of phonemes. Out of the six, only the [tsə] phoneme was used for location validation, while the other five were utilized for classification purposes.

The acquired EEG data were processed employing the LabVIEW software from National Instruments. Initially, a 50

Hz notch filter was applied to remove power-line artifacts. The data were then filtered putting into service a high-order bandpass filter with a sampling frequency of 512 Hz, to extract five distinct frequency bands: 0-4 Hz (Theta wave), 4-8 Hz (Delta wave), 8-12 Hz (Alpha wave), 12-20 Hz (Beta wave), and 70-200 Hz (Gamma wave). The experiments were conducted implementing two types of electrodes: conventional wet Ag/Ag-Cl electrodes and graphene-based dry electrodes [15].

IV. RESULTS AND DISCUSSION

Experimentation was conducted to confirm the location using nine healthy subjects. Graphene-based 6 cm² dry electrodes [18] were selected due to their larger bandwidth compared to wet and other dry electrodes. Power Spectral Density (PSD) was measured, in every iteration, and the average was calculated. PSD was calculated using the Welch method with a Hanning window technique, as illustrated in Tables I-V. The standard deviation indicates that the gamma wave (70-200 Hz) and theta (0-4 Hz) wave have distinct variations compared to other features like delta, alpha, and beta for different prompts. However, the standard deviation for the same prompt across various subjects is insignificant for theta and gamma waves. The bar graphs (Figures 6-9) visualize the results.

The average PSD is highest for all participants at location L1, with minimum deviation for the gamma and theta frequency range. Observations infer that Location 1 provides a more robust output for theta and gamma signals. Location P1 is considered normal with the minimum distance to STG [18], and the standard deviation is lower at the same position. Location P1 exhibits an average PSD higher than the other five locations. Delta, alpha, and beta waves gave inconclusive results because alpha and beta waves arrived with artifacts arising from other activity. The FFT shows a similar gamma frequency range for different participants for the same prompt, and spectrogram features offer similar spectral visualizations for actual classification. The spectrogram of gamma waves (Figure 10) also provides important information about the prompt, which can be further used for regression and classification. Machine learning algorithms like Support Vector Machines and Random Forest, can be utilized to check classification accuracy [19-21].

TABLE I. SD AND AVPSD FOR LOCATION 1

Type of EEG signal	Theta 0-4 Hz	Delta 4-8 Hz	Alfa 8-12 Hz	Beta 12-20 Hz	Gamma 70-200 Hz
Modality	(SD) AvPSD	(SD) AvPSD	(SD) AvPSD	(SD) AvPSD	(SD) AvPSD
Without speech	(0.23) 2.56	(0.28) 2.91	(0.11) 2.15	(0.17) 3.92	(0.07) 0.24
Overt / [tsə]	(0.11) 5.77	(0.21) 2.51	(0.19) 2.62	(0.14) 3.42	(0.08) 3.68
Covert / [tsə]	(0.14) 4.64	(0.13) 1.75	(0.21) 1.68	(0.13) 2.31	(0.06) 3.14

TABLE II. SD AND AVPSD FOR LOCATION 2

Type of EEG signal	Theta 0-4 Hz	Delta 4-8 Hz	Alfa 8-12 Hz	Beta 12-20 Hz	Gamma 70-200 Hz
Modality	(SD) AvPSD	(SD) AvPSD	(SD) AvPSD	(SD) AvPSD	(SD) AvPSD
Without speech	(0.34) 2.89	(0.43) 2.89	(0.29) 2.09	(0.45) 4.67	(0.09) 0.15
Overt / [tsə]	(0.56) 4.12	(0.33) 2.67	(0.34) 2.57	(0.39) 2.57	(0.23) 1.68
Covert / [tsə]	(0.37) 4.19	(0.28) 1.75	(0.36) 1.67	(0.41) 3.12	(0.19) 1.56

TABLE III. SD AND AVPSD FOR LOCATION 3

Type of EEG signal	Theta 0-4 Hz	Delta 4-8 Hz	Alfa 8-12 Hz	Beta 12-20 Hz	Gamma 70-200 Hz
Modality	(SD) AvPSD	(SD) AvPSD	(SD) AvPSD	(SD) AvPSD	(SD) AvPSD
Without speech	(0.23) 2.54	(0.19) 1.23	(0.19) 1.92	(0.36) 4.72	(0.15) 0.42
Overt /[tsə]	(0.36) 4.49	(0.23) 2.76	(0.28) 2.72	(0.31) 3.52	(0.32) 1.43
Covert /[tsə]	(0.21) 2.23	(0.13) 1.87	(0.34) 1.47	(0.29) 2.67	(0.14) 0.78

TABLE IV. SD AND AVPSD FOR LOCATION 4.

Type of EEG signal	Theta 0-4 Hz	Delta 4-8 Hz	Alfa 8-12 Hz	Beta 12-20 Hz	Gamma 70-200 Hz
Modality	(SD) AvPSD	(SD) AvPSD	(SD) AvPSD	(SD) AvPSD	(SD) AvPSD
Without speech	(0.34) 3.21	(0.32) 2.42	(0.38) 2.72	(0.49) 4.65	(0.23) 0.56
Overt /[tsə]	(0.26) 2.72	(0.39) 2.56	(0.29) 1.67	(0.35) 3.25	(0.34) 1.43
Covert /[tsə]	(0.45) 3.56	(0.67) 1.87	(0.69) 1.565	(0.45) 2.23	(0.28) 2.47

TABLE V. SD AND AVPSD FOR LOCATION 5

Type of EEG signal	Theta 0-4 Hz	Delta 4-8 Hz	Alfa 8-12 Hz	Beta 12-20 Hz	Gamma 70-200 Hz
Modality	(SD) AvPSD	(SD) AvPSD	(SD) AvPSD	(SD) AvPSD	(SD) AvPSD
Without speech	(0.21) 3.24	(0.31) 2.75	(0.22) 1.89	(0.45) 4.43	(0.12) 0.56
Overt /[tsə]	(0.26) 3.67	(0.34) 2.45	(0.31) 2.45	(0.32) 2.41	(0.25) 1.39
Covert /[tsə]	(0.26) 3.15	(0.28) 2.76	(0.29) 1.67	(0.27) 2.89	(0.18) 1.12

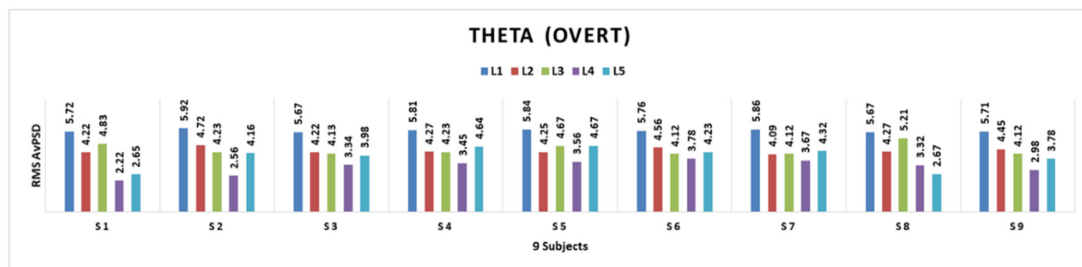


Fig. 6. Column chart showing variation RMSAvPSD theta wave for overt speech

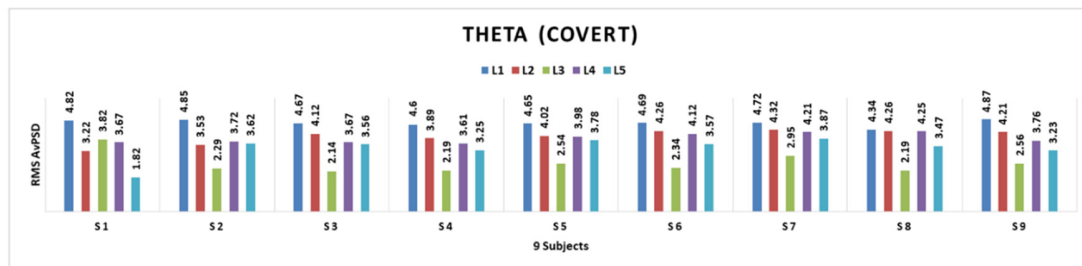


Fig. 7. Column chart showing variation RMSAvPSD theta wave for covert speech.

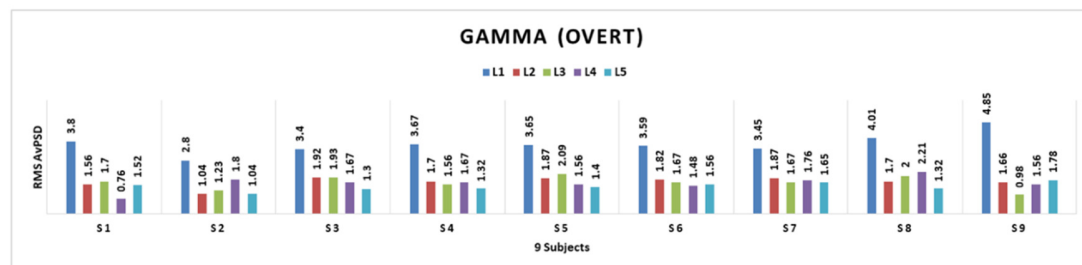


Fig. 8. Column chart showing variation RMSAvPSD gamma wave for overt speech.

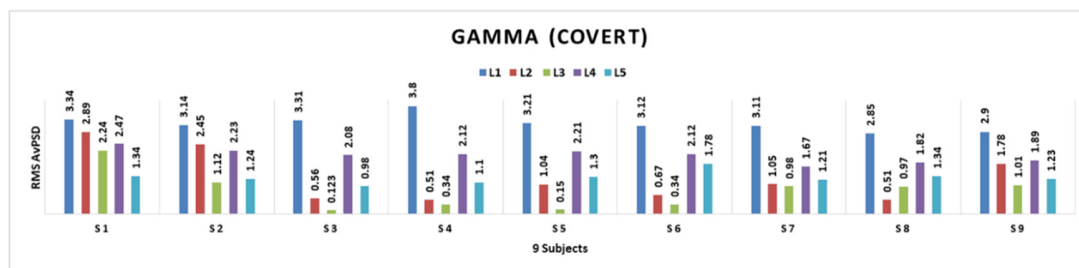


Fig. 9. Column chart showing variation AvRMS PSD theta wave for overt speech.

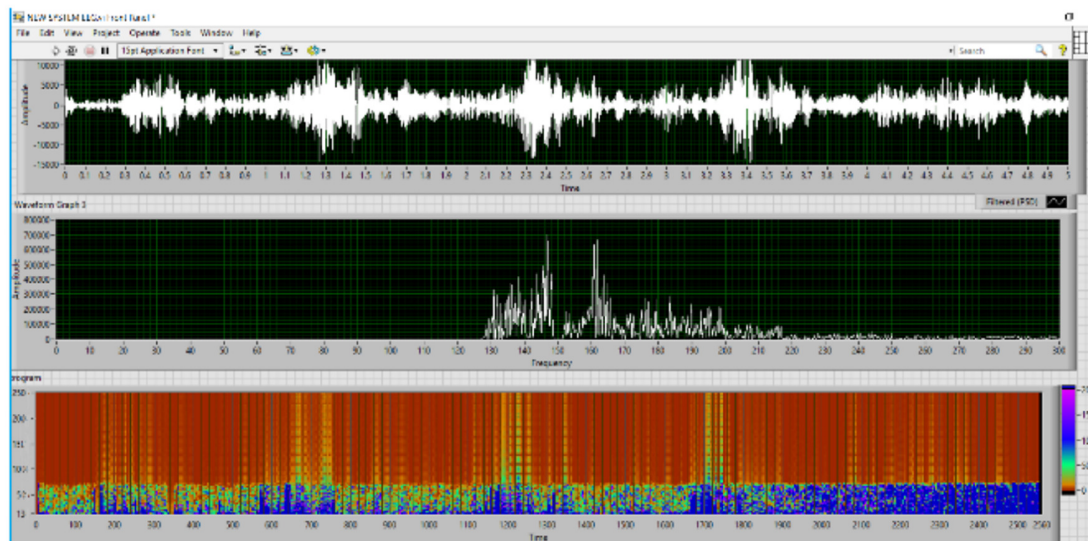


Fig. 10. Time, FFT, and Spectrogram of Covert [tsə] in Marathi.

V. CONCLUSION

This study's experimentation using graphene-based dry electrodes and simultaneous measurements from healthy participants has yielded valuable results. The former reveals that selecting an adequate electrode and identifying the location correctly can map speech-like activity on the brain scalp. This observation provides new insights into the intricate dynamics of brain signal localization at various positions on the scalp. The superior temporal gyrus, one of the areas that participate in speech production, can be mapped on the scalp based on head geometry and the international standard of the 10-20 method. Signals on the scalp, specifically the theta and gamma signals, exhibited a significant contribution to speech decoding. The output obtained from this mapping can enhance classification accuracy, as it provides distinct spectrotemporal features. Our work enhances the understanding of the brain's speech-processing capabilities and holds promise for empowering innovative applications in brain-computer interfaces and speech-assistive devices.

REFERENCES

- [1] S. Martin, C. Mikutta, R. T. Knight, and B. N. Pasley, "Understanding and Decoding Thoughts in the Human Brain," *Frontiers for Young Minds*, <https://kids.frontiersin.org/articles/10.3389/frym.2016.00004>.
- [2] S. Martin *et al.*, "Decoding spectrotemporal features of overt and covert speech from the human cortex," *Frontiers in Neuroengineering*, vol. 7, May 2014, <https://doi.org/10.3389/fneng.2014.00014>.
- [3] P. L. Nunez and R. Srinivasan, *Electric Fields of the Brain: The Neurophysics of EEG*, 2nd ed. Oxford, UK: Oxford University Press, 2005.
- [4] R. E. Weller, "Chapter 27: Two cortical visual systems in Old World and New World primates," in *Progress in Brain Research*, vol. 75, T. P. Hicks and G. Benedek, Eds. Elsevier, 1988, pp. 293–306.
- [5] J. T. Panachakel and A. G. Ramakrishnan, "Decoding Covert Speech From EEG-A Comprehensive Review," *Frontiers in Neuroscience*, vol. 15, Apr. 2021, Art. no. 642251, <https://doi.org/10.3389/fnins.2021.642251>.
- [6] S. R. Synigal, E. S. Teoh, and E. C. Lalor, "Including Measures of High Gamma Power Can Improve the Decoding of Natural Speech From EEG," *Frontiers in Human Neuroscience*, vol. 14, 2020, Art. no. 130, <https://doi.org/10.3389/fnhum.2020.00130>.
- [7] U. Mhapankar and M. M. Shah, "Mapping the Imagined Speech Location on the Brain Scalp Through Magnetoencephalography (MEG)," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 11, Jul. 2022, <https://doi.org/10.35940/ijrte.b7144.0711222>.
- [8] U. Mhapankar and M. M. Shah, "Mapping the Imagined Speech Location on the Brain Scalp Through Magnetoencephalography (MEG)," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 11, Jul. 2022, <https://doi.org/10.35940/ijrte.b7144.0711222>.
- [9] S. Deng, R. Srinivasan, and M. D'Zmura, "Cortical Signatures of Heard and Imagined Speech Envelopes," Aug. 2013, [Online]. Available:

- <https://cnslab.ss.uci.edu/speechattention/content/DengSrinivasanDZmur a2013.pdf>.
- [10] J. Derix, O. Iljina, J. Weiske, A. Schulze-Bonhage, A. Aertsen, and T. Ball, "From speech to thought: the neuronal basis of cognitive units in non-experimental, real-life communication investigated using ECoG," *Frontiers in Human Neuroscience*, vol. 8, Jun. 2014, <https://doi.org/10.3389/fnhum.2014.00383>.
- [11] H. Lu, J. Li, L. Zhang, S. S. M. Chan, L. C. W. Lam, and for the Open Access Series of Imaging Studies, "Dynamic changes of region-specific cortical features and scalp-to-cortex distance: implications for transcranial current stimulation modeling," *Journal of NeuroEngineering and Rehabilitation*, vol. 18, no. 1, Jan. 2021, Art. no. 2, <https://doi.org/10.1186/s12984-020-00764-5>.
- [12] B. Oshri, N. Khandwala, and M. Chopra, "Classifying Syllables in Imagined Speech using EEG Data," [Online]. Available: <https://cs229.stanford.edu/proj2014/Barak%20Oshri,%20Nishith%20Khandwala,%20Manu%20Chopra,%20Classifying%20Syllables%20in%20Imagined%20Speech%20using%20EEG%20Data.pdf>.
- [13] A. Borna *et al.*, "A 20-channel magnetoencephalography system based on optically pumped magnetometers," *Physics in Medicine and Biology*, vol. 62, no. 23, pp. 8909–8923, Nov. 2017, <https://doi.org/10.1088/1361-6560/aa93d1>.
- [14] C. Im and J.-M. Seo, "A review of electrodes for the electrical brain signal recording," *Biomedical Engineering Letters*, vol. 6, no. 3, pp. 104–112, Aug. 2016, <https://doi.org/10.1007/s13534-016-0235-1>.
- [15] D. A. Rojas, L. A. Góngora, and O. L. Ramos, "EEG Signal Analysis Related to Speech Process through Bci Device Emotiv, FFT and Statistical Methods," *ARPJ Journal of Engineering and Applied Sciences*, vol. 11, no. 5, pp. 3074–3080, 2016.
- [16] D. W. Jeong, G. H. Kim, N. Y. Kim, Z. Lee, S. D. Jung, and J.-O. Lee, "A high-performance transparent graphene/vertically aligned carbon nanotube (VACNT) hybrid electrode for neural interfacing," *RSC Advances*, vol. 7, no. 6, pp. 3273–3281, Jan. 2017, <https://doi.org/10.1039/C6RA26836F>.
- [17] "Marathi phonology," *Wikipedia*. Feb. 20, 2024, [Online]. Available: https://en.wikipedia.org/w/index.php?title=Marathi_phonology&oldid=1209133005.
- [18] U. Mhapankar and M. Shah, "The Comparison of the Dry Electrodes to wet Ag/AgCl electrode for Decoding Imagined Speech from the EEG," in *2022 International Conference for Advancement in Technology (ICONAT)*, Goa, India, Jan. 2022, pp. 1–6, <https://doi.org/10.1109/ICONAT53423.2022.9726038>.
- [19] M. B. Ayed, "Balanced Communication-Avoiding Support Vector Machine when Detecting Epilepsy based on EEG Signals," *Engineering, Technology & Applied Science Research*, vol. 10, no. 6, pp. 6462–6468, Dec. 2020, <https://doi.org/10.48084/etasr.3878>.
- [20] G. Anuradha and D. N. Jamal, "Classification of Dementia in EEG with a Two-Layered Feed Forward Artificial Neural Network," *Engineering, Technology & Applied Science Research*, vol. 11, no. 3, pp. 7135–7139, Jun. 2021, <https://doi.org/10.48084/etasr.4112>.
- [21] M. A. Alsawaiet, "Feature Extraction of EEG Signals for Seizure Detection Using Machine Learning Algorithms," *Engineering, Technology & Applied Science Research*, vol. 12, no. 5, pp. 9247–9251, Oct. 2022, <https://doi.org/10.48084/etasr.5208>.

A Low-Profile Electrically Small Serrated Rectangular Patch Antenna for RFID Applications

Naveen Kumar Majji

Department of ECE, KLEF, India | Department of ECE, PSCMR College of Engineering and Technology, India
navvenkumarmajji@gmail.com

Venkata Narayana Madhavareddy

Department of ECE, KLEF, India
mvn@kluniversity.in

Govardhani Immadi

Department of ECE, KLEF, India
govardhanee_ec@kluniversity.in (corresponding author)

Navya Ambati

Department of ECE, KLEF, India
ambatinavya88@gmail.com

Received: 1 February 2024 | Revised: 15 February 2024 | Accepted: 5 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6989>

ABSTRACT

This paper presents the design and analysis of a global system for mobile communication applications, as well as the design and analysis of a compact, bidirectional Electrically Small Antenna (ESA) at 0.9 GHz for Radio Frequency Identification (RFID) applications. In order to attain results at lower frequencies while keeping a compact size, the proposed design consists of a microstrip patch antenna in which an SRR and a semicircular-shaped SRR were subtracted from the ground plane. The dimensions of the FR4 substrate, on which this ESA was designed, were $20 \times 18 \times 1.6$ mm. Ansys HFSS was used for the design and simulation of the antenna. Chemical etching was implemented to fabricate the ESA and MS2037C Anritsu Combinational Analyzer was applied for testing. The simulated results show that at 0.9 GHz, the ESA achieves a bandwidth of 300 MHz (700 MHz-1000 MHz). At the resonance frequency, a bidirectional radiation pattern with a 80% radiation efficiency is obtained in both H and E planes. A 90% agreement between the simulated and the fabricated results has been achieved.

Keywords-electrically small antenna; gain; Chu limit; reflection coefficient; VSWR

I. INTRODUCTION

Electrically Small Antennas (ESAs) play an important role in modern communication engineering as the requirement for miniaturization is growing day by day. An antenna which is enclosed completely in a sphere with radius a is treated as an ESA. For such an antenna $Ka < 1$, where Ka is a wave number equal to $2\pi/\lambda$, where λ is the free space wavelength. Normally, if an antenna has dimensions (including any ground plane) less than $\lambda/4$, it is often referred to as an ESA. ESAs tend to be less efficient and more challenging to design than antennas of normal size. Also, reducing the antenna size, results in larger input reactance, smaller radiation resistance, poor radiation

efficiency, and narrow impedance bandwidth. Being smaller in size, ESAs are easily integrated into smaller devices and have a simple structure, but they suffer from low radiation resistance and high reactance due to their shorter electrical length. Because of the larger wavelength of the EM waves in this frequency band, it is demanded to reduce the antenna size, which in turn requires the use of ESAs. A meander line is placed in order to decrease the size and achieve good impedance matching at 2.4 GHz. The former is designed by employing a characteristic mode theory, according to which this line should have dimensions $8 \text{ mm} \times 8 \text{ mm} \times 1.5 \text{ mm}$ [2]. A low-cost, uniplanar, electrically small structure antenna is proposed in [3] with size $0.154 \times 0.154 \times 0.0004 \lambda^3$ and $Ka = 0.68$,

to achieve minimum return loss and efficient radiation patterns [3]. An electrically small, low-profile antenna supporting both linear and circularly polarized waves plus rectification at 915 MHz in the ISM band with less quality factor for wireless power transfer technologies in the IOT applications is designed in [4]. Authors in [5, 7] proposed a novel ESA using DAM techniques to improve the bandwidth and Q factor of the antenna. Authors in [8] presented a small monopole antenna design for super-wideband (SWB) applications. An antenna was constructed based on a trapezoid form radiator, triangle tapered feed line, and semi-circular ground, to produce good broad frequency bandwidth at the resonating frequency of 1.42 GHz. The reactive impedance problem can be overcome by constructing an antenna with an embedded non-foster circuit for Negative Impedance Conversion (NIC) [9]. In order to optimize and reduce the size of the antenna, active circuits were placed inside the antenna. An ESA for GNSS terminals was proposed in [10], in which a dielectric block with more permittivity was used to achieve miniaturization. In [11], a multi-band, wireless USB dongle ESA with low profile was suggested to operate at LTE 700 and 1.6 GHz. An ESA with a compact size of 4 mm × 4 mm and with a circular shape slot which resulted in an improvement of broad band characteristics was proposed in [13]. An Ultra-High Frequency (UHF) radio frequency identifier that comprised a C-shaped structure with a diameter of 41 mm and thickness of 6.48 mm was introduced in [14]. The proposed Tunable Bio Metric Antenna Array (BMAA) in [15], consisted of two quarter wave closely-spaced monopoles operating at 600 MHz.

From the aforementioned studies, it can be concluded that the suggested ESAs either had complicated structures covering large regions, or they did not cover the 900 MHz frequency range for GSM and RFID applications. The proposed design consists of a microstrip patch antenna in which an SRR and a semicircular shaped SRR were subtracted from the ground plane to obtain results at lower frequencies by maintaining a compact size. Wide impedance bandwidth of 0.3 GHz (0.7 GHz – 1.0 GHz) was obtained for a single band ESA. The return loss of the single band is -23 dB at 0.9 GHz with VSWR < 2. The recommended antenna has 61.7% radiation efficiency and the significant peak gain of 1 dB. Additionally, in both E and H planes, the serrated rectangular ESA offers a bidirectional radiation pattern.

II. ESA DESIGN METHODOLOGY AND GEOMETRY

The serrated rectangular ESA is created by starting with a rectangular patch antenna and applying fundamental microstrip patch antenna design principles, and finally modifying its shape to meet the required parameters. The rectangular patch antenna has good return loss only in a particular frequency range. This simple patch antenna design has undergone a few modifications to get the resonant band at 900 MHz. As a substrate, FR4 epoxy material measuring 1.6 mm in height was used. The proposed antenna has a size of 20 × 18 mm² and it is fed with a 50 Ω microstrip transmission line. In iteration 1, the basic shape of a rectangular shaped SRR is designed with 12 mm length (L1) and width (W1) using the HFSS. Another rectangular shaped patch antenna with L2 = 16 mm and W2 = 15 mm is added in iteration II. In iteration III, along with the

rectangular SRR, a rectangular micro strip and a semi-circular shaped SRR are placed at the bottom of the design. The semi-circular SRR has a radius of 7 mm. In iteration IV, along with the rectangular shaped slot, a stub is added on the left side and in iteration V, rectangular 0.5 mm slots are placed to improve the antenna performance. As per the fundamental limit of ESA, the Ka value for the designed ESA is 0.87. Figures 1-5 represent the iterations I-V of the antennas along with their return loss in dB. The mathematical formulas that are utilized in manufacturing the proposed ESA are shown below. The dimensions of the microstrip patch antenna were calculated by utilizing the concept behind the theory of the transmission lines. The width of the patch can be estimated by:

$$W = \frac{V_0}{2F_r} \sqrt{\frac{2}{\epsilon_r + 1}} \quad (1)$$

The length of the patch can be calculated by (2), where ϵ_{reff} is the effective dielectric constant of the substrate.

$$L = \frac{C}{2F_r \sqrt{\epsilon_{\text{reff}}}} - 2\Delta l \quad (2)$$

$$\Delta l = 0.412h \frac{(\epsilon_{\text{reff}} + 0.03)(w + 0.26h)}{(\epsilon_{\text{reff}} - 0.258)(w + 0.8h)}$$

The dielectric constant of the substrate is:

$$\epsilon_{\text{reff}} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left[1 + \left(\frac{12h}{w} \right) \right]^{-1/2} \quad (3)$$

These models are very small in size and easy to fabricate. The step-by-step design procedure of creating an ESA is shown in Figure 1. The dimensions used in designing the ESA are shown in Table I.

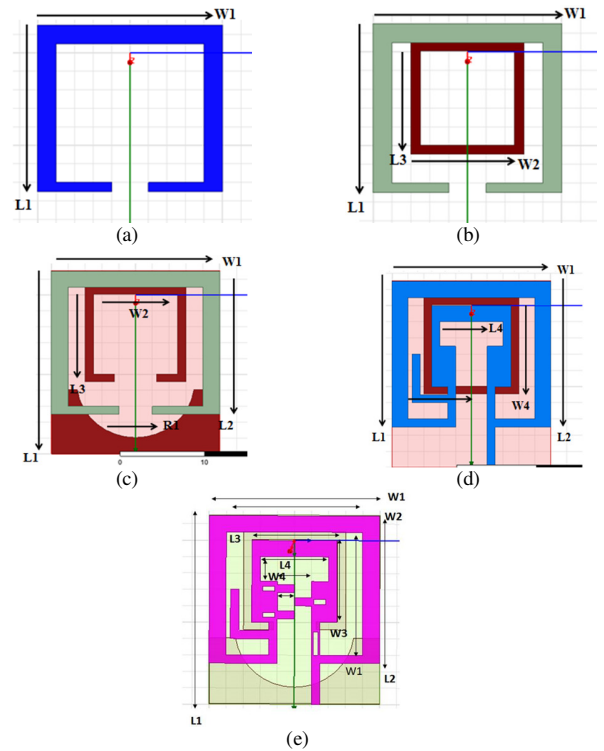


Fig. 1. Iterations of the serrated rectangular ESA.

TABLE I. DIMENSIONS OF DGS SRR

Parameter	Value (mm)	Parameter	Value (mm)
L1	20	L ₃	10
W1	18	W ₃	10
L2	16	L ₄	8
W2	15	W ₄	4
S1, S2	0.5	S3, S4	0.5
		R1	7

The prototype of the serrated rectangular ESA is displayed in Figure 2(a)-(b) and Figure 2(c) depicts the measurement setup for the antenna with MS2037C Anritsu combinational Analyzer. Figure 3 exhibits the setup for gain and radiation pattern in an anechoic chamber.

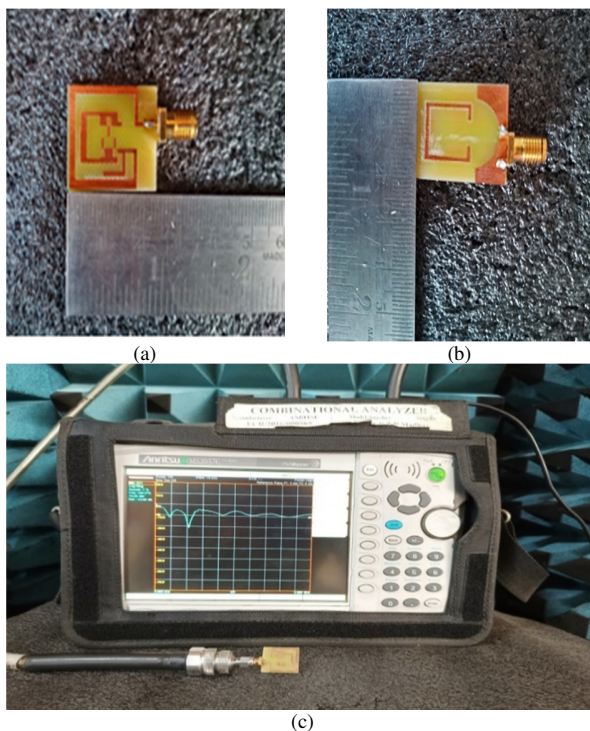


Fig. 2. Serrated Rectangular ESA prototype. (a) Top view, (b) bottom view, (c) measurement setup with the MS2037C Anritsu Combinational Analyser.

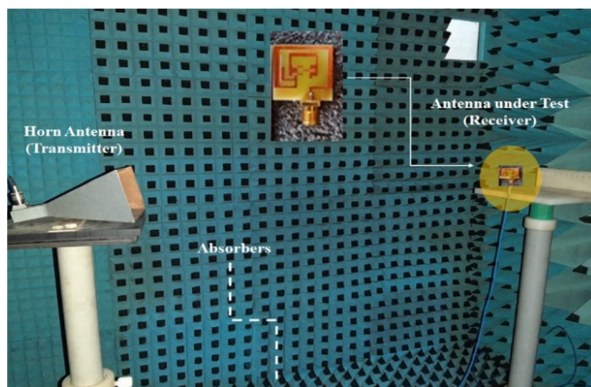


Fig. 3. Gain and radiation pattern measurement in anechoic chamber.

Figure 4 portrays the S11 of the serrated rectangular microstrip antenna. In iteration 1, the ESA operates at 1.13 GHz with a return loss of -21 dB. In iteration 2, the ESA operates at 1 GHz with S11 of -21 dB, in iteration 3 it operates at 1.0 GHz with S11 of -20 dB, in iteration 4, it operates at 0.94 GHz with S11 of -21 dB, and in iteration 5 the serrated rectangular ESA with DGS operates at 0.9 GHz with S11 = -23 dB which is suitable for GSM and RFID applications.

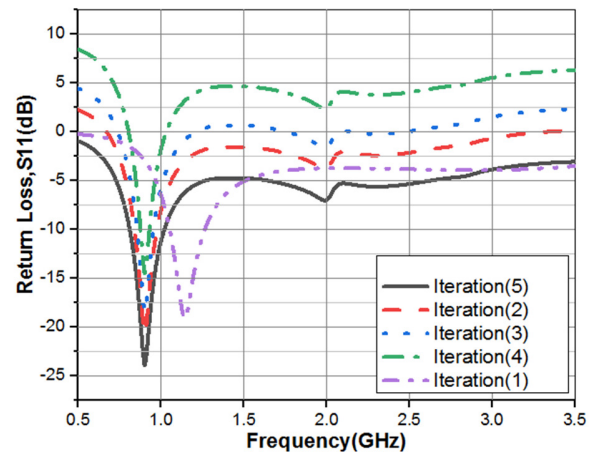


Fig. 4. Return loss of the serrated rectangular ESA for all the iterations.

III. RESULTS AND DISCUSSION

The HFSS-IE employs the Method of Moments (MoM), while Ansys Electronic Desktop's simulation of the serrated rectangular ESA is based on the finite element approach. The boundary box had dimensions 100 mm × 100 mm × 50 mm and the antenna was simulated through a lumped port. A prototype of the serrated rectangular ESA was fabricated (Figure 2) to validate the accuracy of the simulation results. The combinational analyzer MS2037C was used to measure the reflection coefficient S11. The variation of S11 with respect to frequency of the serrated rectangular ESA is displayed in Figure 5. The fabricated antenna's results demonstrate that the antenna exhibits a resonance frequency at 0.9 GHz with reflection coefficient of 23 dB and bandwidth of 0.3 GHz (0.7-1.0 GHz).

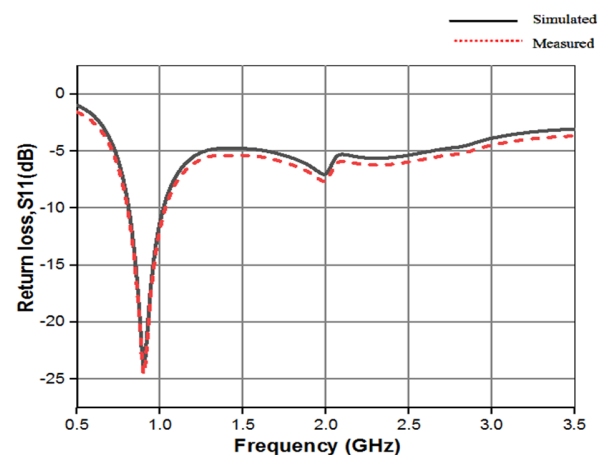


Fig. 5. S11 of the simulated and fabricated serrated rectangular ESA.

The antenna was fabricated by utilizing chemical etching. For the measured antenna, a return loss of -22 dB at 0.92 GHz was observed. The source antenna is utilized as the standard for measurement, and is connected to the receiver end of the proposed antenna to obtain the measured S11. The radiation pattern of the serrated rectangular ESA was acquired in an anechoic chamber. In an anechoic chamber (Figure 3), the radiation pattern and peak gain of the serrated rectangular ESA were computed. Figure 7 illustrates the radiation pattern at the resonant frequency of 0.9 GHz of the simulated and the manufactured antennas. The serrated rectangular ESA is shown to have high radiation pattern performance and achieves a bidirectional pattern in both the H-plane and the E-plane.

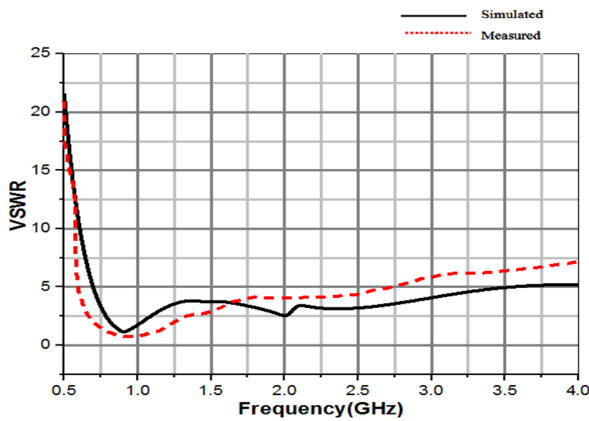


Fig. 6. VSWR of the simulated and fabricated serrated rectangular ESA.

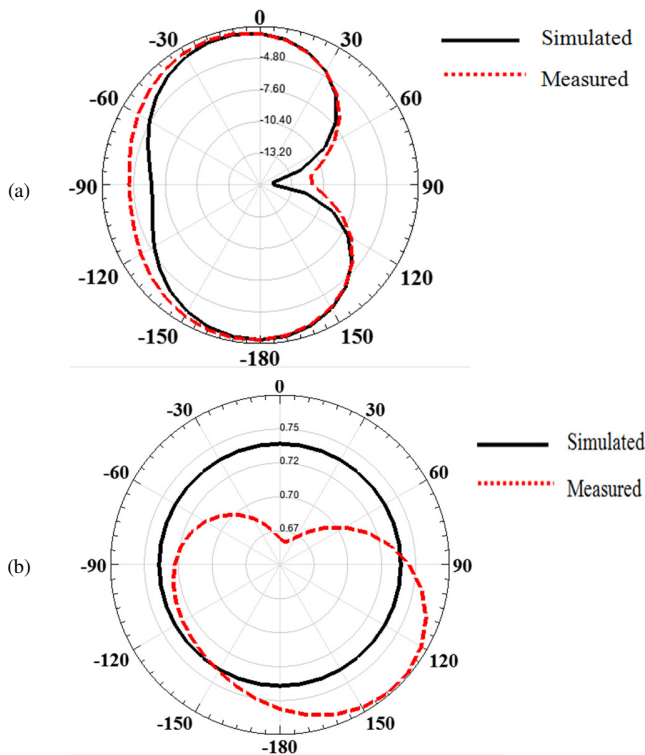


Fig. 7. Serrated rectangular ESA's radiation pattern at 900MHz: (a) E-plane, (b) H-plane.

Figure 8 presents the observed peak gain at 0.9 GHz. In terms of peak gain and radiation efficiency, it is evident that the serrated rectangular ESA works well. Figure 6 depicts the antenna's VSWR, which is less than 2 at 0.9 GHz. Figure 10 displays the surface current distribution at the resonant frequency, where the coupling between the feed line and the patch results in the maximum current density of 170 A/m². We can observe that the current density increases after changing the structure by adding slots and utilizing defective ground because the maximum current density of the rounded rectangular patch without slots and DGS is 81 A/m². At 0.9 GHz, the serrated rectangular ESA achieves a gain of 1 dB, as shown in Figure 7.

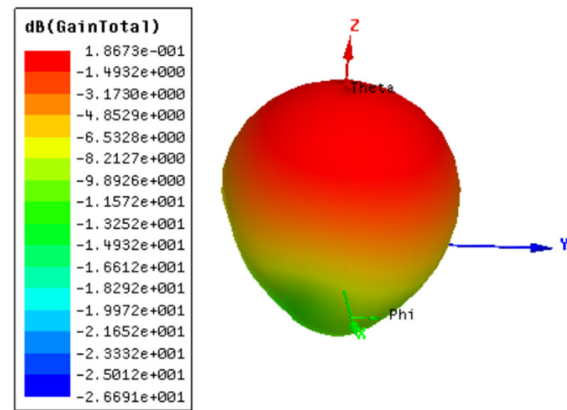


Fig. 8. 3D gain of the serrated rectangular ESA.

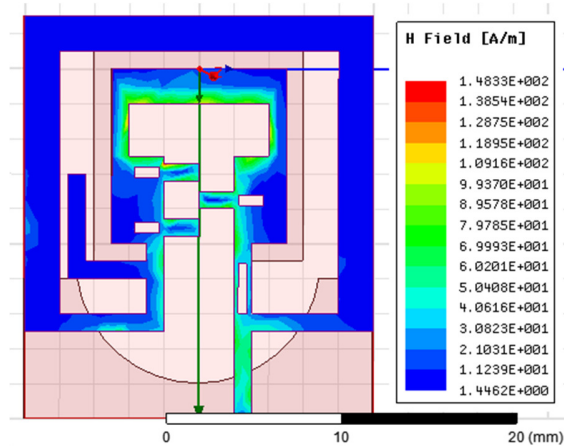


Fig. 9. H field distribution of the serrated rectangular ESA.

Table II analyzes the proposed broadband antenna for RFID applications with a few antennas that have already been published in the literature. The antenna described in [4] has a relatively small return loss compared to the proposed antenna. The large antenna in [6] has a limited impedance bandwidth. The antenna in [8] is simpler to construct and has a lower return loss and VSWR. The antenna in [13] is larger and has a complex architecture printed in Taconic RF-35. The comparison demonstrates that, in terms of impedance bandwidth, antenna size, and gain, the proposed broadband antenna has a number of advantages over the previously reported antennas.

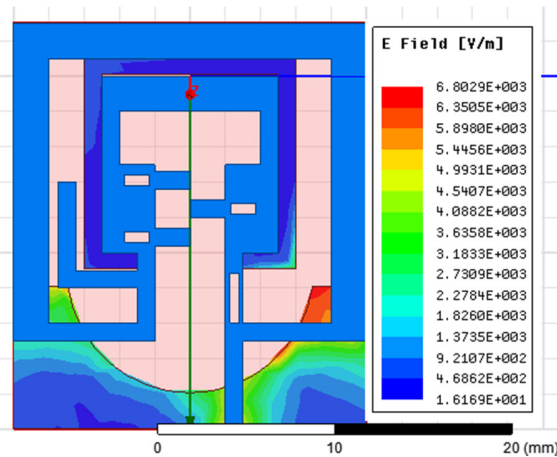


Fig. 10. E field distribution of the serrated rectangular ESA.

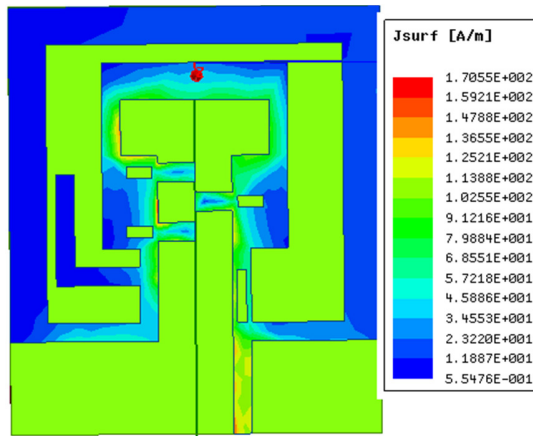


Fig. 11. Surface current distribution of the serrated rectangular ESA.

TABLE II. COMPARISON OF THE SERRATED RECTANGULAR ESA WITH ANTENNAS AVAILABLE IN THE LITERATURE

Ref	Antenna size (mm ³)	Material used	Resonant frequency	Return loss	Gain	VSWR
[4]	60×60×1.6	FR4	870 MHz	-11.1 dB	0.4 dB	1.9
[6]	70×80×3.175	Rogers/RT Duroid 5880	890 MHz	-12 dB	0.8 dB	1.8
[8]	70×70×1.6	FR4	900 MHz	-17 dB	0.2 dB	1.5
[9]	60×60×1.5	Rogers/RT Duroid 5880	870 MHz	-14 dB	0 dB	1.8
[13]	70×20×0.5	Taconic RF-35	1.7 GHz	-18 dB	0.89 dB	1.7
Proposed	20×18×1.6	FR4	900 MHz	-23 dB	1.8 dB	1.25

IV. CONCLUSION

Using the MS2037C Anritsu Combinational Analyzer, a compact, bidirectional ESA resonating at 900 MHz was designed, simulated, and tested. The serrated rectangular ESA has a simulated S11 of -23 dB at 900MHz. An ESA with a circular slot has a 300 MHz measured bandwidth for S11 <-10 dB. In terms of VSWR and S11, the simulated and measured

results match each other well. Since the radiation pattern is bidirectional in both azimuthal and elevation angles, the recommended antenna is appropriate for RFID and mobile communication applications.

REFERENCES

- [1] K. Finkenzeller, *RFID Handbook: Fundamentals and Applications in Contactless Smart Cards, Radio Frequency Identification and Near-Field Communication*, 1st ed. Chichester, West Sussex, UK ; Hoboken, NJ, USA: Wiley, 2010.
- [2] K. V. S. Rao, P. V. Nikitin, and S. F. Lam, "Antenna design for UHF RFID tags: a review and a practical application," *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 12, pp. 3870–3876, Sep. 2005, <https://doi.org/10.1109/TAP.2005.859919>.
- [3] V. P. Plessky and L. M. Reindl, "Review on SAW RFID tags," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 57, no. 3, pp. 654–668, Mar. 2010, <https://doi.org/10.1109/TUFFC.2010.1462>.
- [4] D. V. Kholodnyak, P. A. Turalchuk, A. B. Mikhailov, S. Yu. Dudnikov, and I. B. Vendik, "3D Antenna for UHF RFID Tags with Eliminated Read-Orientation Sensitivity," in *2006 European Microwave Conference*, Manchester, UK, Sep. 2006, pp. 583–586, <https://doi.org/10.1109/EUMC.2006.281459>.
- [5] J. C. Lin, Ed., *Electromagnetic Fields in Biological Systems*, 1st ed. Boca Raton, FL, USA: CRC Press, 2016.
- [6] S. Rao *et al.*, "Miniature implantable and wearable on-body antennas: towards the new era of wireless body-centric systems [antenna applications corner]," *IEEE Antennas and Propagation Magazine*, vol. 56, no. 1, pp. 271–291, Oct. 2014, <https://doi.org/10.1109/MAP.2014.6821799>.
- [7] C. A. Balanis, *Antenna Theory: Analysis and Design*, 3rd ed. Hoboken, NJ, USA: Wiley, 2005.
- [8] D. M. Dopkin, S. M. Weigand, and N. Iye, "Segmented Magnetic Antennas for Near-field UHF RFID," *Microwave Journal*, vol. 50, no. 6, Jun. 2007, Art. no. 96.
- [9] A. L. Popov, O. G. Vendik, and N. A. Zubova, "Magnetic field intensity in near field zone of loop antenna for RFID systems," *Technical Physics Letters*, vol. 36, no. 10, pp. 882–884, Oct. 2010, <https://doi.org/10.1134/S1063785010100032>.
- [10] O. G. Vendik and I. A. Pakhomov, "Electric-and magnetic-field strengths in the Fresnel zone of a microradiator formed by an electric and a magnetic dipole," *Technical Physics*, vol. 50, no. 11, pp. 1479–1484, Nov. 2005, <https://doi.org/10.1134/1.2131958>.
- [11] A. Shvetsov *et al.*, "Choice of quartz cut for sensitive wireless SAW temperature sensor," in *2014 IEEE International Ultrasonics Symposium*, Chicago, IL, USA, Sep. 2014, pp. 1505–1508, <https://doi.org/10.1109/ULTSYM.2014.0372>.
- [12] L. J. Chu, "Physical Limitations of Omni-Directional Antennas," *Journal of Applied Physics*, vol. 19, no. 12, pp. 1163–1175, Dec. 1948, <https://doi.org/10.1063/1.1715038>.
- [13] P. A. Turalchuk, D. V. Kholodnyak, and O. G. Vendik, "A novel low-profile antenna with hemispherical coverage suitable for wireless and mobile communications applications," in *2008 Loughborough Antennas and Propagation Conference*, Loughborough, UK, Mar. 2008, pp. 337–340, <https://doi.org/10.1109/LAPC.2008.4516935>.
- [14] I. Govardhani *et al.*, "Design of high directional crossed dipole antenna with metallic sheets for UHF and VHF applications," *International Journal of Engineering & Technology*, vol. 7, no. 1.5, pp. 42–50, 2018, <https://doi.org/10.14419/ijet.v7i1.5.9120>.
- [15] G. Imamdi, M. Narayan, A. Navya, and A. Roja, "Reflector array antenna design at millimetric (mm) band for on the move applications," *ARNP Journal of Engineering and Applied Sciences*, vol. 13, no. 1, pp. 352–359, Jan. 2018.
- [16] G. Immadi *et al.*, "Analysis of substrateintegrated frequency selective surface antenna for IoT applications," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 2, pp. 875–881, May 2020, <https://doi.org/10.11591/ijeecs.v18.i2.pp875-881>.

- [17] M. N. Kumar *et al.*, "Analysis of a low-profile, dual band patch antenna for wireless applications," *AIMS Electronics and Electrical Engineering*, vol. 7, no. 2, pp. 171–186, 2023, <https://doi.org/10.3934/electreng.2023010>.
- [18] K. H. Reddy, M. V. Narayana, G. Immadi, P. Satyanarayana, K. Rajkamal, and A. Navya, "A Low-profile Electrically Small Antenna with a Circular Slot for Global Positioning System Applications," *Progress In Electromagnetics Research C*, vol. 133, pp. 27–38, 2023, <https://doi.org/10.2528/PIERC23021601>.
- [19] H. R. Katireddy, M. V. Narayana, and G. Immadi, "Innovative Design and Analysis of an Electrically Small Reconfigurable Antenna for GPS and Blue Tooth Applications," *Engineering, Technology & Applied Science Research*, vol. 11, no. 5, pp. 7684–7688, Oct. 2021, <https://doi.org/10.48084/etasr.4465>.
- [20] M. O. Dwairi, "Increasing Gain Evaluation of 2×1 and 2×2 MIMO Microstrip Antennas," *Engineering, Technology & Applied Science Research*, vol. 11, no. 5, pp. 7531–7535, Oct. 2021, <https://doi.org/10.48084/etasr.4305>.
- [21] S. Sarade and S. D. Ruikar, "Development of a Wide Bandwidth Massive Eight Dissimilar Radiating Element Multiband MIMO Antenna for mm-Wave Application," *Engineering, Technology & Applied Science Research*, vol. 12, no. 5, pp. 9166–9171, Oct. 2022, <https://doi.org/10.48084/etasr.5133>.
- [22] M. Venkateswara Rao, B. T. P. Madhav, J. Krishna, Y. Usha Devi, T. Anilkumar, and B. Prudhvi Nadh, "CSRR-loaded T-shaped MIMO antenna for 5G cellular networks and vehicular communications," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 29, no. 8, 2019, Art. no. e21799, <https://doi.org/10.1002/mmce.21799>.
- [23] K. V. Vineetha, P. R. Kumar, A. N. Babu, J. B. Naik, B. T. P. Madhav, and S. Das, "Investigations on Complementary Split Ring Resonator (CSRR) array integrated printed conformal band pass filters for modern wireless communication applications," *Journal of Instrumentation*, vol. 17, no. 10, Jul. 2022, Art. no. P10043, <https://doi.org/10.1088/1748-0221/17/10/P10043>.
- [24] M. Najumunnisa *et al.*, "A Metamaterial Inspired AMC Backed Dual Band Antenna for ISM and RFID Applications," *Sensors*, vol. 22, no. 20, Jan. 2022, Art. no. 8065, <https://doi.org/10.3390/s22208065>.
- [25] S.-M. Chiang, E.-H. Lim, P.-S. Chee, Y.-H. Lee, and F.-L. Bong, "Dipolar Tag Antenna With a Top-Loading Inductive Channel With Broad Range Frequency Tuning Capability," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 3, pp. 1653–1662, Mar. 2022, <https://doi.org/10.1109/TAP.2021.3111093>.
- [26] M. Murugesh, E.-H. Lim, P.-S. Chee, and F.-L. Bong, "Complementarily Coupled C-Shaped Microstrip Patches With Wide-Range Frequency Tuning Capability for Metal-Applicable UHF RFID Tag Design," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 12, pp. 11548–11558, Sep. 2022, <https://doi.org/10.1109/TAP.2022.3209733>.
- [27] T. Althobaiti, A. Sharif, J. Ouyang, N. Ramzan, and Q. H. Abbasi, "Planar Pyramid Shaped UHF RFID Tag Antenna With Polarisation Diversity for IoT Applications Using Characteristics Mode Analysis," *IEEE Access*, vol. 8, pp. 103684–103696, 2020, <https://doi.org/10.1109/ACCESS.2020.2999256>.
- [28] W.-H. Ng, E.-H. Lim, F.-L. Bong, and B.-K. Chung, "Compact Folded Crossed-Dipole for On-Metal Polarization Diversity UHF Tag," *IEEE Journal of Radio Frequency Identification*, vol. 4, no. 2, pp. 115–123, Jun. 2020, <https://doi.org/10.1109/JRFID.2020.2965282>.
- [29] N. Ambati, G. Immadi, M. V. Narayana, K. R. Baredy, M. S. Prapurna, and J. Yanapu, "Parametric Analysis of the Defected Ground Structure-Based Hairpin Band Pass Filter for VSAT System on Chip Applications," *Engineering, Technology & Applied Science Research*, vol. 11, no. 6, pp. 7892–7896, Dec. 2021, <https://doi.org/10.48084/etasr.4495>.

Investigating the Impact of Domestic Sewage on Asphalt Concrete Pavement Strength

Afzal Ahmed

Department of Urban and Infrastructure Engineering, NED University of Engineering and Technology, Pakistan
afzalahmed@neduet.edu.pk

Sajjad Ali

Department of Civil Engineering, NED University of Engineering and Technology, Pakistan
sajjad.ali@cloud.neduet.edu.pk

Ashar Ahmed

Department of Urban and Infrastructure Engineering, NED University of Engineering and Technology, Pakistan
aahmed@cloud.neduet.edu.pk (corresponding author)

Farah Khan

Department of Urban and Infrastructure Engineering, NED University of Engineering and Technology, Pakistan
farah.iftikhark@gmail.com

Received: 10 December 2023 | Revised: 11 January 2024 and 16 February 2024 | Accepted: 18 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6736>

ABSTRACT

This study evaluates the impact of exposing asphalt pavement to sewage and fresh water. In total, 87 samples were prepared, where half of them were immersed in sewage and the others were immersed in freshwater. The Marshall mix design method was adopted for the preparation of samples. Three immersed samples were tested for stability and flow every 24 hours in both fresh water and sewage, comparing the results with a control sample. The samples immersed in fresh water lost their stability and flow after 11 days, while the ones immersed in sewage water lost their stability and flow after 9 days. Furthermore, the loss in stability for samples immersed in fresh water and sewage after 14 days was found to be 38.8 and 55.6%, respectively. The results revealed that sewage water affects asphalt concrete pavement more severely than freshwater. Finally, it was concluded that proper drainage and adequate supplemental sewerage systems are necessary to maintain the desired strength of the pavement throughout its design life.

Keywords: pavement deterioration; Marshall mix design; sewerage overflows; potholes; drainage system

I. INTRODUCTION

Pavements are engineered structures constructed to facilitate the movement of vehicles. With time, pavements deteriorate due to various reasons, including traffic loading, weather, and other external factors. In addition to these factors, one of the primary causes of distress in Hot Mix Asphalt (HMA) is moisture susceptibility and water ponding [1-2]. The accumulation of water on pavements for more than 24 hours results in pavement deterioration, making the road less efficient by decreasing speed and increasing traffic congestion and accidents [3-4]. A cross slope of about 1.5 to 2.0% is used to reduce the chance of possible water accumulation on the pavement and to facilitate water flow through drainage systems

[5-6]. A proper and well-maintained drainage system can increase pavement life by approximately 50% [7-8].

Underdeveloped and developing countries are mainly affected by water pond issues due to improper drainage systems and sewage overflows. Drainage is often blocked due to poor maintenance, leading to water accumulation on the road surface [4, 9]. Poor drainage systems also contribute to rainwater ponding, especially during the monsoon season [10-11]. Water ponding has harmful effects on road pavements, whereas the intensity of the damage depends on the type of rainfall. Heavy pouring for a continuous period results in pavement disintegration and pothole creation [12]. In addition to rainwater, pavements are also prone to sewage water that contains some amount of soaps, shampoos, toilet cleaning

detergents, etc., and can cause emulsification, while a high acid content can cause pH instability [13-14]. Furthermore, toxic chemicals from industrial zones can infiltrate sewage water and flow through roads in the event of an overflow of the sewer system [15-16].

Moisture damage leads to loss of stiffness and strength due to inundation by water on pavements. The presence of moisture severely affects asphaltic pavements by making the asphalt mastic lose its cohesion, thus damaging the bond between the aggregate and the asphalt and disintegrating the aggregates [17-19]. Moisture creates various problems on the pavement, such as fatigue, potholes, rutting, and reduction in compressive strength [20-21]. Aggregate gradation plays an important role in the formation of adhesive bonds between mineral aggregates and the asphalt binder. The loss of adhesion results in more serious damage to the pavement, such as stripping of asphalt [22-23].

Aggregates are mechanically interlocked when asphalt enters the pores and the irregularities present on the aggregate surface. As a consequence, the aggregates are coated and mechanical forces are formed after the mix cools [24]. This mechanical bonding is highly affected by the physical properties of the aggregate in use, that is, an aggregate with a rough texture as well as more pores and angularity can bring about stronger bonding [25]. In [26], the effects of aggregate gradation on moisture damage on pavements were evaluated. Tests were carried out on three different grades of crushed limestone aggregate that were used to prepare the specimens. Stability and tensile strength tests were conducted on the pavement to assess moisture susceptibility, according to the AASHTO T-283. The outcomes showed that the gradation significantly affected the performance of the pavement against moisture. The chemical composition of the water to which the pavement is exposed also influences the damage. In [27], pavement damage due to exposure to seawater was evaluated. Asphalt Cement (AC) of grade 60/70 and crushed limestone were utilized along with hydrated lime as a filler to stiffen the asphalt binder. Marshall specimens were prepared and immersed into seawater, according to AASHTO T-283. The results of the Marshall stability test revealed that the specimens submerged in seawater had lower stability. In [28], the role of the chemical properties of water, such as pH and salt concentration, in the damage to asphalt pavements was investigated. The specimens were prepared employing grade 60/70 asphalt and then immersed into four different types of water solution, including acidic, alkaline, distilled, and saline water. The testing method involved immersing 6 grams of asphalt binder into different volumes and pH of water solutions for 7 days. A visual observation of the immersed samples disclosed that the surfaces became rough and micropits were observed over time. The study concluded that asphalt damage was highest for the alkaline solution, followed by the acidic solution, and then the saline solution. However, this study did not explore the effects of exposure to moisture on the mechanical and physical properties of the asphalt.

Moisture-damaged pavements deteriorate with time, as they withstand traffic and other loads while pounded by water [29]. In [30], the premature failure of asphalt pavements in South

China during the humid and rainy seasons was examined. Field investigations observed the mechanism that occurs on pavements in service by traffic and axial loading in the presence of moisture. Nine highways and expressways were studied, detaching cores for laboratory tests. The results showed that water infiltrates the pavement from the push it receives from the traffic load and, in return, causes the asphalt film to strip. The stripped asphalt then moves upwards due to dynamic hydraulic flow and gathers on the road surface as spots. While the asphalt immigrates upward, the aggregates disintegrate as an outcome of loss of adhesion to the asphalt binder. Also, owing to poor interior drainage conditions, a pumping action occurs, causing the base cement to become a slurry and after that make its way to the surface through the voids in the permeable surface layer. Cracks are generated as a consequence of settlement due to pumping, and deformation arises to a considerable degree.

In Karachi, the most populous city in Pakistan, the sewer system is often blocked with garbage due to poor infrastructure and lack of municipal services, inducing sewage overflows. This problem is combined with the lack of a stormwater drainage system, causing water ponds on the roads for prolonged periods after rains [31-32]. It has been observed that road segments with sewage ponding are damaged earlier than road segments without. Only 7% of the roads are in good condition compared to the 26% of roads in bad condition [33]. This study aimed to evaluate the hypothesis that sewage creates greater damage to asphalt pavements and significantly reduces their serviceability.

II. RESEARCH GAP AND CONTRIBUTION

This study aimed to investigate the damage caused to the asphalt concrete mix due to prolonged exposure of pavements to rain and sewage. There is no study in the existing literature that quantifies the impact of sewage on pavements. In [27], the impact of freshwater and seawater on pavements was investigated by immersing samples for 14 days, carrying out Marshall stability and flow tests only on selected days (1, 3, 7, and 14). In [28], the influence of acidic, alkaline, distilled, and saline water was explored by immersing samples for 7 days. However, this study did not perform any mechanical tests and the deterioration was only visually observed. All in all, this study examines the impact on pavement due to prolonged exposure to sewage using Marshall stability and flow measurements.

III. EXPERIMENTAL FRAMEWORK

Figure 1 describes the research framework of this study.

A. Material Testing and Selection

Physical and mechanical tests were performed for the selected aggregates and asphalt. Aggregates and asphalt were tested according to ASTM/BS/AASHTO standards, while the Marshall mix design method was used to examine the behavior of HMA.

1) Testing of Aggregates

Crushed limestone was selected as the aggregate for this experimental program. The aggregates were sieved in the

required proportions according to the specifications defined by the National Highway Authority (NHA) in Pakistan [34]. Table I displays the physical properties of aggregates with relevant standards and limits. All physical properties of the aggregates were under the limitations specified by NHA. The sieving of aggregates was performed using the NHA specification, as illustrated in Table II. The aggregates were thoroughly washed, oven-dried, and left at ambient temperature overnight before use. All samples followed this gradation, and the average values were taken for the sieves with the specified ranges.

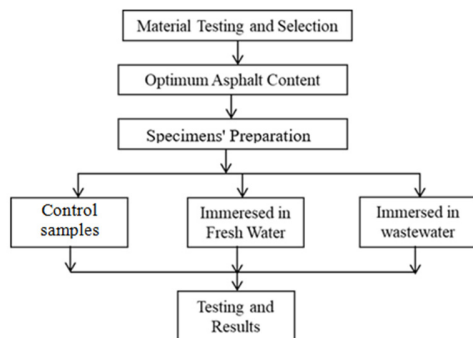


Fig. 1. Research method to evaluate the effect of water ponding.

TABLE I. SUMMARY OF AGGREGATES TESTING

Properties (unit)	Standard	Test result	NHA Specification
Specific Gravity	ASTM C127-07 [35]	2.50	≤ 2.60
Absorption (%)	ASTM C127-07 [35]	0.68	≤ 2.0
LA Abrasion (%)	ASTM C131-06 [36]	2.32	≤ 40
Impact Value (AIV) (%)	BS 812-112 [37]	15.82	-
Crushing Value (ACV) (%)	BS 812-110 [38]	20.52	-

TABLE II. SIEVE SIZES AND PERCENTAGE PASSING AS RECOMMENDED BY NHA

Sieve Size	Passing (%)
1"	100
3/4"	90-100
3/8"	51-75
#4	33-54
#8	23-35
#16	5-12
#200	2-8
Pan	0.0

2) Testing of Asphalt Binder

Conventional asphalt was used as the binder. Table III depicts the properties of the binder with their standards and specifications, which were within the limits defined by NHA.

TABLE III. SUMMARY OF ASPHALT TESTING

Properties (unit)	Standard	Observed	NHA Spec.
Penetration Grade	ASTM: D 5-86 [39]	60-70	60 - 70
Softening point (°C)	ASTM: D 36-8 [40]	43.75	-
Flashpoint (°C)	ASTM: D 92-78 [41]	180	-
Fire point (°C)	ASTM: D 92-78 [41]	300	> 232
Ductility (°C)	ASTM: D 113-86 [42]	105	> 100
Specific Gravity (-)	ASTM: D 70-97 [43]	0.99	-

B. Specimen Preparation

The Marshall mix design procedure was applied to determine the Optimum Asphalt Content (OAC). In total, 36 samples were prepared, 18 compacted and 18 uncompacted. The percentage of asphalt content ranged from 3.5 to 6.0% as per the weight of the specimens. The compacted samples were utilized to determine the volumetric properties of the mix, while the uncompacted samples were used to determine the maximum theoretical specific gravity (GMM) and then the air voids percentage (Va). OAC was determined at 4.0 to 7.0% of the air voids as per NHA specifications. Figure 2 shows the behavior of air voids in HMA with increasing asphalt. The optimum asphalt content percentage was 5.25% at 4.0% air voids.

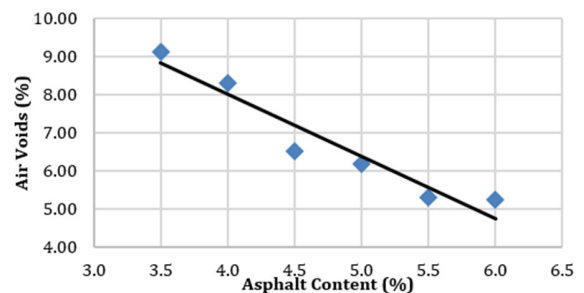


Fig. 2. Variation in air voids vs asphalt contents of mix samples.

Each specimen was prepared with 1200 g of washed and dried aggregates, which were obtained by passing through the sieves as portrayed in Table II. The samples were prepared following the standard procedure described in [44]. About 5.25% of asphalt by weight was added for the preparation of the asphalt mix. The final specimen had an approximate height of 64 mm and a diameter of 102 mm, as noticed in Figure 3.



Fig. 3. Preparation of specimens.

C. Testing Scheme

Based on the estimated OAC, the selected materials were employed to prepare the samples for the described testing scheme. In total, 87 samples were prepared for Marshall testing. Three samples were taken as control (dry) samples and were tested after 24 hours of preparation. All the other samples were divided into two groups of 42, which were kept in dry conditions for 24 hours before immersion in water. One set of samples was immersed into freshwater, while the other one was immersed into sewage. From each set, three random samples

were drawn every day until the 14th day to perform the Marshall stability test. Figure 4 (a) and (b) exhibit the samples immersed into fresh water and sewage, respectively.

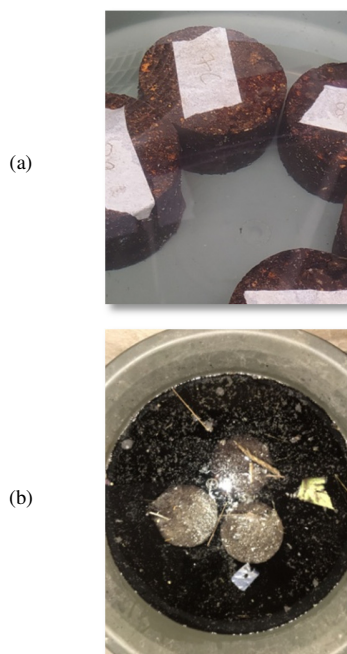


Fig. 4. Specimens immersed into (a) freshwater (a) and (b) sewage.

D. Water Ponding

Two different immersion conditions were created for this experiment. Half of the samples were immersed into fresh water, while the remaining samples were immersed into sewage, taken from one of the main sewer lines in Karachi. These conditions were created because the study area is a city where the cross-slope and longitudinal slope are almost nonexistent. As a result, fresh water and sewage remain on the roads for several days, generating a situation in which the pavement remains completely immersed. The pH of freshwater and sewage water was examined, being 6.8 and 8.1, accordingly.

E. Marshall Testing

The Marshall testing apparatus was put into service to estimate the strength parameters of all specimens, including stability (lb) and flow (in). Based on the observed values of stability and flow, the Marshall quotient, defined as stability per unit of flow, was also determined, which helps in understanding the loss of strength due to prolonged exposure to water. Figure 5 displays the Marshall testing apparatus implemented in this experiment, where the standard procedure for Marshall testing was followed [45]. For the Marshall stability and flow test, the compacted samples were immersed into water for 30 minutes at 60°C. After this, the samples were placed one by one on the breaking head of the Marshall tester.

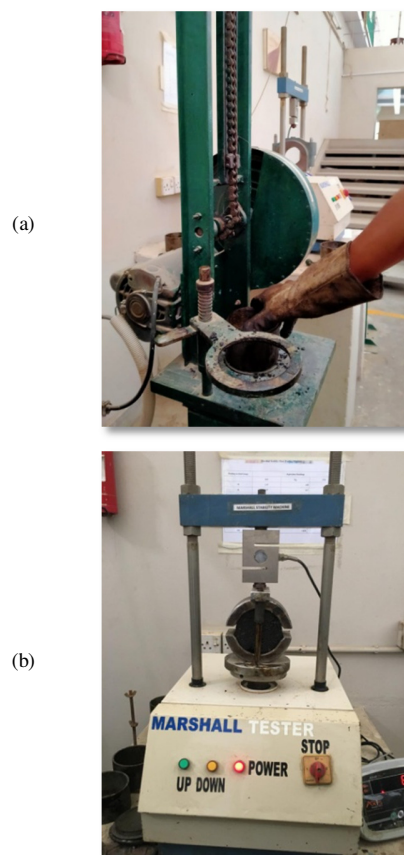


Fig. 5. Marsha (a) compacter and (b) tester.

IV. RESULTS AND DISCUSSION

The stability and flow of the immersed samples were tested every day for 14 days. All samples were kept in dry conditions for 24 hours before Marshall testing. On Day-0, after 24 hours in dry conditions, three samples were tested as control samples while all the others were immersed into water. Three samples were drawn every 24 hours in freshwater and sewage water for Marshall testing. Table IV presents the mean values for stability and flow of the three samples tested each day.

TABLE IV. RESULTS OF MARSHALL TEST OF SPECIMENS

Days	Freshwater		Sewage	
	Flow (0.01 in.)	Stability (lbs)	Flow (0.01 in.)	Stability (lbs)
Day-01	11.74	2221.9	11.91	2206.0
Day-02	12.01	2198.0	12.72	2146.8
Day-03	12.64	2163.4	13.41	1937.8
Day-04	13.18	2051.7	13.62	1930.7
Day-05	13.21	2043.4	13.56	1848.6
Day-06	13.30	1996.2	14.3	1718.5
Day-07	13.36	1984.7	14.38	1700.8
Day-08	13.49	1862.9	14.49	1695.7
Day-09	14.26	1706.4	15.61	1540.2
Day-10	14.74	1684.2	16.42	1403.2
Day-11	15.01	1601.0	17.99	1324.0
Day-12	16.11	1489.4	19.04	1212.0
Day-13	16.24	1411.0	20.04	1198.6
Day-14	16.82	1386.0	22.12	1004.2

A. Loss of Stability

The stability and flow of the control sample, which was not exposed to any kind of immersion, were 2265 lbs and 11.1 in. Figure 6 showcases the loss of stability over time for samples immersed into fresh water and sewage. The stability of these samples was compared with that of the control sample. The Marshall testing results disclosed a decrease in stability for samples immersed into fresh water and sewage. The loss of stability for samples immersed into sewage was significantly greater than that of samples immersed into freshwater. The samples were prepared for heavy traffic with 75 blows on each side. The Asphalt Institute recommends a minimum acceptable stability of 1500 lbs for pavements designed for heavy traffic, implying that a lower value would be subject to failure [46]. The control sample had significantly higher stability, demonstrating the compatibility of the samples with heavy traffic. Therefore, it can be implied that when the sample strength is reduced below 1500 lb, the pavement will not support the traffic load. The samples immersed into freshwater failed after Day-11, as they lost 34.24% of their strength during this period. Samples immersed into sewage failed after Day-09, as they lost 38.04% of their stability. After 14 days, the samples immersed into freshwater lost their stability by 38.81% while those exposed to sewage lost their stability by 55.66%. Therefore, samples exposed to sewage water had a 28% reduced stability on the 14th day than samples exposed to fresh water. Comparison of stability loss with time also revealed that the loss stability rate for samples immersed in sewage was higher than the other set of samples.

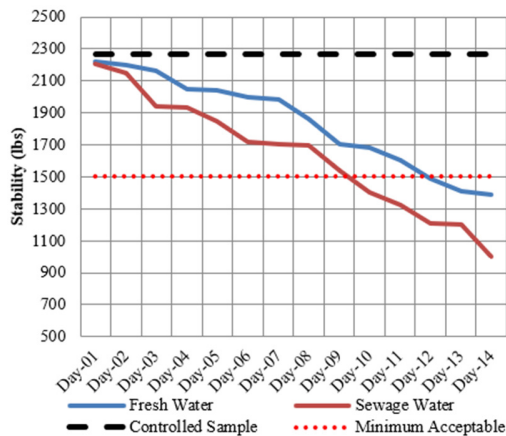


Fig. 6. Loss in stability of samples with time.

The findings exhibit a similar trend to other studies. In [27], the use of seawater displayed a lower pH value in the Marshall stability test, concluding that lower pH and salt concentration in water pose a high disintegration risk for asphaltic pavements. In [28], the alkaline solution affected the asphaltic pavement samples more severely compared to acidic and saline solutions. Therefore, it can be deduced that sewage has a pH in the alkaline range and thus has a more detrimental effect on the pavement compared to rainwater.

B. Loss in Flow

Flow is the measure of the maximum deformation in the asphalt mix when subjected to the maximum load. The flow values were also observed during the Marshall test. Figure 7 portrays the observed flow values for the samples. The Asphalt Institute recommends a range of 8-16 (0.01 in.) as an acceptable value for flow on pavements subjected to heavy traffic [46]. The failure in terms of flow follows a similar pattern to that of stability. Samples immersed into freshwater crossed the acceptable maximum value of flow after Day-11, whereas samples exposed to sewage crossed this limit after the 9th day. The flow of the specimens exposed to fresh water increased by 51.5% after 14 days, while the flow of samples immersed into sewage rose by almost 100%. The change in flow for samples exposed to sewage is significantly higher than for those immersed into freshwater. The comparison of the change in flow and stability during the observation period indicates that the change in flow was more than the change in stability. The samples lost their stability by 38.81%, whereas flow change was 51.5% for the same period. Similarly, the samples placed in sewage lost their stability by 55.66% and increased their flow by 100%.

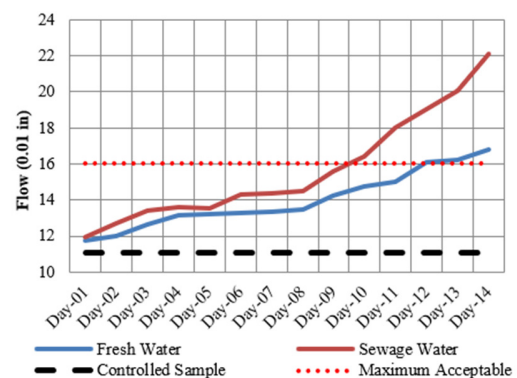


Fig. 7. Flow comparison for samples immersed into fresh and sewage water.

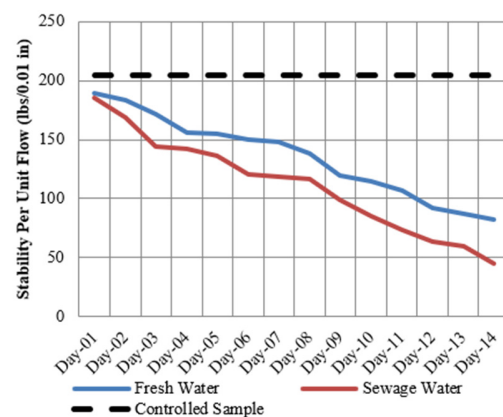


Fig. 8. Comparison of stability per unit flow for the samples immersed into fresh water and sewage.

C. Stability per Unit Flow

Stability per unit flow represents the overall capacity of load handling by the asphalt pavement, combining the effect of loss in stability and increase in flow. Figure 8 depicts the change in stability per unit flow over the observation period. Compared with stability per unit flow measured for the control sample, the samples exposed to freshwater and sewage experienced a reduction of 59.61 and 77.75%, respectively.

V. CONCLUSIONS AND RECOMMENDATIONS

This study highlighted one of the major problems faced by the transport network and commuters in Pakistan and other developing countries. The condition of the pavement of the road network in Karachi and other parts of the developing world is generally deteriorated. One of the contributing factors to the condition of pavements is exposure to water. The impact of water-exposed pavements was evaluated by immersing Marshall specimens under controlled conditions. The Marshall methodology has been widely used in previous studies for measuring flow and loss of stability in pavements [47-50]. The specimens were immersed into fresh water and sewage for 14 days. Three samples were used after each day to perform Marshall stability and flow tests. The test results showed that the stability of the samples fell below the minimum required value, implying that continuous exposure of pavement to ponded water would lead to rapid damage.

Poor rainwater drainage systems and improper sewerage cause excessive rainwater and sewage ponding, especially on local streets and collector roads. Severe damage to the pavement produces potholes that cause traffic congestion, delays, and road accidents. This study evaluated the impact of exposure to water on pavement conditions. The main contribution of this study is the quantification of the influence of sewage on the loss of strength of asphalt mix, in addition to quantifying the effect of extended exposure to water, such as water ponding and runoff. The samples immersed into freshwater had less stability than the minimum required after 11 days of immersion, while the samples immersed into sewage were found to have less stability than the minimum required after 9 days of immersion. The results reveal that sewage causes more damage to the pavement compared to fresh water. This study demonstrated that prolonged exposure of pavements to water can severely damage roads. Therefore, the importance of having a proper rainwater drainage system and efficient sewerage is further highlighted to ensure a road network with pavements that provide good riding quality. By reducing damage to pavements due to water, the cost of frequent road repairs could be decreased along with eliminating excessive delays and road accidents.

This study was limited to measuring the deterioration of the asphalt mix due to reduced stability and increased flow. The testing of such samples can be extended to include more detailed analysis using rutting susceptibility tests and fatigue analysis. Furthermore, the testing scheme should investigate extracted cores from road segments ponded with fresh water and sewage and test real-life conditions.

ACKNOWLEDGMENT

The authors express their gratitude to research assistants Ms. Madiha Rehman, Mr. Abdul Rafay, Mr. Ghusarib Ahmed, and Mr. Muhammad Murtaza for their help in the experimental framework.

REFERENCES

- [1] T. M. Torgyia, I. V. Terparse, and I. E. Terlumun, "The effects of poor drainage system on road pavement: A review," *International Journal for innovative research in multidisciplinary field*, vol. 2, no. 8, pp. 216–223, 2016.
- [2] F. M. Fernandes and J. Pais, "Assessment of moisture in road pavements," in *Proceedings of the 15th International Conference on Ground Penetrating Radar*, Brussels, Belgium, Jun. 2014, pp. 909–912, <https://doi.org/10.1109/ICGPR.2014.6970558>.
- [3] M. M. Zumrawi, "The impacts of poor drainage on road performance in Khartoum," *International Journal of Multidisciplinary and Scientific Emerging Research*, vol. 3, no. 1, pp. 901–907, 2014.
- [4] G. J. Hearn, "Project scoping study for Sierra Leone and Liberia," Africa Community Access Partnership, Scoping Study Report, Sep. 2016.
- [5] H. A. Mahdy and S. A. Mostafa, "Minimum acceptable cross slopes of asphalt roads for drainage consideration," in *The International Conference on Civil and Architecture Engineering*, 2012, vol. 9.
- [6] G. M. Abdel-Aal, M. R. Fahmy, I. Fathy, and A. A. Fathy, "Effect of road's slope on the efficiency of the rain storm drainage networks," *International Journal of Civil Engineering and Technology*, vol. 10, no. 4, pp. 321–332, Apr. 2019.
- [7] S. Rokade, P. K. Agarwal, and R. Shrivastava, "Drainage and flexible pavement performance," *International Journal of Engineering Science and Technology*, vol. 4, no. 4, pp. 1308–1311, 2012.
- [8] P. K. Azad, Y. U. Shah, S. S. Jain, and M. Parida, "Estimating the benefits of improved drainage on pavement performance," *International Journal of Engineering, Science and Technology*, vol. 5, no. 3, pp. 78–92, 2013, <https://doi.org/10.4314/ijest.v5i3.7>.
- [9] R. Diriba, "Causes and Effects of Poor Drainage System on the Asphalt Pavement Layer: a Case Study in Agaro to Bedele Road,," Jimma Institute of Technology, Jimma, Ethiopia, 2016.
- [10] S. Raina and R. Gupta, "A Review Paper on Study on Strengthening and Drainage of Flexible Pavements," *International Research Journal of Engineering and Technology*, vol. 5, no. 8, pp. 1154–1156, Sep. 2018.
- [11] "Road Deterioration in Developing Countries," World Bank, Policy Study, 1988.
- [12] P. V. Hung and N. T. Thanh, "Asphalt Pavement Performance in Extreme Weather Events in Tropical Country," *International Journal of Engineering and Technical Research*, vol. 8, no. 5, pp. 76–82, May 2019, <https://doi.org/10.17577/IJERTV8IS050057>.
- [13] S. Feng, W. Wang, H. Tan, M. Xue, and K. Hög, "Effects of acidic water on hydro-mechanical properties of asphalt concrete," *Case Studies in Construction Materials*, vol. 13, Dec. 2020, Art. no. e00445, <https://doi.org/10.1016/j.cscm.2020.e00445>.
- [14] P. J. T. M. van Puijenbroek, A. H. W. Beusen, and A. F. Bouwman, "Global nitrogen and phosphorus in urban waste water based on the Shared Socio-economic pathways," *Journal of Environmental Management*, vol. 231, pp. 446–456, Feb. 2019, <https://doi.org/10.1016/j.jenvman.2018.10.048>.
- [15] A. L. L. Silveira, "Problems of modern urban drainage in developing countries," *Water Science and Technology*, vol. 45, no. 7, pp. 31–40, Apr. 2002, <https://doi.org/10.2166/wst.2002.0114>.
- [16] A. O. Sojobi and T. Zayed, "Impact of sewer overflow on public health: A comprehensive scientometric analysis and systematic review," *Environmental Research*, vol. 203, Jan. 2022, Art. no. 111609, <https://doi.org/10.1016/j.envres.2021.111609>.
- [17] N. Kringos and A. Scarpas, "Raveling of Asphaltic Mixes Due to Water Damage: Computational Identification of Controlling Parameters," *Transportation Research Record*, vol. 1929, no. 1, pp. 79–87, Jan. 2005, <https://doi.org/10.1177/0361198105192900110>.

- [18] A. Yilmaz and Ş. Sargin, "Water effect on deteriorations of asphalt pavements," *TOJSAT*, vol. 2, no. 1, pp. 1–6, 2012.
- [19] Q. Lu, "Investigation of Conditions for Moisture Damage in Asphalt Concrete and Appropriate Laboratory Test Methods," University of California Transportation Center, University of California Transportation Center, Working Paper, Jan. 2005.
- [20] R. Moraes, R. Velasquez, and H. U. Bahia, "Measuring the Effect of Moisture on Asphalt–Aggregate Bond with the Bitumen Bond Strength Test," *Transportation Research Record*, vol. 2209, no. 1, pp. 70–81, Jan. 2011, <https://doi.org/10.3141/2209-09>.
- [21] D.-W. Cho and K. Kim, "The mechanisms of moisture damage in asphalt pavement by applying chemistry aspects," *KSCE Journal of Civil Engineering*, vol. 14, no. 3, pp. 333–341, May 2010, <https://doi.org/10.1007/s12205-010-0333-z>.
- [22] A. R. Copeland, J. Youtcheff, and A. Shenoy, "Moisture Sensitivity of Modified Asphalt Binders: Factors Influencing Bond Strength," *Transportation Research Record*, vol. 1998, no. 1, pp. 18–28, Jan. 2007, <https://doi.org/10.3141/1998-03>.
- [23] A. Bhasin, D. N. Little, K. L. Vasconcelos, and E. Masad, "Surface Free Energy to Identify Moisture Sensitivity of Materials for Asphalt Mixes," *Transportation Research Record*, vol. 2001, no. 1, pp. 37–45, Jan. 2007, <https://doi.org/10.3141/2001-05>.
- [24] D. E. Packham and C. Johnston, "Mechanical adhesion: were McBain and Hopkins right? An empirical study," *International Journal of Adhesion and Adhesives*, vol. 14, no. 2, pp. 131–135, Apr. 1994, [https://doi.org/10.1016/0143-7496\(94\)90008-6](https://doi.org/10.1016/0143-7496(94)90008-6).
- [25] B. M. Kiggundu, F. L. Roberts, and Auburn University. National Center for Asphalt Technology, "Stripping in HMA Mixtures: State-of-the-Art and Critical Review of Test Methods," National Center for Asphalt Technology, Sep. 1988.
- [26] H. Habeeb, S. Chandra, and Y. Nashaat, "Estimation of moisture damage and permanent deformation in asphalt mixture from aggregate gradation," *KSCE Journal of Civil Engineering*, vol. 18, no. 6, pp. 1655–1663, Sep. 2014, <https://doi.org/10.1007/s12205-014-1356-7>.
- [27] A. E. A. E. M. Behiry, "Laboratory evaluation of resistance to moisture damage in asphalt mixtures," *Ain Shams Engineering Journal*, vol. 4, no. 3, pp. 351–363, Sep. 2013, <https://doi.org/10.1016/j.asej.2012.10.009>.
- [28] L. Pang, X. Zhang, S. Wu, Y. Ye, and Y. Li, "Influence of Water Solute Exposure on the Chemical Evolution and Rheological Properties of Asphalt," *Materials*, vol. 11, no. 6, Jun. 2018, Art. no. 983, <https://doi.org/10.3390/ma11060983>.
- [29] S. Im, S. Sebesta, and Y. Rew, "Methods of Rehabilitating Pavements with Moisture Damaged Asphalt Layers," Texas A&M Transportation Institute, FHWA/TX-18/0-6926-R1, Nov. 2018.
- [30] C. Q. Pang, "Research on the Moisture Damage Occurred on the Asphalt Pavement," *Applied Mechanics and Materials*, vol. 178–181, pp. 1117–1124, 2012, <https://doi.org/10.4028/www.scientific.net/AMM.178-181.1117>.
- [31] W. A. Butt, "Decades-old issue of poor drainage haunts Gujrat," *dawn.com*, 07:07:42+05:00, <https://www.dawn.com/news/1575362>.
- [32] A. Hasan, "Urban flooding: the case of Karachi," <https://www.iied.org/urban-flooding-case-karachi>.
- [33] "Sindh Provincial Road Improvement Project," Asian Development Bank, Sovereign Projects 46377–002, Oct. 2015. [Online]. Available: <https://www.adb.org/projects/46377-002/main>.
- [34] "NHA Code." National Highway Authority of Pakistan, 2005.
- [35] "ASTM C127-07: Standard Test Method for Density, Relative Density (Specific Gravity), and Absorption of Coarse Aggregate." ASTM, 2007, <https://doi.org/10.1520/C0127-07>.
- [36] "ASTM C131-06: Standard Test Method for Resistance to Degradation of Small-Size Coarse Aggregate by Abrasion and Impact in the Los Angeles Machine." ASTM, 2006, <https://doi.org/10.1520/C0131-06>.
- [37] "BS 812-112:1990 Testing aggregates Method for determination of aggregate impact value (AIV)." European Standard, 1990.
- [38] "BS 812-110:1990 Testing aggregates Methods for determination of aggregate crushing value (ACV)." European Standard, 1990.
- [39] "ASTM D5-06: Standard Test Method for Penetration of Bituminous Materials." ASTM, 2006, <https://doi.org/10.1520/D0005-06>.
- [40] "ASTM D36-06: Standard Test Method for Softening Point of Bitumen (Ring-and-Ball Apparatus)." ASTM, 2006, <https://doi.org/10.1520/D0036-06>.
- [41] "ASTM D92-05a: Standard Test Method for Flash and Fire Points by Cleveland Open Cup Tester." ASTM, 2005, <https://doi.org/10.1520/D0092-05A>.
- [42] "ASTM D113-99: Standard Test Method for Ductility of Bituminous Materials." ASTM, 1999, <https://doi.org/10.1520/D0113-99>.
- [43] "ASTM D70-97: Standard Test Method for Specific Gravity and Density of Semi-Solid Bituminous Materials (Pycnometer Method)." ASTM, 1999, <https://doi.org/10.1520/D0070-97>.
- [44] "ASTM D6926-20: Standard Practice for Preparation of Asphalt Mixture Specimens Using Marshall Apparatus." ASTM, 2020, <https://doi.org/10.1520/D6926-20>.
- [45] "ASTM D6927-15: Standard Test Method for Marshall Stability and Flow of Asphalt Mixtures." ASTM, 2015, <https://doi.org/10.1520/D6927-15>.
- [46] "Mix Design Methods for Asphalt Concrete and Other Hot-Mix Types," Asphalt Institute, 1996.
- [47] K. A. Kaaf and V. T. Ibeabuchi, "Marshall Asphalt Mix and Superior Performance Asphalt Mix in Oman: A Comparative Study," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12258–12263, Dec. 2023, <https://doi.org/10.48084/etasr.6206>.
- [48] H. M. A. A. Kareem and A. H. K. Albayati, "The Possibility of Minimizing Rutting Distress in Asphalt Concrete Wearing Course," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 8063–8074, Feb. 2022, <https://doi.org/10.48084/etasr.4669>.
- [49] S. S. Almasoudi and A. H. K. Albayati, "Statistical Analysis of Component Deviation from Job Mix Formula in Hot Mix Asphalt," *Engineering, Technology & Applied Science Research*, vol. 12, no. 5, pp. 9295–9301, Oct. 2022, <https://doi.org/10.48084/etasr.5225>.
- [50] F. Althoey *et al.*, "Prediction models for marshall mix parameters using bio-inspired genetic programming and deep machine learning approaches: A comparative study," *Case Studies in Construction Materials*, vol. 18, Jul. 2023, Art. no. e01774, <https://doi.org/10.1016/j.cscm.2022.e01774>.

A Novel Computational Mathematical Model for Team and Route Selection of the Emergency Response Operations

Dalvana Lopes Ribeiro

Graduate Program in Computational Modeling, Federal University of Rio Grande (FURG), Brazil
ribeiro_dalvana20@furg.br

Andre Andrade Longaray

Graduate Program in Computational Modeling, Federal University of Rio Grande (FURG), Brazil
andrelongaray@furg.br (corresponding author)

Received: 17 January 2024 | Revised: 2 March 2024 | Accepted: 6 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6926>

ABSTRACT

During the daily operations of emergency response, the decision maker is faced with the complex challenge of selecting a team and route in a short time period to respond and attend to the emergency. This study presents the combined use of the Analytic Hierarchy Process (AHP), the Élimination Et Choix Traduisant la Réalité II (ELECTRE II), and the Dijkstra algorithm to deal with such situations. First, the AHP method is implemented to rank the aspects that are most relevant to a given emergency. Subsequently, this ranking is employed in ELECTRE-II to determine which emergency response team is best prepared to provide support. In the last stage of the proposed model, regarding the geographic coordinates of the team and the emergency, the Geographic Information System (GIS) utilizes the Dijkstra algorithm to regulate the most suitable route for assistance.

Keywords-AHP; ELECTRE II; GIS; Dijkstra algorithm; emergency response operations

I. INTRODUCTION

Public safety organizations deal with the most diverse situations every day. Emergency services, such as police, fire, and health services, encounter unprecedented challenges as the nature, form, and pattern of the risks faced by society increase and change [1-2]. The main objective of incident management is to provide emergency services as quickly as possible to the affected areas [3]. To achieve this goal, numerous factors must be assessed and considered in a relatively short decision-making period. Particularly in emergency response, issues such as the number of rescuers, resources, and equipment, as well as the routes to be followed and spatial attributes, are crucial for estimating response times. In these situations, managers often need to be equipped with tools to help with planning and management and promote better assistance [4]. Several incident management tools are available, including Multi-Criteria Decision Making/Aid (MCDM/A).

MCDM/A is a tool that implements the knowledge of experts in computational mathematical procedures to model complex problems and support decision-making [5-6]. MCDM/A approaches differ from each other in terms of data aggregation, normalization, structuring, output type, and application kinds [7]. For example, in emergency operations, rescue agencies must organize staff and materials quickly and coherently according to the emergency needs. However, the

decision criteria for the distribution of emergency resources are often diverse, along with a lack of information acquisition due to the complexity of the event [8]. In this impasse, multi-criteria methods, such as AHP and ELECTRE-II, may contribute to modeling these problems.

In addition to the administrative support management process, reaching the incident site in a short time is also crucial, as it can be a determining factor in the success of the assistance provided. In this process, many factors, such as road width, speed limits, and traffic volume, affect route selection [9]. One strategy that helps decide on an ideal route is the use of the Dijkstra algorithm associated with GIS. The Dijkstra algorithm is based on graph theory, whereas its dynamics correlate with finding the path with the lowest cost from a starting to a destination node by comparing the weights of the paths to identify the most efficient pattern [10]. GIS offers powerful resources for network analysis and urban traffic network management [4]. GIS are systems for processing and storing geospatial data, vectorization, and real-time visualization of objects and events, also providing a robust set of geospatial analytical tools and strategic support for management decisions [11].

Responding to an incident is a complex task, as it involves both administrative and spatial processes. In situations entailing these characteristics, many studies have combined MCDM/A

with a GIS in several ways. In [12], the AHP method was combined with GIS to determine flood risk zones. In [13], these methods were combined to analyze fire risks in factories. In [14], a similar method was used to analyze landslides to plan support routes for an event. In [15], TOPSIS and GIS methods were combined to build a flood-support environment. Despite the great applicability of this combination, there is a scarcity in the necessity for overcoming MCDM/A, in which overcoming relationships are built between alternatives and there are computational limitations due to the large spatial dataset [16-18].

This study combines MCDM/A with GIS in a different approach, using a model that focuses on the dynamic situation of an incident call. In emergencies, the manager must quickly decide to meet the demands of the incident and determine which team and route are the most appropriate for assistance. To assist him, the AHP and ELECTRE-II methods were combined with the Dijkstra algorithm, due to the flexibility and robustness of multicriteria methods combined with the potential of GIS for spatial problems.

II. METHODOLOGY

A. Analytic Hierarchy Process (AHP)

AHP enables a combination of rational and irrational intuitive values in decision-making through a pairwise comparison approach [19-20]. The ability to model the thought processes of experts allows for the decomposition of complex problems into a hierarchical arrangement of criteria and sub-criteria. The dynamics of its application comprise the stages of structuring the model, the comparative judgment of alternatives and criteria, and the synthesis of priorities. As a starting point for the application of AHP, the structuring of the model comprises the decomposition of a complex decision problem into a hierarchical structure of criteria, as shown in Figure 1.

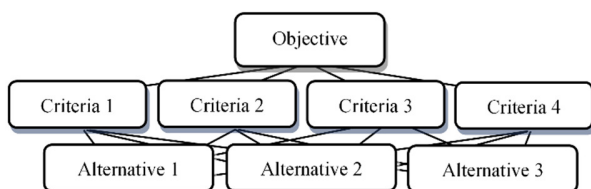


Fig. 1. Hierarchical structure of AHP.

Figure 1 shows the hierarchical structure of the AHP method, showing the arrangement in levels, where the top level establishes the general objective of the problem. The next level includes multiple criteria involved in the solution, and the possible alternatives are at the last level. After structuring, the paired judgment stage follows, where expert knowledge is used based on experience, intuition, and expertise to determine the relevance of one criterion over another with a focus on the main objective, using a numerical scale to transform verbal assessments into numbers. The values derived from the judgments were used to construct comparison matrices. Thus, A can be formulated as an $n \times n$ matrix, for n criteria:

$$A = \begin{matrix} & \begin{matrix} C_1 & C_2 & \dots & C_n \end{matrix} \\ \begin{matrix} C_1 \\ C_2 \\ \dots \\ C_n \end{matrix} & \begin{bmatrix} C_1 & C_2 & \dots & C_n \\ 1 & a_{12} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ 1/a_{1n} & 1/a_{2n} & \dots & 1 \end{bmatrix} \end{matrix} \quad (1)$$

where n is the number of criteria to be evaluated, C_i is the criterion i , and a_{ij} is the weight of criterion i to criterion j . Matrix A must satisfy the condition of reciprocity (i.e., if C_i is a_{ij} more preferable than C_j , then C_j is $1/a_{ij}$ times more preferable than C_i). From the judgment matrix, the importance or relative weight of the criteria can be calculated through normalization as:

$$w_{Cn_i} = \frac{\sum_{j=1}^n a_{ij}}{\sum_{i=1}^n \sum_{j=1}^n a_{ij}} \quad (2)$$

Normalization produces a single estimate of the ratio scale underlying the judgments [21]. In real-world problems, it is impossible to obtain a perfectly consistent judgment matrix after a pairwise comparison [22]. This is due to the possibility of uncertainty on the part of decision-makers when comparing certain elements [23]. In this context, the consistency index was introduced to measure the coherence of judgments [24]:

$$CI = \frac{\lambda_{max} - n}{n - 1} \quad (3)$$

where CI is the consistency index, n is the number of criteria evaluated, and λ_{max} is the eigenvector of A. A consistency relationship (CR) was also proposed to determine whether the value of CI is appropriate. CR is determined by the ratio between the values of the consistency index CI and the random consistency index RI according to:

$$CR = \frac{CI}{RI} < 0.1 \sim 10\% \quad (4)$$

where CI is the consistency index and RI is a random index that relies on the dimension of the matrix being compared. RI is a tabulated value derived from an experiment with a sample size of 500 filled with random values from the scale [25]. The accepted limit for CR is 0.1. If the CR exceeds this value, the evaluation procedure should be repeated to improve consistency. When CR is within the expected range, the method leads to an approximation of the decision-maker's priorities.

B. Method of Elimination and Choice Translation Reality (ELECTRE-II)

ELECTRE-II is part of a family of methods based on the aggregation of preferences [26], which also works with paired comparisons of actions, but with an overcoming approach [27]. To apply this method, A was considered as a set of actions and $g_i(a)$ as the evaluation of any of these actions according to the criterion i ($i = 1, 2, \dots, n$). Applying the overclassification relation to the elements of the set, A can be defined if an alternative a outranks b . This gives aSb , if alternative a is at least as good as alternative b . After evaluating the criteria, the next step is to calculate the agreement and disagreement matrices, which are constructed using the agreement and disagreement indices. The ELECTRE-II method uses concordance indices, which is the extent to which one alternative outperforms the other based on a comparison between the criteria by:

$$C(a, b) = \frac{1}{W} \sum_{j: g_j(a) \geq g_j(b)} w_j \quad (5)$$

where $W = \sum_{j=1}^n w_j$

and disagreement, which is the performance disadvantage of one alternative over another based on a comparison between the criteria, as defined in:

$$D(a, b) = 0, \text{ se } g_j(a) \geq g_j(b) \forall_j \text{ or}$$

$$D(a, b) = \frac{1}{\delta} \max_j [g_j(b) - g_j(a)], n.c$$

$$\text{with } \delta = \max_{c,d,j} [g_j(c) - g_j(d)] \quad (6)$$

where g is the evaluation of the actions and δ is the amplitude of the criteria for each alternative. In ELECTRE-II, strong and weak agreement and disagreement thresholds (c_1 , c_2 , d_1 , d_2) are used to find strong and weak overclassification relations (S^F) and (S^G), respectively, which are calculated using:

$$aS^F b \text{ if } = \begin{cases} C(a, b) \geq c_1 \\ D(a, b) \leq d_1 \\ \sum_{j: g_j(a) > g_j(b)} w_j > \sum_{j: g_j(a) < g_j(b)} w_j \end{cases} \quad (7)$$

$$aS^G b \text{ if } = \begin{cases} C(a, b) \geq c_2 \\ D(a, b) \leq d_2 \\ \sum_{j: g_j(a) > g_j(b)} w_j > \sum_{j: g_j(a) < g_j(b)} w_j \end{cases} \quad (8)$$

Ordering is determined through two preorders constructed using two over-ranking relations (S^F , S^G). In S^F , a ranking is made in descending order of the alternatives, while in S^G the ranking must be in ascending order. Once the two rankings are established, the final step in ELECTRE-II is to order them according to their average rankings. This establishes a complete order of the classification [28].

C. GIS and Dijkstra Algorithm

The spatial nature of emergency response is fundamental for the construction of efficient routes. Road networks in urban areas occasionally suffer from obstructions, whether due to road obstructions, traffic jams, or other types of adversity, causing traffic to be blocked [21-22, 29]. Under these circumstances, the layers of data provided by a GIS and associated with the Dijkstra algorithm can be useful in creating new routes that speed up the service. Determining the shortest path is a fundamental problem in graph theory [30]. Based on the graph theory, Dijkstra's algorithm can search for the shortest path between two vertices. The wide spectrum of its applications ranges from routing problems in communication networks to robot motion planning and highway and power line engineering [31]. This algorithm accepts the input of a graph G that has weights between two interconnected nodes, a starting and an ending node that is the destination of the route [32].

Consider a weighted graph G with n vertices numbered from 1 to n . Let $d[v]$ be the current distance from the source vertex s to vertex v and $w(u, v)$ be the weight of the edge that connects u and v . The algorithm maintains a set S of vertices whose shortest distances from s are known. The algorithm can be understood as follows:

```

1. Initialization:
   d[s]=0
   d[v] = ∞ for v ≠ s
   S = ∅
2. For each vertex v in G:
   S = S ∪ {v}
   For each vertex u adjacent to v in G:
     If d[v] + w(u,v) < d[u]:
       d[u] = d[v] + w(u,v)
3. Remove v from S
4. Repeat step 2 and 3 until S is empty

```

Finally, the vector d will contain the shortest distances from the source vertex s to all other vertices in the graph. The algorithm has the potential to extract complex geographic information for network analysis [33]. To achieve this, the vertices corresponding to geographic coordinates are connected by polylines representing roads or transport routes [34]. Therefore, by linking it to a GIS, the road network can be adapted to source-destination problems by attaching the information that a GIS offers (types of street structure, afforestation, and flow) to the attribution of weights used in the algorithm to build the ideal access route to events.

D. The Proposed Model

The proposed model is an integrated MCDM/A and GIS approach comprising the AHP-ELECTRE II methods and the Dijkstra algorithm, as shown in Figure 2.

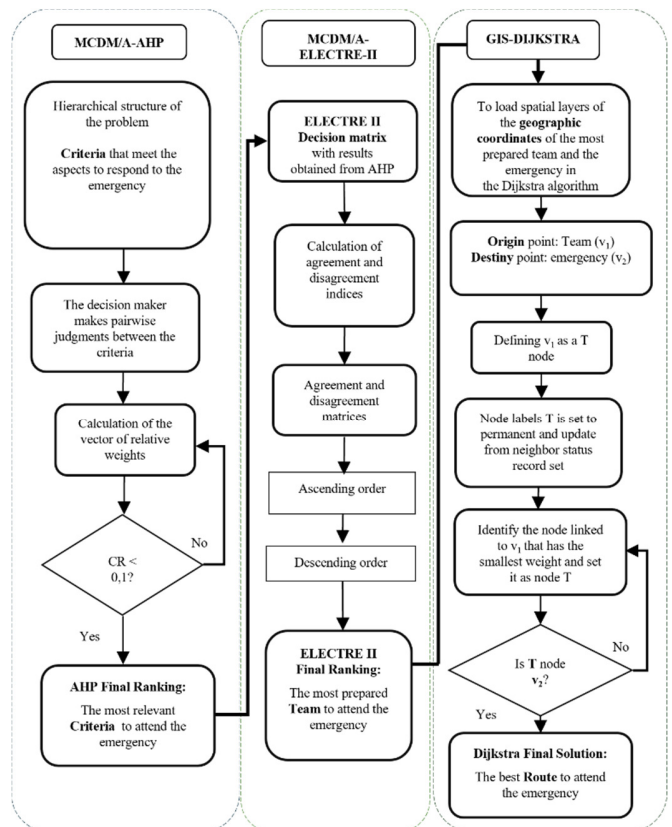


Fig. 2. Proposed model.

First, the most relevant criterion for occurrence is determined using the AHP method. At this stage, the experience of the manager helps determine the importance of each criterion, generating a ranking of the most important criteria [35]. The values of these criteria are used as weights in the ELECTRE-II decision matrix, which determines the team that is most prepared for support. Consequently, the ELECTRE-II method provides an outperforming relationship between teams. This relationship consists of ordering teams according to the assistance conditions. This configuration, in addition to providing the most suitable team, also allows switching to a second or third team, or a team that is more prepared for the situation, thus bringing more agility and dynamism. When considering the coordinates of the support team and its occurrence in a GIS, the Dijkstra algorithm can be used to calculate the access routes.

III. APPLICATION EXAMPLE: EMERGENCY WITH DANGEROUS PRODUCTS

To demonstrate the proposed method, an example of an emergency incident with dangerous products was used. In this context, given the changes over the last few decades, there has been a continuous increase in the rates of accidents with dangerous products, along with associated adverse impacts such as human, environmental, and economic losses [36]. To intervene in these accidents, services such as those provided by fire departments are often requested to inspect and isolate sites [37]. To prepare the model, the criteria presented in [38] were considered for the categories of equipment, accessories, and vehicles. When considering the necessary criteria for this occurrence, an example of decision modeling based on the AHP method can be constructed as shown in Figure 3.

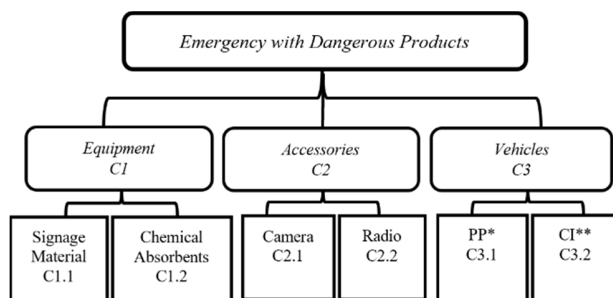


Fig. 3. The hierarchical structure of the example based on AHP (PP*: Vehicle for dangerous products | CI**: Fire fighting vehicle)

Table I shows the results of paired comparisons of the process after the hierarchy involves judging the criteria, evaluating the consistency of the judgments, and synthesizing the priorities using the AHP method.

TABLE I. RESULTS OF PAIRED COMPARISONS OF CRITERIA

Criteria	Equipment	Accessories	Vehicles
Equipment	1	5	1/3
Accessories	1/5	1	1/7
Vehicles	3	7	1

To evaluate consistency, the CR was calculated using (4):

$$CR = \frac{CI}{RI} = 0,05$$

If $CR < 0.1$, the judgments are considered consistent. Following this process, pairwise comparison matrices are constructed for the subcriteria. Table II presents the general classification.

TABLE II. CLASSIFICATION RESULTING FROM THE AHP METHOD

Criterion weight	Sub-criterion weight	Cumulative weight	Classification
C1: Equipment $w = 0.283$	C1.1: Signage Material $w = 0.266$	0.075	4
	C1.2: Chemical absorbents $w = 0.734$	0.207	3
C2: Accessories $w = 0.074$	C2.1: Camera $w = 0.680$	0.053	5
	C2.2: Radio $w = 0.320$	0.023	6
C3: Vehicles $w = 0.643$	C3.1: PP $w = 0.630$	0.405	1
	C3.2: CI $w = 0.370$	0.237	2

As shown in Table II, among the evaluated criteria, the vehicles had the highest weights, followed by the equipment and accessories. The cumulative weights of the criteria from the AHP method were considered for team classification using the ELECTRE-II method, as shown in Table III. At this stage, the method helps to decide which team is the most appropriate to support the incident.

TABLE III. ELECTRE-II METHOD EVALUATION

Alternatives	Equipment		Accessories		Vehicles	
	C1.1	C1.2	C2.1	C2.2	C3.1	C3.2
Team 1	3	5	6	3	2	5
Team 2	6	4	7	4	6	8
Team 3	2	5	3	7	4	6
Weights	0.075	0.207	0.053	0.023	0.405	0.237

Using the agreement index in (5), the concordance matrix shown in Table IV is obtained. The discordance matrix according to (6) is shown in Table V.

TABLE IV. CONCORDANCE MATRIX

	Team 1	Team 2	Team 3
Team 1	0.000	0.207	0.335
Team 2	0.793	0.000	0.776
Team 3	0.872	0.230	0.000

TABLE V. DISCORDANCE MATRIX

	Team 1	Team 2	Team 3
Team 1	0	1	1
Team 2	0.25	0	0.75
Team 3	0.75	0.75	0

Once the agreement and disagreement matrices are obtained, the agreement and disagreement thresholds must be established. Thus, based on the values of the matrices, the thresholds can be defined as: $C_1^* = 0,85$, $D_1^* = 0,50$; $C_2^* = 0,76$; and $D_2^* = 0,65$. Having established limits, overcoming relationships are constructed according to the following rule: If $C(a,b) \geq C^*$, $D(a,b) \leq D^*$, the alternative a is strongly superior to the alternative b . Based on this rule, a dominance matrix (Table VI) was defined to establish strong (S^F) and weak (S^f) relationships.

TABLE VI. DOMINANCE MATRIX

	Team 1	Team 2	Team 3
Team 1	0	—	—
Team 2	S^F	0	S^f
Team 3	S^F	—	0

In Table VI, overclassification relationships can be observed based on established thresholds. In this case, Team 3 S^F Team 1 indicates that Team 3 significantly outperforms Team 1. Considering these relationships, the ordering can be performed through ascending and descending distillation processes. The descending preorder is constructed by taking the set of best alternatives, those that outperform other alternatives, and descending to the worst. In contrast, the ascending preorder is constructed by starting with the set of "worst" alternatives, those that are outperformed by other alternatives, and moving up to the best. Therefore, the final classification is as follows: Team 2 > Team 3 > Team 1, where Team 2 exceeds Team 3 and 1, and Team 3 exceeds Team 1. This configuration helps in situations where the indicated team is occupied, and the ordering of the second team with a greater support capacity can take over the occurrence, contributing to the dynamics of the phenomenon.

By ordering the teams, it is possible to define a route using Dijkstra's algorithm. Therefore, the spatial coordinates of the indicated support teams and their occurrence constitute the problems of origin and destination. In this example, a simulation of Dijkstra's algorithm was performed using the open-source Flutter user interface development kit, based on the Dart programming language [39]. For this simulation, an area covering a graph with 20 nodes was considered, with the initial node containing the geographic coordinates of the support team and the final node containing the geographic coordinates of the incident. This graph consists of streets and accesses. Figure 4 shows the possible access routes and Table VII shows a list of possible routes using the Dijkstra algorithm. According to Table VII, the most efficient route is {B, C, D, I, O} which leads to the incident at the lowest cost. Therefore, in this example, using initially the AHP method, the proposed model allowed the assessment of the criteria necessary to respond to emergency calls involving dangerous products (equipment, accessories, and vehicles). Subsequently, using ELECTRE II, the model made it possible to determine which of the teams was best able to deal with the incident (Team 2). In the last stage of the model, Dijkstra's algorithm determined the best route to respond to the incident (route {B, C, D, I, O}).



Fig. 4. Access routes.

TABLE VII. POSSIBLE ACCESS ROUTES

Route		Cost	Destination
Origin	B, G, M, N, S	$(1+1+3+3+2+1) = 11$	
	B, G, H, I, O	$(1+1+5+3+2+1) = 13$	
	B, C, D, I, O	$(1+3+1+1+2+1) = 9$	
	B, A, F, L, Q, R, S	$(1+2+6+3+2+4+2+1) = 21$	
	B, C, H, N, S	$(1+3+4+3+2+1) = 14$	

IV. CONCLUSION

This paper presented a mathematical computational model based on two multi-criteria methods combined with the Dijkstra algorithm to support emergency operations. Compared to previous studies, this method stands out for its dynamic approach to the incident support process, proposing an arrangement that combines the robustness of multi-criteria methods and the efficiency of the Dijkstra algorithm. One of the purposes of this study was to incorporate the experience of decision-makers into a model that considers their expertise combined with spatial dynamics in responding to incidents. To achieve this, a multi-criteria approach was used that enables the transformation of experience into numerical values in complex decision problems. Due to the complex characteristics of an incident, the AHP and ELECTRE-II methods sought to address the importance of specific values for a given incident and how to use the best resources offered by support teams. In the decision matrix of the ELECTRE-II method, adding the weights of the AHP method, which is used to determine the most important criteria for a given situation, provides a better approach to selecting the team that is best prepared to help.

One of the concerns considered was the search for a method that would bring about an ordering of support teams so that when it is impossible to have one, the manager could have a sequence of teams as an option. In this case, the ELECTRE-II method was quite competent by considering strong and weak overcoming relationships in its final ordering. Furthermore, the choice of this method contributes to the lack of overcoming methods in these scenarios. In addition to the administrative nature of resources, Dijkstra's algorithm and GIS were used to consider the spatial dynamics of the phenomenon in determining routes. In this strategy, spatial information from

the GIS is considered and weighted in the algorithm. Therefore, by combining MCDM/A with GIS, the presented formulation seeks to contribute to a more holistic scope of the incident by adding an administrative aspect and covering the spatial part.

ACKNOWLEDGMENT

The authors would like to thank the Brazilian National Council for Technological Development (CNPq - Process: 310976/2022-7 - DT level II Researcher Scholarship) for the financial support of the project.

REFERENCES

- [1] P. Murphy, P. Wankhade, and K. Lakoma, "The strategic and operational landscape of emergency services in the UK," *International Journal of Emergency Services*, vol. 9, no. 1, pp. 69–88, Jan. 2019, <https://doi.org/10.1108/IJES-12-2018-0062>.
- [2] P. Murphy, P. Eckersley, and L. Ferry, "Accountability and transparency: Police forces in England and Wales", *Public Policy and Administration*, vol. 32, no. 3, pp. 197–213, 2017.
- [3] V. Mali, M. Rao, and S. S. Mantha, "AHP Driven GIS Based Emergency Routing in Disaster Management," in *Advances in Computing, Communication, and Control*, Mumbai, India, 2013, pp. 237–248, https://doi.org/10.1007/978-3-642-36321-4_22.
- [4] S. S. H. Mar and O. H. Ne, "Gis based fire emergency response system for Mandalay," *International Journal of Open Information Technologies*, vol. 7, no. 12, pp. 39–45, 2019.
- [5] A. A. Longaray, L. Ensslin, A. Dutra, S. Ensslin, R. Brasil, and P. Munhoz, "Using MCDA-C to assess the organizational performance of industries operating at Brazilian maritime port terminals," *Operations Research Perspectives*, vol. 6, Jan. 2019, Art. no. 100109, <https://doi.org/10.1016/j.orp.2019.100109>.
- [6] A. Longaray, L. Ensslin, S. Ensslin, G. Alves, A. Dutra, and P. Munhoz, "Using MCDA to evaluate the performance of the logistics process in public hospitals: the case of a Brazilian teaching hospital," *International Transactions in Operational Research*, vol. 25, no. 1, pp. 133–156, 2018, <https://doi.org/10.1111/itor.12387>.
- [7] W. Sałabun, J. Wątróbski, and A. Shekhovtsov, "Are MCDA Methods Benchmarkable? A Comparative Study of TOPSIS, VIKOR, COPRAS, and PROMETHEE II Methods," *Symmetry*, vol. 12, no. 9, Sep. 2020, Art. no. 1549, <https://doi.org/10.3390/sym12091549>.
- [8] Y. Wang, "An Optimization Method for Distributing Emergency Materials Which Balances Multiple Decision Criteria," *Processes*, vol. 10, no. 11, Nov. 2022, Art. no. 2317, <https://doi.org/10.3390/pr10112317>.
- [9] National Authority for Remote Sensing and Space Sciences, S. Ahmed, R. Ibrahim, and H. Hefny, "Mobile-based Routes Network Analysis for Emergency Response Using an Enhanced Dijkstra's Algorithm and AHP," *International Journal of Intelligent Engineering and Systems*, vol. 11, no. 6, pp. 252–260, Dec. 2018, <https://doi.org/10.22266/ijies2018.1231.25>.
- [10] A. T. Murray, "Contemporary optimization application through geographic information systems," *Omega*, vol. 99, Mar. 2021, Art. no. 102176, <https://doi.org/10.1016/j.omega.2019.102176>.
- [11] I. S. Zinovieva *et al.*, "The use of GIS in renewable energy specialist's learning," *Journal of Physics: Conference Series*, vol. 2611, no. 1, Jul. 2023, Art. no. 012016, <https://doi.org/10.1088/1742-6596/2611/1/012016>.
- [12] N. Kumar and R. Jha, "GIS-based Flood Risk Mapping: The Case Study of Kosi River Basin, Bihar, India," *Engineering, Technology & Applied Science Research*, vol. 13, no. 1, pp. 9830–9836, Feb. 2023, <https://doi.org/10.48084/etasr.5377>.
- [13] F. Yilmaz, S. Alp, B. Oz, and A. Alkoc, "Analysis of the Risks arising from Fire Installations in Workplaces using the Ranking Method," *Engineering, Technology & Applied Science Research*, vol. 10, no. 4, pp. 5914–5920, Aug. 2020, <https://doi.org/10.48084/etasr.3646>.
- [14] A. R. A. Rasam, D. Mardie, M. S. M. Said, N. Hashim, and R. Ghazali, "Landslide Susceptibility Index and Network Mapping for Spotting the Affected and Alternative Planning Routes," *Planning Malaysia Journal*, vol. 21, no. 26, Jun. 2023, <https://doi.org/10.21837/pm.v21i26.1266>.
- [15] Z. Zhang *et al.*, "A cyberGIS-enabled multi-criteria spatial decision support system: A case study on flood emergency management," *International Journal of Digital Earth*, vol. 12, no. 11, pp. 1364–1381, Nov. 2019, <https://doi.org/10.1080/17538947.2018.1543363>.
- [16] A. P. Vavatsikos, K. F. Sotiropoulou, and V. Tzingizis, "GIS-assisted suitability analysis combining PROMETHEE II, analytic hierarchy process and inverse distance weighting," *Operational Research*, vol. 22, no. 5, pp. 5983–6006, Nov. 2022, <https://doi.org/10.1007/s12351-022-00706-0>.
- [17] K. F. Sotiropoulou and A. P. Vavatsikos, "A Decision-Making Framework for Spatial Multicriteria Suitability Analysis using PROMETHEE II and k Nearest Neighbor Machine Learning Models," *Journal of Geovisualization and Spatial Analysis*, vol. 7, no. 2, Jul. 2023, Art. no. 20, <https://doi.org/10.1007/s41651-023-00151-3>.
- [18] M. Vojtek, J. Vojteková, and Q. B. Pham, "GIS-Based Spatial and Multi-Criteria Assessment of Riverine Flood Potential: A Case Study of the Nitra River Basin, Slovakia," *ISPRS International Journal of Geo-Information*, vol. 10, no. 9, Sep. 2021, Art. no. 578, <https://doi.org/10.3390/ijgi10090578>.
- [19] A. A. Longaray, J. de D. R. Gois, and P. R. da S. Munhoz, "Proposal for using AHP Method to Evaluate the Quality of Services Provided by Outsourced Companies," *Procedia Computer Science*, vol. 55, pp. 715–724, Jan. 2015, <https://doi.org/10.1016/j.procs.2015.07.083>.
- [20] Y. Liu, C. M. Eckert, and C. Earl, "A review of fuzzy AHP methods for decision-making with subjective judgements," *Expert Systems with Applications*, vol. 161, Dec. 2020, Art. no. 113738, <https://doi.org/10.1016/j.eswa.2020.113738>.
- [21] T. L. Saaty, "Some Mathematical Concepts of the Analytic Hierarchy Process," *Behaviormetrika*, vol. 18, no. 29, pp. 1–9, 1991, https://doi.org/10.2333/bhmk.18.29_1.
- [22] A. A. Longaray, C. F. S. Gomes, T. Elacoste, and C. M. dos S. Machado, "Efficiency Indicators to Evaluate Services in Port Services: A Proposal Using Fuzzy-AHP Approach," *Pesquisa Operacional*, vol. 39, no. 3, pp. 437–456, Dec. 2019, <https://doi.org/10.1590/0101-7438.2019.039.03.0437>.
- [23] T. L. Saaty, "A scaling method for priorities in hierarchical structures," *Journal of Mathematical Psychology*, vol. 15, no. 3, pp. 234–281, Jun. 1977, [https://doi.org/10.1016/0022-2496\(77\)90033-5](https://doi.org/10.1016/0022-2496(77)90033-5).
- [24] R. W. Saaty, "The analytic hierarchy process—what it is and how it is used," *Mathematical Modelling*, vol. 9, no. 3, pp. 161–176, Jan. 1987, [https://doi.org/10.1016/0270-0255\(87\)90473-8](https://doi.org/10.1016/0270-0255(87)90473-8).
- [25] B. Roy, "Classement et choix en présence de points de vue multiples," *Revue française d'informatique et de recherche opérationnelle*, vol. 2, no. 8, pp. 57–75, 1968.
- [26] M. Cinelli, S. R. Coles, and K. Kirwan, "Analysis of the potentials of multi criteria decision analysis methods to conduct sustainability assessment," *Ecological Indicators*, vol. 46, pp. 138–148, Nov. 2014, <https://doi.org/10.1016/j.ecolind.2014.06.011>.
- [27] M. Gershon, L. Duckstein, and R. McNiff, "Multiobjective river basin planning with qualitative criteria," *Water Resources Research*, vol. 18, no. 2, pp. 193–202, 1982, <https://doi.org/10.1029/WR018i002p00193>.
- [28] E. N. Tamatjita and A. W. Mahastama, "Shortest Path with Dynamic Weight Implementation using Dijkstra's Algorithm," *ComTech: Computer, Mathematics and Engineering Applications*, vol. 7, no. 3, pp. 161–171, Sep. 2016, <https://doi.org/10.21512/comtech.v7i3.2534>.
- [29] J. L. Ortega, J. López-Sauceda, J. G. Carrillo, and J. Sandoval, "Método de construcción de dígrafos a partir de redes viales reales en mapas digitales con aplicaciones en la búsqueda de rutas óptimas," *Informes de la Construcción*, vol. 71, no. 556, pp. e320–e320, 2019, <https://doi.org/10.3989/ic.65988>.
- [30] B. Bhowmik and S. N. Chowdhury, "Prograph Based Analysis of Single Source Shortest Path Problem with Few Distinct Positive Lengths," *Engineering, Technology & Applied Science Research*, vol. 1, no. 4, pp. 90–97, Aug. 2011, <https://doi.org/10.48084/etasr.41>.

-
- [31] A. Fitro, O. S. Bachri, A. I. S. Purnomo, and I. Frendianata, "Shortest Path Finding in Geographical Information Systems Using Node Combination and Dijkstra Algorithm," *International Journal of Mechanical Engineering and Technology*, vol. 9, no. 2, pp. 755–760, Feb. 2018.
- [32] Q. Wu, G. Qin, and H. Li, "An Improved Dijkstra's algorithm application to multi-core processors," *Metallurgical & Mining Industry*, vol. 9, no. 1, pp. 76–81, 2015.
- [33] F. Zhao *et al.*, "On-Road Trajectory Planning of Connected and Automated Vehicles in Complex Traffic Settings: A Hierarchical Framework of Trajectory Refinement," *IEEE Access*, vol. 12, pp. 7456–7468, 2024, <https://doi.org/10.1109/ACCESS.2024.3352919>.
- [34] M. Chebila, "Predicting the consequences of accidents involving dangerous substances using machine learning," *Ecotoxicology and Environmental Safety*, vol. 208, Jan. 2021, Art. no. 111470, <https://doi.org/10.1016/j.ecoenv.2020.111470>.
- [35] A. A. Longaray, L. Ensslin, S. R. Ensslin, and I. O. da Rosa, "Assessment of a Brazilian public hospital's performance for management purposes: A soft operations research case in action," *Operations Research for Health Care*, vol. 5, pp. 28–48, Jun. 2015, <https://doi.org/10.1016/j.orhc.2015.05.001>.
- [36] S. Y. Sabirovna, "Fire and Explosion Risk Prevention," *Galaxy International Interdisciplinary Research Journal*, vol. 11, no. 5, pp. 137–142, May 2023.
- [37] C. Zhang, "Analysis of Fire Safety System for Storage Enterprises of Dangerous Chemicals," *Procedia Engineering*, vol. 211, pp. 986–995, Jan. 2018, <https://doi.org/10.1016/j.proeng.2017.12.101>.
- [38] Z. Yu and J. Guan, "Fire and Rescue Combat Technical Training System Construction for Dangerous Chemicals," *Procedia Engineering*, vol. 135, pp. 655–660, Jan. 2016, <https://doi.org/10.1016/j.proeng.2016.01.133>.
- [39] "Flutter - Build apps for any screen." Google, [Online]. Available: [//flutter.dev/](https://flutter.dev/).

Development and Characterization of a PLA Biocomposite reinforced with Date Palm Fibers

Ines Ghanmi

Laboratory LR18ES45, Mathematical Physics, Quantum Modeling and Mechanical Design, Nabeul Preparatory Engineering Institute, Carthage University, Tunisia
ines11ghanmi@gmail.com

Faouzi Slimani

LMPE LR18ES01, National Higher Engineering School of Tunis, University of Tunis, Tunisia
slimani.fauzi@yahoo.fr (corresponding author)

Samir Ghanmi

Laboratory LR18ES45, Mathematical Physics, Quantum Modeling and Mechanical Design, Nabeul Preparatory Engineering Institute, Carthage University, Tunisia
samir.ghanmi@ipein.rnu.tn

Mohamed Guedri

Laboratory LR18ES45, Mathematical Physics, Quantum Modeling and Mechanical Design, Nabeul Preparatory Engineering Institute, Carthage University, Tunisia | National Higher Engineering School of Tunis, University of Tunis, Tunisia
mohamed.guedri@isetn.rnu.tn

Received: 1 February 2024 | Revised: 15 February 2024 and 24 February 2024 | Accepted: 3 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6988>

ABSTRACT

Despite the promising potential of bio-composites derived from plant fibers due to their ecological and economic benefits, challenges persist in their preparation, restricting their commercial applications. These challenges are primarily associated with developing suitable methods, acquiring appropriate equipment for treating plant fibers, and addressing the time constraints in preparation. This study aims to contribute to the development and characterization of a new biocomposite and biodegradable material based on natural fibers produced through hot compression. The newly developed biocomposite comprises commercial biodegradable poly-lactic acid (PLA) as a matrix and untreated fiber fabric extracted from date palms as reinforcement. The use of untreated fiber fabric has successfully overcome the preparation difficulties. Experimental results on the new biocomposite reveal the strong adhesion between its fibers and the matrix, emphasizing the significant impact of choosing the right manufacturing conditions on the developed mechanical properties.

Keywords: biocomposite; polylactic acid (PLA); Date Palm Fibers (DPFs); mechanical characterization

I. INTRODUCTION

In today's environmentally conscious world, industries like aeronautics, energy, nuclear, and civil engineering are increasingly intrigued by the mechanical capabilities of polymer composites reinforced with natural fibers [1-2]. Notably, cellulose-based fibers such as kenaf [3], flax [4], hemp [5], banana [6], coconut [7], and *Posidonia Oceanica* [8] are being explored by several researchers. Beyond their commendable mechanical properties, these composites offer the advantage of low density, enhancing structural performance and longevity. Additionally, they stand out as natural, eco-friendly, biodegradable, and recyclable materials. Authors in

[9-13] delved into date palm derivatives due to their abundance and diverse fiber types (petiole, rachis, leaflets, fibrillium, bunch, pedicels, spathe, and thorns) [14]. Research on the mechanical characteristics of date palm frond stems was conducted in [15, 16], in which a biocomposite material reinforced with two types of date palm fibers (rachis and leaflets fibers) at varying percentages (4%, 7%, and 10%) was investigated using contact molding for the manufacture of the biocomposites. These cellulose-based fibers are actively used as reinforcements in developing innovative biodegradable and recyclable biocomposite materials. In addition to selecting appropriate fibers, choosing the right matrices is also pivotal in

composite manufacturing. These polymers are divided into two classes: thermoplastics and thermosets. Thermoplastics, such as polypropylene (PP) [11, 14] and poly-lactic acid (PLA) [5, 9, 17-19] dominate as matrices for biological fibers, while thermosetting matrices [20] like phenolic, epoxy, and polyester resins are widely employed. Generally, the manufacturing process (method of production, nature of the fibers used, nature of the matrix, etc.) significantly influences the biocomposite performance. There are several manufacturing composite processes for reinforcing with natural fibers, such as injection molding and extrusion, which are instrumental in the plastics industry, involving the careful mixing of raw materials at controlled temperatures. The preparation of natural fibers for composites is a crucial step that requires involving extraction and processing, each with its own associated costs and challenges. The fiber extraction methods vary depending on the nature of the plant, e.g. those for flax, requiring multiple steps such as plucking, retting, scutching, and combing [26]. Special machines, such as those for bamboo and alfa fibers, are used for decortication and are also applicable to banana, sisal, flax, and jute fibers. Date Palm Fibers (DPFs), whether in small pieces or not, necessitate operations like grinding, sieving, cleaning, and drying for reinforcement.

This study presents a contribution to the development of new biocomposite and biodegradable materials at lower cost. Our new biocomposite is composed of biodegradable commercial PLA as matrix and an untreated fiber fabric extracted from date palms as reinforcement. The use of untreated fiber fabric allowed us to overcome the above-mentioned preparation difficulties. The experimental results demonstrate the good adhesion between the fibers and the matrix, and that the right choice of manufacturing conditions has a clear influence on the biocomposite's mechanical properties.

II. MATERIALS AND METHODS

A. Materials

To produce the innovative biocomposite, consisting of PLA as the matrix and DPFs as the reinforcement, we devised and implemented an electromechanical system facilitating its creation through hot compression (Figure 1). The core components of this device include:

- A steel die and punch.
- Dual heating resistors, with one integrated into the die and the other in the punch.
- Twin thermocouples positioned in the die and punch.
- A digital controller tasked with regulating the temperature to the specified set point as needed.

The die, punch, and the upper and lower PLA layers, alongside the DPFs layer are illustrated in Figure 2. The two perforations indicate the positions of the heating resistors and the thermocouples are linked to the temperature controller. The assembly of the die and punch is affixed to an HM-S 200KN-04M4210 traction-compression machine.

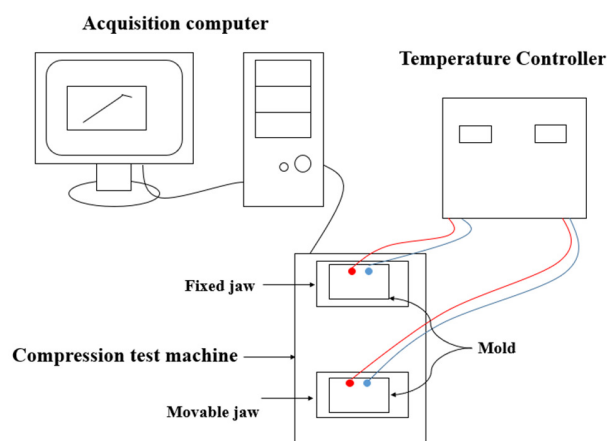


Fig. 1. Schematic diagram of the electromechanical device.

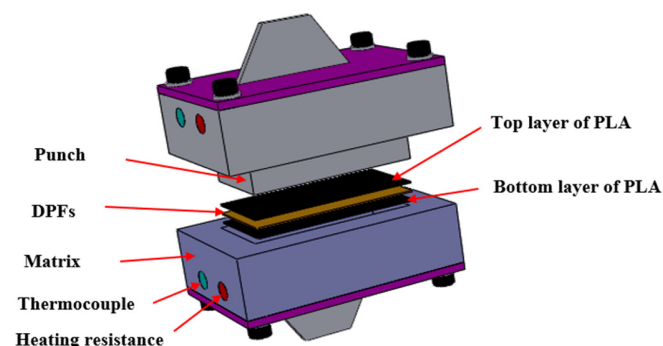


Fig. 2. The experimental device.

B. Methods

Figure 3 outlines the methodology employed in the current research.



Fig. 3. Manufacturing and testing process.

The characterization process for our novel biocomposite is broadly segmented into three primary stages:

1) Step 1

- Preparation of DPFs fabric measuring 100 mm × 50 mm: Manual cleaning and cutting according to the desired

percentage (5% or 10% of the weight of the biocomposite). We note that the fiber fabric has a random structure where some areas may be denser than others, meaning that the distribution of fibers is not uniform over the entire surface. In some places, the fibers may be tighter and closer together, while in others they may be more dispersed. In addition, the yarns that make up the fabric vary in diameter. Some yarns may be thicker, while other, finer yarns may lend additional flexibility and lightness to certain areas of the fiber fabric, as shown in Figure 4.



Fig. 4. Date palm fibers.

- **PLA Preparation:** The PLA is manually cut into small fragments and it is weighed based on the intended biocomposite composition (PLA/DPFs-5 or PLA/DPFs-10).
 - Plates of the new biocomposite through hot compression utilizing the system outlined above and following a predetermined full factorial Design of Experiments (DoE) were generated.
- 2) *Step 2*
- Employ laser cutting on PLA/DPFs specimens.
 - Conduct a series of tensile tests.
- 3) *Step 3*
- Synthesize the findings of this investigation, encompassing the determination of the biocomposite's mechanical properties and the cohesion state of DPFs and PLA.

III. EXPERIMENTS

For the preparation of the novel PLA/DPFs bio-composite, we employed the aforementioned experimental apparatus illustrated in Figures 1-2. This investigation considers five experimental parameters (factors) crucial to the production of the biocomposite: DPFs percentage (5% and 10%), PLA heating temperature T (°C), duration of punch retention in the die during hot compression t (min), cooling duration of the biocomposite plate post-compression t_r (min), and the pressure force required between the die and punch F (daN). These parameters are considered at two levels, as specified in Table I.

To characterize our biocomposite, a series of tensile tests were conducted using a designated apparatus (HM-S 200KN - 04M4210). This machine is equipped with a 10000 daN load cell for recording the applied force, along with pneumatic jaws featuring manual clamping to secure the specimen. The test is displacement-controlled.

TABLE I. EXPERIMENTAL PARAMETERS CONSIDERED FOR THE DEVELOPMENT OF THE NEW BIO-COMPOSITE

Parameter	Description	Level 1	Level 2
T (°C)	PLA heating temperature	160	170
DPF (%)	Percentage of DPFs in the biocomposite	5	10
t (min)	Punch holding time in the die	15	20
t_r (min)	PLA/DPFs plate cooling time	20	30
F (10^3 daN)	Force between die and punch	13	14

The tensile test specimens were laser-cut from the plates of the biocomposite. This selection is justified for two reasons: firstly, laser cutting generally provides precise edges without the need for additional finishing, and secondly, it mitigates issues associated with mechanical cutting. The dimensions of the specimens adhere to NF ISO 527 [27].

IV. RESULTS AND DISCUSSION

In this work, we used PLA as the matrix, whose mechanical properties are: $\rho = 850 \pm 12$ kg/m³, $\sigma_{Max} = 46 \pm 14$ MPa, $\epsilon = 1.26 \pm 0.43$ %, and $E = 3.32 \pm 1.77$ GPa. To characterize our PLA/DPF biocomposite and ascertain its mechanical properties, including Young's modulus (E), maximum stress (σ_{Max}), and corresponding strain (ϵ), a set of tensile tests was conducted on specimens following the ISO 527 standard. These tests facilitated the generation of stress-strain curves depicted in Figures 5 and 6. Figure 5 illustrates the performance of the biocomposite incorporating 5% DPFs (PLA/DPFs-5), while Figure 6 showcases the behavior of the bio-composite comprising 10% DPFs (PLA/DPFs-10).

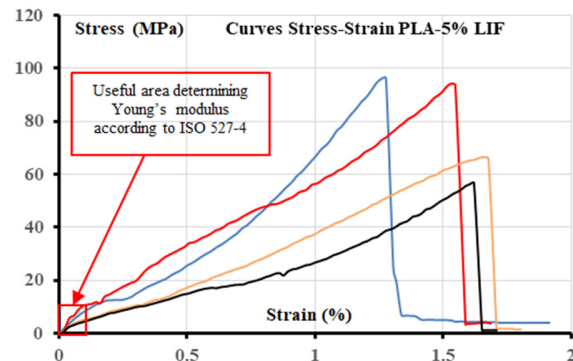


Fig. 5. Tensile curves for samples with 5% natural DPFs.

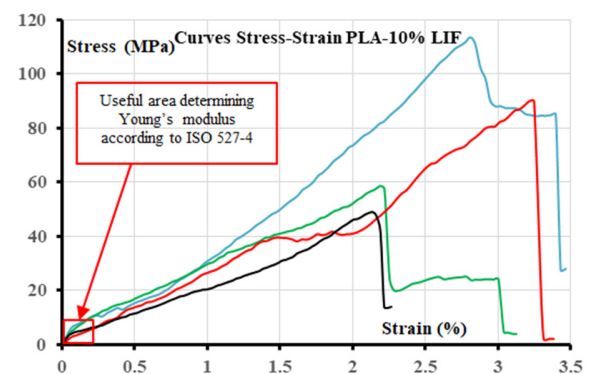


Fig. 6. Tensile curves for samples with 10% natural DPFs.

The tensile behavior of all the specimens is similar and can be broken down into 3 main phases:

The initial phase corresponds to a linear elevation in the applied stress with strain within the [0.05% - 0.25%] strain range, as outlined by ISO 527-4 [28]. This phase corresponds to the elastic region of the stress-strain curve, where the modulus of elasticity can be derived using Hook's law through (1):

$$E = \frac{\Delta\sigma}{\Delta\varepsilon} \quad (1)$$

where E , $\Delta\sigma$, and $\Delta\varepsilon$ represent the modulus of elasticity, the difference in the applied tensile stress between two strain points, and the difference between the two strain points, respectively.

In the second phase, the initial portion of this segment exhibits an almost linear incline, up to the maximum stress value. However, a few peaks within this linear zone are primarily attributed to fiber breaks.

The third and final phase, in which total or degraded stress release is observed, is the phase of total rupture of PLA and DPFs in the bio-composite.

Table II and Figure 7 succinctly present the results of the tensile tests for the two biocomposite types, namely PLA/DPFs-5 and PLA/DPFs-10, as well as standalone PLA. These summaries encompass mean values and standard deviations for the mechanical properties (ρ , σ_{Max} , ε , and E) of the two tensile-stressed biocomposites, comparing them with those of plain PLA. Notably, there is a discernible enhancement in the mechanical property values of PLA/DPFs-5 and PLA/DPFs-10 bio-composites compared to those of plain PLA.

TABLE II. AVERAGE PROPERTIES OF PLA/DPFS BIO-COMPOSITE

Material	ρ (kg/m ³)	σ_{Max} (MPa)	ε (%)	E (GPa)
PLA	850 ± 12	46 ± 14	1.26 ± 0.43	3.32 ± 1.77
PLA-DPFS-5	866 ± 10.5	60 ± 35	1.5 ± 0.2	3.79 ± 2.18
PLA-DPFS-10	871 ± 8.6	75 ± 39	2.8 ± 0.5	3.81 ± 1.91

The results show that the average mechanical properties of PLA/DPFs-10 surpass those of both PLA/DPFs-5 and PLA alone, with substantial increases noted in the majority of cases. Specifically, the average Young's modulus of PLA/DPFs-10 demonstrated a 14.76% augmentation compared to PLA alone and approximately 0.53% when contrasted with PLA/DPFs-5. Similarly, the average maximum strain σ_{Max-10} exhibited a 63.04% increase relative to $\sigma_{Max-PLA}$ and a 25% increase compared to σ_{Max-5} . The average deformation ε_{10} of PLA/DPFs-10 showed a 122.22% increase compared to PLA and an 86.87% increase compared to PLA/DPFs-5. The average density of PLA/DPFs-10 experienced a 2.47% increase compared to PLA and a slight 0.58% increase compared to PLA/DPFs-5. Significant dispersion in the tensile test results was observed for both bio-composites. This is mainly due to:

- The challenging homogeneous PLA distribution between the die and the punch in the produced plates due to the manual placement of PLA fragments in the die.

Consequently, achieving uniform distribution of molten PLA on the plates becomes impractical.

- The DPF fabrics in the biocomposite plates are variable in nature. Indeed, the fibers are of random cross-section, length, orientation, and connection to each other (Figure 4).
- The DPFs have not been physically or chemically treated, which may influence the adhesion of the PLA/DPF interface through the presence of fine impurities on the fibers prior to the hot compression operation in the mold.

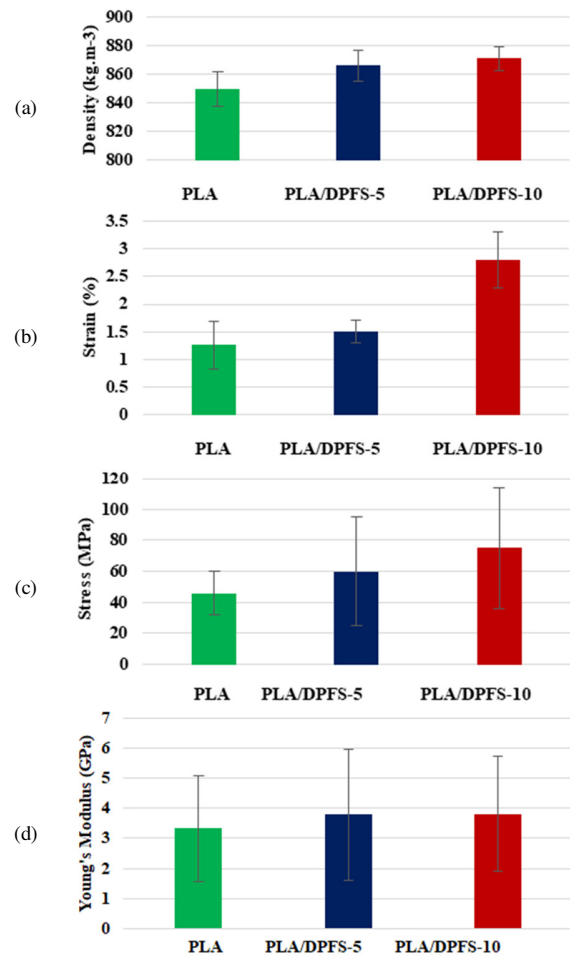


Fig. 7. Variation in bio-composite mechanical properties as a function of DPF percentage.

In order to qualitatively assess the adhesion at the DPF/matrix interface of the new biocomposite, scanning electron microscopy was used to analyze the structure of the new biocomposite. Observations were made on samples cut from damaged specimens. Figure 8 shows an internal view of a plate of the new biocomposite after the hot-pressing operation. It demonstrates the good quality of DPF (10%) and matrix adhesion. This proves that the level of adjustment of the biocomposite's manufacturing parameters is acceptable, as observed on most of the produced plates.



Fig. 8. Adhesion of PLA/DPF-10 interface.

V. CONCLUSION

The primary aim of this study was to propose a cost-effective development and characterization of a novel biocomposite, comprising commercial PLA as the matrix and untreated fiber fabric extracted from date palms as reinforcement. This fabric, abundant in many countries, poses environmental concerns alongside other underutilized waste from date palms and various tree types like olives and almonds. In the initial phase, we engineered a digitally controlled device for producing the new biocomposite through hot compression.

Subsequently, for the characterization of our biocomposite, a series of tensile tests were conducted using a complete factorial experimental design with five factors and two modalities. The outcomes revealed that the average mechanical properties of PLA/DPF-10 surpassed those of PLA/DPF-5 and plain PLA, with significant increases noted in the majority of cases. This underscores the positive influence of augmenting fiber content on the biocomposite's mechanical properties. Furthermore, the results highlighted the high sensitivity of the biocomposite's mechanical properties to slight variations in predetermined manufacturing conditions, such as PLA heating temperature (T), punch retention time in the die during hot compression (t), and the pressure force required between the die and punch of the bio-composite (F). Despite the absence of any physical or chemical treatment, most of the produced biocomposite samples exhibited robust adhesion in all directions between the fiber fabric and the PLA. This was verified by observation using a scanning electron microscope. The current work concerns the development of materials with treated DPFs, 3-point bending tests to characterize the properties of these materials, and the study of the impact of humidity and its effect on the mechanical characteristics of the produced biocomposites.

REFERENCES

- [1] S. Kumar and L. B. Roy, "Investigating the Slope Stability and Factor of Safety Properties of Soil Reinforced with Natural Jute Fibers under Different Rainfall Conditions," *Engineering, Technology & Applied Science Research*, vol. 13, no. 1, pp. 9919–9925, Feb. 2023, <https://doi.org/10.48084/etasr.5481>.
- [2] A. H. Memon, M. H. Peerzada, K. Muhammad, S. A. Memon, S. A. Mangi, and G. Mujtaba, "Recent Eco-Friendly Developments in Personal Protective Clothing Materials for Reducing Plastic Pollution: A Review," *Engineering, Technology & Applied Science Research*, vol. 9, no. 2, pp. 4012–4018, Apr. 2019, <https://doi.org/10.48084/etasr.2674>.
- [3] M. S. Huda, L. T. Drzal, A. K. Mohanty, and M. Misra, "Effect of fiber surface-treatments on the properties of laminated biocomposites from poly(lactic acid) (PLA) and kenaf fibers," *Composites Science and Technology*, vol. 68, no. 2, pp. 424–432, Feb. 2008, <https://doi.org/10.1016/j.compscitech.2007.06.022>.
- [4] C. Poilâne, Z. E. Cherif, F. Richard, A. Vivet, B. Ben Doudou, and J. Chen, "Polymer reinforced by flax fibres as a viscoelastoplastic material," *Composite Structures*, vol. 112, pp. 100–112, Jun. 2014, <https://doi.org/10.1016/j.compstruct.2014.01.043>.
- [5] B. Pietruszka, M. Gołbiewski, and P. Lisowski, "Characterization of Hemp-Lime Bio-Composite," *IOP Conference Series: Earth and Environmental Science*, vol. 290, no. 1, Mar. 2019, Art. no. 012027, <https://doi.org/10.1088/1755-1315/290/1/012027>.
- [6] P. Sivaranjana and V. Arumugaprabu, "A brief review on mechanical and thermal properties of banana fiber based hybrid composites," *SN Applied Sciences*, vol. 3, no. 2, Jan. 2021, Art. no. 176, <https://doi.org/10.1007/s42452-021-04216-0>.
- [7] S. S. Kumar and V. M. Raja, "Processing and determination of mechanical properties of Prosopis juliflora bark, banana and coconut fiber reinforced hybrid bio composites for an engineering field," *Composites Science and Technology*, vol. 208, May 2021, Art. no. 108695, <https://doi.org/10.1016/j.compscitech.2021.108695>.
- [8] F. Slimani, I. Ghanmi, S. Ghanmi, and M. Guedri, "Production of Thermoplastic Composites reinforced with Posidonia Oceanica Fibers," *Engineering, Technology & Applied Science Research*, vol. 14, no. 2, pp. 13243–13247, Apr. 2024, <https://doi.org/10.48084/etasr.6712>.
- [9] S. Awad, T. Hamouda, M. Midani, E. Katsou, and M. Fan, "Polylactic Acid (PLA) Reinforced with Date Palm Sheath Fiber Bio-Composites: Evaluation of Fiber Density, Geometry, and Content on the Physical and Mechanical Properties," *Journal of Natural Fibers*, vol. 20, no. 1, Apr. 2023, Art. no. 2143979, <https://doi.org/10.1080/15440478.2022.2143979>.
- [10] S. W. Ghorri, G. S. Rao, and A. A. Rajhi, "Investigation of Physical, Mechanical Properties of Treated Date Palm Fibre and Kenaf Fibre Reinforced Epoxy Hybrid Composites," *Journal of Natural Fibers*, vol. 20, no. 1, Apr. 2023, Art. no. 2145406, <https://doi.org/10.1080/15440478.2022.2145406>.
- [11] M. Sh. Al-Otaibi, O. Y. Alotman, M. M. Alrashed, A. Anis, J. Naveen, and M. Jawaid, "Characterization of Date Palm Fiber-Reinforced Different Polypropylene Matrices," *Polymers*, vol. 12, no. 3, Mar. 2020, Art. no. 597, <https://doi.org/10.3390/polym12030597>.
- [12] C.-E. Ferrández-García, A. Ferrández-García, M. Ferrández-Villena, J. F. Hidalgo-Cordero, T. García-Ortuño, and M.-T. Ferrández-García, "Physical and Mechanical Properties of Particleboard Made from Palm Tree Prunings," *Forests*, vol. 9, no. 12, Dec. 2018, Art. no. 755, <https://doi.org/10.3390/f9120755>.
- [13] K. Almi, A. Benchabane, S. Lakel, and A. Kriker, "Potential utilization of date palm wood as composite reinforcement," *Journal of Reinforced Plastics and Composites*, vol. 34, no. 15, pp. 1231–1240, Aug. 2015, <https://doi.org/10.1177/0731684415588356>.
- [14] J.-H. Lee *et al.*, "Mechanical Properties of Biocomposites Using Polypropylene and Sesame Oil Cake," *Polymers*, vol. 13, no. 10, Jan. 2021, Art. no. 1602, <https://doi.org/10.3390/polym13101602>.
- [15] H. H. Abdel-Rahman, R. Al-Juruf, F. Ahmad, and I. Alam, "Physical, mechanical and durability characteristics of date palm frond stalks as reinforcement in structural concrete," *International Journal of Cement Composites and Lightweight Concrete*, vol. 10, no. 3, pp. 175–181, Aug. 1988, [https://doi.org/10.1016/0262-5075\(88\)90007-3](https://doi.org/10.1016/0262-5075(88)90007-3).
- [16] T. Djoudi, H. Mabrouk, D. Youcef, and H. Djemai, "Caractérisation Physico-mécanique des Matériaux Bio-Composites Epoxy/Fibres de Palmier Dattier (Material & Processes)," presented at the International Symposium on Mechatronics and Renewable Energies, El-oued, Algeria, Dec. 2018.
- [17] D. G. Zisopol, I. Nae, A. I. Portoaca, and I. Ramadan, "A Theoretical and Experimental Research on the Influence of FDM Parameters on Tensile Strength and Hardness of Parts Made of Polylactic Acid,"

- Engineering, Technology & Applied Science Research*, vol. 11, no. 4, pp. 7458–7463, Aug. 2021, <https://doi.org/10.48084/etasr.4311>.
- [18] N. Maqsood and M. Rimašauskas, "Characterization of carbon fiber reinforced PLA composites manufactured by fused deposition modeling," *Composites Part C: Open Access*, vol. 4, Mar. 2021, Art. no. 100112, <https://doi.org/10.1016/j.jcomc.2021.100112>.
- [19] A. K. Trivedi, M. K. Gupta, and H. Singh, "PLA based biocomposites for sustainable products: A review," *Advanced Industrial and Engineering Polymer Research*, vol. 6, no. 4, pp. 382–395, Oct. 2023, <https://doi.org/10.1016/j.aiepr.2023.02.002>.
- [20] T. Masri, H. Ounis, L. Sedira, A. Kaci, and A. Benchabane, "Characterization of new composite material based on date palm leaflets and expanded polystyrene wastes," *Construction and Building Materials*, vol. 164, pp. 410–418, Mar. 2018, <https://doi.org/10.1016/j.conbuildmat.2017.12.197>.
- [21] C. Baley, Y. Grohens, and I. Pillin, "Etat de l'art sur les matériaux composites biodégradables," *Revue des composites et des matériaux avancés*, vol. 14, no. 2, pp. 135–166, Aug. 2004, <https://doi.org/10.3166/rcma.14.135-166>.
- [22] B. Asaithambi, G. Ganesan, and S. Ananda Kumar, "Bio-composites: Development and mechanical characterization of banana/sisal fibre reinforced poly lactic acid (PLA) hybrid composites," *Fibers and Polymers*, vol. 15, no. 4, pp. 847–854, Apr. 2014, <https://doi.org/10.1007/s12221-014-0847-y>.
- [23] F.-Z. Semlali AouraghHassani, M. E. Achaby, M.-O. Bensalah, D. Rodrigue, R. Bouhfid, and A. E. K. Qaiss, "Injection molding of short fiber thermoplastic bio-composites: Prediction of the fiber orientation," *Journal of Composite Materials*, vol. 54, no. 30, pp. 4787–4797, Dec. 2020, <https://doi.org/10.1177/0021998320938454>.
- [24] M. Zwawi, "A Review on Natural Fiber Bio-Composites, Surface Modifications and Applications," *Molecules*, vol. 26, no. 2, Jan. 2021, Art. no. 404, <https://doi.org/10.3390/molecules26020404>.
- [25] S. Alwekar, P. Yeole, V. Kumar, A. A. Hassen, V. Kunc, and U. K. Vaidya, "Melt extruded versus extrusion compression molded glass-polypropylene long fiber thermoplastic composites," *Composites Part A: Applied Science and Manufacturing*, vol. 144, May 2021, Art. no. 106349, <https://doi.org/10.1016/j.compositesa.2021.106349>.
- [26] E. Tossou, "Développement de nouveaux composites hybrides renforcés par des fibres de carbone et de lin : mise en oeuvre et caractérisation mécanique," Ph.D. dissertation, Normandie Université, 2019.
- [27] ISO 527-2:2012 *Plastiques: Détermination des propriétés en traction. Partie 2: Conditions d'essai des plastiques pour moulage et extrusion*. ISO, 2012.
- [28] ISO 527-4: 2021 *Plastiques: Détermination des propriétés en traction. Partie 4: Conditions d'essai pour les composites plastiques renforcés de fibres isotropes et orthotropes*. ISO, 2021.

Microstructure and Mechanical Properties of Carbon Fiber Phenolic Matrix Composites containing Carbon Nanotubes and Silicon Carbide

Tayyab Subhani

Department of Mechanical Engineering, College of Engineering, University of Ha'il, Saudi Arabia
ta.subhani@uoh.edu.sa (corresponding author)

Received: 14 February 2024 | Revised: 25 February 2024 | Accepted: 27 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7070>

ABSTRACT

A novel class of hybrid composites was prepared containing carbon fibers along with carbon nanotubes and silicon carbide particles in phenolic resin for improved mechanical performance. The loading of carbon fibers was ~60 wt% while carbon nanotubes and silicon carbide particles were reinforced in the fractions of 0.1 wt% and 5 wt%, respectively. Individually reinforced composites containing 0.1 wt% carbon nanotubes and 5 wt% silicon carbide particles were also manufactured for comparison with hybrid composites. Microstructural and mechanical property characterization was performed using electron microscopy and mechanical testing, respectively. Uniform dispersion of nanometer-scale carbon nanotubes and micrometer-scale silicon carbide particles was observed under microscopy. The pooled effect of carbon nanotubes and silicon carbide particles significantly increased the tensile, compressive, and flexural performance of composites while carbon nanotubes offered greater weight fraction value improvement than silicon carbide particles.

Keywords-carbon fibers; phenolic resin; carbon nanotubes; silicon carbide; composite; mechanical properties

I. INTRODUCTION

The combination of carbon fibers with phenolic resin (Cf-Ph) makes a unique composite material ideally suitable for aerospace structural and functional applications [1]. This is due to the outstanding mechanical and thermal properties of this distinctive composite material stemming from its constituents. Carbon not only has a high heat capacity and vaporization energy, but also high strength and stiffness while maintaining a low density value. On the other hand, phenolic resin produces char in large quantities and releases low amounts of evolved gases during applications at high temperatures [2]. In order to further improve the properties of traditional carbon-based composite systems, the scheme of hybrid reinforcements has been investigated [3, 4]. As such, a secondary reinforcement, made of carbon fibers and carbon nanotubes, is incorporated along with the primary carbon-based one. Nanotechnology made possible the use of nanocomposites where the reinforcement at nanometer scale is incorporated in the matrix material [5]. Furthermore, it allowed for the combination of multiscale composites with reinforcements at different scales, i.e. micrometer and nanometer [6]. Carbon based nanomaterials, typically carbon nanotubes [7, 8], nanodiamonds [9, 10] and graphene nanoplatelets [11], have been extensively used in multiscale composites in polymeric [12], ceramic [13] and metallic [14, 15] matrices.

In the present investigation, multiwalled carbon nanotubes (MWCNTs) and silicon carbide (SiC) particles have been incorporated in carbon fiber phenolic resin composites, both individually and combined, to prepare novel hybrid composites. The underlying aim is to investigate the synergistic effect of carbonaceous and ceramic materials, at nanometer and micrometer scales, with particulate or nanotube morphologies, upon the mechanical properties of carbon fiber phenolic resin composites. The effect of the combination of MWCNTs and SiC particles has previously been investigated in a ceramic matrix, i.e. alumina [16].

II. EXPERIMENT

A. Materials

MWCNTs had an average diameter of 35 ± 5 nm and length of 20 ± 2 μ m. SiC particles had an irregular shape with an average diameter of 22 ± 4 μ m. Carbon fibers were used in discontinuous form with an average length of 10 cm and a diameter of 8 μ m. Barium phenolic resin was used as the matrix material having a 70% phenolic content.

B. Manufacturing

Both MWCNT and SiC particles were stirred together in a mixer (Cole-Parmer Model RZ0414905) for 2 h at a speed of 100 rpm. Then, they were placed in the oven at 100 °C for 30 min. Subsequently, the carbon fibers, MWCNT and SiC

particles were combined appropriately with the phenolic resin. The composite constituents were mixed together for 30 min, in a special laboratory mixer (UTAS-0095, UTest, Turkey). Prepregs were, then, prepared by heat-treating the composite mixture at $\sim 75^\circ\text{C}$ for 1 h. Next, the prepregs were compression-molded using a steel die of appropriate shape, and heat-treated at 180°C for 2 h. Finally, the resulting composites were cut into parts with the desired dimensions to be tested. Four types of composites were prepared: (a) reference carbon fiber phenolic resin (Cf-Ph), (b) carbon fiber phenolic resin composites containing 0.1 wt% MWCNTs (MWCNT-Cf-Ph), (c) carbon fiber phenolic resin composites containing 5 wt% SiC particles (SiC-Cf-Ph), and (d) carbon fiber phenolic resin composites containing both 0.1 wt% MWCNTs and 5 wt% SiC particles (MWCNT-SiC-Cf-Ph). All types of composites had 60 wt% carbon fibers.

C. Characterization

The densities of the composite specimens were determined using a densimeter (GF-300, A&D Japan) and verified employing the Archimedes principle. The relative densities were calculated by dividing the actual with the theoretical densities. The latter were determined making use of the rule of mixtures. Microstructural observation on fracture specimens was performed on a scanning electron microscope MIRA3 TESCON at an accelerating voltage of 5 kV. Tensile, compression and flexural tests were performed against ASTM standards D3039, D0695 and D0790 on a universal testing machine (WDW-30, JINAN) at strain rates of 2mm/min, 1mm/min and 1mm/min, respectively.

III. RESULTS AND DISCUSSION

A. Microstructure

Figure 1 shows the SEM images of fractured composites at two different magnifications with and without secondary reinforcements of MWCNT and SiC particles. Figures 1(a) and 1(b) show the carbon fibers in the reference Cf-Ph composite. It can be seen that the carbon fibers inside the composite are randomly oriented. However, locally, they may attain a certain direction, as observed. The phenolic resin is adhered to the surface of the fibers. Being brittle in nature, the fragmented parts of the phenolic resin are, also, visible. Figures 1(c) and 1(d) show composites comprising of MWCNT and carbon fibers. The random orientation of the latter is clearly visible. The presence of MWCNTs has also been indicated, though it is difficult to identify the MWCNTs at the chosen magnification. Figure 2(a) shows the image of MWCNT-Cf-Ph at a higher magnification. Now, they can clearly be observed without an indication of agglomeration. Figures 1(e) and 1(f) show the images of composites containing SiC particles along with carbon fibers. Figures 1(g) and 1(h) show the images of hybrid composites containing both MWCNT and SiC particles along with carbon fibers. A high magnification image of an MWCNT-SiC-Cf-Ph composite is shown in Figure 2(b) where the presence of MWCNTs is indicated.

B. Density

The theoretical, actual, and relative densities of composites are given in Table I. The relative densities of composites are

shown in Figure 3(a). The Cf-Ph reference composite had a relative density of $99.6\pm 0.2\%$. The addition of 0.1 wt% MWCNTs reduced the density to $98.9\pm 0.3\%$. Similarly, adding 5 wt% SiC particles to the Cf-Ph composite also reduced the density to $99.1\pm 0.3\%$. Cavities and pores are produced by micrometer sized particles during packing, if not properly wetted by the resin. They further reduce the density of the composite materials. Generally, composites have relatively lower densities than the reference materials due to the presence of reinforcements which hinder the polymeric matrix flow. It is to be noted that a small fraction of MWCNTs (0.1 wt%) lowered the density more than a large fraction of SiC particles (5 wt%). This may be due to the large aspect ratio and diameters in the range of nanometers for MWCNTs, providing a large surface area with a flexible nature, thus hindering the densification. Moreover, the tendency of nanoparticles to cluster while dispersing in resins, results in porosity in composites. The combined addition of 0.1wt% MWCNTs and 5wt% SiC particles further lowered the density of composites to $98.5\pm 0.4\%$. Nevertheless, the density values are still in an acceptable range allowing for an estimation of the mechanical performance of composites.

TABLE I. THEORETICAL, ACTUAL AND RELATIVE DENSITY OF COMPOSITES

Composite	Theoretical density (g/cm^3)	Actual density (g/cm^3)		Relative density (%)
		Geometric method	Archimedes method	
CF-Ph	1.50	1.49 ± 0.01	1.49 ± 0.02	99.6 ± 0.2
SiC-CF-Ph	1.49	1.47 ± 0.02	1.48 ± 0.03	99.1 ± 0.3
MWCNTs-CF-Ph	1.49	1.47 ± 0.03	1.48 ± 0.04	98.9 ± 0.3
MWCNTs-SiC-CF-Ph	1.48	1.46 ± 0.02	1.47 ± 0.03	98.5 ± 0.4

TABLE II. TENSILE, COMPRESSIVE AND FLEXURAL STRENGTH OF COMPOSITES

Composite	Tensile strength (MPa)	Compressive strength (MPa)	Flexural strength (MPa)
CF-Ph	58 ± 1	62 ± 2	68 ± 2
SiC-CF-Ph	64 ± 3	64 ± 3	72 ± 3
MWCNTs-CF-Ph	74 ± 2	78 ± 2	86 ± 2
MWCNTs-SiC-CF-Ph	80 ± 3	80 ± 4	94 ± 3

C. Tensile Strength

Figure 3(b) shows the effect the addition of MWCNT and SiC particles has on the tensile strength of the composites. The addition of 0.1 wt% MWCNTs increased the tensile strength by 10 % to 64 ± 3 MPa from 58 ± 1 MPa in the reference composite. On the other hand, a rise in tensile strength of 27% was observed when 5 wt% SiC particles were added, reaching a value of 74 ± 2 MPa. Furthermore, when both secondary reinforcements were introduced, the tensile strength rose by 38 % to 80 ± 3 MPa (Table II). At the nanometer scale, the presence of MWCNTs in phenolic resin may impede the progress of cracks by facilitating a bridging effect and crack deflection [6]. The bridging effect between carbon fibers and phenolic resin may be another possible reason, as discussed elsewhere [12]. Conversely, the inherent brittle nature of SiC particles and their

stiff character may produce localized stress regions around the particles thus deflecting the cracks. To be sure, crack pinning may be another possibility. In hybrid composites, however, the combined effect the presence of both reinforcements has, further improves the mechanical performance.

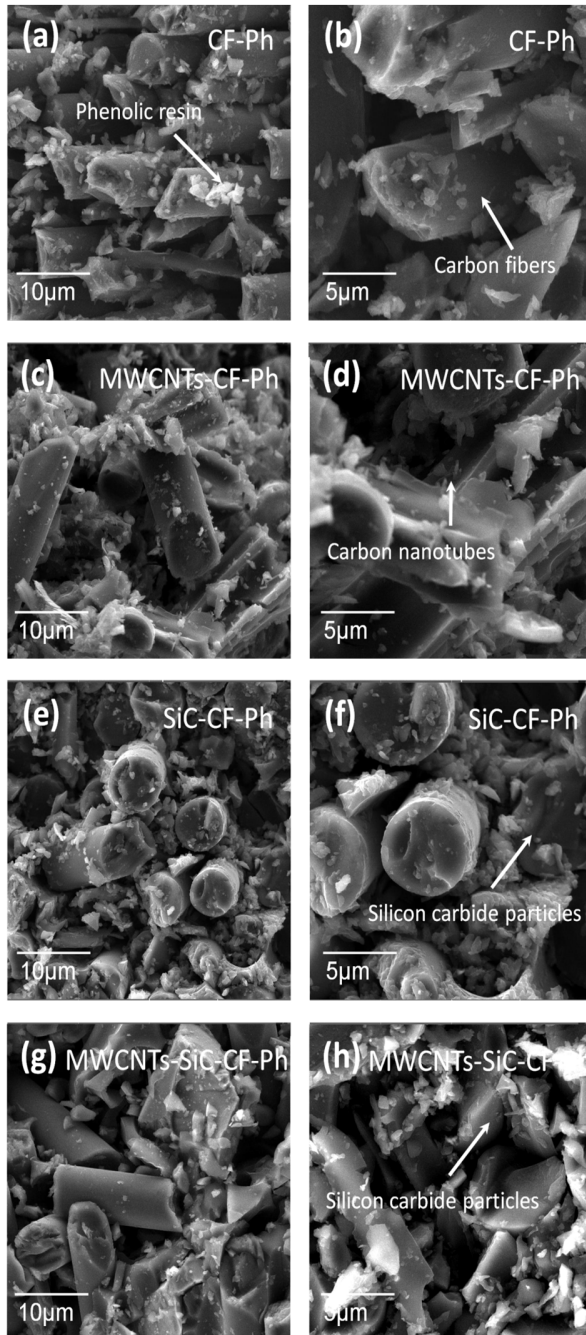


Fig. 1. SEM images of (a-b) reference Cf-Ph, (c-d) MWCNT-Cf-Ph, (e-f) SiC-Cf-Ph, and (g-h) MWCNT-SiC-Cf-Ph composites.

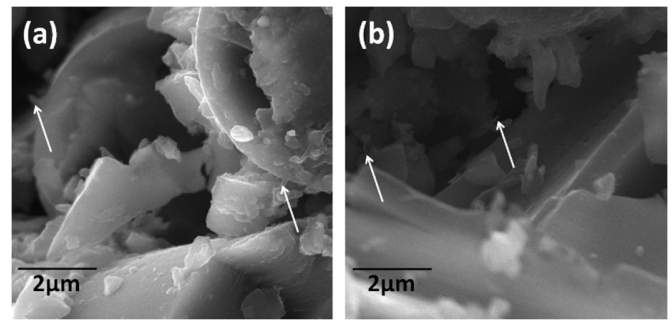


Fig. 2. High magnification SEM images of (a) MWCNTs-Cf-Ph and (b) MWCNT-SiC-Cf-Ph composites.

D. Compressive Strength

Figure 3(c) shows the compressive strength values of composites with and without the addition of secondary reinforcements. Indeed, the behavior of composites under compressive load has been investigated since they can also be used when compressive strength becomes more important than tensile strength. The reference composite shows a compressive strength of 62 ± 2 MPa, which increased by 3% to 64 ± 3 MPa, by adding 0.1 wt% MWCNTs. In comparison, a rise in compressive strength of 26% was observed when 5 wt% SiC particles were added, reaching a value of 78 ± 2 MPa. Furthermore, in hybrid composites, the compressive strength rose by 29 % to 80 ± 4 MPa (Table II). The fact that the tensile strength, when just MWCNTs are added, raised more than the compressive strength will be further discussed below.

E. Flexural Strength

Flexural strength values of the composites are shown in Figure 3(d). A similar trend as before is observed. The reference composite showed the lowest value of 68 ± 2 MPa, which improved to 72 ± 3 MPa by adding 0.1 wt% MWCNTs showing a rise of 6%. On the other hand, the addition of 5 wt% SiC particles improved the value to 86 ± 2 MPa showing a rise of 26%. The combined effect of secondary reinforcements improved the flexural strength to 94 ± 3 MPa resulting in the maximum rise of 38% (Table II). It has been observed [17, 18] that a 66% increase in flexural strength is achieved by adding 1.0 wt% MWCNTs in carbon fiber phenolic matrix composites while a further rise in MWCNT content defects composite centers. In another study, the addition of 1.5 wt% MWCNTs in carbon fiber composites improved the flexural strength up to 12.3% while further increasing the MWCNT contents did not lead to any additional improvement [6].

The presence of MWCNTs plays a vital role in strengthening the composite and changing the nature of fractures. The uniform dispersion of SiC particles and their good interface adhesion improves the strength of composites. The literature indicates a mixed trend in the mechanical properties of polymeric matrix composites containing MWCNTs. Furthermore, the degree of improvement also differs among different studies depending upon the fraction of MWCNTs added and their uniform dispersion [5].

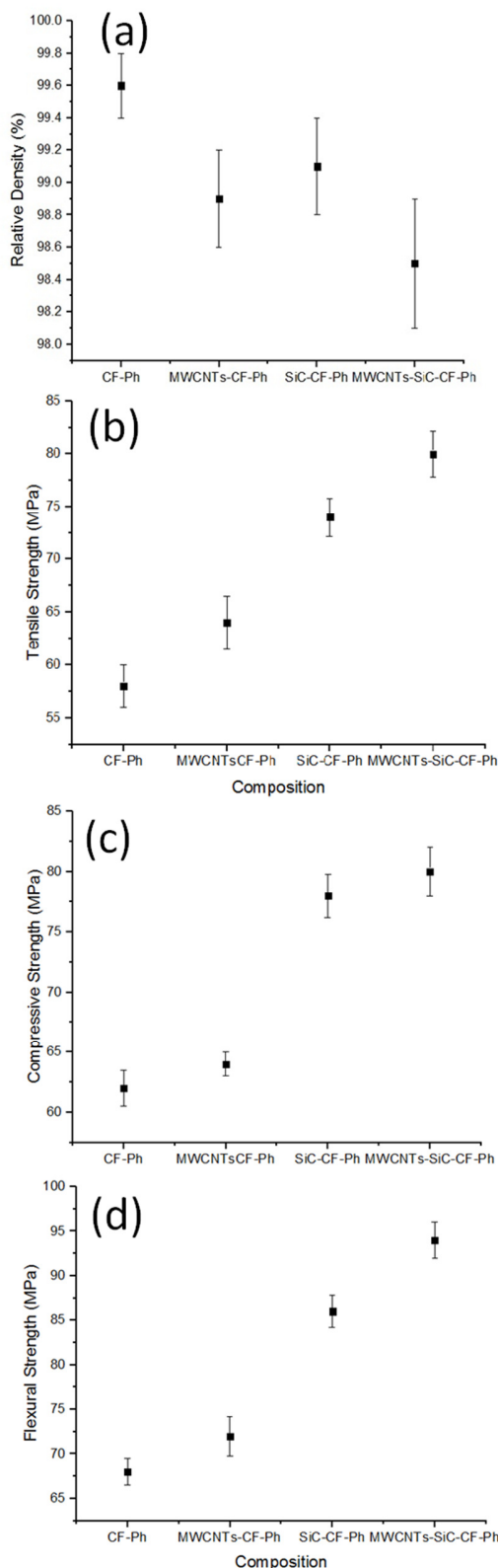


Fig. 3. Graphs showing (a) density, (b) tensile, (c) compressive and (d) flexural strength values of composites.

For example, in [6], 25%, 31% and 40% improvements were observed by adding 0.5 wt%, 1.0 wt% and 1.5 wt% MWCNTs respectively. However, the rising trend reversed after adding 2.0 wt% MWCNTs. It has been thought that the presence of MWCNTs increases the strength as long as they are uniformly dispersed. Upon the development of agglomerates the strength of composites decreases. Due to the flexible nature and curviness of MWCNTs, mechanical keying and interlocking are believed to be some of the key mechanisms by which the mechanical properties of the composites are improved. Further, the MWCNT pull-out is another energy absorbing phenomenon, which increases the strength of composites [6].

Along with strength, the toughness of the composites is also expected to rise. As discussed in a separate study, the presence of secondary reinforcements in the form of particulates and nanotubes, introduces toughening mechanisms such as reinforcement/matrix debonding, pullout of reinforcements, crack deflection and bridging [5].

F. Relative Improvement in the Mechanical Properties

Figure 4(a) shows the percentage relative improvement in tensile, compressive, and flexural strength values of composites containing MWCNTs, SiC particles and both MWCNTs and SiC particles. MWCNTs improved the tensile strength of composites more than the compressive and flexural strengths; Furthermore, the flexural strength improvement was greater than that in compressive strength. This indicates that the addition of MWCNTs has a greater impact on composite properties while under tension than compression. A possible reason may be that CNTs bend or buckle under compression, while under tensile loading CNTs stretch and demand a high load for fracture. Therefore, the addition of MWCNTs offers better tensile than compressive strength. The addition of SiC particles has an almost similar behavior in the three loading modes. Ceramic materials are always strong under compression loading. However, they fracture under tensile loading. The presence, however, of SiC particles, which are strong enough to bear the load coming from the matrix in both tension and compression, hinder the breakage under tension. In flexural testing, both tension and compression loadings are applied, and the presence of SiC in composites improves the strength. In hybrid composites containing both MWCNTs and SiC particles, the compressive values are lower than the tensile and flexural strength ones. This indicates that in hybrid composites, the presence of MWCNTs with their high tensile strength, improves the tensile properties of the composite. Furthermore, the MWCNTs also play a positive role in flexural testing of hybrid composites.

G. Loading Fraction of Secondary Reinforcement and Improvement in Mechanical Properties

Figure 4(b) shows the effect of loading fractions of the secondary reinforcements, to tensile, compressive and flexural strengths. As shown, the addition of only 0.1 wt% of MWCNTs increased tensile, compressive, and flexural strengths by 10%, 3%, and 6% respectively while the addition of 5 wt% SiC particles, improved the tensile, compressive, and flexural strengths by 27%, 26% and 26%, respectively. This indicates that increasing the loading fraction of MWCNTs

should increase the relative improvement in the mechanical properties provided their uniform dispersion and efficient load transfer. This agrees with the results of [19-21]. Moreover, a significant improvement is expected to be observed in tensile strength of composites, as discussed above. Therefore, components in real applications demanding high tensile strength should be manufactured with composites containing CNTs.

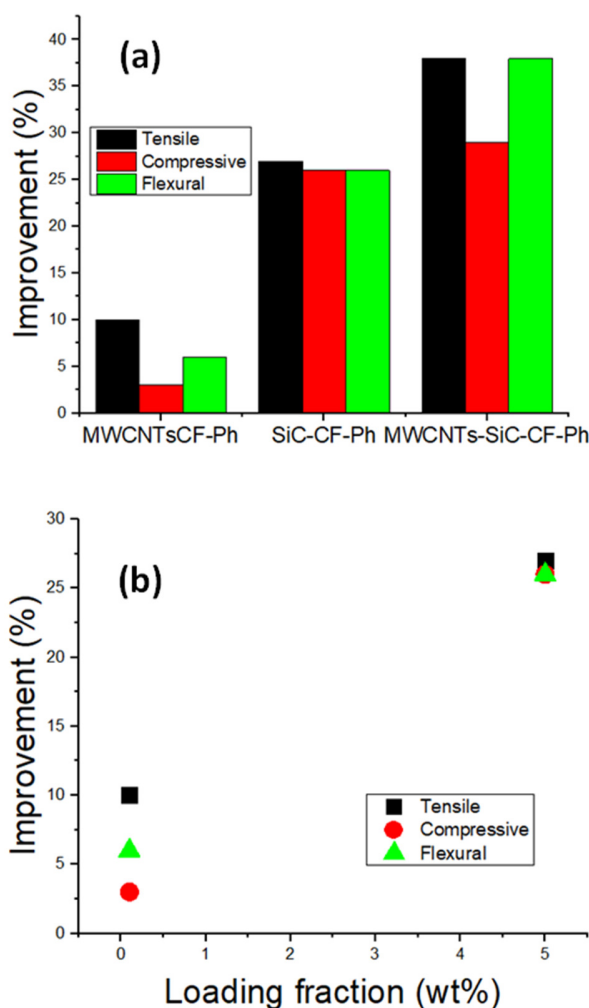


Fig.4. Graphs showing (a) the percentage improvements in tensile, compressive and flexural strength values of composites containing MWCNTs, SiC particles and both MWCNTs and SiC particles and (b) percentage improvement against loading fractions.

IV. CONCLUSION

Carbon fiber phenolic matrix composites were manufactured with nanometer and micrometer sized particulate reinforcements, i.e. carbon nanotubes and silicon carbide particles, so as to investigate the individual and combined effect of the reinforcements on the microstructural evolution and mechanical performance of the composites. The loadings of carbon nanotubes and silicon carbide particles were kept at 0.1 wt% and 5 wt%, respectively, while the content of carbon

fibers was 60 wt%. Microstructural observation revealed the uniform dispersion of both reinforcements without an indication of agglomeration. Individual rises in tensile, compressive and flexural strengths due to the addition of nanotubes and carbide particles were found to be 10%, 3%, 6% and 27%, 26% and 26%, while their combined presence resulted in a rise of 38%, 29% and 38%, respectively.

The present investigation revealed that the high aspect ratio of carbon nanotubes together with their high tensile strength improved the tensile properties of composites both in individually reinforced and hybrid composites. For improved compressive properties, SiC particles are the preferred choice. For combined results, the hybrid approach of manufacturing composites overcomes the limitations of individual incorporation of reinforcements.

Finally, the pooled effect of carbon nanotubes and silicon carbide particles significantly increased the mechanical performance of composites while the addition of carbon nanotubes offered greater value improvement in terms of their weight fractions than silicon carbide particles.

REFERENCES

- [1] G. Pulci, J. Tirillò, F. Marra, F. Fossati, C. Bartuli, and T. Valente, "Carbon-phenolic ablative materials for re-entry space vehicles: Manufacturing and properties", *Composites Part A: Applied Science and Manufacturing*, vol. 41, no. 10, pp. 1483-1490, 2010, <https://doi.org/10.1016/j.compositesa.2010.06.010>.
- [2] A. Saghar, M. Khan, I. Sadiq and T. Subhani, "Effect of carbon nanotubes and silicon carbide particles on ablative properties of carbon fiber phenolic matrix composites", *Vacuum*, vol. 148, pp. 124-126, 2018, <https://doi.org/10.1016/j.vacuum.2017.11.013>.
- [3] M. S. Ahmad, U. Farooq and T. Subhani, "Effect of multiwall carbon nanotubes on the ablative properties of carbon fiber-reinforced epoxy matrix composites", *Arabian Journal for Science and Engineering*, vol. 40, no. 5, pp. 1529-1538, 2015, <https://doi.org/10.1007/s13369-015-1634-9>.
- [4] M. S. Ahmad and T. Subhani, "Thermal and ablative properties of binary carbon nanotube and nanodiamond reinforced carbon fibre epoxy matrix composites", *Plastics, Rubber and Composites*, vol. 44, no. 10, pp. 397-404, 2015, <https://doi.org/10.1179/1743289815Y.0000000036>.
- [5] T. Subhani, M. Latif, I. Ahmad, S. A. Rakha, N. Ali, and A. A. Khurram, "Mechanical performance of epoxy matrix hybrid nanocomposites containing carbon nanotubes and nanodiamonds", *Materials & Design*, vol. 87, pp. 436-444, 2015, <https://doi.org/10.1016/j.matdes.2015.08.059>.
- [6] T. Subhani, B. Shaukat, N. Ali, and A. A. Khurram, "Toward improved mechanical performance of multiscale carbon fiber and carbon nanotube epoxy composites", *Polymer Composites*, vol. 38, no. 8, pp.1519-1528, 2017, <https://doi.org/10.11648/j.nano.20210902.11>.
- [7] M. Khan, A. Rehman, T. Aziz, M. Shahzad, K. Naveed, and T. Subhani, "Effect of inter-cavity spacing in friction stir processed Al 5083 composites containing carbon nanotubes and boron carbide particles", *Journal of Materials Processing Technology*, vol. 253, pp. 72-85, 2018, <https://doi.org/10.1016/j.jmatprotec.2017.11.002>.
- [8] S. A. Rakha, A. A. Khurram, N. Ali, A. Munir, A. Iqbal, T. Subhani, and I. H. Gul, "Preparation of microwave absorbing E-glass/epoxy nanocomposites with low content of binary carbon nanofillers", *Advances in Polymer Technology*, vol. 34, no. 4, 2015.
- [9] U. Zaheer, U. Zulfikar, A. A. Khurram, and T. Subhani, "Improving the performance of conventional glass fiber epoxy matrix composites by incorporating nanodiamonds", *Composite Interfaces*, vol. 25, no. 11, pp. 1005-1018, 2018, <https://doi.org/10.1080/09276440.2018.1454145>.
- [10] A. A. Khurram, A. Khan, I. H. Gul, and T. Subhani, "Glass fiber epoxy matrix composites containing zero and two dimensional carbonaceous

- nano reinforcements", *Polymer Composites*, Vol. 39, No. S4, pp. E2056-E2063, 2018, <https://doi.org/10.1002/pc.24447>.
- [11] I. Ahmad, M. Islam, N. H. Alharthi, H. Alawadhi, T. Subhani, K. S. Munir, S. I. Shah, F. Inam, and Y. Zhu, "Chemical and structural analyses of the graphene nanosheet/alumina ceramic interfacial region in rapidly consolidated ceramic nanocomposites", *Journal of Composite Materials*, vol. 52, no. 3, pp. 417-428, 2018, <https://doi.org/10.1177/0021998317708235>.
- [12] F. S. Awan, M. A. Fakhar, L. A. Khan, U. Zaheer, A. F. Khan, and T. Subhani, "Interfacial mechanical properties of carbon nanotube-deposited carbon fiber epoxy matrix hierarchical composites", *Composite Interfaces*, vol. 25, no. 8, pp. 681-699, 2018, <https://doi.org/10.1080/09276440.2018.1439620>.
- [13] R. Giovanardi, M. Montorsi, G. Ori, J. Cho, T. Subhani, and A. R. Boccaccini, C. Siligardi, "Microstructural characterisation and electrical properties of multiwalled carbon nanotubes/glass-ceramic nanocomposites", *Journal of Materials Chemistry*, vol. 20 no. 2, pp. 308-313, 2010, <https://doi.org/10.1039/B914924D>.
- [14] M. Khan, M. Amjad, A. Khan, R. Ud-Din, I. Ahmad, and T. Subhani, "Microstructural evolution, mechanical profile, and fracture morphology of aluminum matrix composites containing graphene nanoplatelets", *Journal of Materials Research*, vol 32, no. 11, pp. 2055-2066, 2017.
- [15] M. Khan, A. Rehman, T. Aziz, K. Naveed, I. Ahmad, and T. Subhani, "Cold formability of friction stir processed aluminum composites containing carbon nanotubes and boron carbide particles", *Materials Science and Engineering: A*, vol. 696, pp. 552-557, 2017, <https://doi.org/10.1016/j.msea.2017.05.121>.
- [16] I. Ahmad, S. Ahmed, T. Subhani, K. Saeed, M. Islam, N. Wang, and Y. Zhu, "Synergic influence of MWCNTs and SiC nanoparticles on the microstructure and properties of Al₂O₃ ceramic hybrid nanocomposites", *Current Applied Physics*, vol. 16, no. 12, pp. 1649-1658, 2016, <https://doi.org/10.1016/j.cap.2016.10.009>.
- [17] Z. Eslami, F. Yazdani, and M. A. Mirzapour, "Thermal and mechanical properties of phenolic-based composites reinforced by carbon fibres and multiwall carbon nanotubes", *Composites Part A: Applied Science and Manufacturing*, vol. 72, pp. 22-31, 2015, <https://doi.org/10.1016/j.compositesa.2015.01.015>.
- [18] A. Shaheryar, S. Khan, H. Qaiser, A. A. Khurram, and T. Subhani, "Mechanical and thermal properties of hybrid carbon fibre-phenolic matrix composites containing graphene nanoplatelets and graphite powder", *Plastics, Rubber and Composites*, vol. 46, no. 10, 431-441, 2017, <https://doi.org/10.1080/14658011.2017.1385177>.
- [19] Q. A. Kazi, Q. B. Jamali, M. T. Bhatti, U. A. Rajput, I. A. Kazi, and S. H. Solangi, "Development of Vinyl Ester and Reed Composite Material with Weight Fraction of 10% and Reinforcement at Different Orientations", *Engineering, Technology & Applied Science Research*, vol. 9, no. 3, pp. 4306-4309, 2019, <https://doi.org/10.48084/etasr.2710>.
- [20] A. S. Alghamdi, M. Ramadan, K. S. Abdel Halim, and N. Fathy, "Microscopical Characterization of Cast Hypereutectic Al-Si Alloys Reinforced with Graphene Nanosheets", *Engineering, Technology & Applied Science Research*, vol. 8, no. 1, pp. 2514-2519, 2018, <https://doi.org/10.48084/etasr.1795>.
- [21] T. Subhani, "Mechanical Performance of Honeycomb Sandwich Structures Using Three-Point Bend Test", *Engineering, Technology & Applied Science Research*, vol. 9, no. 2, pp. 3955-3958, 2019, <https://doi.org/10.48084/etasr.2597>.

The Influence of Plasma Nitriding Technology Parameters on the Hardness of 18XIT Steel Parts

Nguyen Thai Van

Vinh Long University of Technology Education, Vietnam
vannt@vlute.edu.vn

Le Hong Ky

Vinh Long University of Technology Education, Vietnam
kylh@vlute.edu.vn (corresponding author)

Received: 17 February 2024 | Revised: 4 March 2024 | Accepted: 5 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7089>

ABSTRACT

This article presents the results of the research on the influence of plasma nitriding technology parameters on the working surface hardness of machine parts made of previously hardened 18XIT steel. A total of 27 experiments were conducted on the H4580 Eltrolab instrument. Minitab software was used to process the experimental results. The regression function set up with visual charts was utilized as the basis for analysis of the influence of temperature, time, and gas permeation concentration on the working surface hardness. Analysis of variance (ANOVA) showed that all the nitriding technology parameters influenced the regression function. The permeation temperature TL had the greatest influence on hardness, while the permeation time h and the gas permeation concentration GI had less influence. When the double interaction between the parameters was considered, it was shown that these pairs also had a large influence on the surface hardness, but at different levels.

Keywords-permeation temperature; permeation time; gas permeation concentration; working surface hardness; plasma nitriding; regression equation

I. INTRODUCTION

A permeable layer with high resistance to wear and corrosion and enhanced fatigue resistance can be created on the surface of a part by plasma nitriding. Plasma nitriding is a surface treatment method that uses nitrogen at low pressure and is commonly applied to machine parts made from steel [1]. The infiltration process is performed in a vacuum furnace at low pressure with a gas mixture, composed of H_2 , N_2 , CH_4 , and Ar. [1, 2]. This technology has the benefit of being able to regulate the diffusion layer, which makes it ideal for infiltrating products that have high quality standards [2]. When the parts are exposed to high voltage, the gases become plasma, a state of matter with charged particles. The nitrogen particles in the plasma gain speed when they collide with the parts' surface. This collision causes heat, cleaning, and the formation of a thin and hard layer that resists wear and improves fatigue strength [3]. The permeable layer quality in the nitriding process model depends on various parameters, which are difficult to measure accurately [4-9]. Many researchers have tried to estimate gas nitriding and plasma nitriding processes [5, 6]. Authors in [10] studied how nitride layers form on pure iron during plasma nitriding. They applied Wagner's generalized common model to explain the permeable layer formation and compared it with

the experimental results of plasma nitriding. The predictive models they proposed all follow the rule that the nitrogen content in the permeable layer and the diffusion zone increases in a parabolic way in plasma nitriding [7, 8]. A numerical simulation model for the formation and distribution of nitride layers and nitrogen in the ϵ , γ' , and α phases during plasma nitriding of pure iron was developed in [11]. It was found that the nitride layers are composed of ϵ , γ' , and α phases and that the thickness of each phase grows in a parabolic curve, the same as in previous works [12]. Authors in [13] provided a numerical model of nitrogen diffusion in pure iron, which shows that the diffusion depends on the concentration of nitrogen. They calculated the thickness of the permeable layers to be the square root of the nitriding time at 843 K (570 °C) [7]. Authors in [14] used a gas mixture of NH_3 - H_2 to permeate the gas and form the nitriding layer. They adopted a well-known model based on Fick's law to examine the growth kinetics of the γ' phase during gas nitriding on pure iron. They also developed a simple model to estimate the thickness of the nitride layers in pure iron and constructed a plasma nitriding model using a kinetic model derived from Fick's second law.

The literature on plasma nitriding is abundant, but there is little research on the formation of nitride compounds in the

plasma nitriding layer. Some methods have been developed to calculate and predict the nitrogen concentration and the nitride compounds produced during the nitriding process for pure iron and to measure the depth of the nitrogen diffusion zone in parts after plasma nitriding [4, 5, 14-17].

We did not find any research on plasma nitriding of 18XIT alloy steel parts, especially on hypoid gears. This study uses an ELTROPUL permeation furnace with a pre-generated plasma pulse source that produces stable plasma in all cases and easily penetrates parts with complex geometries.

II. EXPERIMENTAL PART

The passive hypoid gear teeth are made of 18XIT steel according to Russian standards (equivalent to ASTM's SAE 5120). The chemical composition of the elements (%) is: C: 0.17-0.23; Si: 0.17-0.37; Mn: 0.8-1.1; P: < 0.035; S: < 0.035; Cr: 1-1.3; Ni: < 0.3; Ti: 0.03-0.09; Cu: < 0.3; and N: < 0.008 [17]. Surface hardness is often understood as the maximum hardness, although the maximum hardness is found usually a few μm from the surface. The penetration layer depth and the hardness (or micro hardness) are inversely related. The teeth, part of the plasma nitriding sample, are separated by wire cutting and are numbered from 1 to 27, see Figure 1(a).

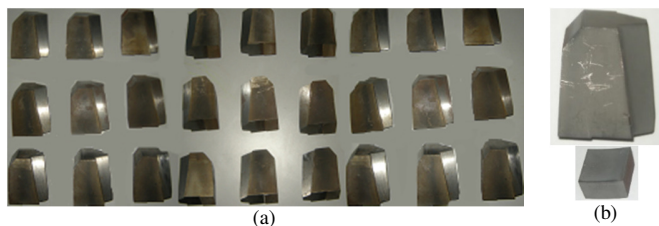


Fig. 1. Sample preparation. (a) The plasma nitriding sample, (b) experimental samples and control sample.

To compare the working hardness after applying the JIS Z 2244-2009 standard, each experiment considered two samples of the same material: one experimental sample and one control sample. The H4580 Eltrolab instrument was used to perform plasma nitriding with a programmable setting. The permeation process is similar for samples of the same material and size, so the voltage V and pressure P are constant [7, 16, 17]. According to the requirements for selecting input factors, the experimental variables were selected, including temperature TL , permeation time h , and gas permeation concentration $G1$. The remaining factors remain unchanged. The influence of plasma nitriding technology parameters on the working surface hardness of the part is described in the general regression equation (1):

$$H = f(h, TL, G1) \quad (1)$$

where H is the hardness of the working surface (HV).

To study how the working hardness of the sample is affected by the plasma nitriding technology parameters, we designed a second-order experiment with 27 trials and used the software of the H4580 Eltrolab device to regulate the nitriding process [16]. The levels of the parameters are shown in Table I.

TABLE I. PARAMETER LEVELS OF THE EXPERIMENTAL DESIGN

Parameter			Level		
Name	Symbol	Unit	(-1)	(0)	(+1)
Temperature	TL	($^{\circ}\text{C}$)	510	530	550
Permeation time	h	h	4	6	8
Gas permeation concentration	$G1$	(l/h)	4	6	8

The power is turned off at high voltage and low pressure to start the surface activation stage, which initiates the first heating process until the temperature reaches 250 to 300 $^{\circ}\text{C}$. The second heating process follows until the temperature reaches the permeation temperature of 510 to 550 $^{\circ}\text{C}$. The permeation period lasts from 4 to 8 h, depending on the experiment parameters, and then the automatic cooling phase begins. Experimental and control samples after permeation in each experiment are shown in Figure 1(b). The Vickers hardness measurement method (HV), according to the JIS Z 2244 - 2009 standard, with the FM-700e instrument, was used to test the microscopic hardness of the experimental sample, because the plasma nitriding layer on its surface is very thin (μm scale). The experiment sample measurement results are shown in Table II. We used the statistical software Minitab 21 to analyze the experimental results (Table II) with a full second order polynomial model [17, 19-21]. The results are shown in Table III. In Table III, we noted that the coefficients of $G1 \cdot G1$ and $h \cdot TL$ have a P-Value greater than the significance level $\alpha=5\%=0.05$ (i.e. confidence interval $CI < 95\%$), which means that these terms have no significant effect in the regression equation, so they were removed, and the experimental results were re-analyzed with other terms. The results are shown in Table IV.

TABLE II. EXPERIMENTAL RESULTS

Trial	h (h)	TL ($^{\circ}\text{C}$)	$G1$ (l/h)	H (HV)
1	4	510	4	632.0
2	4	510	6	644.2
3	4	510	8	654.1
4	4	530	4	670.7
5	4	530	6	676.2
6	4	530	8	678.8
7	4	550	4	699.9
8	4	550	6	700.4
9	4	550	8	694.2
10	6	510	4	638.1
11	6	510	6	641.3
12	6	510	8	653.5
13	6	530	4	670.6
14	6	530	6	677.9
15	6	530	8	676.4
16	6	550	4	703.1
17	6	550	6	701.5
18	6	550	8	693.9
19	8	510	4	652.7
20	8	510	6	652.7
21	8	510	8	659.7
22	8	530	4	684.8
23	8	530	6	681.4
24	8	530	8	684.3
25	8	550	4	717.0
26	8	550	6	706.1
27	8	550	8	701.9

TABLE III. PRELIMINARY COEFFICIENTS OF THE REGRESSION EQUATION

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	674.91	1.14	592.34	0.000	
<i>h</i>	5.006	0.527	9.49	0.000	1.00
<i>TL</i>	27.206	0.527	51.58	0.000	1.00
<i>G1</i>	1.550	0.527	2.94	0.009	1.00
<i>h</i> * <i>h</i>	4.361	0.914	4.77	0.000	1.00
<i>TL</i> * <i>TL</i>	-3.106	0.914	-3.40	0.003	1.00
<i>G1</i> * <i>G1</i>	0.128	0.914	0.14	0.890	1.00
<i>h</i> * <i>TL</i>	-0.358	0.646	-0.55	0.586	1.00
<i>h</i> * <i>G1</i>	-2.758	0.646	-4.27	0.001	1.00
<i>TL</i> * <i>G1</i>	-6.208	0.646	-9.61	0.000	1.00

TABLE IV. COEFFICIENTS OF THE REGRESSION EQUATION AFTER REMOVING THE NON-SIGNIFICANT TERMS

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	674.993	0.920	734.00	0.000	
<i>h</i>	5.006	0.504	9.94	0.000	1.00
<i>TL</i>	27.206	0.504	54.01	0.000	1.00
<i>G1</i>	1.550	0.504	3.08	0.006	1.00
<i>h</i> * <i>h</i>	4.361	0.872	5.00	0.000	1.00
<i>TL</i> * <i>TL</i>	-3.106	0.872	-3.56	0.002	1.00
<i>h</i> * <i>G1</i>	-2.758	0.617	-4.47	0.000	1.00
<i>TL</i> * <i>G1</i>	-6.208	0.617	-10.06	0.000	1.00

Analysis of Variance (ANOVA) was performed for the regression equation of the experimental results, with the option to remove the terms that have non-significant effect on the defined regression equation [17-20, 22, 23]. The analysis of ANOVA table (Table V) has all the P-Values less than $\alpha=5\%$, i.e. the parameters with a significant effect on the regression equation.

TABLE V. ANALYSIS OF VARIANCE

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	7	14542.6	2077.5	454.93	0.000
Linear	3	13816.8	4605.6	1008.53	0.000
<i>h</i>	1	451.0	451.0	98.76	0.000
<i>TL</i>	1	13322.6	13322.6	2917.37	0.000
<i>G1</i>	1	43.2	43.2	9.47	0.006
Square	2	172.0	86.0	18.83	0.000
<i>h</i> * <i>h</i>	1	114.1	114.1	24.99	0.000
<i>TL</i> * <i>TL</i>	1	57.9	57.9	12.67	0.002
2-Way Interaction	2	553.8	276.9	60.64	0.000
<i>h</i> * <i>G1</i>	1	91.3	91.3	19.99	0.000
<i>TL</i> * <i>G1</i>	1	462.5	462.5	101.28	0.000
Error	19	86.8	4.6		
Total	26	14629.4			

By using statistical software, the result table of the regression coefficient evaluation parameters can be determined, as in Table VI. The parameters R^2 (R-squared), R^2 -adj (R²-adjusted), and R^2 -pred (R^2 -predicted) all have values larger than 90%, that is, the regression equation fits perfectly with the experimental data [24, 25]. The regression equation for working hardness *H* is written in natural form as follows:

$$H = 2726 - 6.44h + 10.52TL + 87.17G1 + 1.090h^2 - 0.00776TL^2 - 0.690h \times G1 - 0.1552TL \times G1 \quad (2)$$

TABLE VI. REGRESSION COEFFICIENT EVALUATION PARAMETERS

S	R-sq	R-sq(adj)	R-sq(pred)
2.13697	99.41%	99.19%	98.85%

III. RESULTS AND DISCUSSION

The influence of the input parameters on the output parameters is shown in Figure 2. Regression equation (2) and Figure 2 show that the permeation temperature *TL* has the greatest influence on hardness, while the permeation time *h* and the gas permeation concentration *G1* have less influence.

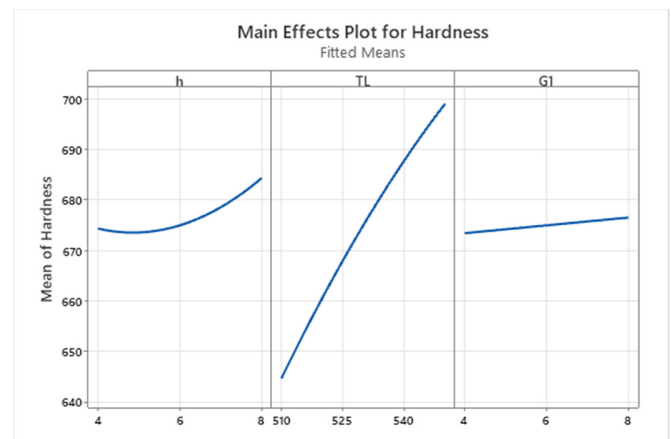


Fig. 2. Chart of the influence of technological parameters on hardness.

In Figure 2, corresponding to the three levels of each parameter, the software calculates the average value of the output response and represents it as a point. Consequentially, in order to find out the technological parameters that most affect the output value, it is possible to make a brief comment on the influence of the input parameters on the output parameter. With permeation time *h*, the effect on hardness is parabolic, with the minimum being near level (0), the effect chart goes from the interval value to the minimum and approaches the maximum value of 685 HV at the level area (+1) of the plan. With the permeation temperature *TL*, the effect on hardness is in the form of parabolic branches, with the smallest value being about 645 HV and the maximum value being about 700 HV, with no extremes in the planning space. Regarding the gas permeation concentration *G1*, the effect on hardness is linear, going from the minimum value at level (-1) and towards the maximum value at level (+1).

The Pareto chart in Figure 3, in addition to visualizing the influence of the coefficients on the result of the regression equation, also shows the influence of the technological parameters on the hardness of the working surface. Accordingly, the permeation temperature *TL* has the greatest influence on hardness, followed by the permeation time *h*, whereas the gas permeation concentration *G1* has the least influence. The double interaction between permeation temperature *TL* and gas permeation concentration *G1* and permeation time *h* also affects the surface hardness.

Figures 4-6 show more clearly the double interaction between the parameters. From these charts, the changing trend of hardness H is known when changing each pair of technological parameters. That is, when the permeation temperature TL and the penetration time h increase together, the working surface hardness H increases, but the effect of the permeation temperature TL is stronger (Figure 4).

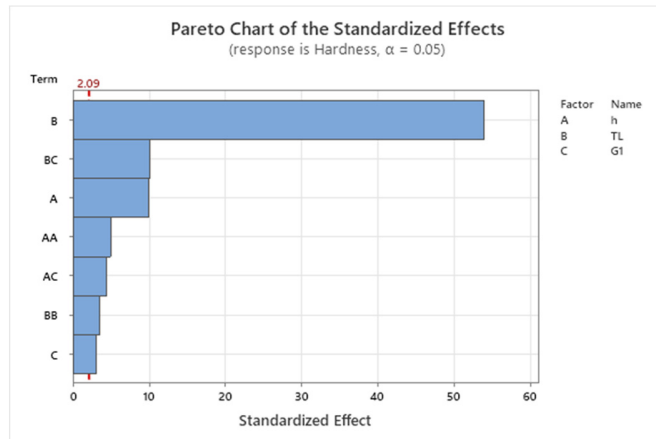


Fig. 3. Pareto chart with input parameters.

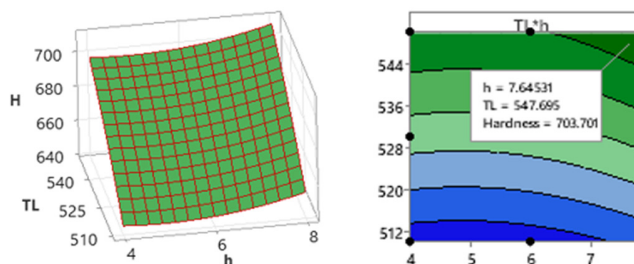


Fig. 4. Charts of interaction between permeation temperature and time to hardness.

Although the permeation time h and the gas permeation concentration $G1$ increase, making the hardness of the working surface H increase, in general, the effect is not much, as shown in (2) and Figure 5.

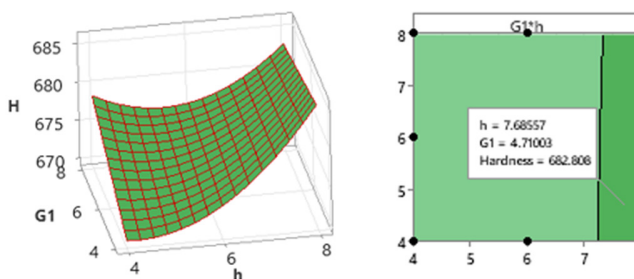


Fig. 5. Chart of interaction between gas permeation concentration and permeation time to hardness.

When the gas permeation concentration $G1$ and the permeation temperature TL increase, they make the hardness of

the working surface H increase, but the effect of the permeation temperature is much larger, as shown in Figure 6.

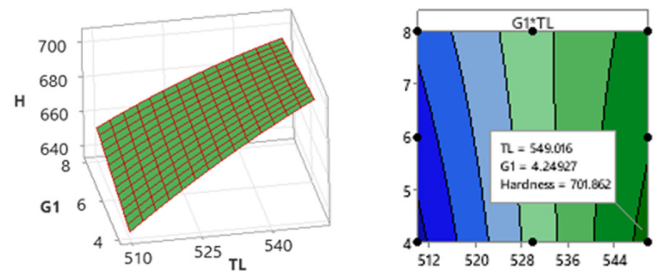


Fig. 6. Chart of interaction between gas permeation concentration and permeation time to hardness.

The above analysis shows that the permeation temperature TL has the greatest influence on hardness. This conclusion is in accordance with the results of [7, 16, 23].

IV. CONCLUSION

Research results on the influence of plasma nitriding technology parameters on the working surface hardness of machine parts made of 18XIT material, specifically, the pair of hypoid gears in the rear axle of light trucks manufactured in Vietnam, have been presented in the article. The regression function is set up with visual charts as a basis for analyzing the influence of temperature, time, and gas permeation concentration on the hardness of the gear working surface. The analysis shows that the permeation temperature TL has the greatest influence on hardness, while the permeation time h and the gas permeation concentration $G1$ have less influence. Specifically, in permeation temperature $TL = 550$ °C, permeation time $h = 8$ h and gas permeation concentration $G1 = 4$ l/h, the surface microhardness reaches 717.0 HV. The results of this study can be used when plasma nitrogen with different machine parts is made of 18XIT material or materials with equivalent chemical, mechanical, and physical properties. In following research, we will investigate how the plasma nitriding technology parameters affect the depth of the plasma nitriding layer of different machine parts made of 18XIT material or similar materials to get optimal results.

REFERENCES

- [1] D. Pye, *Practical Nitriding and Ferritic Nitrocarburizing*. Materials Park, OH, USA: ASM International, 2003.
- [2] ISO/TR 22849:2011. *Design recommendations for bevel gears*. ISO, 2011.
- [3] D. Y. Chung, H. J. Kim, and H. N. Kim, "A Study on the Erosion Characteristics of the Micropulsed Plasma Nitrided Barrel of a Rifle," in *19th International Symposium of Ballistics*, Interlaken, Switzerland, May 2001.
- [4] T. K. Hirsch, A. D. S. Rocha, F. D. Ramos, and T. R. Strohaecker, "Residual stress-affected diffusion during plasma nitriding of tool steels," *Metallurgical and Materials Transactions A*, vol. 35, no. 11, pp. 3523–3530, Nov. 2004, <https://doi.org/10.1007/s11661-004-0189-2>.
- [5] U. Huchel and S. Strömke, *Pulsed Plasma Nitriding of Sintered Parts—Production Experiences*. Baesweiler, Germany: ELTRO GmbH, 2003.
- [6] U. Huchel, *Pulsed Plasma Nitriding of Sintered Parts—Production Experiences*. Baesweiler, Germany: ELTRO GmbH, 2003.

- [7] D. Vu, "Prediction of the Adhesion Strength of Coating in Plasma Spray Deposition," *Engineering, Technology & Applied Science Research*, vol. 13, no. 2, pp. 10367–10371, Apr. 2023, <https://doi.org/10.48084/etasr.5673>.
- [8] L. Maldzinski, W. Liliental, G. Tymowski, and J. Tacikowski, "New possibilities for controlling gas nitriding process by simulation of growth kinetics of nitride layers," *Surface Engineering*, vol. 15, no. 5, pp. 377–384, Oct. 1999, <https://doi.org/10.1179/026708499101516740>.
- [9] O. Ortac, "Microstructural and mechanical characterization of nitrogen ion implanted and plasma ion nitrided plastic injection mould steel," M. S. thesis, Izmir Institute of Technology, Izmir, Turkey, 2003.
- [10] V. I. Dimitrov, G. Knuyt, L. M. Stals, J. D'Haen, and C. Quaeys, "Generalized Wagner's diffusion model of surface modification of materials by plasma diffusion treatment," *Applied Physics A*, vol. 67, no. 2, pp. 183–192, Aug. 1998, <https://doi.org/10.1007/s003390050757>.
- [11] M. Yan, J. Yan, and T. Bell, "Numerical simulation of nitrided layer growth and nitrogen distribution in ϵ -Fe₂-3N, γ '-Fe₄N and α -Fe during pulse plasma nitriding of pure iron," *Modelling and Simulation in Materials Science and Engineering*, vol. 8, no. 4, Apr. 2000, <https://doi.org/10.1088/0965-0393/8/4/307>.
- [12] J. Baranowska, *Plasma Nitriding and Plasma Immersion Ion Implantation*. Szczecin University of Technology, 2008.
- [13] T. Belmonte, M. Gouné, and H. Michel, "Numerical modeling of interstitial diffusion in binary systems. Application to iron nitriding," *Materials Science and Engineering: A*, vol. 302, no. 2, pp. 246–257, Apr. 2001, [https://doi.org/10.1016/S0921-5093\(00\)01830-X](https://doi.org/10.1016/S0921-5093(00)01830-X).
- [14] M. Keddad, "Surface modification of the pure iron by the pulse plasma nitriding: Application of a kinetic model," *Materials Science and Engineering: A*, vol. 462, no. 1, pp. 169–173, Jul. 2007, <https://doi.org/10.1016/j.msea.2006.02.459>.
- [15] M. B. Karamiş, "Some effects of the plasma nitriding process on layer properties," *Thin Solid Films*, vol. 217, no. 1, pp. 38–47, Sep. 1992, [https://doi.org/10.1016/0040-6090\(92\)90603-9](https://doi.org/10.1016/0040-6090(92)90603-9).
- [16] H. K. Le, "A Study on the Influence of Plasma Nitriding Technology Parameters on the Working Surface Deformation of Hypoid Gears," *Engineering, Technology & Applied Science Research*, vol. 12, no. 6, pp. 9760–9765, Dec. 2022, <https://doi.org/10.48084/etasr.5365>.
- [17] T. H. Le, V. B. Pham, and T. D. Hoang, "Surface Finish Comparison of Dry and Coolant Fluid High-Speed Milling of JIS SDK61 Mould Steel," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 8023–8028, Feb. 2022, <https://doi.org/10.48084/etasr.4594>.
- [18] "Марочник сталей и сплавов онлайн - скачать PDF," *Metal Place*. <https://metal.place/ru/wiki/18khgt>.
- [19] D. T. Do and N.-T. Nguyen, "Applying Cocoso, Mabac, Mairca, Eamr, Topsis and Weight Determination Methods for Multi-Criteria Decision Making in Hole Turning Process," *Strojnícky časopis - Journal of Mechanical Engineering*, vol. 72, no. 2, pp. 15–40, Nov. 2022.
- [20] D. D. Trung, "Influence of Cutting Parameters on Surface Roughness in Grinding of 65G Steel," *Tribology in Industry*, vol. 43, no. 1, pp. 167–176, Mar. 2021, <https://doi.org/10.24874/ti.1009.11.20.01>.
- [21] K. P. Bui, T. D. Duc, C. Ngo, and M. N. Dinh, "Research on Optimization of Plunge Centerless Grinding Process using Genetic Algorithm and Response Surface Method," *Research on Optimization of Plunge Centerless Grinding Process using Genetic Algorithm and Response Surface Method*, vol. 4, no. 3, pp. 207–211, 2015.
- [22] V. C. Nguyen, T. D. Nguyen, and D. H. Tien, "Cutting Parameter Optimization in Finishing Milling of Ti-6Al-4V Titanium Alloy under MQL Condition using TOPSIS and ANOVA Analysis," *Engineering, Technology & Applied Science Research*, vol. 11, no. 1, pp. 6775–6780, Feb. 2021, <https://doi.org/10.48084/etasr.4015>.
- [23] B. T. Danh and L. H. Ky, "Optimization of Technological Parameters when Plasma Nitriding the Gear Working Surface," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 11006–11010, Jun. 2023, <https://doi.org/10.48084/etasr.5946>.
- [24] "Model summary table for Fit Regression Model," *Minitab*. <https://support.minitab.com/en-us/minitab/21/help-and-how-to/statistical-modeling/regression/how-to/fit-regression-model/interpret-the-results/all-statistics-and-graphs/model-summary-table/>.
- [25] D. C. Montgomery, *Design and Analysis of Experiments*, 10th ed. Hoboken, NJ, USA: Wiley, 2019.

Innovative Technological Solutions for Environmental Sustainability in Chinese Engineering Practices

Azhar Ud Din

School of Management, Harbin Institute of Technology, China
azhar.marwat@hit.edu.cn (corresponding author)

Yang Yang

School of Management, Harbin Institute of Technology, China
yfield@hit.edu

Muhammad Inam Makki Khan

School of Economics and Management, Chang'an University, China
inamkhan@chd.edu.cn

Waqas Khuram

School of Management, Harbin Institute of Technology, China
waqaskhuram@hit.edu.cn

Received: 20 January 2024 | Revised: 9 February 2024 | Accepted: 24 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6935>

ABSTRACT

The Chinese government announced the clear goal of attaining carbon neutrality by 2060, in order to gradually achieve net-zero carbon dioxide (CO₂) emissions, whose impact on global warming needs to be reduced while also a sustainable industry needs to be promoted. Recognizing the critical role of Green Human Resource Management (GHRM) in supporting green innovation and achieving the carbon neutrality agenda, this study aims to fill a research gap by emphasizing this overlooked nexus. The former examines the influence of GHRM, green innovation, and carbon neutrality on environmental performance by carefully analyzing the current literature on China's achievement of carbon neutrality and its implications for environmentally friendly performances. The current study assesses the planning frameworks of the country, explores the concept of achieving carbon neutrality, and evaluates the practical implications.

Keywords: *green human resource management; carbon neutrality; environmental performance; green innovation; China*

I. INTRODUCTION

Environmental pollution and the necessity for sustainable resource utilization are considered important worldwide issues due to environmental concerns and resource limitations and constitute a top priority for communities, governments, and residents. Businesses must take responsibility to address these challenges. Many businesses adopt environmentally friendly initiatives to gain a competitive edge and enhance their Environmental Performance (EP). These organizations allocate resources to address sustainability challenges and respond to the demands of stakeholders [1, 2]. Implementing Green Human Resource Management (GHRM), Green Innovation (GI), and carbon neutrality policies can help leverage the impact of these difficulties. GHRM refers to adopting

environmentally friendly practices in various aspects of employee management, such as recruitment, training, involvement, leadership, rewards, compensation, performance appraisal, and orientation to environmental corporate social responsibility. These practices aim to assist a company in achieving its environmentally friendly goals [3-5]. Researchers have investigated the impact of GHRM on financial performance [6], EP [7], and workplace practices [8]. It is widely recognized that the environmental effects of GHRM procedures can be considered at every stage of the process because GHRM practices help organizations implement and maintain an Environmental Management System (EMS), which in turn helps individuals achieve better EP [9]. GHRM is essential to company administration for various reasons,

including environmental benefits, professionalism, and enhancing the company's appeal. Human Resource Management (HRM) used to be concentrated on the influence of particular practices on organizational effectiveness [10]. Aligning GHRM policies can significantly impact the environment and organizational performance [11]. As a result, most current research on GHRM focuses on how GHRM practices affect efficiency, since they can potentially improve an organization's financial effectiveness, productivity and hence its economic performance [12-14].

GI is an innovation strategy for green processes and products that seeks to reduce energy use, stop pollution, create environmentally friendly products, recycle waste, and manage the environment [15, 16]. To enhance the performance of a business, green operational practices might be employed to mitigate environmental consequences. These initiatives include implementing GI and utilizing green technology to develop environmentally friendly products. The GI classification entails elements of the manufacturing process and product design. GI in manufacturing can minimize environmental impact throughout product acquisition, production, and delivery, whereas it is able to modify existing product designs to minimize the environmental impact [17].

Carbon neutrality is the process of lowering emissions of greenhouse gases (GHG) by the implementation of various strategies, including energy efficiency, the use of renewable energy sources, carbon offsetting, and carbon sequestration methods [18]. Carbon neutrality currently involves offsetting an equivalent amount of carbon and remarkably reducing gross carbon emissions, resulting in nearly zero impact on the ecosystem. Carbon neutrality is crucial for sustainable financial growth as most governments including China strive for ensuring sustainable EP [19-20]. China is presently the globe's biggest carbon emitter, having contributed 28% of all CO₂ emissions in 2019 due to its rapid economic expansion. Carbon intensity has increased by 45% since the preceding decade, surpassing the emissions produced by the US in 2007, the EU, and the UK collectively [21]. The downward economic development from 2013 to 2016 indicated that China's CO₂ emissions reduced [21, 22]. After that, however, emissions began to increase again [22, 23] suggesting that long-term reduction remains a significant policy issue for China as for the rest of the globe. The fast economic growth of China over the past three decades has resulted in several adverse side effects, such as resource depletion and environmental degradation, making it extremely difficult for the country's economy and society to expand sustainably [24]. Carbon Dioxide Emissions (CDE) are an important contributor to the climate state [25]. China is under intense pressure to reduce emissions because it is the largest producer of CDE in the world and one of the nations which have determined to attain carbon neutrality and eradicate carbon emissions by 2060 [26].

Human resources have long been acknowledged as essential to economic success for developed and developing countries [27]. However, there are always expenses associated with using commodities for the economy, which are more evident for the ecosystem [27, 28].

Eco-design is the implementation of environmental considerations at every stage of the product creation process, aiming to diminish the environmental effects during the whole lifespan of the product [30]. The former is a component of environmentally friendly advancements and has been discovered to positively affect financial productivity [31]. Eco-friendly and good green performance influences vendor and consumer behavior, improving the financial success of worldwide operations [26, 32, 33]. Companies may be able to supply items that are barely damaging to the environment, boosting their EP by monitoring and teaching suppliers [34]. Emphasis on GI that uses eco-friendly manufacturing methods and reduces actual prices might help businesses achieve greater effectiveness while consuming fewer commodities. Nevertheless, the relationship between GHRM, carbon neutrality, GI, and green advantages needs further investigation. In order to look into the inverse relationships between various functions, the way green management is scattered throughout multiple functional areas is studied in [35]. Even though GI and GHRM are widely acknowledged to benefit their surroundings, there has not been much investigation into how these can be combined [36].

Although academics have mainly concentrated on firms' external difficulties, GHRM practice describes the inner procedures. The impact of carbon neutrality and GHRM on the EP or innovation is scarcely highlighted. This domain needs more research to determine how GHRM and GI help the triple bottom line and a company's economic and ecological efficiency. After acknowledging the need for carbon neutrality for sustainable EP, China's path to carbon neutrality should be more thoroughly examined. There is a big research gap, so, a review of the studies in this field is needed. Such a review would assist researchers in identifying significant concerns and research barriers and improve follow-up research. This study explores green management systems' effects on human resources, innovation, carbon neutrality, and green performance in China.

II. RESEARCH METHOD

Given the growing significance of sustainability, businesses must attain higher levels of economic growth and environmental and social performance. This necessity also emerges due to innovations and the critical role that GHRM functions and carbon neutrality can play in the latter. This paper explores the link between GHRM and carbon neutrality on the EP in the presence of GI. A conceptual model that connects these crucial functions and structures with the corresponding elements and dimensions is created. Additionally, this research explicitly utilizes the conceptual framework developed in [11], which combines HRM, i.e. organizational sustainability, and green organizational activities that further modify and extend the EP, which is a crucial aspect of environmental sustainability. Earlier studies have not made connections between the elements of EP and the elements of GHRM and carbon neutrality. The research steps are:

- Step 1: Problem identification by reviewing the existing literature for an integrative model connecting the concepts.

- Step 2: An analysis of specific sources or studies on the topic.
- Step 3: An idea for integrating carbon neutrality, GHRM, and EP based on the traits of GI is presented.
- Step 4: Presentation of the results with recommendations for additional study and their implications for model implementation and knowledge advancement.

The lack of existing research connecting carbon neutrality, GHRM, and EP led to the creation of a theoretical or conceptual model combining the three components of operations management (step 1). The need to approach the topic holistically is reinforced to ensure that the model is helpful for academics and practitioners and has significant future research potential (steps 2–4). This approach should be guided by existing theory and research.

The databases and search engines utilized for this study included Google Scholar, Web of Science, Science Direct, Emerald, Springer, MDPI, and Wiley. Every database needed a unique and specific approach. Our analysis focused on the following basic keyword combinations: "green human resource management," "GI," "EP," and "carbon neutrality." All references were gathered into a Mendeley database. Three different researchers independently assessed the references and chose the final collection of papers for further study. To ascertain the inclusion of documents in this review that encompassed contemporary subjects and concerns within the sectors of interest, the researchers carefully examined the titles and abstracts of each article to determine their suitability for inclusion in the selection. The relevant literature was carefully selected manually. The text was rigorously analyzed to identify themes and terminology related to the current study.

III. GREEN HUMAN RESOURCE MANAGEMENT

GHRM refers to a system that develops and executes HRM practices to emphasize the influence of the actions of an organization on the surrounding environment [37]. GHRM refers to implementing environmentally friendly practices in various aspects of employee management, such as recruitment, training, involvement, leadership, rewards, compensation, performance appraisal, and orientation to environmental corporate social responsibility. These practices aim to assist a company in achieving its environmentally friendly goals [3, 4]. Previous studies have investigated the relationship between GHRM and an entity's economic success [6], sustainability practices [7], and personnel attributes [38–40]. However, there is a lack of research on the organizational social aspect. This research aims to determine the fundamental connection between GHRM practices, GI, and carbon neutrality. The study also tries to fill the gap between GHRM and sustainable innovation literature.

Sustainable development and social consciousness have led to an increased focus on human resource departments, along with the economic, regulatory, and other components of greener manufacturing. Although human resources are increasingly viewed as "the soul of the company," particularly in the context of the environment, finance is still commonly acknowledged as the lifeblood of business [41]. A business can

be rebuilt from the rubble with an identical replica of its previous structure and still function successfully. However, losing crucial human capital may cause a business to slow down. So, human resources serve as a company's knowledge foundation [42], inspiring other businesses to make additional investments in its potential. International employment practices necessitate the creation and managerial staff of global HR proficiency in a way that guarantees the attributes influencing their business strategy endorses the accomplishment of green growth goals. These practices give an enterprise significant dominance in the focus of a competitive environment. All firms will undoubtedly see a boost in their productivity, profitability, and long-term growth and development when they employ optimal resource allocation techniques [43]. Yet, the main issues are locating the benchmarks and implementing best practices. Considering human resources' attitudes, ideas, disposition, and behavior in achieving their institution's objectives related to sustainable development, an important question still has to be answered. The current corporate world regularly uses phrases like going green and sustainability, demanding that every corporation follow the highest eco-friendly and resource-efficient standard practices. The Go-Green Movement has produced green jobs in businesses by advocating the most efficient manufacturing techniques that decrease the adverse consequences of environmental contamination, wastage, and other dangerous substances. According to UBLS (US Bureau of Labor Statistics) (2013), green jobs address to either the manufacturing or the provision of services that contribute to the conservation and preservation of our precious resources. Also, green jobs may refer to a type of employment where the institutional staff is responsible for making their firms' production process more reliable, sustainable and eco-friendly, thereby reducing its negative impact on the planet. Reduced negative ecological impacts have been achieved by investments in green resources and cleaner production methods [44]. The "green wave" [45] across nations has strengthened the significance of producing higher possibilities, especially in businesses with a reliable strategy and management objectives. Subsequently, GHRM appeared as a unique discipline from both theory and practice standpoints due to the rising ecologic concerns [46, 47]. Managers are willing to make investments in maintaining and expanding their understanding in hopes of transforming their organization into one that is economically viable via eco-innovation and individual development [24]. By aiding an organization in sustaining its understanding through optimal eco-friendly practices that produce no emissions of ecological disharmony, GHRM delivers additional economic strategies for accomplishing sustainability objectives [48].

Green rewards and performance administration practices can help align employees' behavior with an institution's operational objectives [49]. Although GI is the element of ecological administration that identifies specific sustainability problems, green performance is an effective source to boost executive environmental responsibility and, as a result, the eagerness to participate in eco-friendly innovation [46]. Moreover, encouraging innovative green products, process ideas, and environmental activities can foster an innovative work atmosphere [50]. According to the relevant research,

HRM systems can positively impact the development of new products or processes [51]. Hence, combining HRM practices with other practices can significantly affect innovation more than using individual techniques alone [52]. Even though GHRM research has attracted a great amount of attention recently, much of the earlier work in this area [53] focuses on how businesses can become aware of, adopt, and apply GHRM strategies [49]. A literature review is necessary to research the potential influence of applying GHRM practices in organizations. An in-depth tool enables categorizing and reducing a sizable quantity of information previously obtained using qualitative assessment methods that use precise and consistent decision standards to find relevant gaps in the literature [54].

IV. GREEN INNOVATION

Innovation refers to the systematic conversion of ideas into novel products, services, or processes that enable a company to gain a competitive advantage in the market [55]. GI (GI) is crucial for promoting green growth [56]. GI includes different aspects of innovation that specifically target energy conservation, recycling of waste materials, prevention of pollution, reduction of waste, creation of eco-friendly products, and implementation of an EMS [57]. The concept of GI aims to mitigate the adverse environmental impacts of production and activities by improving methods, techniques, structures, products, and management strategies [58, 59]. By utilizing these innovations, businesses are better able to encourage green growth and address environmental protection issues. GI can be categorized into three distinct forms: process innovation, product innovation, and Green Management Innovation (GMI) [17, 60]. Green Process Innovation (GPI) uses green process technologies, including clean production, pollution control, pollution prevention, environmental efficiency, and recirculation. These technologies involve the implementation of novel or enhanced activities that contribute to creating goods or services from a sustainability perspective [61, 62]. At the same time, GPI refers to developing goods or services by an enterprise that exhibits novel or significantly enhanced environmental characteristics [63, 64]. On the other hand, GMI involves implementing innovative environmental management practices within a corporation [17, 65]. GMI focuses on green management techniques within organizations, emphasizing novelty-level invention, whether initiated or adopted by other firms. Businesses with GI are remarkably successful and perform better than their rivals [66]. These businesses employ environmentally friendly resources and expertise to efficiently and promptly meet consumers' needs and serve as sources for organizations to increase their EP [1] while providing intangible assets and resources.

The effectiveness of EP is contingent upon organizational innovativeness and creativity, as highlighted in [67, 68]. This implies that EP should go beyond simply complying with established standards and policies [69, 71]. An organization's EP is influenced by various factors, including the ecological impact of its products, processes, resource allocation, and compliance with environmental regulations [72]. According to the literature, the effectiveness of ecological environment performance depends on several factors, which include the

quality of eco-friendly products, implementing green procedures and innovations in merchandise, incorporating nature-consistent elements into organizational operation models, and integrating these elements into merchandise development processes [71-74]. GI is commonly integrated into an organization's environmental management plan to enhance the EP [75, 77]. Authors in [78] emphasized that innovation in the context of environmentally friendly practices involves implementing strategies to minimize waste, mitigate global warming, and optimize the utilization of natural resources. The non-manufacturing/service businesses, such as banks, hotels, and manufacturing industries, can prioritize GIs to ensure their investments are directed toward sustainable resources.

V. CARBON NEUTRALITY

Carbon neutrality is the process of lowering GHG emissions by the implementation of various strategies, including energy efficiency improvements, the use of renewable energy sources, carbon offsetting, and carbon sequestration methods [18, 80]. The state of carbon neutrality is reached when human activities take in the same amount of carbon dioxide (CO₂) they release [79]. Since it is not possible to attain zero CO₂ emissions, it is necessary to achieve a balance between residual emissions and removal emissions through the use of carbon offsetting or renewable energy [81]. The elimination of anthropogenic GHG emissions is referred to as net-zero emissions and the climatic indices that are utilized to monitor these emissions are those that determine the phrase [82]. In general, the objective of carbon neutrality is to reduce carbon emission's impact on the climate by lowering and compensating for those emissions.

VI. PATHWAY TOWARD CARBON NEUTRALITY

A. Negative Emission Technology (NET)

NETs, concerned with ecological security and international climate policy, are crucial for meeting the 1.5 °C temperature reduction goal. Ecological development and environmental impact are key research areas of NET and GHRM. Research studies of NETs focus on the expenses and advantages of technological innovations and the socioeconomic effects of NET deployment. A range of approaches, such as geoinformation research, lifecycle evaluation, and raw material analysis, are employed to study the accessibility of the financial support needed to facilitate the creation of NETs and their external environmental repercussions [83, 86, 87].

B. Zero-Carbon Technology

The carbon peak and neutrality objectives will force previously carbon-intensive firms to innovate and go through many transitions because businesses have historically been the biggest emitters of GHG emissions. Present investigations concentrate on microbial synthetic structures, carbon-neutral methanol innovation, advanced biofuel, photoelectrically degradation, and zero-carbon construction materials. Green manufacturing is an essential first step in predominant view and modernizing traditional industries. Investigators should consider conventional construction and scientific constraints to determine whether these processes have low or zero carbon

emissions. To promote the decarbonization of activities, appropriate carbon emission assessment and accounting practices must be adopted [88, 90].

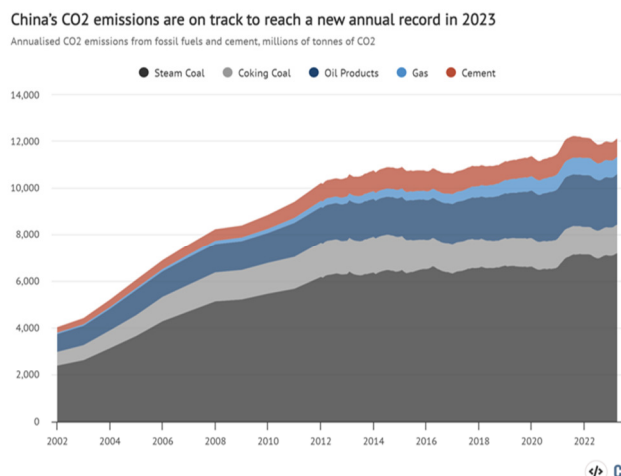


Fig. 1. China's CO₂ emissions, categorized into cement, gas, oil, coking coal, and steam coal. Emissions are calculated using data from the National Bureau of Statistics regarding the production of various fuels and cement, China Customs data on imports and exports, and WIND Information data on inventory changes. The calculations utilize default emission factors provided by the Intergovernmental Panel on Climate Change (IPCC) and annual emissions factors per ton of cement production up until 2019. The monthly readings are converted to yearly fuel consumption data in the Statistical Communiques and the annual Yearbooks published by the National Bureau of Statistics. Chart created by Carbon Brief [87].

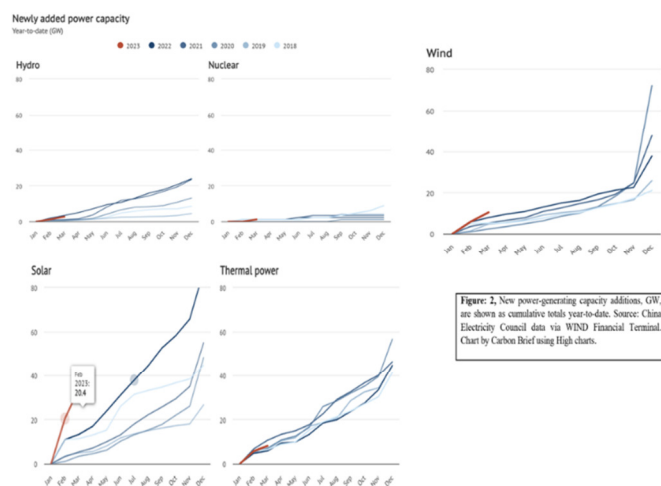


Fig. 2. Total number of new power-generating capacity additions, measured in GW. Source: China Electricity Council data via WIND Financial Terminal. Chart created by Carbon Brief [87].

C. Carbon Pricing Mechanism

There is a strong record of using economic incentives to decrease CO₂ emissions. By 2020, 61 carbon pricing schemes, which will account for 22% of global GHG emissions, were expected to be globally implemented. Due to the novel features of every state, different pricing techniques can be applied to various economic sectors in different ways. Implementing

carbon pricing plans would be costly, whereas the fairness of the upcoming regional and industrial growth is a concern. Determining the types of organizations that must reduce emissions and validating and maintaining a watchful eye on carbon emission data are essential to efficient corporate carbon management. The guiding principles of market mechanism and design should ensure that market forces reduce emissions regardless of the pricing mechanism and efficiency [91, 94].

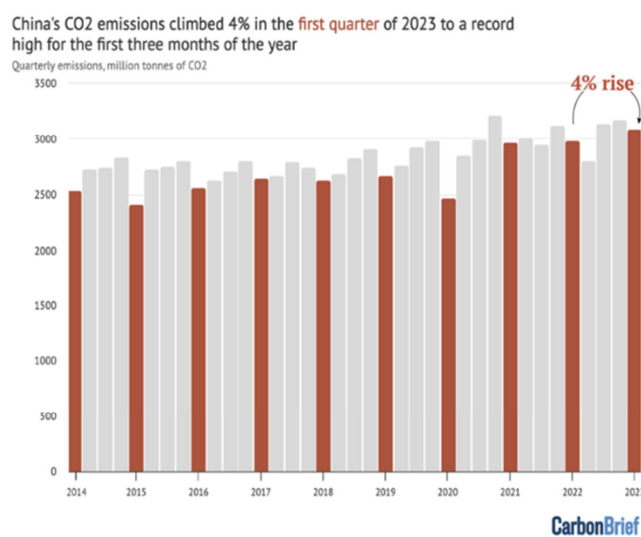


Fig. 3. China's quarterly emissions from fossil fuels and cement, measured in million tons of CO₂. Emissions are calculated using data from the National Bureau of Statistics regarding fuel and cement production, China Customs data on imports and exports, and WIND Information data on inventory changes. The calculations utilize default emissions factors provided by the IPCC and annual emissions factors per ton of cement production until 2019. The monthly statistics are adjusted to match the yearly fuel consumption data in the Statistical Communiques and the National Bureau of Statistics' annual Yearbooks. Chart created by Carbon Brief [87].

D. Green Performance under a Scenario of Carbon Neutrality

The term green performance refers to the impact on an entity's, such as a business, community, or country environment and future sustainability. Green performance is essential in reaching carbon neutrality, and tries to balance the removal of GHG and their emission from the atmosphere [95]. Numerous aspects of green performance have the potential to contribute towards achieving carbon neutrality, such as energy efficiency, renewable energy utilization, circular economy adaptation, Sustainable Transportation, and carbon offset and removal. Improving energy efficiency is a critical component of lowering carbon emissions. Energy-efficient technologies and practices can help businesses reduce their energy use and, as a result, their carbon footprint. For example, research has shown that energy-efficient buildings can dramatically reduce GHG emissions [96]. Green performance can be improved by implementing energy management systems, renewable energy sources, and energy-saving techniques [97, 98]. To become carbon neutral, we must switch to renewable energy sources. Renewable energy sources like solar, wind, and hydropower do not increase GHG emissions when producing energy. A circular economy encourages material reduction, reuse,

recycling, and recovery to lessen the demand for raw materials and the accompanying emissions from extraction and production. Circular manufacturing, waste management, and consumption practices can improve EP and contribute to carbon neutrality [99]. The transportation sector is also an essential contributor to carbon emissions.

The adoption of Electric Vehicles (EVs) and the enhancement of public transportation infrastructure have the potential to facilitate the attainment of carbon neutrality. EVs exhibit a notable reduction in emissions compared to vehicles powered by internal combustion engines. Research has demonstrated that promoting EVs and enhancing public transportation systems can significantly reduce GHG emissions [100]. Achieving 100% carbon neutrality may not always be possible in the short term. As a result, carbon offset techniques are frequently employed. Carbon offset allows businesses to invest in projects that reduce or eliminate GHG emissions, effectively offsetting their total emissions. For example, investing in forestry initiatives or carbon capture and storage can help achieve carbon neutrality [101].

To summarize, in a carbon-neutral situation, green performance is important. Energy efficiency, renewable energy, circular economy, sustainable transportation, and carbon offsetting and removal are all critical components in reaching carbon neutrality and lowering GHG emissions.

VII. PRACTICAL IMPLICATIONS

The findings of this research will enable public and private firms in China to understand how to handle their ethical responsibility to the environment. According to these results, it can be assumed that GHRM practices, GI, and carbon neutrality positively impact EP and promote an environmental management-based philosophy. Environmental friendly practices of GHRM, GI, and sustainable performance would improve community wellness and enhance company reputation. They would also save money and promote company social responsibility. Although the outcomes from the present research show a positive relationship with EP, it is essential to examine the various interactions between carbon neutrality, GHRM, GI, and EP. Therefore, the former involve important practical implications for various stakeholders, including organizations, governments, and researchers committed to promoting sustainable initiatives, mitigating carbon emissions, and fostering a more environmentally friendly and sustainable future. The primary and most immediate advantage of implementing green practices and attaining carbon neutrality is to reduce the environmental impact, such as the rapidly increasing rate of global warming. Similarly, achieving carbon neutrality would reduce the risk of natural disasters. The immediate effect would be to maintain the current social structure, while an indirect effect would be to encourage the development of society. Carbon neutrality will promote a transition in economic growth towards environmentally friendly, low-carbon, and sustainable development. It will also substantially influence emerging technology trends, including decarbonization, energy efficiency, recycling, new power systems, energy storage, and reduced emission technologies. In addition, new approaches will likely replace particular industries or enterprises. For instance, the current coal industry

and its associated infrastructure, manufacturing, and service sectors are expected to experience more job losses. Carbon neutrality can lead to job growth in clean energy, carbon-free energy, and renewable energy sectors.

VIII. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

Green practices and carbon neutrality are essential for the future ecological development. GHRM and GI assist low-carbon growth, shifting the emissions graph, and may lead to low-carbon neutrality to improve EP. China tries to become carbon neutral by 2060. However, there are many problems to solve along this journey. China is continuously trying to advance scientific research and developments to solve the challenges faced by businesses and people due to carbon emissions. For this purpose, China has been striving either by itself or by internationally collaborating with different countries and world organizations to bring carbon neutrality through a low-carbon strategy, efforts to save energy, use renewable energy, etc. As a result of the large-scale deployment of renewable power generation, mainly hydro, wind, and solar, the share of fossil fuel has reached an all-time low, and coal usage has plateaued. Similarly, several mitigation initiatives have been implemented to acknowledge how crucial China is in stabilizing the planet's climate. To meet these objectives, in December 2020, President Xi improved China's Intended Nationally Determined Contributions (INDC) targets to a >65% decrease in carbon intensity by 2030 from 2005 levels. In September 2020, the long-term mitigation goal of carbon neutrality by 2060 was proposed. Nevertheless, it is still challenging to understand the trends and trajectories of carbon emissions, given the lack of uncertainty around the global economy and technological advancements.

Achieving net-zero carbon emissions in China before 2060 will facilitate the promotion of sustainable and high-quality growth. Integrating carbon neutrality into national social and economic strategies would speed up the efficiency and greening of industrial and energy frameworks while fostering the advancement of new technologies within low-carbon industries. China is globally the top producer of green technologies since it has the highest installation capacity in the world for wind, solar, and hydro power. Each of these qualities has the potential to facilitate China's ability to transform and enhance its various businesses, hence fostering the creation of additional employment opportunities within green industries.

This article concludes by offering an in-depth review of the relationships among GHRM, GI, and EP within the framework of China's efforts toward achieving carbon neutrality. The study highlighted several reasons in favor of adopting GHRM practices and promoting GIs, stressing their potential to significantly enhance a firm's environmental performance and ultimately help achieve the carbon neutrality goals. The study's findings emphasize the importance of fostering a culture of sustainability within organizations through the strategic integration of HRM practices that embrace environmental values, such as green recruitment and selection, green training and development, green compensation, and green employee engagement. These sustainable approaches to HRM help employees gain environmental awareness, knowledge, skills,

and attitudes, promoting GI, technology developments, and environmentally friendly corporate strategies. Additionally, the research emphasizes the significance of fostering an encouraging corporate environment and integrating GI frameworks to assist the successful adoption of GHRM practices. It is pointed out that organizations should embrace a proactive and participatory stance in promoting eco-innovation. This involves fostering green creativity and facilitating the exchange of ideas and knowledge across various organizational levels and departments. Enhancing environmental performance requires the construction of collaborative networks, the formation of green teams, and the allocation of resources for innovation.

The study also provides information on China's distinct background and commitment to carbon neutrality. China's significant economic growth and enormous environmental issues make it a critical case study for understanding the dynamics of GHRM, GIs, and EP. The country's efforts to a transition from a carbon-intensive economy to a low-carbon and sustainable future are underlined and the necessity for the Chinese government, organizations, and policymakers to prioritize GHRM practices and promote GIs is accentuated. This study makes a valuable contribution to the current literature by illustrating the significant impact of GHRM and GI on the EP, aiding China's efforts toward achieving carbon neutrality.

REFERENCES

- [1] S. K. Singh, M. D. Giudice, R. Chierici, and D. Graziano, "Green innovation and environmental performance: The role of green transformational leadership and green human resource management," *Technological Forecasting and Social Change*, vol. 150, Jan. 2020, Art. no. 119762, <https://doi.org/10.1016/j.techfore.2019.119762>.
- [2] H. Zameer, Y. Wang, H. Yasmeen, and S. Mubarak, "Green innovation as a mediator in the impact of business analytics and environmental orientation on green competitive advantage," *Management Decision*, vol. 60, no. 2, pp. 488–507, Jan. 2020, <https://doi.org/10.1108/MD-01-2020-0065>.
- [3] M. Aboramadan, "The effect of green HRM on employee green behaviors in higher education: the mediating mechanism of green work engagement," *International Journal of Organizational Analysis*, vol. 30, no. 1, pp. 7–23, Jan. 2020, <https://doi.org/10.1108/IJOA-05-2020-2190>.
- [4] U. Ahmed, W. A. Umrani, A. Yousaf, M. A. Siddiqui, and M. H. Pahi, "Developing faithful stewardship for environment through green HRM," *International Journal of Contemporary Hospitality Management*, vol. 33, no. 10, pp. 3115–3133, Jan. 2021, <https://doi.org/10.1108/IJCHM-09-2020-1066>.
- [5] M. Ahmed, Q. Guo, M. A. Qureshi, S. A. Raza, K. A. Khan, and J. Salam, "Do green HR practices enhance green motivation and proactive environmental management maturity in hotel industry?," *International Journal of Hospitality Management*, vol. 94, Apr. 2021, Art. no. 102852, <https://doi.org/10.1016/j.ijhm.2020.102852>.
- [6] W. O'Donohue and N. Torugsa, "The moderating effect of 'Green' HRM on the association between proactive environmental management and financial performance in small firms," *The International Journal of Human Resource Management*, vol. 27, no. 2, pp. 239–261, Jan. 2016, <https://doi.org/10.1080/09585192.2015.1063078>.
- [7] Y. J. Kim, W. G. Kim, H.-M. Choi, and K. Phetvaroon, "The effect of green human resource management on hotel employees' eco-friendly behavior and environmental performance," *International Journal of Hospitality Management*, vol. 76, pp. 83–93, Jan. 2019, <https://doi.org/10.1016/j.ijhm.2018.04.007>.
- [8] R. Hameed, N. Rehaman, M. Shoaib, and M. Ibtisam, "Promoting pro-environmental behavior among one belt one road firms' employees through the lens of green human resource practices," *Work*, vol. 76, no. 4, pp. 1373–1384, Jan. 2023, <https://doi.org/10.3233/WOR-211442>.
- [9] Y. M. Yusoff, M. Nejati, D. M. H. Kee, and A. Amran, "Linking Green Human Resource Management Practices to Environmental Performance in Hotel Industry," *Global Business Review*, vol. 21, no. 3, pp. 663–680, Jun. 2020, <https://doi.org/10.1177/0972150918779294>.
- [10] F. H. Awan, L. Dunnann, K. Jamil, and R. F. Gul, "Stimulating environmental performance via green human resource management, green transformational leadership, and green innovation: a mediation-moderation model," *Environmental Science and Pollution Research*, vol. 30, no. 2, pp. 2958–2976, Jan. 2023, <https://doi.org/10.1007/s11356-022-22424-y>.
- [11] P. Benevene and I. Buonomo, "Green Human Resource Management: An Evidence-Based Systematic Literature Review," *Sustainability*, vol. 12, no. 15, Jan. 2020, Art. no. 5974, <https://doi.org/10.3390/su12155974>.
- [12] O. M. Okeyinka and O. J. Idowu, "Assessment of the Suitability of Paper Waste as an Engineering Material," *Engineering, Technology & Applied Science Research*, vol. 4, no. 6, pp. 724–727, Dec. 2014, <https://doi.org/10.48084/etasr.485>.
- [13] Y. Kassem, H. Gokcekus, and A. Aljatlawe, "Utilization of Solar Energy for Electric Vehicle Charging and the Energy Consumption of Residential Buildings in Northern Cyprus: A Case Study," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11598–11607, Oct. 2023, <https://doi.org/10.48084/etasr.6142>.
- [14] J. S. Khan, S. Khoso, Z. Iqbal, S. Sohu, and M. A. Keerio, "An Outlook of Ozone Air Pollution through Comparative Analysis of Artificial Neural Network, Regression, and Sensitivity Models," *Engineering, Technology & Applied Science Research*, vol. 8, no. 5, pp. 3387–3391, Oct. 2018, <https://doi.org/10.48084/etasr.1944>.
- [15] D. Agustia, Y. Permatasari, H. Fauzi, and M. N. A. Sari, "Research and Development Intensity, Firm Performance, and Green Product Innovation," *Journal of Security and Sustainability Issues*, vol. 9, no. 3, pp. 1039–1049, Mar. 2020, [https://doi.org/10.9770/jssi.2020.9.3\(27\)](https://doi.org/10.9770/jssi.2020.9.3(27)).
- [16] M. Novitasari and D. Agustia, "Competitive advantage as a mediating effect in the impact of green innovation and firm performance," *Business: Theory and Practice*, vol. 24, no. 1, pp. 216–226, May 2023, <https://doi.org/10.3846/btp.2023.15865>.
- [17] E. R. Lestari and N. M. S. Sunyoto, "Fostering green innovation in achieving sustainable performance," *Natural Resources Forum*, vol. 47, no. 3, pp. 413–434, 2023, <https://doi.org/10.1111/1477-8947.12293>.
- [18] M. Liu, J. Wu, Z. Lang, and X. Meng, "Long-term energy-environment-economic programming under carbon neutrality target: a study on China's regional energy transition pathways and CO2 mitigation strategies," *Energy Sources, Part B: Economics, Planning, and Policy*, vol. 18, no. 1, Dec. 2023, Art. no. 2229321, <https://doi.org/10.1080/15567249.2023.2229321>.
- [19] L. Chen *et al.*, "Strategies to achieve a carbon neutral society: a review," *Environmental Chemistry Letters*, vol. 20, no. 4, pp. 2277–2310, Aug. 2022, <https://doi.org/10.1007/s10311-022-01435-8>.
- [20] K. H. Shaltout, M. T. Ahmed, S. A. Alrumman, D. A. Ahmed, and E. M. Eid, "Evaluation of the carbon sequestration capacity of arid mangroves along nutrient availability and salinity gradients along the Red Sea coastline of Saudi Arabia," *Oceanologia*, vol. 62, no. 1, pp. 56–69, Jan. 2020, <https://doi.org/10.1016/j.oceano.2019.08.002>.
- [21] D. Pappas, K. J. Chalvatzis, D. Guan, and A. Ioannidis, "Energy and carbon intensity: A study on the cross-country industrial shift from China to India and SE Asia," *Applied Energy*, vol. 225, pp. 183–194, Sep. 2018, <https://doi.org/10.1016/j.apenergy.2018.04.132>.
- [22] Z. Liu *et al.*, "Challenges and opportunities for carbon neutrality in China," *Nature Reviews Earth & Environment*, vol. 3, no. 2, pp. 141–155, Feb. 2022, <https://doi.org/10.1038/s43017-021-00244-x>.
- [23] S. Shan, S. Y. Genç, H. W. Kamran, and G. Dinca, "Role of green technology innovation and renewable energy in carbon neutrality: A sustainable investigation from Turkey," *Journal of Environmental Management*, vol. 294, Sep. 2021, Art. no. 113004, <https://doi.org/10.1016/j.jenvman.2021.113004>.
- [24] X. Huang, S. Chang, D. Zheng, and X. Zhang, "The role of BECCS in deep decarbonization of China's economy: A computable general

- equilibrium analysis," *Energy Economics*, vol. 92, Oct. 2020, Art. no. 104968, <https://doi.org/10.1016/j.eneco.2020.104968>.
- [25] T. S. Adebayo, E. N. Udemba, Z. Ahmed, and D. Kirikkaleli, "Determinants of consumption-based carbon emissions in Chile: an application of non-linear ARDL," *Environmental Science and Pollution Research*, vol. 28, no. 32, pp. 43908–43922, Aug. 2021, <https://doi.org/10.1007/s11356-021-13830-9>.
- [26] N. Li, S. Bao, S. Naseem, M. Sarfraz, and M. Mohsin, "Extending the Association Between Leader-Member Exchange Differentiation and Safety Performance: A Moderated Mediation Model," *Psychology Research and Behavior Management*, vol. 14, pp. 1603–1613, Oct. 2021, <https://doi.org/10.2147/PRBM.S335199>.
- [27] S. P. Nathaniel, Md. S. Alam, M. Murshed, H. Mahmood, and P. Ahmad, "The roles of nuclear energy, renewable energy, and economic growth in the abatement of carbon dioxide emissions in the G7 countries," *Environmental Science and Pollution Research*, vol. 28, no. 35, pp. 47957–47972, Sep. 2021, <https://doi.org/10.1007/s11356-021-13728-6>.
- [28] M. Mohsin, K. Jamil, S. Naseem, M. Sarfraz, and L. Ivascu, "Elongating Nexus Between Workplace Factors and Knowledge Hiding Behavior: Mediating Role of Job Anxiety," *Psychology Research and Behavior Management*, vol. 15, pp. 441–457, Feb. 2022, <https://doi.org/10.2147/PRBM.S348467>.
- [29] S. K. Mousa and M. Othman, "The impact of green human resource management practices on sustainable performance in healthcare organisations: A conceptual framework," *Journal of Cleaner Production*, vol. 243, Jan. 2020, Art. no. 118595, <https://doi.org/10.1016/j.jclepro.2019.118595>.
- [30] M.-F. Vernier, N. Arfaoui, S. Challita, P. Lanoie, and S. Plouffe, "Factors Influencing Profitability in Eco-design: Lessons from European and Canadian Firms," *Journal of Innovation Economics & Management*, vol. 39, no. 3, pp. 141–172, 2022, <https://doi.org/10.3917/jie.pr1.0117>.
- [31] E. Dostatni, D. Mikolajewski, J. Dorozynski, and I. Rojek, "Ecological Design with the Use of Selected Inventive Methods including AI-Based," *Applied Sciences*, vol. 12, no. 19, Jan. 2022, Art. no. 9577, <https://doi.org/10.3390/app12199577>.
- [32] X. Ye, S. Cai, X. Li, and Z. Wang, "How and when top management green commitment facilitates employees green behavior: a multilevel moderated mediation model," *Chinese Management Studies*, vol. 17, no. 5, pp. 970–990, Jan. 2022, <https://doi.org/10.1108/CMS-11-2021-0493>.
- [33] X. Long, C. Sun, C. Wu, B. Chen, and K. A. Boateng, "Green innovation efficiency across China's 30 provinces: estimate, comparison, and convergence," *Mitigation and Adaptation Strategies for Global Change*, vol. 25, no. 7, pp. 1243–1260, Oct. 2020, <https://doi.org/10.1007/s11027-019-09903-3>.
- [34] M. Sarfraz, M. Mohsin, S. Naseem, and A. Kumar, "Modeling the relationship between carbon emissions and environmental sustainability during COVID-19: a new evidence from asymmetric ARDL cointegration approach," *Environment, Development and Sustainability*, vol. 23, no. 11, pp. 16208–16226, Nov. 2021, <https://doi.org/10.1007/s10668-021-01324-0>.
- [35] X. Lei, Y. Wang, D. Zhao, and Q. Chen, "The local-neighborhood effect of green credit on green economy: a spatial econometric investigation," *Environmental Science and Pollution Research*, vol. 28, no. 46, pp. 65776–65790, Dec. 2021, <https://doi.org/10.1007/s11356-021-15419-8>.
- [36] G. Rana and V. Arya, "Green human resource management and environmental performance: mediating role of green innovation – a study from an emerging country," *Foresight*, vol. 26, no. 1, pp. 35–58, Jan. 2023, <https://doi.org/10.1108/FS-04-2021-0094>.
- [37] T. T. Kim, W. G. Kim, S. Majeed, and K. Haldorai, "Does green human resource management lead to a green competitive advantage? A sequential mediation model with three mediators," *International Journal of Hospitality Management*, vol. 111, May 2023, Art. no. 103486, <https://doi.org/10.1016/j.ijhm.2023.103486>.
- [38] N. T. Pham, Z. Tuckova, and Q. P. T. Phan, "Greening human resource management and employee commitment toward the environment: An interaction model," *Journal of Business Economics and Management*, vol. 20, no. 3, pp. 446–465, Apr. 2019, <https://doi.org/10.3846/jbem.2019.9659>.
- [39] Q. Mohi Ud Din and L. Zhang, "Unveiling the Mechanisms through Which Leader Integrity Shapes Ethical Leadership Behavior: Theory of Planned Behavior Perspective," *Behavioral Sciences*, vol. 13, no. 11, Nov. 2023, Art. no. 928, <https://doi.org/10.3390/bs13110928>.
- [40] Q. M. U. Din, L. Zhang, S. Karim, M. A. Golo, and A. U. Din, "Ethical leadership as a catalyst for sustainable training and green performance," in *2nd International Conference on Sustainable Technology and Management*, Dongguan, China, Jul. 2023, vol. 12804, pp. 442–452, <https://doi.org/10.1117/12.3004393>.
- [41] F. Haider Alvi and K. Ulrich, "Innovation finance ecosystems for entrepreneurial firms: A conceptual model and research propositions," *Journal of Business Research*, vol. 156, Feb. 2023, Art. no. 113450, <https://doi.org/10.1016/j.jbusres.2022.113450>.
- [42] Q. Iqbal, "The Era of Environmental Sustainability: Ensuring That Sustainability Stands on Human Resource Management," *Global Business Review*, vol. 21, no. 2, pp. 377–391, Apr. 2020, <https://doi.org/10.1177/0972150918778967>.
- [43] G. Anwar and N. N. Abdullah, "The Impact of Human Resource Management Practice on Organizational Performance," *International Journal of Engineering, Business and Management*, vol. 5, no. 1, pp. 35–47, Jan. 2021, <https://doi.org/10.22161/ijebm.5.1.4>.
- [44] S. Ahmad, "Green Human Resource Management: Policies and practices," *Cogent Business & Management*, vol. 2, no. 1, Dec. 2015, Art. no. 1030817, <https://doi.org/10.1080/23311975.2015.1030817>.
- [45] D. Roche *et al.*, "The Green Wave: Adding value through net zero energy strategy," RACE for 2030 CRC, 2021.
- [46] D. W. S. Renwick, T. Redman, and S. Maguire, "Green Human Resource Management: A Review and Research Agenda*," *International Journal of Management Reviews*, vol. 15, no. 1, pp. 1–14, 2013, <https://doi.org/10.1111/j.1468-2370.2011.00328.x>.
- [47] M. Shah, "Green human resource management: Development of a valid measurement scale," *Business Strategy and the Environment*, vol. 28, no. 5, pp. 771–785, 2019, <https://doi.org/10.1002/bse.2279>.
- [48] F. Amjad *et al.*, "Effect of green human resource management practices on organizational sustainability: the mediating role of environmental and employee performance," *Environmental Science and Pollution Research*, vol. 28, no. 22, pp. 28191–28206, Jun. 2021, <https://doi.org/10.1007/s11356-020-11307-9>.
- [49] M. Guerci, A. Longoni, and D. Luzzini, "Translating stakeholder pressures into environmental performance – the mediating role of green HRM practices," *The International Journal of Human Resource Management*, vol. 27, no. 2, pp. 262–289, Jan. 2016, <https://doi.org/10.1080/09585192.2015.1065431>.
- [50] X. Liu, X. Yu, and W. Xing, "Green Product Innovation via Green Transformational Leadership and employees' OCB: The Moderating Role of Green Organizational Climate– Empirical Evidence from China' Manufacturing Enterprises," *Polish Journal of Environmental Studies*, vol. 31, no. 5, pp. 4487–4498, Sep. 2022, <https://doi.org/10.15244/pjoes/150015>.
- [51] J. Chowhan, "Unpacking the black box: understanding the relationship between strategy, HRM practices, innovation and organizational performance," *Human Resource Management Journal*, vol. 26, no. 2, pp. 112–133, 2016, <https://doi.org/10.1111/1748-8583.12097>.
- [52] S. F. Rasool, M. Samma, M. Wang, Y. Zhao, and Y. Zhang, "How Human Resource Management Practices Translate Into Sustainable Organizational Performance: The Mediating Role Of Product, Process And Knowledge Innovation," *Psychology Research and Behavior Management*, vol. 12, pp. 1009–1025, Nov. 2019, <https://doi.org/10.2147/PRBM.S204662>.
- [53] S. Ren, K. Jiang, and G. Tang, "Leveraging green HRM for firm performance: The joint effects of CEO environmental belief and external pollution severity and the mediating role of employee environmental commitment," *Human Resource Management*, vol. 61, no. 1, pp. 75–90, 2022, <https://doi.org/10.1002/hrm.22079>.
- [54] C. T. Nolan and T. N. Garavan, "Human Resource Development in SMEs: A Systematic Review of the Literature," *International Journal of Management Reviews*, vol. 18, no. 1, pp. 85–107, 2016, <https://doi.org/10.1111/ijmr.12062>.

- [55] S. Roper, J. H. Love, and K. Bonner, "Firms' knowledge search and local knowledge externalities in innovation performance," *Research Policy*, vol. 46, no. 1, pp. 43–56, Feb. 2017, <https://doi.org/10.1016/j.respol.2016.10.004>.
- [56] A. Razaq, T. Fatima, and M. Murshed, "Asymmetric effects of tourism development and green innovation on economic growth and carbon emissions in top 10 GDP countries," *Journal of Environmental Planning and Management*, vol. 66, no. 3, pp. 471–500, Feb. 2023, <https://doi.org/10.1080/09640568.2021.1990029>.
- [57] N. Soewarno, B. Tjahjadi, and F. Fithrianti, "Green innovation strategy and green innovation: The roles of green organizational identity and environmental organizational legitimacy," *Management Decision*, vol. 57, no. 11, pp. 3061–3078, Jan. 2019, <https://doi.org/10.1108/MD-05-2018-0563>.
- [58] S. Asadi *et al.*, "Investigating influence of green innovation on sustainability performance: A case on Malaysian hotel industry," *Journal of Cleaner Production*, vol. 258, Jun. 2020, Art. no. 120860, <https://doi.org/10.1016/j.jclepro.2020.120860>.
- [59] Y. Huang, S. Li, X. Xiang, Y. Bu, and Y. Guo, "How can the combination of entrepreneurship policies activate regional innovation capability? A comparative study of Chinese provinces based on fsQCA," *Journal of Innovation & Knowledge*, vol. 7, no. 3, Jul. 2022, Art. no. 100227, <https://doi.org/10.1016/j.jik.2022.100227>.
- [60] Y.-S. Chen, "The Driver of Green Innovation and Green Image – Green Core Competence," *Journal of Business Ethics*, vol. 81, no. 3, pp. 531–543, Sep. 2008, <https://doi.org/10.1007/s10551-007-9522-1>.
- [61] P. Kivimaa and P. Kautto, "Making or breaking environmental innovation? Technological change and innovation markets in the pulp and paper industry," *Management Research Review*, vol. 33, no. 4, pp. 289–305, Jan. 2010, <https://doi.org/10.1108/01409171011030426>.
- [62] H. Tian and Z. H. Wang, "Chinese green process innovation in automotive painting: the strategic niche management perspective," *International Journal of Environmental Science and Technology*, vol. 17, no. 2, pp. 993–1010, Feb. 2020, <https://doi.org/10.1007/s13762-019-02530-0>.
- [63] X. Xie, J. Huo, and H. Zou, "Green process innovation, green product innovation, and corporate financial performance: A content analysis method," *Journal of Business Research*, vol. 101, pp. 697–706, Aug. 2019, <https://doi.org/10.1016/j.jbusres.2019.01.010>.
- [64] X. Xie, Q. Zhu, and R. Wang, "Turning green subsidies into sustainability: How green process innovation improves firms' green image," *Business Strategy and the Environment*, vol. 28, no. 7, pp. 1416–1433, 2019, <https://doi.org/10.1002/bse.2323>.
- [65] Y. Ma, G. Hou, Q. Yin, B. Xin, and Y. Pan, "The sources of green management innovation: Does internal efficiency demand pull or external knowledge supply push?," *Journal of Cleaner Production*, vol. 202, pp. 582–590, Nov. 2018, <https://doi.org/10.1016/j.jclepro.2018.08.173>.
- [66] G. Albort-Morant, J. Henseler, A. Leal-Millan, and G. Cepeda-Carrion, "Mapping the Field: A Bibliometric Analysis of Green Innovation," *Sustainability*, vol. 9, no. 6, Jun. 2017, Art. no. 1011, <https://doi.org/10.3390/su9061011>.
- [67] R. Y. K. Chan, "Does the Natural-Resource-Based View of the Firm Apply in an Emerging Economy? A Survey of Foreign Invested Enterprises in China*," *Journal of Management Studies*, vol. 42, no. 3, pp. 625–672, 2005, <https://doi.org/10.1111/j.1467-6486.2005.00511.x>.
- [68] T. Chin, Y. Shi, S. K. Singh, G. K. Agbanyo, and A. Ferraris, "Leveraging blockchain technology for green innovation in ecosystem-based business models: A dynamic capability of values appropriation," *Technological Forecasting and Social Change*, vol. 183, Oct. 2022, Art. no. 121908, <https://doi.org/10.1016/j.techfore.2022.121908>.
- [69] S. A. R. Khan, A. Razaq, Z. Yu, and S. Miller, "Retracted: Industry 4.0 and circular economy practices: A new era business strategies for environmental sustainability," *Business Strategy and the Environment*, vol. 30, no. 8, pp. 4001–4014, 2021, <https://doi.org/10.1002/bse.2853>.
- [70] S. U. Rehman, S. Kraus, S. A. Shah, D. Khanin, and R. V. Mahto, "Analyzing the relationship between green innovation and environmental performance in large manufacturing firms," *Technological Forecasting and Social Change*, vol. 163, Feb. 2021, Art. no. 120481, <https://doi.org/10.1016/j.techfore.2020.120481>.
- [71] Y.-S. Chen, C.-H. Chang, S.-L. Yeh, and H.-I. Cheng, "Green shared vision and green creativity: the mediation roles of green mindfulness and green self-efficacy," *Quality & Quantity*, vol. 49, no. 3, pp. 1169–1184, May 2015, <https://doi.org/10.1007/s11135-014-0041-8>.
- [72] R. Dubey, A. Gunasekaran, and S. Samar Ali, "Exploring the relationship between leadership, operational practices, institutional pressures and environmental performance: A framework for green supply chain," *International Journal of Production Economics*, vol. 160, pp. 120–132, Feb. 2015, <https://doi.org/10.1016/j.ijpe.2014.10.001>.
- [73] F. L. Oliva *et al.*, "Innovation in the main Brazilian business sectors: characteristics, types and comparison of innovation," *Journal of Knowledge Management*, vol. 23, no. 1, pp. 135–175, Jan. 2018, <https://doi.org/10.1108/JKM-03-2018-0159>.
- [74] N. Darnall, G. J. Jolley, and R. Handfield, "Environmental management systems and green supply chain management: complements for sustainability?," *Business Strategy and the Environment*, vol. 17, no. 1, pp. 30–45, 2008, <https://doi.org/10.1002/bse.557>.
- [75] A. Adegbile, D. Sarpong, and D. Meissner, "Strategic Foresight for Innovation Management: A Review and Research Agenda," *International Journal of Innovation and Technology Management*, vol. 14, no. 4, Aug. 2017, Art. no. 1750019, <https://doi.org/10.1142/S0219877017500195>.
- [76] D. Kammerer, "The effects of customer benefit and regulation on environmental product innovation.: Empirical evidence from appliance manufacturers in Germany," *Ecological Economics*, vol. 68, no. 8, pp. 2285–2295, Jun. 2009, <https://doi.org/10.1016/j.ecolecon.2009.02.016>.
- [77] Y.-S. Chen, S.-B. Lai, and C.-T. Wen, "The Influence of Green Innovation Performance on Corporate Advantage in Taiwan," *Journal of Business Ethics*, vol. 67, no. 4, pp. 331–339, Sep. 2006, <https://doi.org/10.1007/s10551-006-9025-5>.
- [78] S. Kraus, S. U. Rehman, and F. J. S. Garcia, "Corporate social responsibility and environmental performance: The mediating role of environmental strategy and green innovation," *Technological Forecasting and Social Change*, vol. 160, Nov. 2020, Art. no. 120262, <https://doi.org/10.1016/j.techfore.2020.120262>.
- [79] Y. H. Chuang, Y.-C. Tsao, and W. Y. Chen, "Fossil fuel and renewable energy allocation policy toward carbon neutrality under climate change," *Energy & Environment*, Jun. 2023, Art. no. 0958305X231185911, <https://doi.org/10.1177/0958305X231185911>.
- [80] C. Ziozias, E. Kontogianni, and L. Anthopoulos, "Carbon-Neutral City Transformation with Digitization: Guidelines from International Standardization," *Energies*, vol. 16, no. 15, Jan. 2023, Art. no. 5814, <https://doi.org/10.3390/en16155814>.
- [81] J. Pan, Y. Li, Y. Zhang, and Z. Ji, "China's Revolutionary Transition towards Carbon Neutrality and Prevention of Relevant Risks," *Chinese Journal of Urban and Environmental Studies*, vol. 11, no. 1, Mar. 2023, Art. no. 2350003, <https://doi.org/10.1142/S2345748123500033>.
- [82] N. Z. Khanna, J. Zhang, H. Lu, W. Feng, M. Johnson-Wang, and N. Zhou, "Conceptualizing demand-side technological and social innovations in modeling pathways to carbon neutrality," *Energy Research & Social Science*, vol. 100, Jun. 2023, Art. no. 103115, <https://doi.org/10.1016/j.erss.2023.103115>.
- [83] O. Rueda, J. M. Mogollon, A. Tukker, and L. Scherer, "Negative-emissions technology portfolios to meet the 1.5 °C target," *Global Environmental Change*, vol. 67, Mar. 2021, Art. no. 102238, <https://doi.org/10.1016/j.gloenvcha.2021.102238>.
- [84] J. Fuhrman, H. McJeon, P. Patel, S. C. Doney, W. M. Shobe, and A. F. Clarens, "Food–energy–water implications of negative emissions technologies in a +1.5 °C future," *Nature Climate Change*, vol. 10, no. 10, pp. 920–927, Oct. 2020, <https://doi.org/10.1038/s41558-020-0876-z>.
- [85] S. Rajbhandari and B. Limmeechokchai, "Assessment of greenhouse gas mitigation pathways for Thailand towards achievement of the 2°C and 1.5°C Paris Agreement targets," *Climate Policy*, vol. 21, no. 4, pp. 492–513, Apr. 2021, <https://doi.org/10.1080/14693062.2020.1857218>.
- [86] A. J. Melara, U. Singh, and L. M. Colosi, "Is aquatic bioenergy with carbon capture and storage a sustainable negative emission technology?"

- Insights from a spatially explicit environmental life-cycle assessment," *Energy Conversion and Management*, vol. 224, Nov. 2020, Art. no. 113300, <https://doi.org/10.1016/j.enconman.2020.113300>.
- [87] L. Myllyvirta, "Analysis: China's CO2 emissions hit Q1 record high after 4% rise in early 2023," *Carbon Brief*, May 11, 2023. <https://www.carbonbrief.org/analysis-chinas-co2-emissions-hit-q1-record-high-after-4-rise-in-early-2023/>.
- [88] J. Y. Y. Loh, N. P. Kherani, and G. A. Ozin, "Persistent CO2 photocatalysis for solar fuels in the dark," *Nature Sustainability*, vol. 4, no. 6, pp. 466–473, Jun. 2021, <https://doi.org/10.1038/s41893-021-00681-y>.
- [89] X. Yao, X. Zhang, and Z. Guo, "The tug of war between local government and enterprises in reducing China's carbon dioxide emissions intensity," *Science of The Total Environment*, vol. 710, Mar. 2020, Art. no. 136140, <https://doi.org/10.1016/j.scitotenv.2019.136140>.
- [90] M. Rumayor, A. Dominguez-Ramos, and A. Iribien, "Environmental and economic assessment of the formic acid electrochemical manufacture using carbon dioxide: Influence of the electrode lifetime," *Sustainable Production and Consumption*, vol. 18, pp. 72–82, Apr. 2019, <https://doi.org/10.1016/j.spc.2018.12.002>.
- [91] D. J. C. MacKay, P. Cramton, A. Ockenfels, and S. Stoft, "Price carbon - I will if you will," *Nature*, vol. 526, no. 7573, pp. 315–316, Oct. 2015, <https://doi.org/10.1038/526315a>.
- [92] B.-J. Tang, C.-J. Ji, Y.-J. Hu, J.-X. Tan, and X.-Y. Wang, "Optimal carbon allowance price in China's carbon emission trading system: Perspective from the multi-sectoral marginal abatement cost," *Journal of Cleaner Production*, vol. 253, Apr. 2020, Art. no. 119945, <https://doi.org/10.1016/j.jclepro.2019.119945>.
- [93] M. L. Weitzman, "How a Minimum Carbon Price Commitment Might Help to Internalize the Global Warming Externality." National Bureau of Economic Research, Apr. 2016, <https://doi.org/10.3386/w22197>.
- [94] K. Hubacek, G. Baiocchi, K. Feng, R. Munoz Castillo, L. Sun, and J. Xue, "Global carbon inequality," *Energy, Ecology and Environment*, vol. 2, no. 6, pp. 361–369, Dec. 2017, <https://doi.org/10.1007/s40974-017-0072-9>.
- [95] C. Guo and G. Song, "Research on the Impact on ESG Performance on Green Innovation Under the Background of Carbon Emissions Peak-Based on the Data of Listed Companies in Shandong Province," in *International Conference on Environmental Pollution and Governance*, Shanghai, China, Sep. 2022, pp. 671–678, https://doi.org/10.1007/978-3-031-25284-6_71.
- [96] M. D. Ibrahim, D. C. Ferreira, S. Daneshvar, and R. C. Marques, "Transnational resource generativity: Efficiency analysis and target setting of water, energy, land, and food nexus for OECD countries," *Science of The Total Environment*, vol. 697, Dec. 2019, Art. no. 134017, <https://doi.org/10.1016/j.scitotenv.2019.134017>.
- [97] C. Iris and J. S. L. Lam, "A review of energy efficiency in ports: Operational strategies, technologies and energy management systems," *Renewable and Sustainable Energy Reviews*, vol. 112, pp. 170–182, Sep. 2019, <https://doi.org/10.1016/j.rser.2019.04.069>.
- [98] S. Saha, G. Saini, S. Mishra, A. Chauhan, and S. Upadhyay, "A comprehensive review of techno-socio-enviro-economic parameters, storage technologies, sizing methods and control management for integrated renewable energy system," *Sustainable Energy Technologies and Assessments*, vol. 54, Dec. 2022, Art. no. 102849, <https://doi.org/10.1016/j.seta.2022.102849>.
- [99] S. Sasmoko *et al.*, "How Do Industrial Ecology, Energy Efficiency, and Waste Recycling Technology (Circular Economy) Fit into China's Plan to Protect the Environment? Up to Speed," *Recycling*, vol. 7, no. 6, Dec. 2022, Art. no. 83, <https://doi.org/10.3390/recycling7060083>.
- [100] K. J. Shah *et al.*, "Green transportation for sustainability: Review of current barriers, strategies, and innovative technologies," *Journal of Cleaner Production*, vol. 326, Dec. 2021, Art. no. 129392, <https://doi.org/10.1016/j.jclepro.2021.129392>.
- [101] Q. Ma, M. Murshed, and Z. Khan, "The nexuses between energy investments, technological innovations, emission taxes, and carbon emissions in China," *Energy Policy*, vol. 155, Aug. 2021, Art. no. 112345, <https://doi.org/10.1016/j.enpol.2021.112345>.

Short Empirical Insight: Leadership and Artificial Intelligence in the Pharmaceutical Industry

Chunjia Hu

School of Economics and Management, Harbin Institute of Technology, Weihai, Shandong, China
drchunjiahu@hit.edu.cn

Qaiser Mohi Ud Din

School of Management, Harbin Institute of Technology, China
qaiser.mohi-ud-din@outlook.com, qaiser@hit.edu.cn (corresponding author)

Li Zhang

School of Management, Harbin Institute of Technology, China
zhanglihit@hit.edu.cn (corresponding author)

Received: 6 February 2024 | Revised: 23 February 2024 | Accepted: 6 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7025>

ABSTRACT

This study aims to analyze the importance of the emerging idea of green talent management and its effect on employees' innovative work behavior. In addition, the study examines how ethical leadership and artificial intelligence influence Pakistan's pharmaceutical industry. Four hundred and seven (407) survey forms were gathered from the management departments of five pharmaceutical industries in the twin cities of Pakistan (Islamabad and Rawalpindi). The data collected were analyzed using PLS-SEM with the help of Smart PLS. The empirical evidence presented in this study supports the notion that green talent management significantly affects employees' innovative work behavior. Furthermore, the results reveal that ethical leadership and artificial intelligence are crucial in regulating the connection between green talent management and innovative work behavior. This study provides managerial and theoretical implications derived from its results. These implications can help leaders in pharmaceutical industries effectively leverage green talent management to stimulate innovative work behaviors of their employees and attain a competitive edge in their respective marketplaces. Several studies focused on addressing the difficulties faced by organizational leaders in cultivating and maintaining people who can make valuable contributions to their companies and help gain a competitive edge in their markets. However, studies that investigate these risks are limited.

Keywords-*ethical leadership; green talent management; artificial intelligence; work behavior; competitive advantage; pharmaceutical industry*

I. INTRODUCTION

In recent years, there has been a growing focus on organizational performance. This ever-increasing concern can be linked to the dynamic and competitive development of the global market [1]. According to [2], organizations aim for organizational performance to gain a competitive edge in their market. The former seek compelling advantages, prompting researchers to explore innovative approaches, which will motivate positive employee outcomes [3-4]. Effective and efficient workforces are crucial to organizational success and require ongoing training and development of human capital [5]. Human capital is vital for organizations to achieve competitive advantage. Retaining talent remains a global challenge, particularly as organizations need technologically compliant

employees to remain competitive [6-7]. Talent management practices should incorporate green human capital development programs to attract, nurture, maintain, and deploy talent for competitive advantages [8]. The concept of green talent management, as discussed in [9], is defined as a strategic approach employed by organizational leaders to effectively invite, improve, and retain brilliant employees in the context of the organizational competitive advantage.

This study analyzes the importance of various critical variables in the Pakistani manufacturing industry. Ethical leadership involves organizational leaders who employ ethical ideas and values in their decision-making to cultivate a positive workplace culture [10]. Green Talent Management (GTM) involves strategies and methods focused on recruiting,

nurturing, and keeping individuals who are environmentally aware and capable of promoting sustainable innovation in companies [1]. Artificial Intelligence (AI) uses sophisticated algorithms and technology to automate tasks, increase decision-making processes, and improve organizational efficiency [11]. Work behavior refers to visible acts and behaviors that individuals exhibit at work, including their Innovative Work Behavior (IWB) which entails the creation and execution of new ideas to enhance organizational processes and results [12].

There is a prevailing view that innovation plays a crucial role in enabling companies to endure in a highly competitive and globalized environment [13-14]. The existing literature lacks clarity on the elements that determine IWBs in employees. More research is needed to discover the mechanisms that enhance "innovative behavior" and examine their relationships with other factors within organizational settings, particularly in the GTM setting. In [15], four distinct categories of variables were identified, namely leadership, work, group and network, and institute. Similarly in [16], it was found that worker and administrative individualities significantly promote IWB. Meanwhile, the outcomes derived from each category display inconclusive findings. Consequently, leadership style has been indicated to be of considerable importance in modern innovation [17]. In [18], it was argued that competent leadership can effectively fulfill essential roles in advancing work operations, simplifying complicated activities, and facilitating knowledge sharing, which can boost team productivity and ultimately provide organizations with a competitive advantage.

According to [10], ethical leadership (EL) is believed to be the style of leadership that most frequently stimulates employee creative thinking. EL entails setting an exemplary example, acting with justice and honesty, and being committed to doing what is right [19]. Ethical leaders demonstrate a commitment to addressing the needs and interests of their employees, stakeholders, and the wider community rather than solely prioritizing their own interests or those of the organization. EL has been found to have a favorable influence on the promotion of IWBs among employees [20]. The impact of GTM on IWB, in the context of manufacturing industries and specifically in pharmaceuticals, has not been thoroughly investigated. This study area is fascinating, as it provides valuable information on the predictive capabilities of GTM, EL, and AI with the innovative behavior of employees. Achieving a leading position in the market is highly influenced by the performance of employees, and this also applies to academic institutions [21]. This is because the performance of human resources within the manufacturing industry, and particularly of the top-level managers who are vital participants in research activities, significantly affects the overall effectiveness of the organization [22-24]. This study focuses on the pharmaceutical sector because of its significant role in promoting medical research and development. IWB is essential to encourage breakthroughs and enhance healthcare results with the assistance of AI [25-26]. Furthermore, considering the increasing ambiguity around the volatility of digitalization, it is necessary to study the influence of leadership style and AI on the connection between GTM and IWB of employees [2, 13, 27-28]. After analyzing data and literature, it is evident that

there is a significant lack of research in the pharmaceutical industry on the evaluation of GTM [1] and its correlation with IWB [29], EL [30], and AI [27]. Previous studies have explored these concepts separately in different sectors, but there is a shortage of studies that precisely analyze how they interact in the pharmaceutical industry.

The existing literature mainly focuses on GTM [1] or on innovative behavior [29] individually, with minimal consideration given to their confluence within particular industries, such as pharmaceuticals. This gap highlights the need for a thorough investigation of how GTM techniques impact innovative behavior in pharmaceutical companies, considering the distinct constraints and opportunities of this industry. This study fills a knowledge gap and offers valuable insights for pharmaceutical companies that use GTM techniques to promote innovation and sustainable growth. In addition, it connects theoretical frameworks with practical implications, providing actionable suggestions for industry practitioners and policymakers.

II. LITERATURE REVIEW

GTM is a vital technique for companies seeking to incorporate environmental sustainability into their human resource procedures [6]. Many studies have illustrated that GTM methods, such as environmentally friendly recruitment, training, and development, correlate favorably with organizational sustainability performance and innovation [31]. Furthermore, GTM has been shown to increase employee engagement and dedication, resulting in enhanced organizational outcomes [32]. Ethical leadership influences business culture and employee behavior toward ethical and sustainable practices [10]. Ethical leadership benefits employee attitudes and behaviors, such as job satisfaction, organizational commitment, and ethical decision-making [33]. Ethical leadership in sustainability promotes trust and transparency, motivating employees to participate in ecologically responsible actions [34].

AI transforms the industry through automation, predictive analytics, and optimization of production processes [35]. AI technologies, such as predictive maintenance and intelligent manufacturing systems, improve industrial operations' productivity, quality, and safety [1]. Furthermore, AI assists in making data-driven decisions, allowing companies to pinpoint the chances of lasting innovation and efficient resource management [36]. IWB is crucial for firms looking to adapt to dynamic market conditions and maintain competitive advantage [27]. Studies indicate that employees' desire to participate in innovative actions is affected by elements, like organizational climate, leadership support, and individual traits [37]. Creating an environment that promotes innovation is essential in the manufacturing industry to facilitate advances in products and processes, increase efficiency, and sustain growth [38].

III. HYPOTHESES

This study examines the difficulties of talent management techniques with an environmental focus on promoting innovation among employees in a diverse workplace. The

former aims to understand how EL affects the innovation of manufacturing employees. Given the growing use of AI in pharmaceutical workplaces, the way AI affects the IWB of employees is investigated. The relationships among GTM, EL, and AI are specifically examined to be determined whether EL can moderate the impact of talent management strategies on innovation. The current study aims to illuminate the complex linkages and moderating effects that lead to a complete understanding of IWB. Five hypotheses were proposed according to this conceptual framework:

- H1: Employees' IWB is significantly affected by GTM.
- H2: The IWB of employees is directly influenced by EL.
- H3: AI directly influences IWB.
- H4: EL acts as a moderator among GTM and IWB.
- H5: AI moderates the connection between GTM and IWB.

IV. METHODOLOGY

The constructs examined in this study are AI, EL, IWB, and GTM. The GTM measurement consists of five components adapted from [1]. IWB was measured using five questions adapted from [29], while AI and EL were assessed using a set of four and five items, respectively, developed and modified from [27, 30]. A five-point Likert scale was used for each construct, ranging from strongly agree to strongly disagree. The study sample consisted of management staff members from pharmaceutical industries in Rawalpindi and Islamabad, Pakistan. The selection of the sample size was based on [39], which allowed the achievement of an appropriate sample size. Before distributing the questionnaire, participants received assurance about the confidentiality of their information, communicating that their participation was voluntary while actively encouraging it. As a result, 700 forms were distributed among management staff. Of these, 450 questionnaires were returned by participants. After careful consideration, 407 questionnaires were selected for the final analysis.

V. DATA ANALYSIS

A. Descriptive and Measurement Model Analysis

Data preparation and demographic analysis of the respondents were conducted using IBM's SPSS statistical tool. The respondents were 69% male and 31% female. PLS-SEM was employed to build the relationships suggested in the model, utilizing the SmartPLS software. Given the complexity of real-life scenarios, PLS-SEM is thought to effectively evaluate the connection between concepts and result predictions [39-40]. Additionally, it is effective when dealing with a large sample because it is independent of data normality. Table I portrays the analysis of the model's variables. The findings indicate that the all-factor loading items related to GTM, IWB, EL, and AI exceeded the value of 0.5. Furthermore, the p-values linked with these factor loadings were statistically significant at a confidence level of less than 1%. These findings suggest that the tool utilized to assess the constructs shows consequential convergent validity. Furthermore, the Cronbach alpha and composite reliability coefficients exhibited values exceeding the predefined criterion

of 0.7, as suggested in [40]. This finding provides evidence that the measurement instrument has a high level of reliability. Furthermore, all variables have an Average Variance Extracted (AVE) greater than the threshold value of 0.5 [41].

TABLE I. FACTOR LOADINGS, CRONBACH'S α , AVERAGE VARIANCE EXTRACTED, AND COMPOSITE RELIABILITY VALUES

Constructs	Loadings	α	AVE	CR
Green Talent Management		0.788	0.857	0.600
1. GTM	0.727			
2. GTM	0.755			
3. GTM	0.795			
4. GTM	0.818			
Ethical Leadership		0.892	0.918	0.692
1. EL	0.820			
2. EL	0.841			
3. EL	0.835			
4. EL	0.824			
5. EL	0.839			
Artificial Intelligence		0.911	0.928	0.721
1. AI	0.740			
2. AI	0.869			
3. AI	0.917			
4. AI	0.868			
5. AI	0.841			
Innovative Work Behavior		0.855	0.896	0.634
1. IWB	0.791			
2. IWB	0.814			
3. IWB	0.855			
4. IWB	0.771			
5. IWB	0.746			

α : Cronbach's alpha, AVE: Average Variance Extracted, CR: Composite Reliability.

To improve the constructs' discriminant validity evaluation, HTMT was created in response to the deficiencies of the Fornell-Larcker criterion [42]. According to [43], an acceptable criterion for the discriminant validity of a model construct is an HTMT value significantly less than 0.90. Table II displays the results of the model assessment, demonstrating that all values were smaller than 0.90, supporting the discriminant validity of the model variables. Variance Inflation Factors (VIFs) are preferable to fundamental correlation values to measure collinearity. According to [44], VIFs are advantageous in identifying which coefficients are influenced by collinearity. Typically, a variable β_i with a VIF below 10 indicates no multicollinearity. However, a tighter criterion was established in [45], requiring a VIF of less than 3.3 for β_i . Table III reveals that all values were below 3.3.

TABLE II. HTMT - DISCRIMINATING VALIDITY

No	Constructs	GTM	EL	IWB	AI	AI×GTM
1	GTM					
2	EL	0.482				
3	IWB	0.482	0.469			
4	AI	0.406	0.34	0.117		
5	AI × GTM	0.22	0.169	0.218	0.269	
6	EL × GTM	0.179	0.09	0.16	0.213	0.344

Common Method Variance (CMV) is a notable concern in survey-based research, particularly when based on an individual respondent. All variable items were inserted into a one-way ANOVA test in SPSS using this method. The initial

component accounted for only 36% of the variance, falling short of the 50% threshold, verifying that there are no issues with CMV in the dataset.

TABLE III. COLLINEARITY STATISTICS (VIF)

No	Constructs	VIF
1	GTM \rightarrow IWB	1.280
2	EL \rightarrow IWB	1.240
3	AI \rightarrow IWB	1.138
4	AI \times GTM \rightarrow IWB	1.204
5	EL \times GTM \rightarrow IWB	1.153

B. Structural Model Assessment - Hypothesis Testing

The results in Table IV disclose that the GTM coefficients are positive and significant at the 5% significance level. As a consequence, H1 was supported, indicating that GTM significantly affects the creative behavior of the employees. Furthermore, a statistically significant positive correlation was found between EL and IWB ($b = 0.33$, $p < 0.001$). The findings demonstrate the considerable impact of EL on employees' IWB, thus supporting H2. Similarly, the evaluation of the direct influence of AI on employees' IWB revealed a positive and significant association ($b = -0.292$, $p < 0.001$). Therefore, the results support H3, leading to the conclusion that AI has a favorable impact on employees' IWB. Regarding the moderating role of EL, the results indicate that it increases the effect of GTM on IWB, validating the H4 hypothesis. The results on the hypothesized moderating role of AI indicated that it dampens the influence of GTM on IWB, supporting the H5 hypothesis.

TABLE IV. HYPOTHESIS TESTING

No	Hypothesis	b-value	SD	T-statistic	p-values	Status
H1	GTM \rightarrow IWB	0.375	0.05	7.509	0.000	Accepted
H2	EL \rightarrow IWB	0.33	0.041	8.046	0.000	Accepted
H3	AI \rightarrow IWB	-0.292	0.062	4.686	0.000	Accepted
H4	EL \times GTM \rightarrow IWB	0.212	0.034	6.207	0.000	Accepted
H5	AI \times GTM \rightarrow IWB	-0.191	0.045	4.246	0.000	Accepted

b-value: Beta coefficients, SD: Standard Deviation, T-statistic: Hypothesis test statistic, p-value: Probability value.

VI. DISCUSSION AND CONCLUSION

This study examined the growing idea of GTM and its application within the pharmaceutical industry to achieve a competitive advantage by fostering innovative employee behavior. The purpose of this study was twofold: first, to investigate the connection between GTM and employees' IWB, and second, to offer insights into how companies could enhance their performance and gain a lead in their respective markets. Empirical evidence is presented supporting GTM's crucial and positive impact on employees' IWB. The findings are consistent with [1] and [46]. Various studies have shown the influence of GTM on employee withdrawal and work outcomes [47]. GTM positively affects employee innovation [6, 22, 47]. GTM involves increasing commitment through effective communication and supporting talent comfort and welfare. These characteristics foster an environment that encourages employee creativity and tacit knowledge, leading to

a competitive advantage for the organization. The influence of EL on IWB was additionally examined, highlighting the role of leadership in promoting novel thinking and creating a conducive work atmosphere for acquiring advanced knowledge and technology. The findings indicate that EL fosters employees' innovative behavior. Furthermore, this study identified AI as a noteworthy factor for neglecting employees' creative work behavior. This outcome aligns with [48-49], which suggested that employees' use of digital tools facilitates enhanced communication and knowledge sharing, promoting IWB.

Furthermore, it was discovered that EL and AI affect the relationship between GTM and IWB. According to [6], EL that demonstrates empathy, flexible control, effective communication, and inclusivity can mitigate the challenges associated with GTM and empower employees to express their innovative ideas. This study provides evidence that AI affects the influence of GTM on employees' creative behavior. AI has been recognized in previous studies [48, 50] to improve productivity and performance. The current study demonstrates that AI can mitigate the impact of GTM on IWB.

A unique perspective on how integrating GTM, EL, and AI might forecast employees' IWB within manufacturing industries, such as pharmaceuticals, is presented. The research was carried out within the geographical setting of Pakistan, which is situated in the Asian continent. This particular context has not been explored in earlier studies, thus providing an opportunity to offer novel and significant insights that hold relevance for both theoretical and practical applications. The EL phenomenon and its implications for organizations and the future of employment are also examined. It was found that GTM positively affects employee innovation. In the debate over AI's potential, evidence is provided to enhance the understanding of its roles. It is projected that AI directly affects employees' innovative behavior, while GTM will positively influence it. Insights from emerging fields in literature are provided, complementing and expanding the current understanding of GTM, EL, AI, and employees' IWB.

The conclusions of this study also have significant implications for practitioners and policymakers in developing guidelines to advance GTM in manufacturing sector settings. For a company to effectively cultivate and retain talents to enhance its competitive advantage, leaders and practitioners must prioritize the implementation of the values inherent in GTM. Organizations should consider employees with exceptional talent and dedication as valuable assets that require support to foster the development and dissemination of competitive advantages throughout the institution.

LIMITATIONS

By including additional pertinent indicators, the model created in this study opens opportunities for researchers and practitioners to develop models for examining the results of GTM or other antecedents of the innovative behavior of employees that are more complex, holistic, and comprehensive. However, a potential disadvantage is the current study's focus on a specific industry, which may restrict the ability to generalize its findings. Hence, it would be exciting to apply the

model within a different sector. Furthermore, this research explicitly examined managerial staff, limiting the potential for comprehensive insights that can be gained from a wider range of participants within the pharmaceutical industry. Therefore, further research is needed on this matter. Furthermore, it is suggested that future research should be a longitudinal investigation, encompassing a wide range of manufacturing and considering national and cultural factors. This approach would enhance the establishment of interconnection and the generalizability of this conceptual model.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support from the following funding agencies:

Dr. Chunjia Hu:

- Natural Science Foundation of Shandong Province (Grant No. ZR2023QG128)
- Shandong Philosophy and Social Science Research Project (Grant No. 23CGLJ01)

Dr. Li Zhang:

- National Natural Science Foundation of China (Grant No. 72272043)
- Heilongjiang Province Philosophy and Social Sciences Research Planning Project (Grant No. 22JYB221)

The funding provided by these institutions was instrumental in the completion of this research.

ETHICAL CONSIDERATIONS

This study was conducted following the Declaration of Helsinki and was approved by the Institutional Review Board (IRB) of the School of Management, Harbin Institute of Technology, China. Informed consent was obtained from all participants involved in this study. Data were collected online and participants were allowed to anonymously contribute their information voluntarily. The authors ensure that the collected data will not be shared with anyone and will be kept strictly confidential throughout this study and subsequent analysis.

REFERENCES

- [1] S. Ogbeibu, C. J. Chiappetta Jabbar, J. Burgess, J. Gaskin, and D. W. S. Renwick, "Green talent management and turnover intention: the roles of leader STARA competence and digital task interdependence," *Journal of Intellectual Capital*, vol. 23, no. 1, pp. 27–55, Jan. 2021, <https://doi.org/10.1108/JIC-01-2021-0016>.
- [2] P. Ye, L. Liu, and J. Tan, "The influence of organisational justice and ethical leadership on employees' innovation behaviour," *European Journal of Innovation Management*, vol. 26, no. 4, pp. 1129–1149, Jan. 2022, <https://doi.org/10.1108/EJIM-08-2021-0421>.
- [3] J. T. Vullings, A. H. B. De Hoogh, D. N. Den Hartog, and C. Boon, "Ethical and Passive Leadership and Their Joint Relationships with Burnout via Role Clarity and Role Overload," *Journal of Business Ethics*, vol. 165, no. 4, pp. 719–733, Sep. 2020, <https://doi.org/10.1007/s10551-018-4084-y>.
- [4] A. Kumar, R. K. Singh, and S. Modgil, "Influence of data-driven supply chain quality management on organizational performance: evidences from retail industry," *The TQM Journal*, vol. 35, no. 1, pp. 24–50, Jan. 2020, <https://doi.org/10.1108/TQM-06-2020-0146>.
- [5] R. Abraham, "Emotional intelligence in organizations: A conceptualization," *Genetic, Social, and General Psychology Monographs*, vol. 125, no. 2, pp. 209–224, 1999.
- [6] B. B. Gardas, S. K. Mangla, R. D. Raut, B. Narkhede, and S. Luthra, "Green talent management to unlock sustainability in the oil and gas sector," *Journal of Cleaner Production*, vol. 229, pp. 850–862, Aug. 2019, <https://doi.org/10.1016/j.jclepro.2019.05.018>.
- [7] A. B. Sieglings, C. Nielsen, and K. V. Petrides, "Trait emotional intelligence and leadership in a European multinational company," *Personality and Individual Differences*, vol. 65, pp. 65–68, Jul. 2014, <https://doi.org/10.1016/j.paid.2014.01.049>.
- [8] U. Bamel, V. Pereira, M. Del Giudice, and Y. Temouri, "The extent and impact of intellectual capital research: a two decade analysis," *Journal of Intellectual Capital*, vol. 23, no. 2, pp. 375–400, Jan. 2020, <https://doi.org/10.1108/JIC-05-2020-0142>.
- [9] E. Farndale, M. Thite, P. Budhwar, and B. Kwon, "Deglobalization and talent sourcing: Cross-national evidence from high-tech firms," *Human Resource Management*, vol. 60, no. 2, pp. 259–272, 2021, <https://doi.org/10.1002/hrm.22038>.
- [10] M. E. Brown, L. K. Treviño, and D. A. Harrison, "Ethical leadership: A social learning perspective for construct development and testing," *Organizational Behavior and Human Decision Processes*, vol. 97, no. 2, pp. 117–134, Jul. 2005, <https://doi.org/10.1016/j.obhdp.2005.03.002>.
- [11] P. Kumar, S. Chauhan, and L. K. Awasthi, "Artificial Intelligence in Healthcare: Review, Ethics, Trust Challenges & Future Research Directions," *Engineering Applications of Artificial Intelligence*, vol. 120, Apr. 2023, Art. no. 105894, <https://doi.org/10.1016/j.engappai.2023.105894>.
- [12] A. Battistelli, C. Odoardi, C. Vandenberghe, G. Di Napoli, and L. Piccione, "Information sharing and innovative work behavior: The role of work-based learning, challenging tasks, and organizational commitment," *Human Resource Development Quarterly*, vol. 30, no. 3, pp. 361–381, 2019, <https://doi.org/10.1002/hrdq.21344>.
- [13] H. M. Gelaidan, A. K. Al-Swidi, and M. A. Al-Hakimi, "Servant and authentic leadership as drivers of innovative work behaviour: the moderating role of creative self-efficacy," *European Journal of Innovation Management*, Jan. 2023, <https://doi.org/10.1108/EJIM-07-2022-0382>.
- [14] N. C. Eli-Chukwu, J. M. Aloha, and C. O. Ezeagwu, "A Systematic Review of Artificial Intelligence Applications in Cellular Networks," *Engineering, Technology & Applied Science Research*, vol. 9, no. 4, pp. 4504–4510, Aug. 2019, <https://doi.org/10.48084/etasr.2788>.
- [15] W. Zhou and V. K. Velamuri, "Key contextual success factors for employee innovative behavior: A study in a foreign manufacturing subsidiary in China," *Cogent Business & Management*, vol. 5, no. 1, Jan. 2018, Art. no. 1471770, <https://doi.org/10.1080/23311975.2018.1471770>.
- [16] Anna. C. Bos-Nehles and A. A. R. Veenendaal, "Perceptions of HR practices and innovative work behavior: the moderating effect of an innovative climate," *The International Journal of Human Resource Management*, vol. 30, no. 18, pp. 2661–2683, Oct. 2019, <https://doi.org/10.1080/09585192.2017.1380680>.
- [17] K. Teymournejad and R. Elghaei, "Effect of Transformational Leadership on the Creativity of Employees: An Empirical Investigation," *Engineering, Technology & Applied Science Research*, vol. 7, no. 1, pp. 1413–1419, Feb. 2017, <https://doi.org/10.48084/etasr.765>.
- [18] W. Li, T. A. Bhutto, W. Xuhui, Q. Maitlo, A. U. Zafar, and N. Ahmed Bhutto, "Unlocking employees' green creativity: The effects of green transformational leadership, green intrinsic, and extrinsic motivation," *Journal of Cleaner Production*, vol. 255, May 2020, Art. no. 120229, <https://doi.org/10.1016/j.jclepro.2020.120229>.
- [19] Z. ShiYong, S. Mirza, R. Ali, S. Mehar, M. Hussain, and S. Basharat, "The mediation mechanism of ethical climate in the relationship of ethical leadership and internal whistleblowing intention," *Journal of Contemporary Issues in Business and Government*, vol. 28, no. 04, pp. 817–833, 2022.
- [20] T. Yidong and L. Xinxin, "How Ethical Leadership Influence Employees' Innovative Work Behavior: A Perspective of Intrinsic

- Motivation," *Journal of Business Ethics*, vol. 116, no. 2, pp. 441–455, Aug. 2013, <https://doi.org/10.1007/s10551-012-1455-7>.
- [21] Z. Mahmood, W. Iftikhar, D. M. Vistro, and H. I. Tariq, "An Examination Of The Use Of Knowledge Management/ Knowledge Management Systems In Achieving And Sustaining Competitive Advantage In Higher Education Institutes: A Case Study Of Asia Pacific University, Malaysia," *International Journal of Scientific & Technology Research*, vol. 9, no. 04, pp. 3653–3658, 2020.
- [22] C. S. Fashami and M. Babaei, "A Behavioral Maturity Model to Establish Knowledge Management in an Organization," *Engineering, Technology & Applied Science Research*, vol. 7, no. 3, pp. 1725–1731, Jun. 2017, <https://doi.org/10.48084/etasr.987>.
- [23] F. Alfawaire and T. Atan, "The Effect of Strategic Human Resource and Knowledge Management on Sustainable Competitive Advantages at Jordanian Universities: The Mediating Role of Organizational Innovation," *Sustainability*, vol. 13, no. 15, Jan. 2021, Art. no. 8445, <https://doi.org/10.3390/su13158445>.
- [24] O. Fawehinmi, M. Y. Yusliza, Z. Mohamad, J. Noor Faezah, and Z. Muhammad, "Assessing the green behaviour of academics: The role of green human resource management and environmental knowledge," *International Journal of Manpower*, vol. 41, no. 7, pp. 879–900, Jan. 2020, <https://doi.org/10.1108/IJM-07-2019-0347>.
- [25] S. Noreen, S. Saif, and S. Haider, "Ethics determinant of commitment: An impacts of organizational culture and ethos based leadership," *The Explorer Islamabad: Journal of Social Sciences*, vol. 2, no. 6, pp. 132–139, 2016.
- [26] J. Zeng and G. Xu, "Linking ethical leadership to employee voice: The role of trust," *Social Behavior and Personality: an international journal*, vol. 48, no. 8, pp. 1–12, Aug. 2020, <https://doi.org/10.2224/sbp.9200>.
- [27] J. A. Odugbesan, S. Aghazadeh, R. E. Al Qaralleh, and O. S. Sogoke, "Green talent management and employees' innovative work behavior: the roles of artificial intelligence and transformational leadership," *Journal of Knowledge Management*, vol. 27, no. 3, pp. 696–716, Jan. 2022, <https://doi.org/10.1108/JKM-08-2021-0601>.
- [28] X. Li, S. Nosheen, N. U. Haq, and X. Gao, "Value creation during fourth industrial revolution: Use of intellectual capital by most innovative companies of the world," *Technological Forecasting and Social Change*, vol. 163, Feb. 2021, Art. no. 120479, <https://doi.org/10.1016/j.techfore.2020.120479>.
- [29] M. A. Opoku, S. B. Choi, and S. W. Kang, "Servant Leadership and Innovative Behaviour: An Empirical Analysis of Ghana's Manufacturing Sector," *Sustainability*, vol. 11, no. 22, Jan. 2019, Art. no. 6273, <https://doi.org/10.3390/su11226273>.
- [30] K. Kalshoven, D. N. Den Hartog, and A. H. B. De Hoogh, "Ethical leadership at work questionnaire (ELW): Development and validation of a multidimensional measure," *The Leadership Quarterly*, vol. 22, no. 1, pp. 51–69, Feb. 2011, <https://doi.org/10.1016/j.leaqua.2010.12.007>.
- [31] S. Y. Malik, Y. Cao, Y. H. Mughal, G. M. Kundi, M. H. Mughal, and T. Ramayah, "Pathways towards Sustainability in Organizations: Empirical Evidence on the Role of Green Human Resource Management Practices and Green Intellectual Capital," *Sustainability*, vol. 12, no. 8, Jan. 2020, Art. no. 3228, <https://doi.org/10.3390/su12083228>.
- [32] S. Umair, U. Waqas, and B. Mrugalska, "Cultivating sustainable environmental performance: The role of green talent management, transformational leadership, and employee engagement with green initiatives," *Work*, Jan. 2023, <https://doi.org/10.3233/WOR-230357>.
- [33] Q. Mohi Ud Din and L. Zhang, "Unveiling the Mechanisms through Which Leader Integrity Shapes Ethical Leadership Behavior: Theory of Planned Behavior Perspective," *Behavioral Sciences*, vol. 13, no. 11, Nov. 2023, Art. no. 928, <https://doi.org/10.3390/bs13110928>.
- [34] Q. M. U. Din, L. Zhang, S. Karim, M. A. Golo, and A. U. Din, "Ethical leadership as a catalyst for sustainable training and green performance," in *Second International Conference on Sustainable Technology and Management (ICSTM 2023)*, Dongguan, China, Oct. 2023, vol. 12804, pp. 442–452, <https://doi.org/10.1117/12.3004393>.
- [35] D. Gursoy, O. H. Chi, L. Lu, and R. Nunkoo, "Consumers acceptance of artificially intelligent (AI) device use in service delivery," *International Journal of Information Management*, vol. 49, pp. 157–169, Dec. 2019, <https://doi.org/10.1016/j.ijinfomgt.2019.03.008>.
- [36] N. Malik, S. N. Tripathi, A. K. Kar, and S. Gupta, "Impact of artificial intelligence on employees working in industry 4.0 led organizations," *International Journal of Manpower*, vol. 43, no. 2, pp. 334–354, Jan. 2021, <https://doi.org/10.1108/IJM-03-2021-0173>.
- [37] A. M. Elsayed, B. Zhao, A. E. Goda, and A. M. Elsetouhi, "The role of error risk taking and perceived organizational innovation climate in the relationship between perceived psychological safety and innovative work behavior: A moderated mediation model," *Frontiers in Psychology*, vol. 14, Jan. 2023, <https://doi.org/10.3389/fpsyg.2023.1042911>.
- [38] C. Li, H. U. R. Makhdoom, and S. Asim, "Impact of Entrepreneurial Leadership on Innovative Work Behavior: Examining Mediation and Moderation Mechanisms," *Psychology Research and Behavior Management*, vol. 13, pp. 105–118, Jan. 2020, <https://doi.org/10.2147/PRBM.S236876>.
- [39] R. V. Krejcie and D. W. Morgan, "Determining Sample Size for Research Activities," *Educational and Psychological Measurement*, vol. 30, no. 3, pp. 607–610, Sep. 1970, <https://doi.org/10.1177/001316447003000308>.
- [40] N. Kock, "Common Method Bias in PLS-SEM: A Full Collinearity Assessment Approach," *International Journal of e-Collaboration (IJeC)*, vol. 11, no. 4, pp. 1–10, Oct. 2015, <https://doi.org/10.4018/ijec.2015100101>.
- [41] J. Hair, C. L. Hollingsworth, A. B. Randolph, and A. Y. L. Chong, "An updated and expanded assessment of PLS-SEM in information systems research," *Industrial Management & Data Systems*, vol. 117, no. 3, pp. 442–458, Jan. 2017, <https://doi.org/10.1108/IMDS-04-2016-0130>.
- [42] J. Henseler, G. Hubona, and P. A. Ray, "Using PLS path modeling in new technology research: updated guidelines," *Industrial Management & Data Systems*, vol. 116, no. 1, pp. 2–20, Jan. 2016, <https://doi.org/10.1108/IMDS-09-2015-0382>.
- [43] G. Albort-Morant, J. Henseler, A. Leal-Millán, and G. Cepeda-Carrión, "Mapping the Field: A Bibliometric Analysis of Green Innovation," *Sustainability*, vol. 9, no. 6, Jun. 2017, Art. no. 1011, <https://doi.org/10.3390/su9061011>.
- [44] M. O. Akinwande, H. G. Dikko, and A. Samson, "Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis," *Open Journal of Statistics*, vol. 5, no. 7, 2015, Art. no. 62189, <https://doi.org/10.4236/ojs.2015.57075>.
- [45] A. Diamantopoulos and J. A. Siguaw, "Formative Versus Reflective Indicators in Organizational Measure Development: A Comparison and Empirical Illustration," *British Journal of Management*, vol. 17, no. 4, pp. 263–282, 2006, <https://doi.org/10.1111/j.1467-8551.2006.00500.x>.
- [46] L. T. T. Bui and Y. Chang, "Talent management and turnover intention: focus on Danang city government in Vietnam," *International Review of Public Administration*, vol. 23, no. 4, pp. 219–236, Oct. 2018, <https://doi.org/10.1080/12294659.2018.1552403>.
- [47] P.-A. Wikström, "Sustainability and organizational activities – three approaches," *Sustainable Development*, vol. 18, no. 2, pp. 99–107, 2010, <https://doi.org/10.1002/sd.449>.
- [48] S. Mahlasela and W. T. Chinyamurindi, "Technology-related factors and their influence on turnover intentions: A case of government employees in South Africa," *The Electronic Journal of Information Systems in Developing Countries*, vol. 86, no. 3, 2020, Art. no. e12126, <https://doi.org/10.1002/isd2.12126>.
- [49] S. Berraies, R. Lajili, and R. Chtioui, "Social capital, employees' well-being and knowledge sharing: does enterprise social networks use matter? Case of Tunisian knowledge-intensive firms," *Journal of Intellectual Capital*, vol. 21, no. 6, pp. 1153–1183, Jan. 2020, <https://doi.org/10.1108/JIC-01-2020-0012>.
- [50] C. Hughes *et al.*, "Artificial Intelligence, Employee Engagement, Fairness, and Job Outcomes," in *Managing Technology and Middle- and Low-skilled Employees*, Emerald Publishing Limited, 2019, pp. 61–68.

AUTHORS PROFILE

Chunjia Hu is a lecturer at the School of Economics and Management, Harbin Institute of Technology, Weihai campus, Shandong Province, China.

She received her Ph.D. from Harbin Institute of Technology. Her scholarship focuses on big data resources, innovation management, market orientation, strategic planning, and statistical methods.

Qaiser Mohi Ud Din is a Ph.D. scholar under a Chinese Government Scholarship at the School of Management of Harbin Institute of Technology, Harbin, China. He completed his M.S. in Business Administration from the National College of Business Administration & Economics, Sub Campus Bahawalpur, Pakistan. His research interests are related to the areas of leadership and applied knowledge of human resource management. He has published in reputed journals and has continuously been reviewing articles for leadership and organizational development journals.

Li Zhang is a professor at the School of Management of Harbin Institute of Technology, China. She has been studying as a postdoctoral fellow in Organizational Psychology, at the University of Guelph, Canada. Her research interest is focused on organizational behavior, human resource management, leadership, and work-family relationships.

Particle Swarm Optimization for Wireless Sensor Network Lifespan Maximization

Souad Kamel

College of Computer Science and Engineering, University of Jeddah, Saudi Arabia
skamel@uj.edu.sa (corresponding author)

Abeer Al Qahtani

College of Computer Science and Engineering, University of Jeddah, Saudi Arabia
abera.stu@uj.edu.sa

Abdullah Saad Musaed Al-Shahrani

College of Computer Science and Engineering, University of Jeddah, Saudi Arabia
aslshahrani@uj.edu.sa

Received: 14 December 2023 | Revised: 1 March 2024 | Accepted: 9 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6752>

ABSTRACT

Despite the deployment of wireless sensor networks in diverse fields (health, environment, military applications, etc.) for tracking or monitoring, several challenges, such as extending the lifetime of the network under energy constraints, still need to be resolved. Lifetime is the operational time of the network during which it can perform dedicated tasks and satisfy the application requirements. The energy constraints dictate that the energy consumption of sensors should be minimized since in most cases the sensors are battery-powered. Various methods have been proposed to work around this problem using scheduling approaches. In this paper, particle swarm optimization-based scheduling was designed and implemented to maximize the lifetime of wireless sensor networks formulated as a Non-Disjoint Sets Cover (NDSC) problem. The experimental findings show that the proposed approach is extremely competitive to the state-of-the-art algorithms, as it is able to find the optimal and best-known solutions in the instances investigated.

Keywords-*scheduling; target coverage problem; non-disjoint set covers; wireless sensor networks; lifespan; particle swarm optimization*

I. INTRODUCTION

Wireless Sensor Networks (WSNs) are a type of wireless and ad hoc network. WSNs combine sensing, processing, and networking over miniaturized sensor nodes (often hundreds or thousands). They are typically deployed to monitor large or hazardous areas [1]. WSNs offer some significant advantages, including the fact that they are less expensive to deploy than wired networks. Sensor nodes can be added and withdrawn with ease. Furthermore, the node's location can be modified without rewiring. Finally, WSNs can be configured into different network topologies (star, tree, mesh, etc.). Despite the distinct features of WSNs, particularly their simplicity and effective cost, they have an anomalous character related to their resource restrictions in terms of computing power and energy. In fact, in most cases, the sensor nodes are powered by batteries. Manual configuration, maintenance and battery replacement are often impossible. So, to overcome these energy constraints and to save energy consumption and therefore extend the network lifetime, many methods have been proposed, such as optimal deployment, clustering, multi-hop

routing, data aggregation, energy harvesting, and sleep-wake scheduling [2-4]. Much research is conducted in the direction of scheduling sensor activities. The deployed sensors are divided into a number of sensor sets each of which can cover all the targets and can send all the sensed data to the base station. These sensor sets can be disjoint or non-disjoint and are activated successively one by one. In each round, only one set is active. Only sensors in the active set are used to collect data from the surrounding environment and to relay them to the sink, whereas all the other sensors go into energy-saving sleep mode. Thus, each node or a cover of nodes should save as much power as possible by turning off the radio transmission when there is nothing to transmit. The process of clustering sensor nodes into cover sets and scheduling them to maximize the lifespan of the network belongs to the NP-hard problems. This kind of problem can be solved using either exact methods, like linear programming and branch-bound or metaheuristics, namely genetic, invasive weed optimization, and Particle Swarm Optimization (PSO) algorithms [4-8]. The former approach requires an exponential computation time depending on the size of the problem to be solved, while the latter tries to

obtain good solutions (not necessarily optimal) in a reasonable time. We focused on utilizing PSO since it has been proved to be effective in tackling a plethora of problems and it has not been investigated enough in extending the lifetime of WSNs.

In this paper, a PSO algorithm is used to solve WSN lifespan maximization problem, formulated as a non-disjoint set cover problem.

II. RELATED WORKS

This section reviews the related work on scheduling sensor activities to maximize network lifetime under application requirements that include target coverage. Lifetime is the operational time of the network during which it is able to perform the dedicated tasks while the coverage is a measure of the physical space the sensors are able to observe. Authors in [9] presented a greedy algorithm for solving the maximum lifetime coverage problem and the energy consumed in the sleep/active schedule. An empty blacklist is created when constructing a new set cover. Iteratively, a minimal energy sensor, which is not present in the blacklist and covers at least a new target is chosen. When there are many sensors with minimal energy, the selected sensor will be the one that covers the largest number of new targets. Unfortunately, this approach is not suitable for homogeneous networks. Authors in [10] suggested an improved cuckoo search algorithm that partitions sensors into a maximum number of NDSC with a sensing range that can be adjustable. Each cover provides k -target coverage to maximize the lifetime of the WSN. After calculating the upper bounds of the maximum number of NDSC, the authors find the covers that will be scheduled and activated one by one. Unfortunately, the k -target coverage consumes more energy than the single-target coverage. Authors in [11] proposed the bidirectional mutation hybrid Genetic Algorithm (GA) to find the maximum NDSC that completely covers all the targets and maximizes the lifetime of the WSN. This algorithm differs from the traditional GA in the chromosomes representation and in the use of a greedy technique for initialization rather than the random initialization. Also, a novel bidirectional mutation was utilized to speed up the convergence. The proposed algorithm is suitable for heterogeneous WSNs that contain sensors with different levels of initial energy. Authors in [12] recommended an exact method (LP formulation) and two polynomial-time greedy heuristics for target coverage problem in Directional Sensor Networks (DSNs). Their aim was to balance between maximizing DSN lifetime and fault tolerance. They provide optimal solutions when using linear programming but with high cost. However, at a lower cost, with the use of heuristics, the solutions obtained were suboptimal. Authors in [13] proposed two methods to extend the lifetime of WSNs adopting the NDSC approach rather than DSC. First, they used the binary coverage relations matrix to sort the randomly deployed sensors nodes and to find the binary relations that link sensors and targets to construct a maximum number of non-disjoint set covers. These obtained covers are scheduled utilizing an exact method and a GA with a novel gene coding to get a near-optimal solution in reasonable time. The exact method gives the optimal solution, but requires a lot of computational time. Also, the metaheuristic employed is not efficient in a network that has a large number of sensors. Authors in [14] suggested a

column generation method that finds the maximum NDSC to extend the lifetime of WSN. It is an exact algorithm that seeks to discover valid covers by a new integer linear programming model. The process is repeated iteratively and stopped only if there is no column with a positive reduced cost. The computation cost is decreased through the use of a branch-and-cut method. This reaches the best solution, but it is costly. Authors in [15] recommended a novel scheduling called Energy-Efficient Connected Coverage (EECC). EECC increased the level of coverage and connectivity of the sensors. It considers the remaining energy of each sensor and tries to avoid redundant coverage of critical points in the monitoring area. Unluckily, EECC was only effective with homogeneous WSNs containing sensors with the same initial energy. Authors in [16] proposed a novel local wake-up scheduling based on ant colony optimization. They constructed a first layer which contains a set of active sensors that completely cover the targets. Then, they manufactured multiple successor sets to mitigate the problem that some sensors in the first layer set run out of energy. This approach is effective, but it is suitable only for small and medium-scale networks.

Complexity theory classifies most scheduling problems as NP-hard [5]. This justifies the use of meta-heuristics (approximate methods), which provide acceptable (not necessarily optimal) solutions in a reasonable time. In this paper, PSO is used to solve the problem of randomly deployed sensors network lifespan maximization formulated as a scheduling problem. PSO has been proved to be an effective method for many optimization problems, and in some cases, it does not suffer from the difficulties experienced by other metaheuristics.

III. PROBLEM FORMULATION

The issue of WSN lifetime optimization based on the maximum number of NDSCs and their optimal scheduling belongs to the NP-hard family [5]. It presents two difficulties. The first one is the assignment of each sensor to a cover set, and the second one is the scheduling of these cover sets to optimize the network lifetime. Data, constraints, and objectives of WSN scheduling problem are defined below.

A. Data

- S represents a set of n sensor nodes. A sensor node is labeled as s_i ($i = 1, \dots, n$).
- T represents a set of m targets. A target is labeled as t_j ($j = 1, \dots, m$).
- s_i can monitor a subset of targets $T(s_i) \in T$ located in its coverage range r .
- Each target t_j could be monitored by a subset of sensors $S(t_j) \in S$.
- A collection of elements of S denoted C_l is a cover for the subset of targets denoted $T(C_l)$ if it can sense all the targets of $T(C_l) \in T$.
- C_l is considered as a cover if $T(C_l) = T$.
- Each sensor node S_i has an initial energy E_i .

- $E_i(k)$ is the total energy consumed by S_i during a period of time k .

B. Constraints

- Each target t_j is covered by at least one sensor of S .
- A sensor can be included into at most one set cover in the case of Disjoint Set Cover (DSC).
- A sensor can be included into more than one cover in the case of Non-Disjoint Set Covers (NDSC).

C. Criteria

- We have to maximize the lifespan, which is the time elapsed until all the available sensor nodes do not succeed to satisfy the targeted requirements. It can be expressed as [17]:

$$L = k \times \sum_{i=1}^q y_i \quad (1)$$

D. Example

To illustrate the benefit of using NDSC instead of DSC, let us assume that the three targets (Target₁, Target₂, and Target₃) in Figure 1 can be monitored by three sensor nodes (Sensor₁, Sensor₂, and Sensor₃).

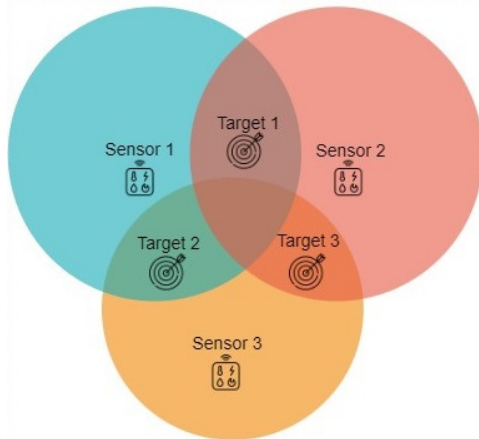


Fig. 1. Example of topology of three targets monitored by three sensors.

Sensor₁ covers Target₁ and Target₂. Sensor₂ covers Target₁ and Target₃, and Sensor₃ covers Target₂ and Target₃. If all the sensor nodes were to be activated simultaneously, then the network lifetime would be equal to the standard lifetime h of a single sensor. By dividing the sensors into disjoint sets, the resulting network lifetime would still be h , since for this topology a disjoint algorithm can only produce one cover set (e.g. $Cover_1 = \{Sensor_1, Sensor_2\}$ or $Cover_2 = \{Sensor_1, Sensor_3\}$, or $Cover_3 = \{Sensor_2, Sensor_3\}$). However, if a sensor node in Figure 1 can be part of two cover sets, then the network lifetime can be extended. By creating three non-disjoint cover sets (see Figure 2), each one activated for $0.5 \times h$ hours, the total network lifetime can be extended to $1.5 \times h$ (see Figure 3), assuming that the energy consumption during the sleep mode is negligible.

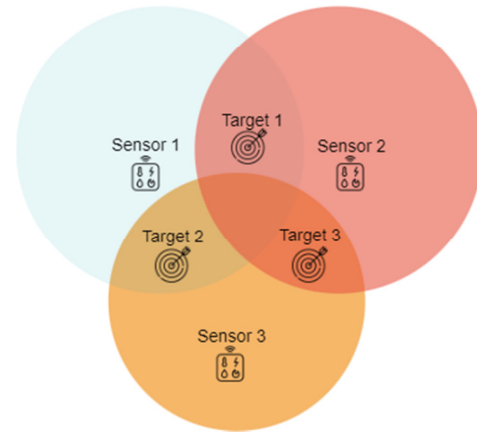
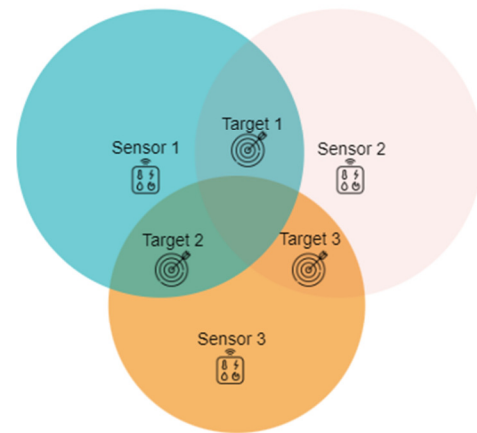
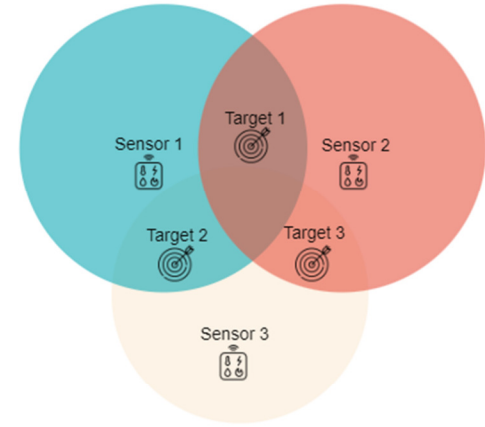


Fig. 2. Three obtained non-disjoint cover sets.

From the above observation, a sensor node can spend part of its energy within one cover and another part within another. So, finding the optimal lifespan requires solving two sub-problems: (1) finding the optimal number of NDSC and (2) maximizing lifespan by scheduling. This problem belongs to the NP-hard family and that is why metaheuristics are considered in solving it. This paper aims to develop a new PSO-based method that can efficiently find the maximum number of NDSC for a set of sensors that can be effectively scheduled to prolong the WSN lifespan.

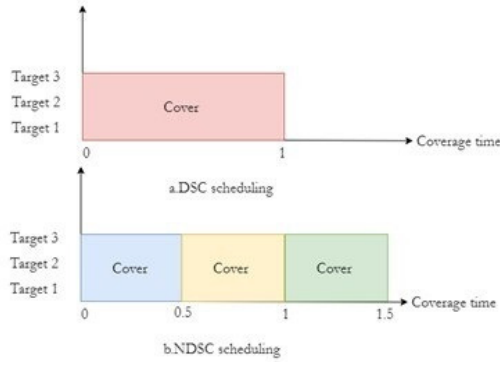


Fig. 3. GANTT in case of DSC and NDSC.

IV. PROPOSED APPROACH

Given that metaheuristic algorithms are competent in finding the best solution in an acceptable amount of time for especially hard problems, the following NDSC-based PSO is proposed to figure out the optimal number of non-disjoint cover sets and then to schedule them in order to maximize the lifespan of the network.

A. Particle Swarm Optimization

PSO was inspired by the social behavior of fish and birds. Each particle of the swarm (individual solution in a population), which is a candidate solution for the optimization problem, performs an individual search (cognitive search) and a global search to minimize an error function. PSO is built on the concept of progressively evolving a swarm of possible solutions to an optimization problem. Each particle has a position denoted by $X^i \in R^n, i = 1, \dots, n$, where n is the number of particles (swarm size), and a velocity to determine and move to the next particle denoted by $V^i \in R^n$. Every particle has also a fitness value to evaluate its quality. During the search process and over each iteration t , the particles move and update their velocities and positions according to (2) and (3) [18]:

$$V_{t+1}^i = \omega V_t^i + C_1 r_1 (p_t^i - X_t^i) + C_2 r_2 (g_t^i - X_t^i) \quad (2)$$

$$X_{t+1}^i = X_t^i + V_{t+1}^i \quad (3)$$

where ω is the inertia weight, C_1 and C_2 are the cognition learning and the social learning rates, respectively, r_1 and r_2 are uniformly random numbers in the range of $[0,1]$, p_t^i and g_t^i are the best personal and global solutions, equivalently, X_{t+1}^i is the modified position of the i^{th} particle, and t is the current iteration.

PSO can be used to solve problems with continuous and discrete variables. It is also applicable to multi-objective and constraint satisfaction problems. More details on PSO, its variants and applications, are referred to [16-26].

1) Coding

The first step in using PSO to solve any optimization problem is to map between the particle position of the swarm evolution concept and the special case of the investigated problem. For a set S of sensors used to monitor a set T of targets, the input of the PSO algorithm is the coverage relation

matrix, which explains the targets $T(i)$ from T that are covered by each sensor s_i from S . The sum of the monitoring periods will be equal to the network lifespan. Figure 4 represents the structure of a given particle.



Fig. 4. A particle coding.

2) Initialization

The particle swarm (or population) can be either generated using a random initialization or can be issued from an output of any other heuristic method.

3) Position and Velocity Update

At each iteration t , the position and velocity of every particle (i.e. a potential solution of the considered optimization problem) are updated according to (2) and (3).

4) Fitness Evaluation

For each particle, the fitness value is the sum of the whole scheduled period y_l multiplied by the period k . It can be mathematically formulated as in (1).

5) Termination

The simplest condition of termination is to let the PSO simulation stop once it reaches a maximum number of iterations, which is predefined before starting the optimization process.

V. NUMERICAL RESULTS

The algorithms were coded implementing python programming language on Windows 10 and with i5 8th gen processor.

A. PSO-based Method for finding the NDSC

This study simulated a set of randomly deployed sensors $S = 10, 20, 30, \dots, 500$, which have a constant sensing range equal to 3 and are used to monitor 5 targets in a 10×10 area. The hyperparameters of PSO are $C_1 = C_2 = 2$ and $\omega = 0.75$. A different number of iterations and population sizes were investigated to find the NDSCs. Table I presents the number of NDSCs found engaging different small numbers of sensors utilized to monitor 5 targets, with number of iterations = 20 and population size = 500. It can be clearly noticed that the number of non-disjoint cover sets increases when the number of deployed sensors increases.

TABLE I. NDSCS FOR A SMALL NUMBER OF SENSOR NODES

Sensors	10	20	30	40	50
NDSCs	1397	2965	5556	5985	8224

Table II depicts the number of NDSCs for a large number of sensors used to monitor 5 targets, number of iterations = 20, and population size = 500. Through the previous table it can be observed that the number of non-disjoint cover sets increases when the number of deployed sensors raises and especially when this number exceeds the range of hundreds (100, 200, 300, ..., 500).

TABLE II. NDSCS FOR A LARGE NUMBER OF SENSOR NODES

Sensors	100	200	300	400	500
NDSCs	9778	9999	10000	10000	10000

Table III presents the effect of the population size on the results (using the same number of iterations = 20). It can be detected that the number of non-disjoint cover sets increases when the number of particles used in the swarm raises. This can lead to the conclusion that the number of the non-disjoint cover sets can be maximized, but this can influence the execution time of PSO.

TABLE III. NDSCS WITH DIFFERENT POPULATION SIZE

Population size	100	200	300	400	500	600
NDSCs	275	600	817	1101	1379	1613

Table IV provides the running time (ms) for different numbers of sensors with a different number of iterations. It can be spotted that for a fixed size of population, runtime raises when the number of deployed sensors increases. In the same direction, when the number of sensors were fixed and the size of the swarm was varied, the runtime increases.

TABLE IV. INSTANCES RUNTIME

Population Iterations	100 sensors	200 sensors	300 sensors	400 sensors
500	10	13	17.51	20.613
1000	19.84	25	34.25	41.185
1500	29.67	41.4	50.54	61.01
2000	40.5	53.6	69.09	78.92

B. PSO-based Method for finding the Optimal Scheduling

Table V compares the results obtained by the proposed NDSC-PSO approach and those provided by NDSC-GA and the exact method (ILP) on the same eight instances [13].

TABLE V. COMPARISON OF ILP, NDSC-GA, AND NDSC-PSO ALGORITHMS

Instances $S, T, r, E_i, E_t(k)$	Lifetime		
	ILP	GA	PSO
5,5,3,160,16	20	20	20
5,5,3,160,8	40	40	40
5,5,3,160,4	80	80	80
5,5,3,160,2	160	160	160
10,5,3,160,8	30	30	30
10,5,3,160,4	60	59	60
10,5,3,160,2	120	117	120
10,5,3,160,1	240	238	240

In Table V, it can be observed that the results obtained by the proposed NDSC-PSO approach are better than those acquired with the NDSC-GA method. Indeed, the findings of NDSC-PSO reached the optimal solution for all the considered instances.

VI. DISCUSSION

After the conduction of many experiments, some conclusions can be drawn:

- To improve the acquired results, the swarm size could be enlarged, at the cost of execution speed.

- The second finding concerns the effect of the number of sensors deployed on the complexity of the problem being addressed.
- NDSC-PSO outperforms other effective approaches known in the literature.

The first and second observations can be considered obvious. The third one is the main contribution and the most important finding of the current study. In fact, in all instances tested in this work, the NDSC-PSO reaches the maximum lifetime. This result can be further generalized by testing other new instances that address the same kind of problem (WSN), but with different constraints and objectives (e.g. sensor type: fixed, mobile, random, homogeneous, heterogeneous, fully/partially connected, mono/multi criteria problem, etc.).

VII. CONCLUSION

In this paper, a NDSC-based PSO was investigated and integrated into a scheduling approach to maximize the lifespan of WSNs. The experimental results were implemented in the case of random deployment of sensor nodes. The obtained outcomes were compared with those derived through an exact method, namely the ILP model, and those acquired via a metaheuristic, namely NDSC-GA, known in the literature. For all the cases considered, the NDSC-PSO approach outperformed NDSC-GA and provided results close to those of ILP. The experimental findings are very encouraging as the proposed method is able to discover the best network lifetime for the studied instances. A more comprehensive study should be conducted on a larger number of cases. Further research could focus on additional constraints of the sensor network, such as the mobility and heterogeneity of nodes. Another future aim would be to test the usage of a fully adaptive PSO algorithm called TRIBES or the hybridization of PSO and other heuristics.

ACKNOWLEDGMENT

This work was funded by the University of Jeddah, Jeddah, Saudi Arabia, under grant No. (UJ-20-119-DR). The authors thank the University of Jeddah for the technical and financial support.

REFERENCES

- [1] H. T. T. Binh and N. Dey, Eds., *Soft Computing in Wireless Sensor Networks*, 1st ed. Boca Raton, FL, USA: CRC Press, 2018.
- [2] Y. Emami and R. Javidan, "An Energy-efficient Data Transmission Scheme in Underwater Wireless Sensor Networks," *Engineering, Technology & Applied Science Research*, vol. 6, no. 2, pp. 931–936, Apr. 2016, <https://doi.org/10.48084/etasr.629>.
- [3] S. P. Singh and S. C. Sharma, "A Novel Energy Efficient Clustering Algorithm for Wireless Sensor Networks," *Engineering, Technology & Applied Science Research*, vol. 7, no. 4, pp. 1775–1780, Aug. 2017, <https://doi.org/10.48084/etasr.1277>.
- [4] A. Rajab, "Genetic Algorithm-Based Multi-Hop Routing to Improve the Lifetime of Wireless Sensor Networks," *Engineering, Technology & Applied Science Research*, vol. 11, no. 6, pp. 7770–7775, Dec. 2021, <https://doi.org/10.48084/etasr.4484>.
- [5] V. Kesavan, R. Kamalakannan, R. Sudhakarapandian, and P. Sivakumar, "Heuristic and meta-heuristic algorithms for solving medium and large scale sized cellular manufacturing system NP-hard problems: A comprehensive review," *Materials Today: Proceedings*, vol. 21, pp. 66–72, Jan. 2020, <https://doi.org/10.1016/j.matpr.2019.05.363>.

- [6] S. Mekni, B. Chaar, and M. Ksouri, "A Novel Particle Swarm Optimization Approach for Multiobjective Flexible Job Shop Scheduling Problem," in *ICINCO-ICSO 2008*, 2008.
- [7] S. Mekni and B. C. Fayeche, "A Modified Invasive Weed Optimization Algorithm for MultiObjective Flexible Job Shop Scheduling Problems," *Computer Science & Information Technology*, vol. 10, pp. 51–60, Nov. 2014, <https://doi.org/10.5121/csit.2014.41106>.
- [8] S. Mekni and B. Chaar Fayeche, "Multiobjective Flexible Job Shop Scheduling Using A Modified Invasive Weed Optimization," *International Journal on Soft Computing*, vol. 6, no. 1, pp. 25–36, Feb. 2015, <https://doi.org/10.5121/ijsc.2015.6103>.
- [9] N. T. Hanh, H. T. T. Binh, N. V. Son, N. T. Trang, and P. N. Lan, "Optimizing wireless sensor network lifetime through K-coverage maximization and memetic search," *Sustainable Computing: Informatics and Systems*, vol. 40, Dec. 2023, Art. no. 100905, <https://doi.org/10.1016/j.suscom.2023.100905>.
- [10] D. Arivudainambi, R. Pavithra, and P. Kalyani, "Cuckoo search algorithm for target coverage and sensor scheduling with adjustable sensing range in wireless sensor network," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 24, no. 4, pp. 975–996, May 2021, <https://doi.org/10.1080/09720529.2020.1753301>.
- [11] J. Li, Z. Luo, and J. Xiao, "A Hybrid Genetic Algorithm With Bidirectional Mutation for Maximizing Lifetime of Heterogeneous Wireless Sensor Networks," *IEEE Access*, vol. 8, pp. 72261–72274, 2020, <https://doi.org/10.1109/ACCESS.2020.2988368>.
- [12] V. R. Ekhlas, M. Hosseini Shirvani, A. Dana, and N. Raeisi, "Discrete grey wolf optimization algorithm for solving k -coverage problem in directional sensor networks with network lifetime maximization viewpoint," *Applied Soft Computing*, vol. 146, Oct. 2023, Art. no. <https://doi.org/10.1016/j.asoc.2023.110609>.
- [13] Y. E. E. Ahmed, "Modeling, Scheduling and Optimization of Wireless Sensor Networks lifetime," Ph.D. dissertation, Université de Lorraine, Lorraine, France, 2016.
- [14] M. E. Keskin, "A column generation heuristic for optimal wireless sensor network design with mobile sinks," *European Journal of Operational Research*, vol. 260, no. 1, pp. 291–304, Jul. 2017, <https://doi.org/10.1016/j.ejor.2016.12.006>.
- [15] D. K. Sah, S. Srivastava, R. Kumar, and T. Amgoth, "An energy efficient coverage aware algorithm in energy harvesting wireless sensor networks," *Wireless Networks*, vol. 29, no. 3, pp. 1175–1195, Apr. 2023, <https://doi.org/10.1007/s11276-022-03125-3>.
- [16] M. Suresh Kumar and G. A. Sathish Kumar, "Enhanced ant colony optimization algorithm for packet delivery with improved energy efficiency in wireless sensor networks," *Journal of Intelligent & Fuzzy Systems*, vol. 44, no. 5, pp. 7909–7917, Jan. 2023, <https://doi.org/10.3233/JIFS-221856>.
- [17] Y. E. E. Ahmed, K. H. Adjallah, R. Stock, I. Kacem, and S. F. Babiker, "NDSC based methods for maximizing the lifespan of randomly deployed wireless sensor networks for infrastructures monitoring," *Computers & Industrial Engineering*, vol. 115, pp. 17–25, Jan. 2018, <https://doi.org/10.1016/j.cie.2017.09.049>.
- [18] M. Clerc, *Particle Swarm Optimization*, 1st ed. Hoboken, New Jersey, USA; London, UK: Wiley-ISTE, 2006.

Digital Image Forensics: An Improved DenseNet Architecture for Forged Image Detection

Ahmed Alzahrani

Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia
aalzahrani9@kau.edu.sa (corresponding author)

Received: 7 February 2024 | Revised: 28 February 2024 and 4 March 2024 | Accepted: 5 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7029>

ABSTRACT

Images sent across internet platforms are frequently subject to modifications, including simple alterations, such as compression, scaling, and filtering, which can mask possible changes. These modifications significantly limit the usefulness of digital image forensics analysis methods. As a result, precise classification of authentic and forged images becomes critical. In this study, a system for augmented image forgery detection is provided. Previous research on identifying counterfeit images revealed unexpected outcomes when using conventional feature encoding techniques and machine learning classifiers. Deep neural networks have been also utilized in these efforts, however, the gradient vanishing problem was ignored. A DenseNet model was created to tackle limitations inherent in typical Convolutional Neural Networks (CNNs), such as gradient vanishing and unnecessary layer requirements. The proposed DenseNet model architecture, which is composed of densely connected layers, is designed for precise discrimination between genuine and altered images. A dataset of forged images was implemented to compare the proposed DenseNet model to state-of-the-art deep learning methods, and the results showed that it outperformed them. The recommended enhanced DenseNet model has the ability to detect modified images with an astonishing accuracy of 92.32%.

Keywords-image forgery detection; convolutional neural networks; digital image forensics; deep learning; DenseNet

I. INTRODUCTION

With the advent of digital image browsing on individuals' devices and their widespread distribution via the World Wide Web, severe security concerns regarding the authenticity of images have emerged [1]. Due to the availability of powerful and user-friendly photo altering software, digital photographs are often susceptible to being altered [2]. Digital images frequently undergo deliberate aberrations during processing, the most typical of which are copy-move, interpolation, and merging. Each of these techniques is employed to mask or edit details in a digital image [3]. Forensic imaging methods are required to determine an image's authenticity, processing history, and originality [4].

A. Research Motivation

The motivation for this research stems from the urgent necessity to tackle the rising problem of picture fraud and manipulation in today's digital world. The advent of sophisticated image editing software has increased the possibility of coming across false and counterfeit images. Traditional counterfeit detection systems frequently fail to detect subtle alterations and produce reliable results [5]. Researchers have proposed numerous strategies for uncovering

image forgeries [4, 6-8]. Traditional forgery detection approaches focus on seeing numerous artifacts inside modified images, such as changes in lighting, contrast, compression, sensor disturbances, and reflections. In the current state of research [1-4], supervised Machine Learning (ML) and Deep Learning (DL) algorithms are used in image forgery detection. Conventional methods for spotting image forgeries have accuracy and robustness limits, especially when dealing with sophisticated modifications. DL, a cutting-edge technique, offers a way to address these issues by automatically learning and extracting detailed information from images [4]. Considerable performance loss can be attributed to improper neural network parameter and layer selection. The above discussion emphasizes the DenseNet-based DL approach's potential as a feasible tool for effectively identifying forged images. While Convolutional Neural Networks (CNNs) show great promise, applying innovative architectures could improve their results. Previous investigations were limited by the lack of training data points and some other challenges, such as gradient fading and the requirement for a large layer count. This study, on the other hand, makes extensive use of training samples. As a result, the primary motivation of the former is the employment of an advanced CNN model with adequate training data for reliable forged image classification.

B. Problem Statement and Discussion

This study focuses on the issue of classifying images with the goal of creating a system that can detect and categorize types of image manipulation. The main challenge lies in spotting forging artifacts, which can undermine the trustworthiness and genuineness of visual material across various fields. To address this problem, we propose the use of DenseNet, a DL framework renowned for its ability to capture details and patterns within images. By enhancing the precision and dependability of algorithms, for detecting image tampering with the features of DenseNet, this study aims to overcome the shortcomings encountered by CNN and ML classifiers, which struggle with gradient disappearance and require complex layering to effectively identify changes in images. The goal is to boost the effectiveness of image manipulation detection, thereby safeguarding the authenticity of the content.

C. Research Contribution

The main raised Research Objectives (ROs) are:

- RO1: Apply a DenseNet-based DL model to detect image forgeries in digital forensics.
- RO2: Compare the results of the proposed DL model to those of other ML and DL strategies.
- RO3: Evaluate the effectiveness of the proposed model related to other comparable techniques.

When answering those ROs, the following important contributions were made:

- The proposed system employs a newly adopted DenseNet-based DL algorithm to carry out digital image forensics for the purpose of identifying image forgery.
- A suitable number of layers are incorporated into the suggested DL approach.
- Extensive experimental evaluations were used to evaluate the DenseNet's performance across a variety of measures, including precision, recall, f1-score, and accuracy.
- The study's originality lies in establishing the efficacy of the DenseNet121 CNN as a superior pre-trained model.
- The study demonstrates the effectiveness of using a lighter model architecture to achieve equivalent results with computationally more complex models.
- Evaluation of the suggested model's competence in the light of comparable research.

II. LITERATURE REVIEW

Several strategies based on ML and pattern identification have been explored in an effort to improve the detection efficiency of counterfeit photographs. To detect manipulative operations and the operator's sequencing for two operators in specific, authors in [2] presented a strong medium CNN. The limiting solution deepens the system and reduces processing. To maximize information exchange and prevent the layer fitting problem, a global average max-pooling was used. Authors in [4] showcased a strong DL system for detecting

image counterfeiting in dual compression techniques. Their model was trained by comparing genuine and recompressed photographs. An end-to-end trainable BusterNet-based DL framework for duplicate image fraud prevention and detection was proposed in [5]. Authors in [6] employed DL methods to detect image forgeries in real-world datasets. Their compact model outperformed the baseline methods. The research conducted in [10] presents a revolutionary DL-based strategy for pinpointing copy-move forgeries. The method uses modified dense peak clustering to split images into patches, attention-based DenseNet121 to extract features, and adaptive chimp patch matching to match patches. Even in compressed or modified images, the method finds forged regions with an improved accuracy and less processing time. Prior research has demonstrated DenseNet's advantage over other CNN models in image classification, due to its low parameter count. Notably, DenseNet has displayed success in image forensics' analysis, particularly in spotting manipulated images.

Despite the efficiency and widespread usage of simple ML and DL approaches in image forgery detection, most previous efforts struggled to improve classification accuracy. Performance degrades due to improper neural network model parameter and layer selection. Furthermore, despite the widespread implementation of standard ML methods, only a few researchers applied DenseNet121 to categorize real and forged images. As a result, difficulties such as improving model accuracy while simplifying the model by using less parameters, layers, depth, runtime, and model size, persist. The proposed DenseNet model detects digital image counterfeiting using several layers and hyper parameters. The proposed model was also compared with benchmark research.

III. THE PROPOSED TECHNIQUE

The primary goal of this study is to use DL technique to conduct digital forensics by identifying instances of image forgery. The proposed procedure works as follows: First, a benchmark dataset is needed to evaluate performance; second, preprocessing methods are employed to clean the data of unwanted noise, and third, a DL system, namely DenseNet, is deployed to reach the ultimate prediction (Figure 1).

A. Data Set

The publically available CASIA V2.0 image tampering detection evaluation database [7] was utilized, as outlined in Table I. This accessible for research purposes database allows for the comparison and assessment of tampering detection methods. It is specifically created for identifying forgeries and is divided into two groups; altered/fake and genuine/original. The dataset consists of a total of 12,323 images with 7,200 authentic and 5,123 altered images.

TABLE I. INFORMATION ABOUT THE CASIA.2.0 DATASET

Real	Fake	Dimensions	Format
7200	5123	From 320x240 to 800x600 color images	JPEG,BMP,TIFF

B. Input Image Preprocessing

The ImageDataGenerator class and API [8] was put into service to perform some preliminary processing on the images

in Keras. The `imageDataGenerator` class was deployed, resizing both training and testing images to 1./255, applying a shear intensity of 0.2, and randomly varying the zoom range by 0.2. It is essential to arrange the data in such a way that they can be processed by the CNN after they have been preprocessed, since the CNN can only comprehend the numerical description of the picture data. Consequently, this is why the convolutional layer's "input shape" parameter displays information as a vector. In an array, numbers represent the values of individual pixels. For instance, the network interprets an input piece of data as a 64×64×3 numeric array (where 64 represents height, 64 represents width, and 3 represents depth) [4]. Following the generation of the raw image matrix, the CNN processes the data.

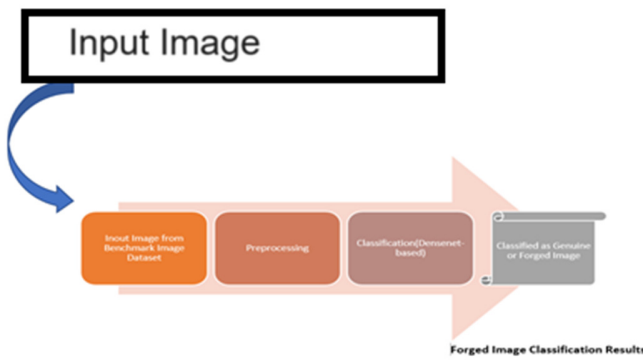


Fig. 1. Block diagram of the proposed system.

C. Existing DenseNet Architecture

The baseline DenseNet architecture [10] is a robust CNN architecture for visual object recognition that is noted for providing top-tier performance with fewer parameters. It is closely related to ResNet with some important modifications. DenseNet uses a concatenation attribute to integrate the output of previous and subsequent layers, whereas ResNet utilizes an additive attribute to combine previous and future levels [11, 12]. Traditional CNNs compute the output of the l^{th} layer by subjecting the previous layer's output to a nonlinear transformation designated as $H_l(\cdot)$.

$$x_l = H_l(x_{l-1}) \quad (1)$$

DenseNet differs from the standard practice of merely adding layer output feature maps to inputs by concatenating them. DenseNet's distinct methodology provides a smooth communication paradigm, improving the flow of information across layers. In particular, the inputs in the l^{th} layer are taken from the properties of all preceding layers:

$$x_l = H_l(\llbracket x_0, x_1, x_2, \dots, x_{l-1} \rrbracket) \quad (2)$$

The tensor $[x_0, x_1, x_2, \dots, x_{l-1}]$ is created by concatenating the output maps from the previous layers [10]. $H_l(\cdot)$ is a nonlinear transformation function. This function consists of three primary operations: Batch Normalization (BN), activation (ReLU), and a pooling and convolution combination. The growth rate, indicated as k , contributes to the generalization of the l^{th} layer as follows:

$$k_l = k_0 + k \times (l - 1) \quad (3)$$

where k_0 represents the array of the channels.

D. Improved Densenet Architecture for Forged Image Detection

The proposed model is a modified version of the DenseNet121 framework that has been enhanced with additional layers that have been painstakingly tailored to the unique requirements of identifying forged and genuine image data. More particular, the model makes use of the DenseNet121 neural network's intrinsic densely interconnected topology. However, as shown in Figure 2, it incorporates tailored changes to improve performance with respect to the targeted dataset.

1) How it Works

The proposed DenseNet architecture is made up of dense blocks with different repetitions of convolutional layers. Each dense block has two layers (1×1 and 3×3), with the former acting as a bottleneck layer to minimize input channels before convolution. Transition layers include 1×1 convolution and 2×2 average pooling (stride 2). The initial layers in the architecture include a 7×7 stride-2 convolutional layer and a 3×3 stride-2 max pooling layer. It then has three thick blocks (with 6, 12, and 24 repeats) followed by corresponding transition layers. The last dense block has 16 repetitions and leads to global average pooling for classification and the output layer. Because of its dense connections and bottleneck layers, DenseNet121 excels in computer vision applications. The proposed extension (Figure 3) improves model performance and training efficiency with fewer parameters [12]. The configuration parameters of the suggested approach are depicted in Table II.

TABLE II. DENSENET ARCHITECTURE

Layers	Output size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	110×110	5×5 Conv, stride 2			
Pooling	40×40	4×4 max pool, stride 2			
Dense block-1	40×40	$\begin{bmatrix} 1 & 1 \\ 4 & 4 \end{bmatrix} \times 8$	$\begin{bmatrix} 1 & 1 \\ 4 & 4 \end{bmatrix} \times 8$	$\begin{bmatrix} 1 & 1 \\ 4 & 4 \end{bmatrix} \times 8$	$\begin{bmatrix} 1 & 1 \\ 4 & 4 \end{bmatrix} \times 8$
Transition layer-1	40×40	1×1 Conv			
	20×20	2×2 average pool, stride 2			
Dense block-2	20×20	$\begin{bmatrix} 1 & 1 \\ 4 & 4 \end{bmatrix} \times 16$	$\begin{bmatrix} 1 & 1 \\ 4 & 4 \end{bmatrix} \times 16$	$\begin{bmatrix} 1 & 1 \\ 4 & 4 \end{bmatrix} \times 16$	$\begin{bmatrix} 1 & 1 \\ 4 & 4 \end{bmatrix} \times 16$
Transition layer-2	20×20	1×1 Conv			
	15×15	2×2 average pool, stride 2			
Dense block-3	15×15	$\begin{bmatrix} 1 & 1 \\ 4 & 4 \end{bmatrix} \times 32$	$\begin{bmatrix} 1 & 1 \\ 4 & 4 \end{bmatrix} \times 48$	$\begin{bmatrix} 1 & 1 \\ 4 & 4 \end{bmatrix} \times 64$	$\begin{bmatrix} 1 & 1 \\ 4 & 4 \end{bmatrix} \times 82$
Transition layer-3	15×15	1×1 Conv			
	5×5	2×2 average pool, stride 2			
Dense block-4	5×5	$\begin{bmatrix} 1 & 1 \\ 4 & 4 \end{bmatrix} \times 24$	$\begin{bmatrix} 1 & 1 \\ 4 & 4 \end{bmatrix} \times 48$	$\begin{bmatrix} 1 & 1 \\ 4 & 4 \end{bmatrix} \times 48$	$\begin{bmatrix} 1 & 1 \\ 4 & 4 \end{bmatrix} \times 64$
Classification layer	1×1	2×2 global average pool, stride 2			
		1000D fully connected soft max			

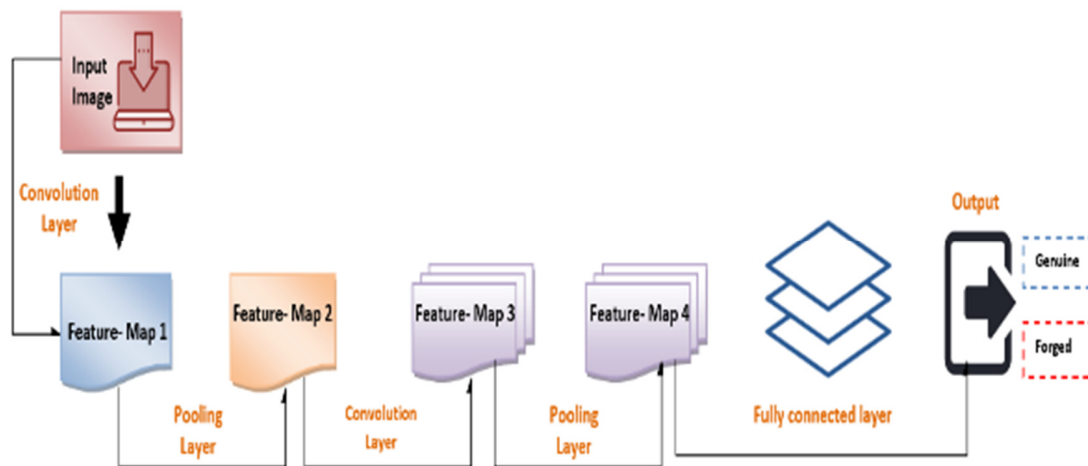


Fig. 2. Improved DenseNet's proposed architecture.

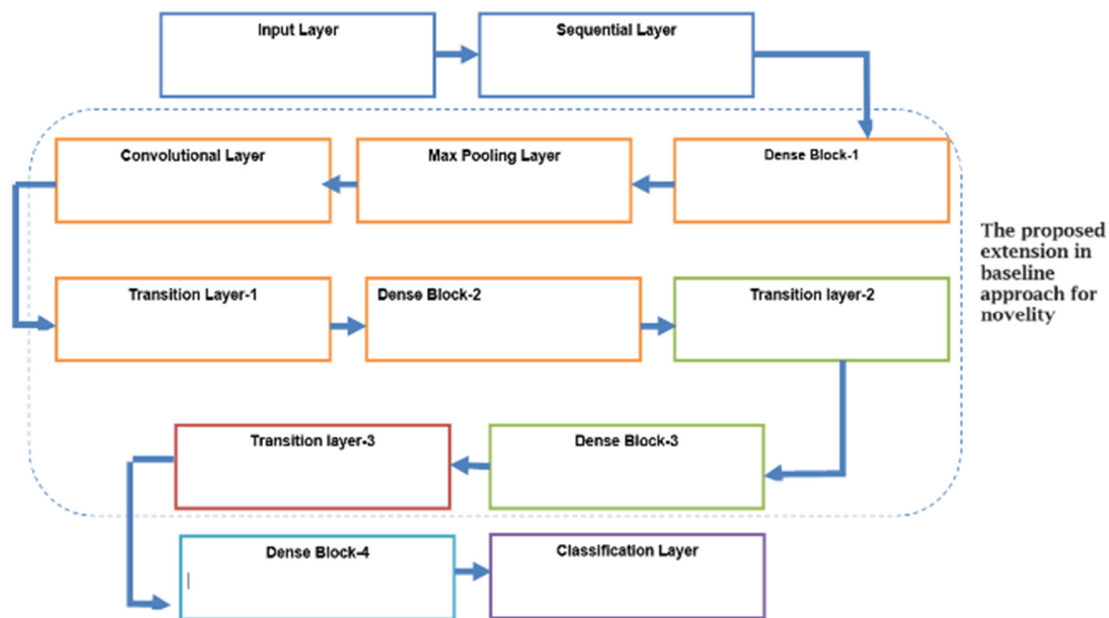


Fig. 3. The proposed improved DenseNet architecture for forged image detection.

The following algorithm describes the pseudocode steps of the proposed system:

```
Data Preprocessing:
def preprocess_data(data_path):
    #Load the CASIA 2.0 dataset
    real_images, fake_images =
    load_dataset(C://CASIA2.0dataset//images)
    #Image resizing and normalisation
    real_images =
    resize_and_normalize(real_images)
    fake_images =
    resize_and_normalize(fake_images)
    #Combine actual and forged images with
    their respective labels
    data =
    np.concatenate((real_images, fake_images))
```

```
labels =
np.concatenate((np.zeros(len(real_images))
,
np.ones(len(fake_images))))
#Shuffle data and labels To avoid biases
data, labels = shuffle(data, labels)
Model Definition:
def build_model():
    #Load pre-trained DenseNet-121 model with
    ImageNet weights
    model =
    DenseNet121(weights="imagenet", include_top
    =False)
    #Freeze pre-trained layers
    for layer in model.layers:
        layer.trainable = False
```

```
#Add custom layers for binary
classification
x = model.output
x = Flatten()(x)
x = Dense(1,activation="sigmoid")(x)
#Compile the model with appropriate loss
function,optimizer,and metrics
model =
Model(inputs=model.input,outputs=x)
model.compile(loss="binary_crossentropy",
optimizer="adam",metrics=["accuracy"])
return model
Model Training:
deftrain_model(model,data,labels,epochs):
#Define data augmentation for improved
generalizability (optional)
datagen = ImageDataGenerator(...)
#Train the model with the specified
data,labels,and epochs
model.fit(datagen.flow(data,labels,batch_s
ize=32),epochs=epochs)
Image Prediction:
defpredict_image(model,image_path):
#Load and preprocess the image
image =
load_and_preprocess_image(image_path)
#Predict the class probability
prediction =
model.predict(np.expand_dims(image,axis=0)
)[0][0]
#Determine and print the result
if prediction < 0.5:
print("Image is classified as Real")
else:
print("Image is classified as Fake")
Evaluation and Analysis:
defevaluate_model(model,data,labels):
#Evaluate model performance on test data
loss,accuracy =
model.evaluate(data,labels)
print(f"Test loss: {loss:.4f}")
print(f"Test accuracy: {accuracy:.4f}")
#Generate and analyze Class Activation
Maps (CAMs) for insights into model's
decision
```

2) Applied Example

A two-dimensional grid of pixel values was the first representation of the digital image 2620.jpg, which was stored on a hard drive. Red, Green, and Blue (RGB) values were used to create a 3-tuple that represented the color information for each pixel. It was crucial to convert this 2D image into a 3D array with 220×220×3 dimensions in order to prepare it for inclusion into a DL model. The RGB values of each pixel were divided into several channels and the image was enlarged to 224×224 pixels. For model compatibility, the RGB values were scaled down from their original range of 0 to 255 to a range of 0 to 1 by dividing each value by 255. Figure 4 provides a

graphic breakdown of the procedure [13]. The common method of pooling was implemented to optimize the input image for use in DenseNet. The supplied image's dimensions were reduced from 224×224×3 to 7×7×3 with this technique. Pooling minimizes the quantity of the input data while preserving key characteristics. The input image is divided into more manageable, non-overlapping portions, and a representative statistic is computed for all sections. In this illustration, the preprocessed 3D array image was separated among two segments of 2×2 size, and the highest value inside each segment was kept. The 3D array's dimensions were effectively decreased to 7×7×3 with this step.

After the pooling operation, a flatten layer, a crucial component in many DL models, processed the resulting 3D array. This layer combined all array values along a single dimension, resulting in a vector spanning of 147 units, and converted the 3D array into a single 1D array. This flatten layer's function is to transmute the output of the preceding layer. It may have various dimensions, inside a vector of fixed length suitable for the next layers. This procedure is essential for creating a model input structure that is consistent and matches the predicted insertion data shape. This 1D array incorporates, in a condensed and standardized structure, all the appropriate information that was taken from the input image and has been preprocessed and pooled [14]. The model's later layers can analyze data more quickly thanks to this format. Figure 5 illustrates the entire series of actions outlined above.

Following the preprocessing and flattening of the input image, the ensuing vector/ray is directed via the concealed layers of the suggested DL model. The flattened vector, spanning 147 units, is initially inputted into a concealed layer harboring 50 neurons. Within this layer, each neuron takes the input vector (x) and conducts a multiplication with a distinctive weight matrix (w) designated for that neuron. Subsequently, the outcome of this multiplication is augmented by a distinct bias term (b) allocated to the same neuron. These actions yield a scalar value known as z for every neuron residing in the layer.

Rectified Linear Unit (ReLU) activation function transforms the z -values, adding a nonlinear component to each neuron's output. The output of the layer, 50 dimensional vectors, has been sent to the next layer that is hidden with 30-neurons. Within this layer, a similar process takes place and results in a vector with a length of 30. Two neurons are housed in the output layer of the model. A softmax activation function is encountered by the z -values in this layer, normalizing the data and generating a probability distribution between the two classes. This leads to an output 2D vector in which every component depicts the probability that the given class will handle the input image. The two considered classes are genuine and forged. The input image classification is affected by the softmax result being measured against a threshold. The output of the model is shown as a $[0, 1]$ vector, denoting the image's classification. Figure 6 provides a clear illustration of the process involving the connection of the flattened vector to the layers in the suggested DL model and the operations in the hidden and output layers.

3) Implementation of the Proposed Model

Python programming language and appropriate libraries designed for the DenseNet121 architecture were employed to implement the proposed DenseNet model. A DL framework like TensorFlow or PyTorch allows to build and train the model by streamlining the computation of the architecture's

parameters and layers. Table III offers a brief summary of the key features of the DenseNet model. This summary is extremely helpful in understanding the model's structural composition and provides a starting point for comparison with similar models in the field of natural image identification.

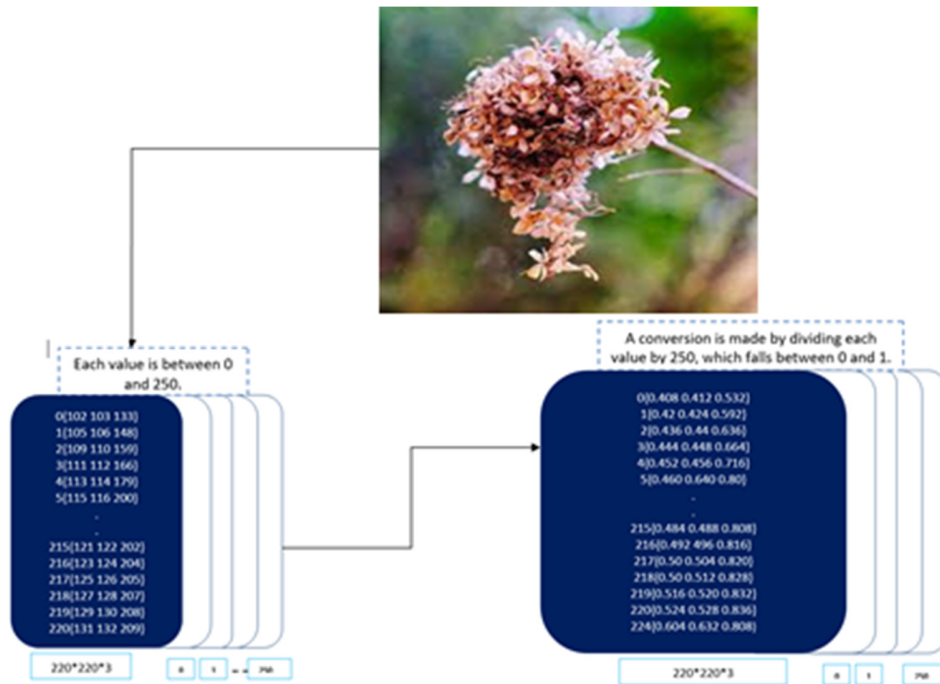


Fig. 4. Arrays of integers from an image.

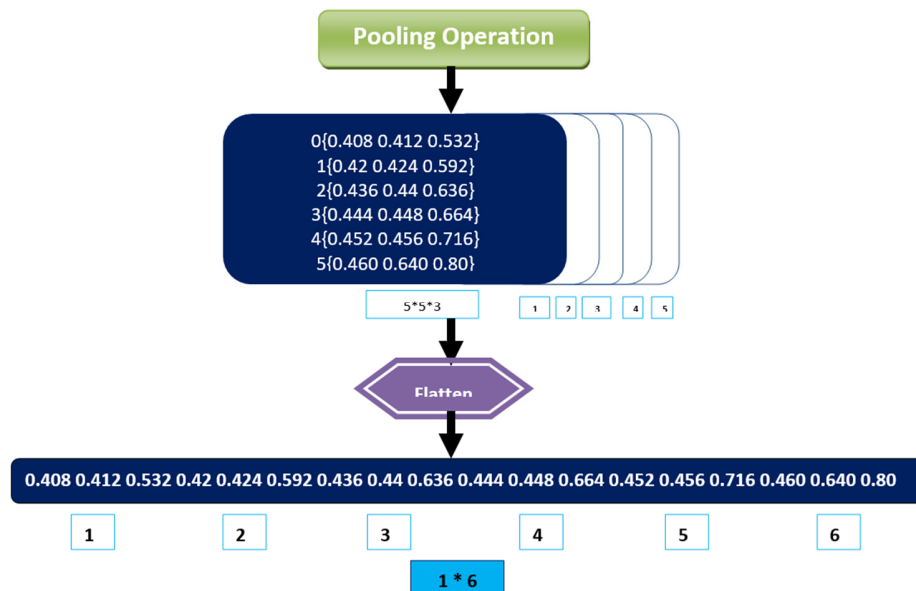


Fig. 5. Pooled and flattened array.

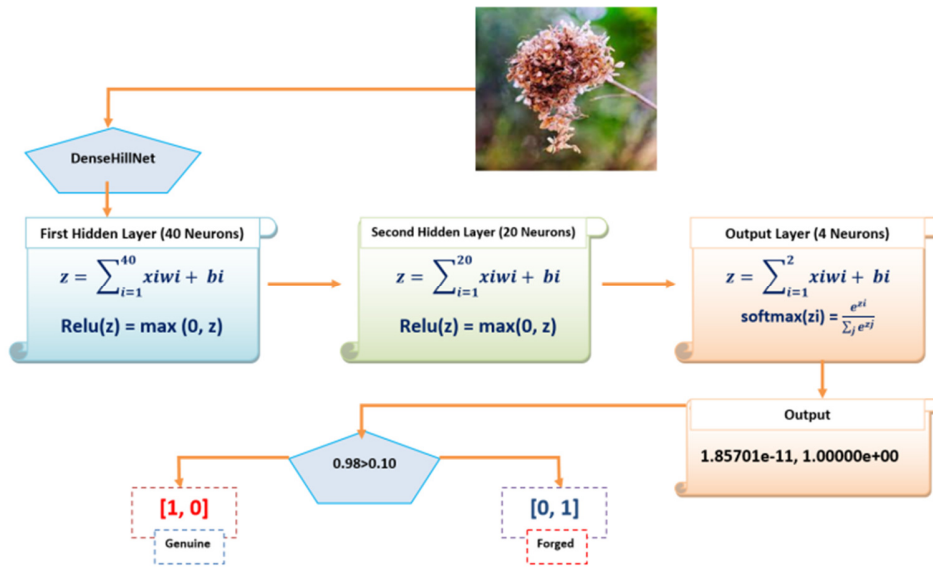


Fig. 6. Using DenseNet to classify an input image.

TABLE III. SUMMARY OF THE PROPOSED MODEL'S PARAMETERS

Layer (type)	Output shape	Param #	Connected to
5conv_16block_con (Concatena)	(False, 6, 6, 1024)	0	5conv_15block_con[0][0]
5conv_16block_2_conv[0][0]			
B_N (BatchNormalization)	(False, 6, 6, 1024)	4090	5conv_16block_con[0][0]
ReLU (Activation)	(False, 6, 6, 1024)	0	B_N[0][0]
flatten (Flatten)	(False, 50170)	0	ReLU [0][0]
den (Dense)	(False, 40)	2508840	flatten[0][0]
den_1 (Dense)	(False, 20)	2520	den[0][0]
den_2 (Dense)	(False, 2)	102	den_1[0][0]
Total params: 9,549,006			
Trainable params: 2,511,502			
Non-trainable params: 7,037,504			

DenseNet use DenseBlocks, a modular framework that maintains feature map size inside each block while changing filter numbers. The transition layers, which exist between DenseBlocks, cut channels in half. This improves the information flow across network layers, which helps with tasks like detecting faked images. DenseNet is created by adding an additional layer to the DenseNet blocks, as portrayed in Figure 7.

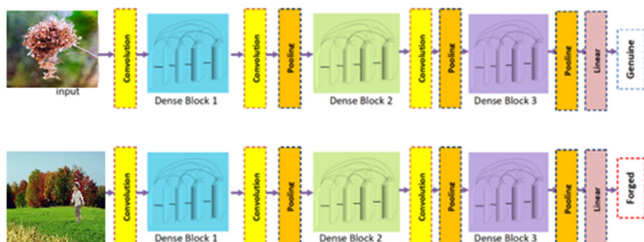


Fig. 7. DenseNet model implementation plan.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Addressing ROI

Numerous settings were deployed, which led to the evaluation of many DenseNet-based DL classifiers on the task

of separating images into two distinct categories: genuine and forged. This study experimented with a variety of parameters and laboratory settings, described as follows:

To ensure an unbiased evaluation of the proposed DenseNet model's performance, the dataset was randomly divided into distinct training and testing subsets, each containing half of the original data. The model was trained with the training subset. Training was done in batches of 19 images over the course of 10 epochs. Table IV provides a thorough overview of the successes the suggested model experienced on the training dataset. The results shown in Table V can be closely examined to provide important insights into the strengths and weaknesses of the proposed DenseNet model. These discoveries might serve as a roadmap for upcoming developments and improvements in the field of digital forensics.

A comparison of the expected and projected results for true forged and true genuine image was carried out. These results were represented by corresponding vectors, [0,1] for the latter and [1,0] for the former. The projected values exhibited in Table VI have two elements [15]. The vector materializes as [1,0] if the first constituent exceeds the other, otherwise, it has the form [0,1]. Table VI displays the determined class that was derived from the values presented in Table V.

TABLE IV. PROPOSED MODEL'S ACCURACY, TEST LOSS, AND TRAINING DURATION

Epochs (E)	Time (s)	Training loss	Training accuracy	Testing loss	Testing accuracy
E-1	87	0.95	0.65	0.50	0.81
E-2	75	0.38	0.84	0.55	0.78
E-3	75	0.36	0.86	0.71	0.72
E-4	80	0.25	0.91	0.41	0.85
E-5	77	0.24	0.89	0.451	0.82
E-6	79	0.20	0.94	0.452	0.83
E-7	100	0.18	0.94	0.450	0.84
E-8	104	0.12	0.970	0.50	0.85
E-9	108	0.10	0.971	0.481	0.86
E-10	103	0.11	0.96	0.483	0.87

B. Results on Testing Data

Table V showcases the results of using the DenseNet model on a selection of 5 initial forged images and 5 initial genuine images of the testing dataset.

TABLE V. VALUES PREDICTED FROM TESTING DATA

Image	Predicted value
Forged-345	(7.1218686e-03, 9.8983681e-01)
Forged-346	(0.02381727, 0.9763827)
Forged-348	(0.02158027, 0.9784198)
Forged-355	(0.01666011, 0.9833399)
Forged-356	(0.05612681, 0.9438732)
Genuine-930	(9.9943285e-01, 4.725437e-04)
Genuine-938	(0.94824463, 0.05175542)
Genuine-941	(0.997224, 0.00277604)
Genuine-946	(9.9983799e-01, 1.6193767e-04)
Genuine-947	(0.41195405, 0.5880459)

Image Genuine-547, despite being genuine in reality, obtained an incorrect forged prediction in the context of this experiment. As a result, it is referred to as a False forged. This situation highlights the value of accurate categorization and identification of natural features while also highlighting the limitations of predictive algorithms [16-18].

The proposed model's prediction results were: 240 images were False genuine (i.e. identified as genuine whereas they were forged), 226 images were True genuine (genuine images predicted as genuine), 230 images were False forged (genuine images predicted as forged), and 127 images were True forged (forged images predicted as forged). Table VII shows the total results, along with the calculated accuracy, recall, and precision values. Accuracy, measures the total prediction score. Precision assesses the accuracy of identifying forged images while recall gauges the models ability to detect all forgeries. The F1 score combines precision and recall to give an assessment of the model performance. High precision reduces alarms while high recall ensures most forgeries are caught. Understanding the F1 score is crucial for grasping the models effectiveness in datasets with distributions of classes. These metrics collectively provide an assessment of how well the model performs, highlighting its strengths and areas that need improvement.

TABLE VI. IDENTIFYING THE PREDICTED CLASSES

Image	Actual (vector)	Actual (class)	Predicted (vector)	Predicted (class)	Result
Forged-221.jpg	[0,1]	Forged	[0,1]	Forged	True forged
Forged-229.jpg	[0,1]	Forged	[0,1]	Forged	True forged
Forged-291.jpg	[0,1]	Forged	[0,1]	Forged	True forged
Forged-340.jpg	[0,1]	Forged	[0,1]	Forged	True forged
Forged-799.jpg	[0,1]	Forged	[0,1]	Forged	True forged
Genuine - 191.jpg	[1,0]	Genuine	[1,0]	Genuine	True genuine
Genuine - 533.jpg	[1,0]	Genuine	[1,0]	Genuine	True genuine
Genuine - 534.jpg	[1,0]	Genuine	[1,0]	Genuine	True genuine
Genuine - 537.jpg	[1,0]	Genuine	[1,0]	Genuine	True genuine
Genuine - 547.jpg	[1,0]	Genuine	[0,1]	Forged	False forged

TABLE VII. CONFUSION MATRIX FOR THE PROPOSED MODEL'S BINARY CATEGORIZATION

Class	Precision	Recall	F1-score
Genuine	0.87	0.86	0.88
Forged	0.91	0.92	0.92
Average	0.89	0.90	0.91
Accuracy	92.32		

The proposed DenseNet model's excellent accuracy in recognizing real and forged images for digital image forensics analysis with the CASIA 2.0 benchmark collection can be attributed to several factors:

- **Dense Connectivity:** By directly connecting each layer to all subsequent levels, DenseNet's architecture encourages efficient information flow and feature reuse. When compared to standard CNNs, which depend purely on feedforward connections, this allows the model to acquire richer and more discriminative characteristics.
- **Bottleneck Layers:** These layers reduce computational complexity and overfitting by limiting the amount of input channels before each 3x3 convolution. This is especially useful for smaller datasets, such as CASIA 2.0, where overfitting is a typical concern.
- **CASIA 2.0 Dataset Properties:** CASIA 2.0 dataset contains a wide range of image forgery types, like copy-paste, splicing, and compression. Because of the variety and difficulty of these modifications, it serves as a solid standard for testing image forgery detection models.
- **Preprocessing approaches:** Using appropriate image normalization and de-noising techniques has considerably improved model performance.

In conclusion, the proposed DenseNet's excellent accuracy in distinguishing authentic and forged images for digital image forensics analysis can be attributed to its efficient design, data augmentation approach, and the nature of the CASIA 2.0

dataset. More advancements can be made by optimizing training techniques, experimenting with different feature engineering methods, and integrating many models using ensemble learning.

C. Complexity of the Proposed Algorithm

The following factors affect the algorithmic complexity:

- Image size input: Since there are more pixels and calculations are required, difficulty rises with greater image sizes.
- Network depth: Compared to shallower networks, DenseNet's 21 Dense Blocks add more layers and connections to the network, increasing its complexity.
- Growth rate: Within the Dense Blocks, DenseNet uses a growth rate parameter to regulate how many feature mappings are added per layer. Complexity rises with increasing growth rate.
- Computational operations: The network's layers carry out a variety of computations, including activation functions, pooling, and convolutions. The particular operations and their parameters determine the level of complexity.

A breakdown of the complexity in big O notation follows:

- Time Complexity: $O(N^2 \times K^2 \times D)$ is the formula for convolution operations, where N is the size of the input image, K is the size of the kernel, and D is the total number of input and output channels.
- Pooling operation: $O(N^2)$
- Activation function: $O(N^2)$
- Dense connections have a size of $O(L \times D \times N^2)$, where L is the network's layer count.
- Space Complexity: features Map: $O(L \times D \times N^2)$ and weights and biases: $O(L \times D \times K^2)$.

As a consequence, the entire complexity of the proposed DenseNet algorithm for digital image forensics analysis can be expressed as:

$$\text{Time complexity: } O(N^2 \times D \times K^2 \times L) + O(N^4)$$

$$\text{Space complexity: } O(L \times D \times N^2) + O(L \times D \times K^2)$$

D. Addressing RO2

To answer RO2, the performance of the proposed DenseNet model was compared with those of different ML and DL classifiers. The results are listed in Table VIII. Established metrics were used to evaluate how well it can distinguish genuine from manipulated images.

E. Addressing RO3

The findings of the baseline methods and the suggested method are compared in Table IX, with the proposed model outperforming the baseline methods. The suggested model performed best in forge picture recognition (92.32% accuracy).

TABLE VIII. COMPARATIVE RESULTS

ML/DL model	Precision	Recall	F1 score	Accuracy
Xception	0.77	0.74	0.73	76.51
Mobilenetv3-small	0.78	0.75	0.77	72.21
Resnet-50	0.59	64	0.63	69.24
LSTM	0.61	0.63	0.62	63.31
RNN	0.71	0.71	0.71	70.64
Improved DenseNet (proposed)	0.89	0.9	0.91	92.32

TABLE IX. PROPOSED MODEL VS. RELATED RESEARCH

Work	Method	Results (accuracy)
[2]	Forged image recognition with DL	0.8542
[10]	Forged image detection with DL	0.882
Proposed	A novel model based on the DenseNet121 architecture, a CNN version with densely connected layers	0.9232

The proposed CNN is contrasted with the CNN model from [2], with four fully connected layers, which recognized falsified images. The quad convolution CNN architecture had lesser effectiveness (accuracy: 85.42%) than the proposed CNN model (92.32%). The CNN different parameter settings and insufficient layer count are held responsible for the baseline model's inadequacy, thus, the effectiveness of the CNN degrades whenever the number of concealed layers is increased by even more than 2. The performance of the proposed CNN model driven by DL was also contrasted with the research conducted in [10]. The recommended method surpassed the benchmark work across a variety of parameter combinations of layers in a deep neural network, including filters, step size, block size, number of iterations, and various measures like precision, recall, and F1-score. A significant difference is revealed when contrasting the current study with the findings in [10]. The latter used a CNN model and was limited to categorizing photos of glaciers as "yes" or "no," with an accuracy rate of 72%. The proposed study, on the other hand, adopted a more thorough methodology by including both forged and real photos in the training data. This method produced a higher accuracy of 84% by adopting a less complicated model architecture.

This comparison emphasizes how essential it is to take a greater variety of natural factors into account when developing ML models that correctly categorize images. The proposed work also shows how a simpler model design can produce outcomes that are equivalent to those of more complex, computationally expensive models. The DenseNet model suggested provides significant advantages in various areas of digital image forensics. In media the former can counteract the proliferation of fake photos and videos, playing a crucial role in combating misinformation and protecting users from online scams. The journalism industry can benefit from the model's capacity to authenticate images used in news reports ensuring correctness and preventing the spread of fake news. In cybersecurity and law enforcement, the proposed model may assist in investigating crimes involving manipulated images, such as identity theft and cyber fraud, by identifying photos and deepfake identities. Moreover, the medical field can be

avored from the model's ability to guarantee the accuracy of diagnoses and treatments. These instances highlight the model's adaptability and its potential to improve image integrity and security across various sectors. With the advancements in DL technology, it is anticipated that DenseNets applications will broaden further as it strengthens efforts, against image manipulation.

V. CONCLUSION AND FUTURE WORK

In conclusion, identifying fake photographs is crucial for a number of industries, including transportation and outdoor recreation. Convolutional Neural Networks (CNNs) have emerged as a popular method for classifying and recognizing images. While earlier studies focused on a variety of image categories, the topic of forged image detection is still largely untapped. The presented DenseNet-based picture forgery recognition model performs well at differentiating genuine and altered photos. Three stages are involved in applying the DenseNet model: data gathering, data preprocessing, and model deployment. The current study presents a novel model based on the DenseNet121 architecture, a version of CNNs with densely connected layers. This model effectively distinguishes fabricated and genuine photos. The model was evaluated with a dataset of 12,614 photos, both forged and genuine, and remarkable results with an accuracy rate of 92.32% were obtained. This accomplishment places the recommended model among the best in forged picture detection tasks, demonstrating its high performance.

The proposed model is a modified version of the DenseNet121 architecture, with extra layers added to meet the special requirements of forged picture detection. This architectural arrangement consists of several layers, beginning with an initial convolutional layer with 64 7×7 filters and a stride of 2. Following that, dense blocks and transition layers are incorporated into the architecture. The final dense block consists of 16 repetitions, ending in the output layer with a global average pooling layer that consolidates all network feature maps for classification. Despite its achievements, the suggested model has some limitations. These include using one dataset and a specific DenseNet design, without testing trained models. Suggestions for research involve broadening the types of studied images, integrating various datasets, and trying out pre-trained neural network models like AlexNet, VGG, and ResNet. Moreover, there are plans to investigate neural network setups, such as CNN+BiLSTM and CNN+BiGRU, to conduct a thorough image forgery analysis. This effort aims to improve the model's accuracy in various scenarios.

REFERENCES

- [1] A. Khattak, M. Z. Asghar, M. Ali, and U. Batool, "An efficient deep learning technique for facial emotion recognition," *Multimedia Tools and Applications*, vol. 81, no. 2, pp. 1649–1683, Jan. 2022, <https://doi.org/10.1007/s11042-021-11298-w>.
- [2] S.-H. Cho, S. Agarwal, S.-J. Koh, and K.-H. Jung, "Image Forensics Using Non-Reducing Convolutional Neural Network for Consecutive Dual Operators," *Applied Sciences*, vol. 12, no. 14, Jan. 2022, Art. no. 7152, <https://doi.org/10.3390/app12147152>.
- [3] A. Kuznetsov, "Digital image forgery detection using deep learning approach," *Journal of Physics: Conference Series*, vol. 1368, no. 3, Aug. 2019, Art. no. 032028, <https://doi.org/10.1088/1742-6596/1368/3/032028>.
- [4] S. S. Ali, I. I. Ganapathi, N.-S. Vu, S. D. Ali, N. Saxena, and N. Werghi, "Image Forgery Detection Using Deep Learning by Recompressing Images," *Electronics*, vol. 11, no. 3, Jan. 2022, Art. no. 403, <https://doi.org/10.3390/electronics11030403>.
- [5] Y. Wu, W. Abd-Almageed, and P. Natarajan, "BusterNet: Detecting Copy-Move Image Forgery with Source/Target Localization," in *Computer Vision – ECCV 2018*, Cham, 2018, pp. 170–186, https://doi.org/10.1007/978-3-030-01231-1_11.
- [6] S. I. S. M. Shazuli and A. Saravanan, "Improved Whale Optimization Algorithm with Deep Learning-Driven Retinal Fundus Image Grading and Retrieval," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11555–11560, Oct. 2023, <https://doi.org/10.48084/etasr.6111>.
- [7] J. Dong, W. Wang, and T. Tan, "CASIA Image Tampering Detection Evaluation Database," in *2013 IEEE China Summit and International Conference on Signal and Information Processing*, Beijing, China, Jul. 2013, pp. 422–426, <https://doi.org/10.1109/ChinaSIP.2013.6625374>.
- [8] K. Team, "Keras documentation: Image data loading," *Keras*. https://keras.io/api/data_loading/image/.
- [9] "Architecture of DenseNet-121," *OpenGenus IQ: Computing Expertise & Legacy*, Aug. 26, 2021. <https://iq.opengenus.org/architecture-of-densenet121/>.
- [10] R. Rajkumar, "Deep Learning Feature Extraction Using Attention-Based DenseNet 121 for Copy Move Forgery Detection," *International Journal of Image and Graphics*, vol. 23, no. 05, Sep. 2023, Art. no. 2350042, <https://doi.org/10.1142/S0219467823500420>.
- [11] S. Alotaibi, "A Fairness-based Cell Selection Mechanism for Ultra-Dense Networks (UDNs)," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11524–11532, Oct. 2023, <https://doi.org/10.48084/etasr.6106>.
- [12] K. H. Hingrajiya and C. Patel, "An Approach for Copy-Move and Image Splicing Forgery Detection using Automated Deep Learning," in *2023 International Conference on Emerging Smart Computing and Informatics (ESCI)*, Pune, India, Mar. 2023, pp. 1–5, <https://doi.org/10.1109/ESCI56872.2023.10100202>.
- [13] R. Zhang and J. Ni, "A Dense U-Net with Cross-Layer Intersection for Detection and Localization of Image Forgery," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, Feb. 2020, pp. 2982–2986, <https://doi.org/10.1109/ICASSP40776.2020.9054068>.
- [14] V. Verma, D. Singh, and N. Khanna, "Block-level double JPEG compression detection for image forgery localization," *Multimedia Tools and Applications*, vol. 83, no. 4, pp. 9949–9971, Jan. 2024, <https://doi.org/10.1007/s11042-023-15942-5>.
- [15] C.-C. Hsu, Y.-X. Zhuang, and C.-Y. Lee, "Deep Fake Image Detection Based on Pairwise Learning," *Applied Sciences*, vol. 10, no. 1, Jan. Art. no. 370, 2020, <https://doi.org/10.3390/app10010370>.
- [16] D. Alghazzawi, O. Bamasag, A. Albeshri, I. Sana, H. Ullah, and M. Z. Asghar, "Efficient Prediction of Court Judgments Using an LSTM+CNN Neural Network Model with an Optimal Feature Set," *Mathematics*, vol. 10, no. 5, Jan. 2022, Art. no. 683, <https://doi.org/10.3390/math10050683>.
- [17] M. M. H. Milu, M. A. Rahman, M. A. Rashid, A. Kuwana, and H. Kobayashi, "Improvement of Classification Accuracy of Four-Class Voluntary-Imagery fNIRS Signals using Convolutional Neural Networks," *Engineering, Technology & Applied Science Research*, vol. 13, no. 2, pp. 10425–10431, Apr. 2023, <https://doi.org/10.48084/etasr.5703>.
- [18] I. Sahib and T. A. A. AlAsady, "Deep learning for image forgery classification based on modified Xception net and dense net," *AIP Conference Proceedings*, vol. 2547, no. 1, Dec. 2022, Art. no. 060003, <https://doi.org/10.1063/5.0112143>.

Optimizing Solar PV Placement for Enhanced Integration in Radial Distribution Networks using Deep Learning Techniques

Mohamed Ali Zdiri

CEM Laboratory, Engineering School of Sfax, Tunisia
mohamed-ali.zdiri@enis.tn (corresponding author)

Bilel Dhouib

CEM Laboratory, Engineering School of Sfax, Tunisia
bilel.dhouib@enis.tn

Zuhair Alaas

Department of Electrical Engineering, Faculty of Engineering, Jazan University, Saudi Arabia
zalaas@jazanu.edu.sa

Hsan Hadj Abdallah

CEM Laboratory, Engineering School of Sfax, Tunisia
hsan.hajabdallah@enis.tn

Received: 26 December 2023 | Revised: 15 January 2024, 1 February 2024, and 24 February 2024 | Accepted: 10 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6818>

ABSTRACT

This study introduces a highly effective technique to address the load flow challenge in Radial Distribution Networks (RDNs). The proposed approach leverages two matrices derived from the topological features of distribution networks to provide an optimal solution to handle load flow challenges. To assess the efficacy of this technique, simulations were executed on an IEEE 33-bus radial distribution system using MATLAB. Deep Learning (DL) has become a powerful artificial intelligence technique that excels at interpreting power grid datasets. Thus, a data-driven methodology is presented that incorporates an advanced Long-Short-Term-Memory (LSTM) network. Employing the Recurrent Neural Network with the LSTM (RNN-LSTM) technique based on these simulations, the study precisely identifies the optimal placement of an integrated PV generator within the radial network. The application of DL techniques, specifically LSTM networks, exemplifies the potential of data-driven approaches in enhancing decision-making processes. The results of this study highlight the potential of RNN-LSTM for the optimal integration of PV generators and for ameliorating the reliability of RDNs.

Keywords-RDN; PV; load flow; DL; RNN-LSTM

I. INTRODUCTION

RDNs featuring main feeders and lateral distributors are gaining popularity due to their simplicity and cost-effectiveness. Power flow analysis is crucial in designing efficient distribution networks, considering factors, such as maximum feeder currents, voltage dips, energy loss, and reliability [1]. Optimization of these networks relies on algorithms that require multiple power flow runs, with uncertainties arising from estimated inputs, namely load forecasts and network parameters. In complex distribution systems, practical challenges in data collection contribute to non-statistical uncertainties. The following characteristics distinguish electric distribution networks: radial or weakly

mesh topologies, unbalanced operation with scattered loads, a large number of buses and branches, varying resistance and reactance values, as well as operation in numerous phases [2].

Conventional power flow techniques, like Newton-Raphson and rapidly decoupled methods, are effective in managing power systems. However, difficulties arise when applying these methods to faulty or improperly initialized systems [3-4]. Additionally, the Gauss-Seidel method, although robust, demonstrates inefficiency when dealing with large power systems [5]. Distribution networks, characterized by their ill-conditioned nature, pose challenges due to the presence of diverse resistance and reactance values and their radial structure. As renewable energy sources become more

integrated, it becomes imperative to address these challenges in distribution systems. Therefore, it is crucial to adapt the load flow method to overcome these challenges [6].

The incorporation of renewable PV energy sources in power distribution networks is rapidly expanding [7]. The swift advances in Photovoltaic Generator (PG) technology and their integration into distribution networks present a multitude of benefits. These advantages encompass the reduction of the highest loads, greater system security, improved dependability, increased voltage stability, strengthened grid resilience, incurred peak operating expenses, and diminished network losses [8-9]. This study proposes an approach that uses Machine Learning (ML) techniques, specifically an Artificial Neural Network (ANN) to determine the optimal placement of PGs [10]. Deep Learning (DL) is viewed as a progression from ML, integrating algorithms that can learn from data to perform tasks without the need for explicit programming [11]. Its ability to extract advanced features from extensive input data, known as feature engineering, sets it apart from ML. Consequently, DL is increasingly favored for its groundbreaking applications in natural language processing, computer vision, and predictive modeling [12]. Various types of DL modeling approaches have been presented, including vector space models, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrid neural networks [13].

This study employed a DL technique, specifically RNN-LSTM, to enhance the placement of PV generators. This approach effectively addresses load flow challenges by utilizing two matrices derived from the topology of the distribution network. The efficiency of the proposed technique is demonstrated through simulations conducted on an IEEE 33-bus RDN using MATLAB. The prediction procedure exhibits significant improvements in the efficacy and performance of the proposed technique. The findings demonstrate the potential of the proposed RNN-LSTM technique for optimizing PG placement in distribution networks, leading to enhanced network performance and efficiency.

II. PHOTOVOLTAIC INTEGRATION IN RDNS

The control of load flow in an RDN is achieved by incorporating Bus Injections into the Branch Current (BIBC) matrix and applying comparable current injections [14]. The model based on current injection, especially beneficial for distribution networks, is employed. The formula for the load apparent power at the node ci is given by:

$$S_{ci} = P_{ci} + s \cdot Q_{ci} \quad (1)$$

where P_{ci} represents the active load power, Q_{ci} represents the reactive load power for each bus, and i takes values from 1 to n . The ii -th current injection can be expressed similarly as:

$$I_{ci}^{ii} = I_{ci}^r (V_{ci}^{ii}) + s \cdot I_{ci}^i (V_{ci}^{ii}) = ((P_{ci} + jQ_{ci}) / V_{ci}^{ii}) * \quad (2)$$

At the ii th iteration, the bus voltage and current injection are symbolized as V_c^{ii} and I_c^{ii} , respectively, and $s^2 = -1$. The real and imaginary parts of the current injection at this iteration are denoted by I_{ci}^r and I_{ci}^i , respectively.

A. Development of Relationship Matrices

Figure 1 illustrates a basic RDN.

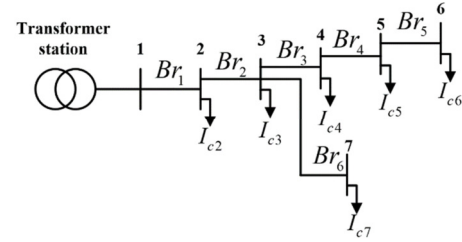


Fig. 1. Simplified RDN.

Equation (2) is used to compute the injected currents, and Kirchhoff's current law is applied to determine the branch currents in the RDN. As a result, the branch currents can be expressed as functions of their respective current injections:

$$\begin{cases} Br_1 = I_{c2} + I_{c3} + I_{c4} + I_{c5} + I_{c6} + I_{c7} \\ Br_2 = I_{c3} + I_{c4} + I_{c5} + I_{c6} + I_{c7} \\ Br_3 = I_{c4} + I_{c5} + I_{c6} \\ Br_4 = I_{c5} + I_{c6} \\ Br_5 = I_{c6} \\ Br_6 = I_{c7} \end{cases} \quad (3)$$

Therefore, the branch currents can be obtained as follows:

$$\begin{bmatrix} Br_1 \\ Br_2 \\ Br_3 \\ Br_4 \\ Br_5 \\ Br_6 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} I_{c2} \\ I_{c3} \\ I_{c4} \\ I_{c5} \\ I_{c6} \\ I_{c7} \end{bmatrix} \quad (4)$$

Equation (4) may be rewritten as:

$$[B] = [BIBC] \cdot [I_c] \quad (5)$$

The BIBC matrix is characterized by an upper triangular structure, with its elements exclusively taking values of 0 or 1.

B. Development of BIBC Matrices

The process for forming the BIBC matrix, as delineated (4), can be outlined as shown in [15].

C. Development of the Bus Voltage

To determine the receiving-end bus voltages, a forward sweep is performed across the ladder network utilizing the following generalized equations:

$$V(cm2) = V(cm1) - Br(jj) Z(jj) \quad (6)$$

In this context, the symbols $cm1$ and $cm2$ represent the transmitting and receiving ends, respectively, and $Br(jj)$ denotes the branch current number. The following equation can be used to calculate the correlation between branch currents and bus voltages:

$$\begin{cases} V_{c2} = V_{c1} - Br_1 Z_{12} \\ V_{c3} = V_{c2} - Br_2 Z_{23} \\ V_{c4} = V_{c3} - Br_3 Z_{34} \\ V_{c5} = V_{c4} - Br_4 Z_{45} \\ V_{c6} = V_{c5} - Br_5 Z_{56} \\ V_{c7} = V_{c3} - Br_6 Z_{37} \end{cases} \quad (7)$$

These relationships can be illustrated in a matrix form:

$$[\Delta V] = [BCBV] \cdot [Br] \quad (8)$$

with:

$$[BCBV] = \begin{bmatrix} Z_{12} & 0 & 0 & 0 & 0 & 0 \\ Z_{12} & Z_{23} & 0 & 0 & 0 & 0 \\ Z_{12} & Z_{23} & Z_{34} & 0 & 0 & 0 \\ Z_{12} & Z_{23} & Z_{34} & Z_{45} & 0 & 0 \\ Z_{12} & Z_{23} & Z_{34} & Z_{45} & Z_{56} & 0 \\ Z_{12} & Z_{23} & 0 & 0 & 0 & Z_{37} \end{bmatrix}$$

The Branch Current to Bus Voltage (BCBV) matrix facilitates the computation of equivalent bus voltage fluctuations corresponding to variations in branch currents. The expression below can be employed to articulate the correlation between bus voltages and current injections:

$$[V_c] = [V_{c1}] - [BCBV] \cdot [BIBC] \cdot [I_c] \quad (9)$$

The updated voltage values are employed for a new top-down iteration. The next equation can be deployed to calculate the connection between branch currents and bus voltages. Additionally, the actual and reactive power loss of the branch ij , with $j=i+1$, can be determined using:

$$\begin{cases} p_{loss} = |Br(ij)|^2 \cdot |R(ij)| \\ q_{loss} = |Br(ij)|^2 \cdot |X(ij)| \end{cases} \quad (10)$$

with $R(ij)$ and $X(ij)$ being the resistance and reactance of the line section ij .

D. Development of PV Bus

The most favorable PG location on an RDN, in terms of minimizing losses, is near the line's termination point. The maximum power capacity of a solar plant is expressed by [16]:

$$P_{PG} = P_1 \cdot E_c \cdot [1 + P_2 \cdot (E_c - E_{cref}) + P_3 \cdot (T_{jc} - T_{jcref})] \quad (11)$$

where E_c represents the panel insolation (W/m^2), E_{cref} and T_{jcref} are equal to 1000 W/m^2 and 25°C , and P_1 , P_2 , and P_3 are fixed values. This simplified model enables the calculation of the maximum PG power at a given panel irradiation and temperature, with only three constant variables and a straightforward equation. The active power at bus i varies according to the following factors:

$$P_{ci} = P_{ci0} - P_{PG} \quad (12)$$

The initial power consumption at bus i , before the injection of power from the PG, is denoted as P_{ci0} .

III. RNN DL PV PLACEMENT

A. DL Techniques

DL modeling techniques facilitate the acquisition of feature representations in data through the utilization of multiple

processing layers and various levels of abstraction [17]. Advanced DL models, founded on ANNs [18], demonstrate proficiency across diverse domains. Despite their effectiveness, ANNs exhibit drawbacks, such as the absence of guaranteed convergence to an optimal solution and susceptibility to overfitting in the training data. The term "deep" in DL reflects the numerous processing layers traversed by data within the network. A DL model comprises stacked layers, as shown in Figures 2 and 3. The initial layer (input) consists of units with values distributed to neurons in the first hidden layer, leading to the final layer where predicted results emerge. The number of units in the final layer corresponds to the desired output classes. Hidden layers positioned between the input and output layers apply weights to inputs, passing them through an activation function that introduces non-linearity, facilitating the learning of complex data relationships. The backpropagation algorithm computes the error between predicted results and the desired output class and then adjusts the weights in the hidden layer to minimize loss. This iterative process continues until the output achieves sufficient accuracy for practical use [19]. Given the aforementioned neural network concepts, numerous DL modeling techniques have been explored [20].

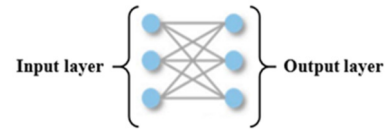


Fig. 2. Conventional neural network.

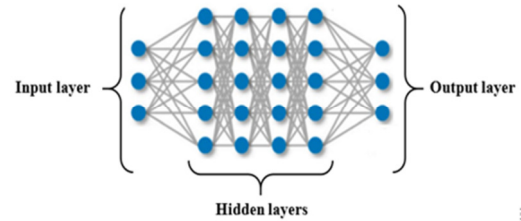


Fig. 3. DL neural network.

B. RNN with LSTM

RNNs have exhibited promising results in various natural language processing tasks, excelling in areas, such as sentiment classification [21], image captioning [22], and language translation [23]. Many scenarios involve data sequences that inherently convey the essence of the information, as noticed in tasks like language modeling where the meaning is derived from the sequential arrangement of words. Traditional neural networks assume no interdependence between input and output, but in cases where sequential order matters, a network that incorporates prior information becomes essential for meaningful comprehension of the data. RNNs address this need by performing the same computation for each element in a sequence, establishing a connection to previous information. This recurrent nature enables RNNs to maintain a memory that retains information from prior computations [24].

An LSTM network consists of distinct memory blocks known as cells. These cells are formed by gates that regulate

the information flow, including forget, input, and output gates. The forget gate eliminates information from the cell configuration, while the input gate incorporates newly entered data into the cell. The input gate governs the pace at which new data is introduced to the cell, and the output gate restricts the data within the cell, determining the output activation of the LSTM unit. The equations below can be used to define the gating mechanism in an LSTM network:

$$f_t = \sigma(W_{ff} \cdot [h_{t-1}, x_t] + b_{ff}) \quad (13)$$

$$i_t = \sigma(W_{ii} \cdot [h_{t-1}, x_t] + b_{ii}) \quad (14)$$

$$\tilde{C}_t = \tanh(W_{CC} \cdot [h_{t-1}, x_t] + b_{CC}) \quad (15)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (16)$$

$$o_t = \sigma(W_{oo} \cdot [h_{t-1}, x_t] + b_{oo}) \quad (17)$$

$$h_t = o_t * \tanh(C_t) \quad (18)$$

where:

- f_t is the forget gate output at the time step t . This element decides the extent to which the information from the previous cell state should be disregarded.
- i_t is the input gate output at t , dictating the proportion of the candidate cell state to be added to the existing cell state.
- \tilde{C}_t is the candidate cell state at t , referring to a novel piece of information that can be incorporated into the cell state.
- C_t is the current cell state at t . This signifies the memory aspect of the LSTM cell.
- o_t is the output gate output at t . This factor decides the proportion of the current cell state to be produced as the output.
- h_t is the hidden state output at t . This is the result or output produced by the LSTM cell.
- W_{ff}, W_{ii}, W_{CC} , and W_{oo} are matrices representing the weights for the forget gate, input gate, candidate cell state, and output gate, respectively.
- b_{ff}, b_{ii}, b_{CC} , and b_{oo} are bias vectors.
- σ denotes the sigmoid activation function.
- $*$ denotes element-wise multiplication.

These equations describe the gating mechanism in an LSTM network. The forget gate (13) determines which information should be discarded from the cell state, while the input gate (14) decides the amount of new information that should be added. The output gate (15) controls the information flow from the cell to the output activation of the LSTM unit. LSTM networks can be used to effectively model the interdependencies and sequential patterns in the input data, allowing the capture of relevant information for the problem at hand. These insights are incorporated into the objective function to ensure that the input signals are appropriately represented.

IV. SIMULATION RESULTS

The proposed algorithm was implemented on the IEEE 33-bus RDN, both with and without a PG. The system consists of 33 buses and 32 lines, as depicted in Figure 4. By applying the proposed algorithm to this real-world distribution system, it is possible to assess its performance and evaluate its applicability in several scenarios. The IEEE 33-bus system is widely utilized as a benchmark in power system analysis, making it an ideal test case for validating the capabilities of the proposed algorithm. The Advance Power API-M370 PV module is well-suited to meet current requirements, and its characteristics are illustrated in Figure 5 and Table I.

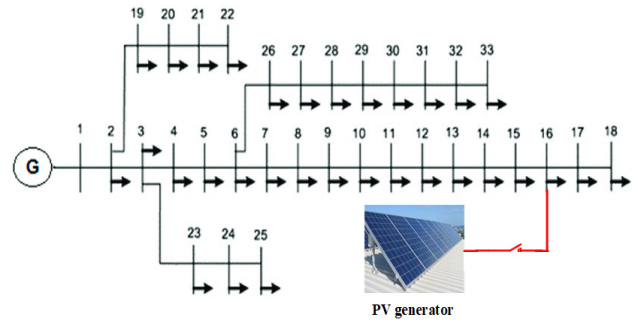


Fig. 4. Configuration of distribution network incorporating PG.

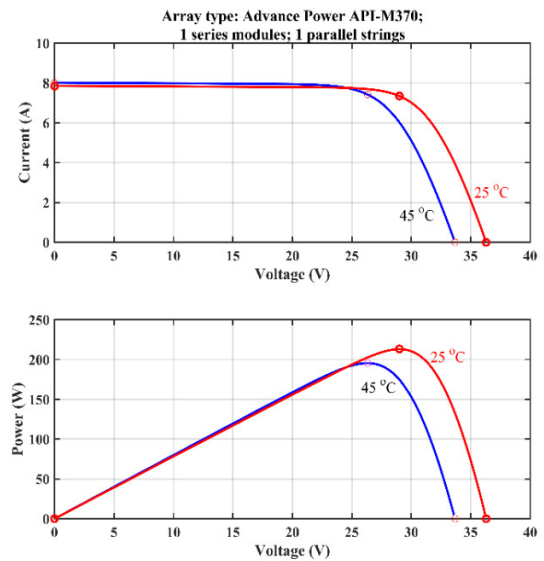


Fig. 5. API-M370 characteristics.

TABLE I. API-M370 PARAMETERS

Maximum power (W)	370
MPP Voltage (V)	38.8
Current at MPP (A)	9.54
Open-circuit voltage (V)	47.8
Short-circuit current (A)	10.17

Figure 6 portrays the voltage load flow solution obtained from the radial network analysis conducted on an IEEE 33-bus distribution system in the absence of a PG. The load flow solution offers important information on the steady-state

operating conditions of the distribution system. Analyzing the load flow solution provides valuable insights into voltage profiles and power flows across the network. These data aid in evaluating system stability, detecting possible voltage violations, and comprehending the overall performance of the distribution system under standard operating conditions. Table II presents Active Power Loss (APL) and Reactive Power Loss (RPL) in the case where a PG is integrated in nodes 2 to 33 of the distribution system, providing a comprehensive overview of the power losses associated with the integration of the PG and quantifying the impact of renewable energy integration on the overall power losses of the system. Specifically, the PG has a power capacity of 500 kW, representing a renewable energy integration rate of 10.77% into the electrical grid. By analyzing the active and reactive power losses in each case, the effectiveness of PG integration in reducing power losses and improving system efficiency can be assessed.

TABLE II. PG- IEEE 33 BUS RDN'S POWER LOSS SOLUTION

Actual PG location nodes	APL (p.u.)	RPL (p.u.)
2	0.2092	0.1421
3	0.2004	0.1375
4	0.1959	0.1351
5	0.1913	0.1327
6	0.1816	0.1246
7	0.1804	0.1214
8	0.1728	0.1159
9	0.1696	0.1135
10	0.1667	0.1115
11	0.1662	0.1113
12	0.1654	0.111
13	0.163	0.1091
14	0.1622	0.1082
15	0.162	0.108
16	0.1619	0.108
17	0.1623	0.1085
18	0.1629	0.1089
19	0.209	0.142
20	0.2085	0.1415
21	0.2085	0.1415
22	0.2089	0.142
23	0.1986	0.1363
24	0.1954	0.1338
25	0.1943	0.133
26	0.1807	0.1241
27	0.1794	0.1234
28	0.1752	0.1198
29	0.1723	0.1174
30	0.1709	0.1166
31	0.1694	0.1153
32	0.1693	0.1151
33	0.1696	0.1156

The results in Table II highlight the importance of renewable energy integration in mitigating power losses and its potential as a means to enhance the sustainability and economic viability of the electrical grid. Table III illustrates the power loss solution for branches 15-16, 16-17, and 17-18 in the distribution system, offering valuable information on the power losses that occur in these specific branches. Tables II and III indicate that the PG installed on node 16 plays a significant role in supplying a substantial amount of energy, highlighting its contribution to reducing power losses and enhancing the

overall performance of the system. Furthermore, nodes 17 and 18, which are part of branches 15-16, 16-17, and 17-18, consume the highest amount of energy. This demonstrates their importance in the distribution system and emphasizes the need for efficient energy management in these nodes. As moving away from node 16 and traverse the branches, the current flowing through the branches gradually decreases. This reduction in current leads to a decrease in power losses along the branches. Consequently, node 16 emerges as the optimal location for integrating the PG system, as it exhibits the lowest active and reactive power losses among the nodes considered. Positioning the PG system at node 16 can effectively minimize power losses and improve the overall performance of the distribution system. This information serves as a guide for the optimal placement of renewable energy sources, such as PG systems, to maximize their impact and achieve greater energy efficiency in the network.

TABLE III. PG- IEEE 33 BUS RDN'S BRANCH POWER LOSS SOLUTION

Branch	Without PG		With PG (in node 16)	
	APL (p.u.)	RPL (p.u.)	APL (p.u.)	RPL (p.u.)
17-18	0.054	0.043	0.051	0.040
16-17	0.257	0.343	0.241	0.322
15-16	0.287	0.210	0.226	0.165

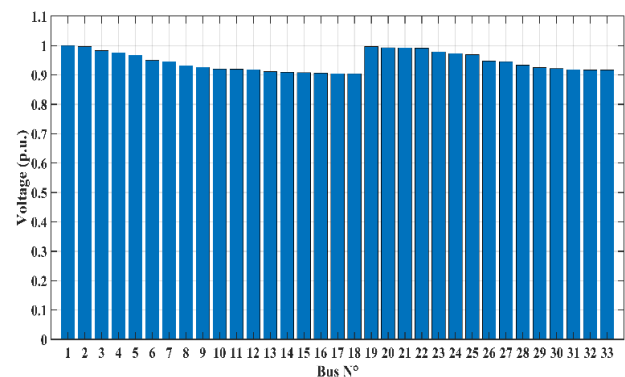


Fig. 6. Voltage load flow solution.

Figure 7 showcases the active and reactive power losses in the IEEE 33 bus network with and without PG. Case 1 represents the scenario in which PG is injected at node 16, while Case 2 represents the absence of PG. These results indicate a significant reduction in both active and reactive power losses compared to previous works. The reductions achieved were 23.27% in APL and 24.48% in RPL, highlighting the strong performance of the proposed approach. This decrease is achieved even without the need for additional shunt capacity, effectively reducing system installation costs. Furthermore, when the integrated PG case at node 16 was compared with other cases, it was found that it offers a more reliable voltage profile. This suggests that integrating the PG at node 16 not only reduces power losses, but also expands the overall voltage stability and reliability of the system.

The RNN-LSTM approach was used to train a collection of localized PV generator data points, leveraging simulation results for various injected PG power datasets. Figure 8

illustrates the results obtained from the RNN-LSTM model, indicating that the actual and predicted trajectories are closely aligned and displaying the high level of performance and accuracy achieved by this technique. The RNN-LSTM method demonstrates remarkable performance, achieving an impressive accuracy level of up to 97%, as observed in Figure 9. This accuracy level provides compelling evidence of the superiority of the RNN-LSTM approach compared to traditional and ANN algorithms commonly employed in similar tasks [10, 25].

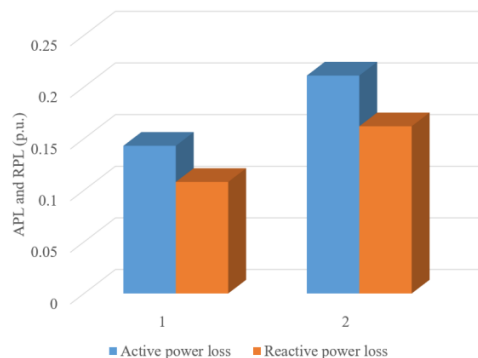


Fig. 7. APL and RPL for the two cases of the IEEE 33 bus network.

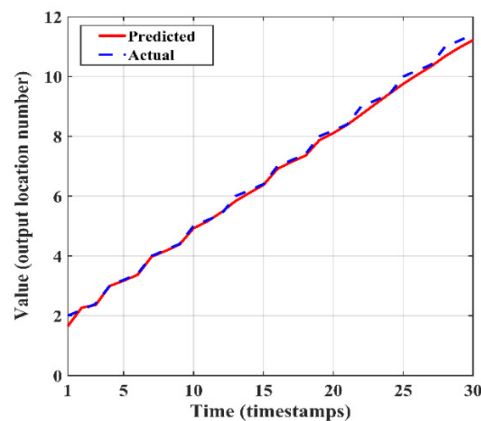


Fig. 8. RNN-LSTM prediction results.

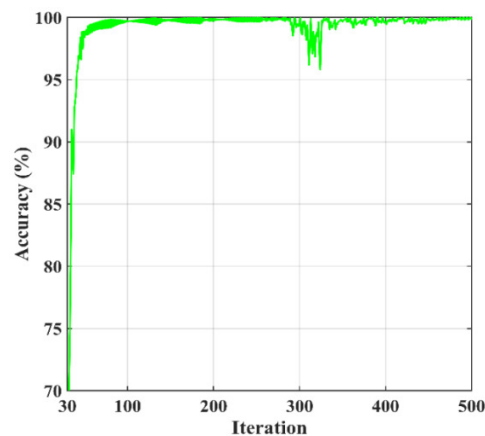


Fig. 9. Prediction accuracy as a function of iterations.

Comparing the proposed RNN-LSTM method with algorithms, such as Artificial Bee Colony (ABC), Water Wave Optimization (WAO), Genetic Algorithm/Particle Swarm Optimization (GA/PSO), Genetic Algorithm (GA), and Firefly Algorithm (FA) can provide valuable insights into its performance and potential advantages [26-28]. By surpassing other methods in terms of accuracy, the RNN-LSTM technique proves its effectiveness in precisely predicting and analyzing data. This exceptional accuracy underscores the robustness and reliability of the RNN-LSTM approach, establishing it as a highly valuable tool in various applications and domains. The evaluation considered key metrics, like the calculation time, the convergence rate, and the accuracy of the optimal solutions. In terms of calculation time, the RNN-LSTM technique demonstrated notable efficiency, significantly reducing the time required for training and prediction compared to the existing algorithms. Additionally, the convergence rate of the RNN-LSTM technique was superior, requiring fewer iterations to achieve a satisfactory solution. Furthermore, the accuracy of the optimal solutions (97%) obtained from the RNN-LSTM technique surpassed that of existing algorithms. This comparative analysis allows for a comprehensive assessment of the deep learning method's performance compared to other established algorithms, as it facilitates a better understanding of its potential benefits, namely improved convergence, accuracy, and robustness, while also identifying any areas where other algorithms may outperform it. The proposed RNN-LSTM approach for optimal PV placement in the distribution network offers several advantages and some disadvantages. Table IV entails some key points to consider.

TABLE IV. ADVANTAGES AND DISADVANTAGES OF THE PROPOSED METHOD.

Advantages	Disadvantages
Accuracy Flexibility Adaptability Automation	Computational Complexity Data Requirements

V. CONCLUSION

This study presented a proficient method to address the load flow problem in RDNs. By harnessing the topological characteristics of distribution networks, two matrices were derived that facilitated an optimal solution to the load flow challenge. The BIBC matrix, computed through the application of Kirchhoff's current law, and the BCBV matrix, illustrating the correlation between bus voltages and branch currents, were instrumental in resolving the load flow. Simultaneous utilization of these matrices presents a direct and effective method to address load flow issues. The proposed method demonstrated excellent performance in practical applications. The simulation results showcased its effectiveness in dealing with widespread RDNs. In addition, considering the integration of PG into the radial system under investigation, simulation results were presented and discussed concerning voltage amplitudes, current branches, and active and reactive power losses. These results helped identify the optimal location for integrating a PG, which was determined to be the farthest node providing reliable voltage and minimizing power loss. Furthermore, the RNN-LSTM technique was applied to

accurately locate a PG that had already been incorporated into the RDN. This prediction procedure caused a significant improvement in the efficacy and performance of this technique, as evidenced by the results obtained.

ABBREVIATIONS

RDN:	Radial Distributed Network
DL:	Deep Learning
PV:	Photovoltaic
APL:	Active Power Loss
RPL:	Reactive Power Loss
RNN:	Recurrent Neural Network
LSTM:	Long Short-Term Memory
PG:	Photovoltaic Generator
ANN:	Artificial Neural Network
BIBC:	Branch Injection into Branch Current
BCBV:	Branch Current to Bus Voltage

REFERENCES

- [1] J. A. M. Rupa and S. Ganesh, "Power Flow Analysis for Radial Distribution System Using Backward/Forward Sweep Method," *International Journal of Electrical and Computer Engineering*, vol. 8, no. 10, pp. 1628–1632, Jan. 2015.
- [2] M. R. Shakarami, H. Beiranvand, A. Beiranvand, and E. Sharifipour, "A recursive power flow method for radial distribution networks: Analysis, solvability and convergence," *International Journal of Electrical Power & Energy Systems*, vol. 86, pp. 71–80, Mar. 2017, <https://doi.org/10.1016/j.ijepes.2016.10.002>.
- [3] B. de Nadai Nascimento, A. C. Zambroni de Souza, J. G. de Carvalho Costa, and M. Castilla, "Load shedding scheme with under-frequency and undervoltage corrective actions to supply high priority loads in islanded microgrids," *IET Renewable Power Generation*, vol. 13, no. 11, pp. 1981–1989, 2019, <https://doi.org/10.1049/iet-rpg.2018.6229>.
- [4] M. A. Zdiri, A. S. Alshammari, A. A. Alzamil, M. B. Ammar, and H. H. Abdallah, "Optimal Shedding Against Voltage Collapse Based on Genetic Algorithm," *Engineering, Technology & Applied Science Research*, vol. 11, no. 5, pp. 7695–7701, Oct. 2021, <https://doi.org/10.48084/etasr.4448>.
- [5] P. S. Bhowmik, S. P. Bose, D. V. Rajan, and S. Deb, "Power flow analysis of power system using Power Perturbation method," in *2011 IEEE Power Engineering and Automation Conference*, Sep. 2011, vol. 3, pp. 380–384, <https://doi.org/10.1109/PEAM.2011.6135117>.
- [6] R. Yan, T. K. Saha, N. Modi, N. A. Masood, and M. Mosadeghy, "The combined effects of high penetration of wind and PV on power system frequency response," *Applied Energy*, vol. 145, pp. 320–330, May 2015, <https://doi.org/10.1016/j.apenergy.2015.02.044>.
- [7] S. M. Ghania, K. R. M. Mahmoud, and A. M. Hashmi, "A Reliability Study of Renewable Energy Resources and their Integration with Utility Grids," *Engineering, Technology & Applied Science Research*, vol. 12, no. 5, pp. 9078–9086, Oct. 2022, <https://doi.org/10.48084/etasr.5090>.
- [8] V. V. Lakshmi, G. V. Naga, and A. J. Laxmi, "Optimal Allocation and Sizing of Multiple Distributed Generators Using Genetic Algorithm," in *International Conference on Advances in Communication, Network, and Computing*, 2014, pp. 305–312.
- [9] O. Mohamed, M. Mohamed, and A. Kansab, "Optimal Placement and Sizing of Distributed Generation Sources in Distribution Networks Using SPEA Algorithm," *International Journal on Electrical Engineering and Informatics*, vol. 11, no. 2, pp. 326–340, Jun. 2019, <https://doi.org/10.15676/ijeii.2019.11.2.7>.
- [10] M. A. Zdiri, B. Dhouib, Z. Alaas, F. B. Salem, and H. H. Abdallah, "Load Flow Analysis and the Impact of a Solar PV Generator in a Radial Distribution Network," *Engineering, Technology & Applied Science Research*, vol. 13, no. 1, pp. 10078–10085, Feb. 2023, <https://doi.org/10.48084/etasr.5496>.
- [11] J. Schmidhuber, "Deep Learning," *Scholarpedia*, vol. 10, no. 11, Nov. 2015, Art. no. 32832, <https://doi.org/10.4249/scholarpedia.32832>.
- [12] J. Ahmad, H. Farman, and Z. Jan, "Deep Learning Methods and Applications," in *Deep Learning: Convergence to Big Data Analytics*, M. Khan, B. Jan, and H. Farman, Eds. Singapore: Springer, 2019, pp. 31–42.
- [13] S. F. Ahmed *et al.*, "Deep learning modelling techniques: current progress, applications, advantages, and challenges," *Artificial Intelligence Review*, vol. 56, no. 11, pp. 13521–13617, Nov. 2023, <https://doi.org/10.1007/s10462-023-10466-8>.
- [14] H. Li, A. Zhang, X. Shen, and J. Xu, "A load flow method for weakly meshed distribution networks using powers as flow variables," *International Journal of Electrical Power & Energy Systems*, vol. 58, pp. 291–299, Jun. 2014, <https://doi.org/10.1016/j.ijepes.2014.01.015>.
- [15] J.-H. Teng and C. Y. Chang, "A novel and fast three-phase load flow for unbalanced radial distribution systems," *IEEE Transactions on Power Systems*, vol. 17, no. 4, pp. 1238–1244, Nov. 2002, <https://doi.org/10.1109/TPWRS.2002.805012>.
- [16] M. A. Fares, L. Atik, G. Bachir, and M. Aillerie, "Photovoltaic panels characterization and experimental testing," *Energy Procedia*, vol. 119, pp. 945–952, Jul. 2017, <https://doi.org/10.1016/j.egypro.2017.07.127>.
- [17] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, <https://doi.org/10.1038/nature14539>.
- [18] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, Jan. 2015, <https://doi.org/10.1016/j.neunet.2014.09.003>.
- [19] A. Oka, N. Ishimura, and S. Ishihara, "A New Dawn for the Use of Artificial Intelligence in Gastroenterology, Hepatology and Pancreatology," *Diagnostics*, vol. 11, no. 9, Sep. 2021, Art. no. 1719, <https://doi.org/10.3390/diagnostics11091719>.
- [20] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to Sequence Learning with Neural Networks," in *Advances in Neural Information Processing Systems*, 2014, vol. 27.
- [21] X. Wang, W. Jiang, and Z. Luo, "Combination of Convolutional and Recurrent Neural Network for Sentiment Analysis of Short Texts," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, Osaka, Japan, Sep. 2016, pp. 2428–2437, [Online]. Available: <https://aclanthology.org/C16-1229>.
- [22] T. Yao, Y. Pan, Y. Li, and T. Mei, "Incorporating Copying Mechanism in Image Captioning for Learning Novel Objects," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 5263–5271, <https://doi.org/10.1109/CVPR.2017.559>.
- [23] J. Li, D. Xiong, Z. Tu, M. Zhu, M. Zhang, and G. Zhou, "Modeling Source Syntax for Neural Machine Translation." arXiv, May 02, 2017, <https://doi.org/10.48550/arXiv.1705.01020>.
- [24] T. Mikolov, S. Kombrink, L. Burget, J. Černocký, and S. Khudanpur, "Extensions of recurrent neural network language model," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011, pp. 5528–5531, <https://doi.org/10.1109/ICASSP.2011.5947611>.
- [25] V. Vita, "Development of a Decision-Making Algorithm for the Optimum Size and Placement of Distributed Generation Units in Distribution Networks," *Energies*, vol. 10, no. 9, Sep. 2017, Art. no. 1433, <https://doi.org/10.3390/en10091433>.
- [26] S. Katyara *et al.*, "Leveraging a Genetic Algorithm for the Optimal Placement of Distributed Generation and the Need for Energy Management Strategies Using a Fuzzy Inference System," *Electronics*, vol. 10, no. 2, Jan. 2021, Art. no. 172, <https://doi.org/10.3390/electronics10020172>.
- [27] R. Deshmukh and A. Kalage, "Optimal Placement and Sizing of Distributed Generator in Distribution System Using Artificial Bee Colony Algorithm," in *2018 IEEE Global Conference on Wireless Computing and Networking (GCWCN)*, Lonavala, India, Nov. 2018, pp. 178–181, <https://doi.org/10.1109/GCWCN.2018.8668633>.
- [28] M. A. Ali, A. R. Bhatti, A. Rasool, M. Farhan, and E. Esenogho, "Optimal Location and Sizing of Photovoltaic-Based Distributed Generations to Improve the Efficiency and Symmetry of a Distribution Network by Handling Random Constraints of Particle Swarm Optimization Algorithm," *Symmetry*, vol. 15, no. 9, Sep. 2023, Art. no. 1752, <https://doi.org/10.3390/sym15091752>.

Feature Imputation using Neutrosophic Set Theory in Machine Learning Regression Context

Yamen El Touati

Department of Computer Science, Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
yamen.touati@nbu.edu.sa (corresponding author)

Walid Abdelfattah

Department of Mathematics, College of Arts and Science, Northern Border University, Saudi Arabia
walid.abdelfattah@nbu.edu.sa

Received: 11 February 2024 | Revised: 28 February 2024 | Accepted: 11 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7052>

ABSTRACT

The prediction context of machine learning aims to discern the underlying patterns that dictate the characteristics needed to forecast the output. This prediction, however, lacks precision when the input data are not accurate or precise. The current study focuses on feature imputation through the application of the neutrosophic set theory. The primary concept involves substituting the feature data, which may exhibit accuracy and correctness issues, with neutrosophic variables considering the degrees of truth, indeterminacy, and falsity to produce more precise and resilient predictions. The proposed method was implemented in a specific case study, and the results are analyzed.

Keywords-neutrosophic set theory; machine learning; prediction; features imputation

I. INTRODUCTION

In the field of machine learning [1], specifically in the context of regression models [2], the correctness of data characteristics is of utmost importance. The efficacy of a regression model depends on its ability to identify patterns and correlations within the input data, therefore establishing the accuracy of these attributes crucial in the overall prediction process [3-6]. Flawed or untrustworthy data characteristics might add interference and hinder the model's ability to make generalizations, resulting in unreliable predictions and impaired effectiveness. Furthermore, the collaboration among specialists in a certain field and professionals specialized in data analysis is crucial in enhancing data feature precision. Experts in the field can provide useful insights into the intricacies of the data, enhancing the comprehension of the latter's traits and their possible influence on the model's performance. Consistent surveillance and verification of the data during the model's lifespan are equally crucial, as the dataset attributes may change over time.

Experts face significant obstacles when evaluating and formalizing the correctness and precision of data, especially when it comes to aspects that directly affect the accuracy of predictions in machine learning and regression [3, 7]. A significant obstacle arises from the ever-changing nature of real-world data, where inaccuracies and imprecisions might

originate from many different sources, such as human fallibility, measurement devices, or environmental influences. Experts must confront the complexities of detecting and reducing these mistake sources, which often necessitate a sophisticated understanding of both the particular field and the subtleties of the machine learning algorithms used. Furthermore, there is also difficulty in reassuring the presence of ongoing monitoring and adjustment, as the data environment changes over time. Achieving a suitable equilibrium between the aspiration for the highest accuracy and the pragmatic limitations of collecting and analyzing data is still another complex facet. Collaboration between experts in a specific field and specialized data analysis professionals becomes crucial to address these challenges. This is because the former utilizes the combined expertise to improve the accuracy and precision of characteristics, ultimately strengthening the dependability of machine learning and regression models to make well-informed predictions. The task of developing a formal system to handle imprecise data in machine learning becomes increasingly challenging when dealing with a wide range of sources [8]. Inclusion of diverse datasets brings about complexity, necessitating the use of advanced statistical models to reconcile the differences between them. The need to measure inaccuracy and resolve uncertainty in feature values requires a well-balanced and flexible formal framework.

The neutrosophic logic concept [9-18] is an important notion that addresses, among other ones, the issue of data inaccuracies. Neutrosophic logic is a theoretical framework that expands on classical and fuzzy logic to incorporate indeterminacy, uncertainty, and ambiguity in greater depth. It is especially significant when it comes to addressing data inaccuracies, and it allows for the representation of not only binary values (true and false), but also values that are uncertain or undefined, recognizing the inherent inaccuracies in real-world data. Therefore, it improves the reliability of decision-making processes, particularly in areas, like machine learning and data analysis, by offering a systematic method to deal with uncertainty. This is particularly important since a precise representation of uncertainties is vital in these domains. The shift in paradigm towards a more refined depiction of truth values provides a valuable means of addressing the difficulties emerged by imperfections in datasets, thus enhancing the accuracy and dependability of outcomes in data-driven applications. Within this specific context, this study aimed to present a strategy that utilizes neutrosophic theory to address data imprecision and ameliorate prediction models in real-world scenarios.

II. MACHINE LEARNING AND REGRESSION TECHNIQUES

Machine learning is at the forefront of a data analysis and predictive modeling revolution, providing deep insights in diverse disciplines, namely finance, healthcare, marketing, and environmental sciences. Regression approaches are highly effective tools in machine learning to understand and forecast continuous results. These tools allow us to see patterns in the data, build connections between variables, and predict future events with exceptional precision. Regression models serve as the foundation for analyzing various phenomena, such as forecasting stock market patterns, estimating patient recovery durations, and predicting customer buying schemes. This section focuses on five distinct regressors, each of them having a distinct ability to model intricate relationships within the data. Consequently, these regressors improve decision-making processes and strategic planning by providing more comprehensive insights.

A. Linear Regression

Linear Regression [19] is a fundamental and traditional regression method that seeks to represent the correlation between a dependent variable and one or more autonomous ones by adjusting a linear equation to the collected data. The equation can be expressed as $Y = aX + b$, where Y is the dependent variable, X is the independent variable, a represents the slope, and b is the intercept. Linear Regression is most effective when there is a linear relationship between the independent and the dependent variables, where the data should be relatively free from noise. It is also efficient when the amount of features is not excessively large, as it can be susceptible to overfitting situations in areas with a high number of dimensions.

B. Ridge Regression

Ridge regression is a form of linear regression that incorporates a regularization term, often known as the L2

penalty, and is obtained by taking the squared magnitude of the coefficients [20]. Multiple regression is used when the data exhibit multicollinearity, which refers to a situation where the independent variables are strongly correlated. Ridge Regression is effective in situations where it is necessary to prevent overfitting and when the number of predictor variables surpasses the number of observations.

C. Support Vector Regression (SVR)

SVR is a machine-learning approach that is derived from the Support Vector Machine (SVM) algorithm [21]. It identifies the optimal line, or hyperplane in higher dimensions, which includes the highest number of points within a specific threshold margin. SVR demonstrates efficacy in handling nonlinear interactions by utilizing kernel tricks. Furthermore, SVR exhibits proficiency in both small and large datasets. In addition, it performs exceptionally well when there is a distinct separation between the data points and any outliers present.

D. Random Forest Regression

Random Forest Regression is an ensemble learning technique that utilizes decision tree regressors [22]. The system functions by generating numerous decision trees throughout the training process and producing the average prediction of each individual tree. This technique demonstrates great accuracy when the dataset comprises a combination of category and numerical variables and when the data exhibit intricate structures that are beyond the capability of basic models, such as linear regressors, to capture. Additionally, it displays resilience against overfitting, particularly when dealing with data that contain noise.

E. Gradient Boosting Regression

The Gradient Boosting Regressor is an ensemble technique that uses decision trees as its foundation [20]. Unlike Random Forest, which constructs trees simultaneously, Gradient Boosting constructs one tree at a time in a sequential manner, with each subsequent tree aiming to rectify the mistakes produced by the previously trained ones. Optimizing an arbitrary differentiable loss function enhances the performance of the model. Gradient Boosting is particularly effective in situations where the data are uneven and the main focus is on achieving high predicted accuracy. It is highly efficient for a broad spectrum of regression problems, particularly when the connection between the input and target variables is intricate and nonlinear. Nevertheless, meticulous adjustment of its parameters is necessary, as it has the potential to overfit in datasets with noise if it is not adequately regularized.

Each of these regressors possesses distinct advantages and optimal scenarios for usage. Linear Regression is highly competent in generating rapid and straightforward models and interpretations, particularly when the data exhibit linearity. Ridge Regression is a preferred method for addressing multicollinearity. SVR is proficient at accurately representing nonlinear relationships. Random Forest and Gradient Boosting are adept at identifying complex correlations within the data, but they demand substantial computational resources and may exhibit overfitting tendencies if not appropriately fine-tuned. The accuracy of a regression model is determined by the degree to which its predictions align with the observed data. It can be

affected by the selection of model parameters, the accuracy of the input data, and the degree to which the assumptions of the regression model align with the actual relationships in the data. It is advisable to evaluate the model's performance on data that have not been seen before and adjust the model's hyperparameters to optimize its prediction accuracy.

III. BACKGROUND ON NEUTROSOPHIC SET THEORY

In the realm of neutrosophic logic [6-15], the concepts of truth, indeterminacy, and falsity degrees form the foundational pillars for understanding and analyzing information that is imprecise, incomplete, or inconsistent. These degrees provide a more nuanced approach to reasoning under uncertainty compared to traditional binary logic. The truth degree represents the extent to which a statement is considered genuine. Unlike binary logic, where a statement can only be true or false, the truth degree allows for varying levels of truthfulness, acknowledging that knowledge about the world is often partial or probabilistic. The indeterminacy degree measures the level of uncertainty or lack of information regarding the truth value of a statement. This is particularly important in scenarios where information is missing, ambiguous, or the existing knowledge about a situation is incomplete, making it difficult to fully ascertain the veracity of a statement. The falsity degree quantifies the extent to which a statement is considered false. Just as with the truth degree, the falsity degree embraces the complexity of real-world information, where statements may not be entirely false but contain elements of falsehood. These three degrees are not mutually exclusive and are designed to capture the multifaceted nature of truth in the real world, where statements can be partially true, partially false, and partially indeterminate at the same time. Thus, the neutrosophic set theory is seen as a generalized extension of fuzzy and intuitionistic fuzzy theories. Furthermore, it exhibits a higher degree of resemblance to human cognition, since it more accurately replicates human decision-making processes by considering evidence that involves uncertainty. Following the first introduction of the concept of neutrosophic sets [9], other extensions have been suggested, the most renowned being the single-valued neutrosophic set [12]. Single-valued neutrosophic numbers are a subset of single-valued neutrosophic sets and play a crucial role in neutrosophic multiattribute decision-making situations, as they accurately represent an uncertain quantity [14]. Neutrosophic sets and their extensions have been applied in many sectors [11].

Triangular neutrosophic numbers are a significant extension of neutrosophic theory [17]. This triangle structure enables a more comprehensive representation of ambiguity and incomplete truths, which frequently arise in real-life situations where facts and human perceptions are rarely absolute. Triangular neutrosophic numbers offer a strong instrument for intricate decision-making procedures that require a subtle approach to ambiguity and subjectivity [16]. A triangular neutrosophic number is defined as $\tilde{a} = \langle (a^1, a, a^2), t_{\tilde{a}}, i_{\tilde{a}}, f_{\tilde{a}} \rangle$, incorporating lower bound a^1 , median value a , and upper bound a^2 besides measures of truth, indeterminacy, and falsity [15]. For de-neutrosophication, the variation degree of \tilde{a} is:

$$\theta_{\tilde{a}(\alpha, \beta, \gamma)} = \frac{1}{4} \left[\frac{\alpha}{t_{\tilde{a}}} + 2 \frac{(1-\beta)}{1-i_{\tilde{a}}} + \frac{(1-\gamma)}{1-f_{\tilde{a}}} \right], \theta_{\tilde{a}(\alpha, \beta, \gamma)} \in [0, 1] \quad (1)$$

where α denotes the minimal degree of acceptance, $\alpha \in [0, t_{\tilde{a}}]$, β denotes the maximal degree of indeterminacy, $\beta \in [d_{\tilde{a}}, 1]$, and γ denotes the maximal degree of rejection, $\gamma \in [f_{\tilde{a}}, 1]$. The proposed approach posits that a triangular neutrosophic number, represented as $\tilde{a} = \langle (a^1, a, a^2), t_{\tilde{a}}, i_{\tilde{a}}, f_{\tilde{a}} \rangle$ can be simplified to its equivalent interval value in the following manner:

- When the acceptance degree α is at its minimal level, and the indeterminacy and rejection degrees β and γ are at their greatest levels, the value of \tilde{a} will fall within the range defined by its lower and upper bounds, thus $\tilde{a} = [a^1, a^2]$.
- When the acceptance degree α is at its highest level and the indeterminacy and rejection degrees β and γ are at their lowest levels, the value of \tilde{a} will tend to be a specific crisp value, which is logically considered to be equal to its median value. Therefore, $\tilde{a} = [a, a] = a$.

The neutrosophic theory is a notable expansion of classical and fuzzy logic that addresses the inherent ambiguity and uncertainty of real-world data. The mathematical framework offers a means of representing uncertain, inconsistent, and incomplete information by utilizing its key elements: truth, indeterminacy, and falsity. Triangular neutrosophic numbers are a specific form of representation in this theory, distinguished by a three-part structure that measures the extent of truth, indeterminacy, and falsehood for each element. This structure is highly skilled at representing the intricate range of real-life situations where data frequently deviate from binary or fuzzy logic.

IV. USING NEUTROSOPHIC VARIABLES TO DEAL WITH DATA IMPRECISION

The adoption of a neutrosophic paradigm allows for a more accurate depiction of fine details in empirical data. Utilizing neutrosophic variables is a deliberate and calculated reaction to the widespread existence of uncertain and contradictory features within datasets. It expands the range of analysis, allowing predictive systems to understand and combine the inherent uncertainties of the data. This enhances the decision-making process by providing a variety of insights that cannot be shown by traditional binary logic.

Consider $A = (a_1, \dots, a_n)$ to be a vector of reals that represents a feature. Consider a_i as a component of A that represents one experiment value of the feature A . a_i^1 and a_i^2 are proposed to be computed as follows.

$$a_i^1 = a_i(1 - |CVar(A)|)$$

$$a_i^2 = a_i(1 + |CVar(A)|)$$

where $CVar(A) = \frac{|\bar{X}(A)|}{\delta_A}$, $\bar{X}(A)$ represents the mean of A , and δ_A represents the standard deviation. The coefficient of variation $CVar(A)$ is a normalized metric that quantifies the extent of variability in probability or frequency distributions, and it is commonly used to assess the level of variation among distinct datasets, even if their means differ significantly.

The proposed method involves using $CVar(A)$ to modify each component a_i of the feature vector A , resulting in the creation of two additional components, a_i^1 and a_i^2 . This adjustment accounts for the fluctuation in the characteristic compared to its average value. A higher coefficient of variation corresponds to larger modifications made to a_i , indicating a greater dispersion of data. On the other hand, a low coefficient of variation results in smaller adjustments, indicating that the data points are closer to the mean. Each component a_i is then represented by the triangular neutrosophic variable as follows:

$$\tilde{a}_i = \langle (a_i^1, a_i, a_i^2), t_{a_i}, i_{\tilde{a}_i}, f_{\tilde{a}_i} \rangle$$

where t_{a_i} is the truth degree related to a_i , $i_{\tilde{a}_i}$ is the indeterminacy degree related to a_i , and $f_{\tilde{a}_i}$ is the falsity degree related to a_i . By integrating $CVar(A)$ into the neutrosophic variable representation, the relative dispersion in the data, which may improve the model's ability to manage and interpret feature fluctuations, can be considered. The model has the potential to enhance accuracy in regression analysis by acknowledging and incorporating the inherent variability in the data into the forecasting process. It is crucial to include this aspect in datasets where the variability greatly affects the behavior of the dependent variable being modeled.

To comprehend each crisp component a_i or its substitute (a_i^1, a_i, a_i^2) in a dataset, the former is linked to a triplet $(t_{a_i}, i_{a_i}, f_{a_i})$ that denotes its levels of truth, indeterminacy, and falsity, respectively. To make this procedure more efficient, one might select a predetermined set of configurations that accurately represent the core characteristics of each variable. This collection was carefully chosen by an expert who has a thorough understanding of the subject matter. It is therefore ensured that the given values for truth, uncertainty, and falsehood are perfectly linked with the variable's real-world meanings and behaviors. This is particularly important when data are collected from many sources that differ in terms of quality and precision. Neutrosophic logic extends the basic notions of truth, indeterminacy, and falsehood by introducing the ideas of minimal degree of acceptance α , maximal degree of indeterminacy β , and maximal degree of rejection γ . These concepts help to enhance the examination of assertions in situations of uncertainty.

The notion of a minimal degree of acceptance establishes the minimum level at which a proposition is deemed sufficiently truthful within a specific context. The purpose of this criterion is to exclude irrelevant or unreliable information and prioritize content that passes a minimum standard of reliability or relevance. The maximal degree of indeterminacy refers to the highest level of uncertainty that can be accepted in a decision-making process. It clearly defines the border between acceptable and excessive vagueness, guaranteeing that judgments are made with the highest level of knowledge feasible. The maximal degree of rejection sets the upper limit for a statement to be considered false. It acts as a standard for filtering out mostly erroneous information, allowing for a more precise differentiation between helpful and misleading information.

The expression θ (1) reflects the overall assessment of the variable within the neutrosophic context. Integrating truth, indeterminacy, and falsity degrees with the minimal degree of acceptance, maximal degree of indeterminacy, and maximal degree of rejection provides a robust framework for navigating the complexities of real-world information. By quantifying the nuances of truthfulness, uncertainty, and falseness, neutrosophic logic allows for a more sophisticated analysis of statements, facilitating informed decision-making even in the presence of incomplete or contradictory information. This holistic approach underscores the adaptability of neutrosophic logic in addressing the challenges of uncertainty, making it a valuable tool in fields ranging from artificial intelligence and data science to philosophy and theoretical mathematics. The final step involves adjusting the original variable a_i using the neutrosophic assessment to obtain two de-neutrosophied variables a_i^l and a_i^u as follows:

$$a_i^l = a_i^1 + \theta_{\tilde{a}_i}(a_i - a_i^1)$$

$$a_i^u = a_i^2 + \theta_{\tilde{a}_i}(a_i^2 - a_i)$$

The de-neutrosophied variables a_i^l and a_i^u serve as more nuanced representations that consider the variable's corresponding levels of truth, indeterminacy, and falsity. In other words, these two variables represent the lower and upper bounds, respectively, that could be considered projections of \tilde{a} via the proposed de-neutrosophication procedure. By incorporating neutrosophic factors into predictive models, the accuracy is expected to be enhanced as these variables enable a more comprehensive depiction of data. Neutrosophic logic captures additional intricacies of the real world that could be overlooked by classical precise and crisp values. By carefully evaluating several factors, this approach allows the model to recognize and adapt to the underlying complexities and uncertainties in the data, resulting in predictions that better reflect the intricate nature of reality.

V. CASE STUDY: HAPPINESS INDICATOR

The dataset used for this case study was derived from the annual happiness report and the country statistics report published by the UN [21]. The former consists of a variety of measures that evaluate the welfare and contentment of inhabitants in different countries. This dataset encompasses various measures, including Gross Domestic Product (GDP) per capita, Economy Services, Agricultural Production Index, Infant Mortality, Health Total Expenditure, CO2 Emission, Energy Production, and several more indicators (a total of 38 indicators). These metrics jointly contribute to the calculation of a happiness score for each country. The specific dataset also classifies countries according to their regions and provides demographic context to augment the examination of happiness levels.

This study focused on the GDP metric as one of the most important key features of happiness. The proposed method was applied using three configurations, as shown in Table I, related to the truth, indeterminacy, and falsity values. These configurations were assigned to each measurement of GDP, mainly according to its belonging region.

TABLE I. USED CONFIGURATIONS OF TRUTH, INDETERMINACY, AND FALSITY

	Truth	Indeterminacy	Falsity
Configuration 1	0.9	0.2	0.2
Configuration 2	0.6	0.4	0.3
Configuration 3	0.2	0.7	0.4

By incorporating neutrosophic logic, traditional GDP values are transformed into a pair of neutrosophic variables that capture the inherent uncertainty and variability within the economic data. New variables, represented as GDP^I and GDP^F , were derived from the original GDP values while adjusting for the degree of acceptance, indeterminacy, and rejection ascribed to each data point by an expert. In the first version of the analysis, traditional regression methods were employed to predict happiness scores using the original dataset, which included GDP as a conventional, crisp variable. Linear Regression, Ridge Regression, SVR, Random Forest regression, and Gradient Boosting regression were among the techniques used. These regressors drew from the raw data, aiming to capture the direct relationship between GDP and the reported levels of happiness in different countries. This approach is grounded in classical statistical analysis, where the impact of each predictor is considered in its original form, with the assumption that the relationships between variables are linear and directly observable.

For the second version, the GDP variable was substituted with two new de-neutrosophied values derived from the neutrosophic logic-based equations. Not only did these new values aim to encapsulate the economic output, but also the relative uncertainty and perception variations in economic well-being that GDP figures might imply. The same regressors were then applied to this transformed dataset. The prediction models could potentially account for the nuanced influences of economic factors on happiness by considering these broader implications. Figure 1 illustrates the R^2 scores obtained with both experiments.

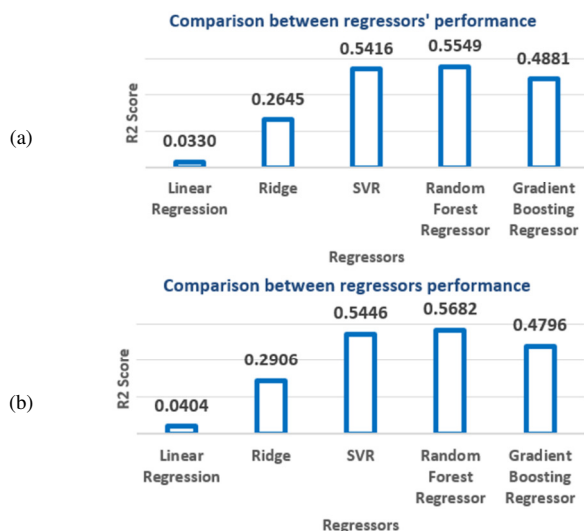


Fig. 1. R^2 score with (a) original dataset, and (b) updated dataset with substituted GDP with GDP^I and GDP^F .

This technique demonstrates a more dynamic and possibly meaningful understanding of the data, suggesting that the overall welfare of a country's inhabitants may not be adequately represented solely by conventional economic measures. The purpose was to determine whether these augmented data could improve the accuracy of the predictions and provide a more profound understanding of the factors that influence happiness. The varying prediction performance using the original dataset and the updated dataset is portrayed in Tables II and III, respectively.

TABLE II. PERFORMANCE WITH ORIGINAL DATASET

Applied Regressor	MSE	RMSE	R^2
Linear Regression	1.31506912	1.14676463	0.03297209
Ridge	1.00025007	1.00012503	0.26447232
SVR	0.62337761	0.78954266	0.54160314
Random Forest	0.60534793	0.77804108	0.55486116
Gradient Boosting	0.69617557	0.83437136	0.48807162

TABLE III. PERFORMANCE WITH UPDATED DATASET

Applied Regressor	MSE	RMSE	R^2
Linear Regression	1.30499936	1.14236569	0.04037682
Ridge Regression	0.96465932	0.98217072	0.29064376
SVR	0.61924131	0.78691887	0.54464475
Random Forest	0.58726705	0.76633351	0.56815682
Gradient Boosting	0.70768057	0.84123752	0.47961149

When comparing the results of the two experiments, traditional prediction models using original data versus models employing neutrosophic logic-enhanced variables, a clear distinction in performance emerges. The first experiment, which deployed original GDP data, demonstrated a range of effectiveness across different types of regression. Linear Regression, constrained by its simplicity and assumption of linearity, lagged in performance. Ridge Regression offered moderate improvements, possibly due to its penalization of coefficient size, which helps to manage multicollinearity. SVR, Random Forest, and Gradient Boosting, known for handling nonlinear patterns and complex relationships, performed better. In particular, the Random Forest and SVR models stood out, indicating their robustness in handling the dataset's complexity. The second experiment introduced the neutrosophic logic-enhanced GDP variables, aiming to capture the uncertainty and vagueness inherent in economic indicators. This transformation was intended to provide richer, more nuanced input for the regressors. The results disclosed that the Random Forest regressor using the new variables had an improvement in the R^2 score, suggesting a better model fit. The robustness of the Random Forest to noise and its ability to handle nonlinear relationships possibly contributed to this improved performance. Conversely, Linear Regression's minimal change in performance reaffirmed its limitations when dealing with complex, transformed datasets. Gradient Boosting and SVR also maintained high performance, indicating their ability to capitalize on nuanced data, although they did not show significant improvement over the traditional method. This could suggest that, while the new variables added information, they did not substantially alter the predictive relationships these models were able to capture.

All in all, the introduction of neutrosophic logic-based variables appeared to bolster the performance of models capable of handling complexity and nonlinearity. The enhanced performance of the Random Forest regressor points to its suitability in scenarios where the independent variables are enriched with additional interpretative layers. Linear Regression's underperformance underscores the importance of model selection in predictive analytics, as simpler models may not suffice when variables capture more profound economic and psychological constructs, such as those implied by neutrosophic logic.

VI. CONCLUSION

In the realm of predictive analytics, accurately modeling and forecasting subjective measures is a complex problem. Traditional linear models that utilize crisp data often fail to capture the multifaceted nature of such phenomena. This study aimed to address the limitations discussed above by introducing a novel approach that incorporates neutrosophic logic into the predictive modeling process, transforming a crisp variable into a pair of de-neutrosophied values. This novel approach was applied on a happiness assessment dataset based on multiple parameters, including GDP. The GPD crisp values were modified adopting the proposed de-neutrosophication technique. Not only was the purpose of this process to encompass economic productivity, but also to incorporate the relative unpredictability and variability in the perception of economic welfare. The results of this method were promising. Compared to traditional regression models, the neutrosophic method showed an improvement in model performance, particularly with the Random Forest regressor, which demonstrated a significant increase in the R^2 score. This indicates a stronger correlation between the predictive model and the actual happiness scores, suggesting that neutrosophic variables provide a more accurate and nuanced reflection of the underlying economic factors that contribute to happiness.

The conclusions drawn from these experiments are twofold. First, the application of neutrosophic logic in predictive modeling can offer a meaningful improvement in the accuracy of predictions, especially in complex and subjective contexts, such as happiness indices. Second, the choice of regressor is crucial. Models that can leverage the complexity and nuances introduced by neutrosophic variables, like the Random Forest regressor, tend to perform better. Although these findings demonstrate the potential of neutrosophic logic, caution should be exercised in generalizing these conclusions to the broader field of machine learning prediction models. There is a need for further investigation and exploration in diverse domains to validate the applicability of neutrosophic set theory across a wider range of predictive models.

ACKNOWLEDGMENT

The authors gratefully acknowledge the approval and support of this research from the Deanship of Scientific Research, Northern Border University, Arar, Kingdom of Saudi Arabia by the CSCR-2022-11-1243 grant.

REFERENCES

- [1] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, Jul. 2015, <https://doi.org/10.1126/science.aaa8415>.
- [2] R. Trinchero and F. Canavero, "Machine Learning Regression Techniques for the Modeling of Complex Systems: An Overview," *IEEE Electromagnetic Compatibility Magazine*, vol. 10, no. 4, pp. 71–79, 2021, <https://doi.org/10.1109/MEMC.2021.9705310>.
- [3] C. J. M. Maas and J. J. Hox, "Robustness issues in multilevel regression analysis," *Statistica Neerlandica*, vol. 58, no. 2, pp. 127–137, 2004, <https://doi.org/10.1046/j.0039-0402.2003.00252.x>.
- [4] Y. E. Touati, "Deadline Verification for Web Services Using Timed Automata," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 8013–8016, Feb. 2022, <https://doi.org/10.48084/etasr.4611>.
- [5] U. Khan, K. Khan, F. Hassan, A. Siddiqui, and M. Afaq, "Towards Achieving Machine Comprehension Using Deep Learning on Non-GPU Machines," *Engineering, Technology & Applied Science Research*, vol. 9, no. 4, pp. 4423–4427, Aug. 2019, <https://doi.org/10.48084/etasr.2734>.
- [6] B. Trstenjak, D. Donko, and Z. Avdagic, "Adaptable Web Prediction Framework for Disease Prediction Based on the Hybrid Case Based Reasoning Model," *Engineering, Technology & Applied Science Research*, vol. 6, no. 6, pp. 1212–1216, Dec. 2016, <https://doi.org/10.48084/etasr.753>.
- [7] R. Chen and I. C. Paschalidis, "A Robust Learning Approach for Regression Models Based on Distributionally Robust Optimization," *Journal of Machine Learning Research*, vol. 19, no. 13, pp. 1–48, 2018.
- [8] E. Hüllermeier, "Learning from imprecise and fuzzy observations: Data disambiguation through generalized loss minimization," *International Journal of Approximate Reasoning*, vol. 55, no. 7, pp. 1519–1534, Oct. 2014, <https://doi.org/10.1016/j.ijar.2013.09.003>.
- [9] F. Smarandache, *A Unifying Field in Logics: Neutrosophic Logic*. American Research Press, 1999.
- [10] F. Smarandache, *A Unifying Field in Logics. Neutrosophy: Neutrosophic Probability, Set and Logic*. American Research Press, 1999.
- [11] F. Smarandache and S. Pramanik, *New Trends in Neutrosophic Theory and Applications*. Pons Editions, 2016.
- [12] H. Wang, Florentin Smarandache, Y. Zhang, and R. Sunderraman, "Single valued neutrosophic sets," in *Collected Papers. Volume XIV: Neutrosophics and other topics*, Global Knowledge, 2022.
- [13] C. H. Wang, C. C. Chuang, and C. C. Tsai, "A fuzzy DEA–Neural approach to measuring design service performance in PCM projects," *Automation in Construction*, vol. 18, no. 5, pp. 702–713, Aug. 2009, <https://doi.org/10.1016/j.autcon.2009.02.005>.
- [14] I. Deli and Y. Şubaş, "A ranking method of single valued neutrosophic numbers and its applications to multi-attribute decision making problems," *International Journal of Machine Learning and Cybernetics*, vol. 8, no. 4, pp. 1309–1322, Aug. 2017, <https://doi.org/10.1007/s13042-016-0505-3>.
- [15] W. Abdelfattah, "Data envelopment analysis with neutrosophic inputs and outputs," *Expert Systems*, vol. 36, no. 6, 2019, Art. no. e12453, <https://doi.org/10.1111/exsy.12453>.
- [16] M. Abdel-Basset, M. Mohamed, Y. Zhou, and I. Hezam, "Multi-criteria group decision making based on neutrosophic analytic hierarchy process," *Journal of Intelligent & Fuzzy Systems*, vol. 33, no. 6, pp. 4055–4066, Jan. 2017, <https://doi.org/10.3233/JIFS-17981>.
- [17] M. Abdel-Basset, M. Mohamed, A. N. Hussien, and A. K. Sangaiah, "A novel group decision-making model based on triangular neutrosophic numbers," *Soft Computing*, vol. 22, no. 20, pp. 6629–6643, Oct. 2018, <https://doi.org/10.1007/s00500-017-2758-5>.
- [18] M. Abdel-Basset, M. Mohamed, and F. Smarandache, "An Extension of Neutrosophic AHP–SWOT Analysis for Strategic Planning and Decision-Making," *Symmetry*, vol. 10, no. 4, Apr. 2018, Art. no. 116, <https://doi.org/10.3390/sym10040116>.
- [19] G. A. F. Seber and A. J. Lee, *Linear Regression Analysis*. John Wiley & Sons, 2012.

-
- [20] G. C. McDonald, "Ridge regression," *WIREs Computational Statistics*, vol. 1, no. 1, pp. 93–100, 2009, <https://doi.org/10.1002/wics.14>.
- [21] M. Awad and R. Khanna, "Support Vector Regression," in *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, M. Awad and R. Khanna, Eds. Berkeley, CA, USA: Apress, 2015, pp. 67–80.
- [22] V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, and M. Chica-Rivas, "Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines," *Ore Geology Reviews*, vol. 71, pp. 804–818, Dec. 2015, <https://doi.org/10.1016/j.oregeorev.2015.01.001>.
- [23] S. Peter, F. Diego, F. A. Hamprecht, and B. Nadler, "Cost efficient gradient boosting," in *Advances in Neural Information Processing Systems*, 2017, vol. 30.
- [24] Elai, "Elaishalev/Countries_Happiness." Mar. 17, 2020, [Online]. Available: https://github.com/Elaishalev/Countries_Happiness.

The Role of Machine Learning in Managing and Organizing Healthcare Records

Ahmed Mohammed Alghamdi

Department of Software Engineering, College of Computer Science and Engineering, University of Jeddah, Saudi Arabia
amalghamdi@uj.edu.sa

Mahmoud Ahmad Al-Khasawneh

School of Computing, Skyline University College, University City Sharjah, 1797, Sharjah, UAE | Applied Science Research Center. Applied Science Private University, Amman, Jordan
mahmoud@outlook.my (corresponding author)

Ala Abdulsalam Alarood

College of Computer Science and Engineering, University of Jeddah, Saudi Arabia
aasoleman@uj.edu.sa

Eesa Alsolami

College of Computer Science and Engineering, University of Jeddah, Saudi Arabia
eaalsulami@uj.edu.sa

Received: 7 February 2024 | Revised: 25 February 2024 and 6 March 2024 | Accepted: 7 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7027>

ABSTRACT

With the exponential growth of medical data, Machine Learning (ML) algorithms are becoming increasingly important to the management and organization of healthcare information. This study aims to explore the role that ML can play in optimizing the management and organization of healthcare records, by identifying the challenges, advantages, and limitations associated with this technology. Consequently, the current study will contribute to the understanding of how ML might be applied to the healthcare industry in a variety of circumstances. Using the findings of this study, healthcare professionals, researchers, and policymakers will be able to make informed decisions regarding the adoption and implementation of ML techniques for regulating healthcare records. The findings of this paper revealed that ML can play an important role in efficiently directing and classifying healthcare records using different perspectives.

Keywords-machine learning; healthcare records; literature review methodology

I. INTRODUCTION

Today, healthcare management systems store a plethora of patient data regarding medical records and reports, body scans, treatments, medications, hospital records, etc. These data are increasingly complicated to handle. Inefficient data management can lead to bad decision-making [1-3]. Such huge data sets require some smart technology to be processed and extracted in a useful manner. It is necessary to have an accurate diagnosis system, especially when it comes to diseases like cancer, so that better treatment can be provided. A combination of human knowledge and computer derived assistance could be used to accomplish this. Various fields have already proven that ML is a fast and precise means of obtaining valid results; therefore, physicians can benefit from employing this intelligent technology to make better decisions [4, 5]. Doctors

and hospitals can regulate this immense amount of data by applying ML techniques, keeping electronic health records of patients accessible from anywhere at any time, and thus being able to provide better treatment. The use of ML has become integral to our daily lives. Several domains, such as transport and manufacturing have already proven their value, but now they are gaining popularity thanks to their excellent results in healthcare fields, like lung disease classification [6, 7], body organ recognition by medical images [8], lung nodule detection [9], reconstruction of medical images [10], and segmentation of brain tumors [11, 12]. To examine patients, the use of intelligent systems that are based on ML is increasing [13]. Combining mobile and cloud technology with ML provides more predictive outputs that can create human-like intelligent systems. At the same time, using ML in administering human resources has faced several challenges in privacy, security, and

data quality. Accordingly, this study explores how ML can organize and manage healthcare records. Therefore, the findings of this study will clarify how ML may affect the healthcare industry. A key benefit of this study is that it enlightens healthcare professionals, researchers, and policymakers, allowing them to make informed decisions about the implementation of ML techniques in arranging patients' healthcare records. This research is eventually anticipated to promote the evolution of healthcare practices by taking advantage of the ML capacity to assist in more effectively disposing patient data, which will accordingly lead to improved patient outcomes and medical services.

II. METHODOLOGY

To review the studies conducted on healthcare, ML areas, and the role ML plays in revolutionizing the healthcare world, this study adopted the literature review method [14-17]. A literature review plays a crucial role in research designs, granting a full assessment and a combination of the offered literature. It accomplishes several goals, involving recognizing gaps, discovering and integrating data, directing research, and significantly assessing existing work [14]. Figure 1 illustrates the adopted methodology, which involved the following stages:

- **Identification:** This stage entailed identifying the search terms, i.e. language, publication year, and keywords utilized in the search process. By employing six search engines, namely Google Scholar, IEEE Xplore, Science Direct, Scopus, Springer, and Web of Science (WoS), papers written in English and published from 2015 to 2024 about healthcare records and ML were collected. Then, papers related to healthcare records and ML were selected based on their subjects. Our search procedure used the following keywords: "healthcare records" and "machine learning". Out of the total 4,885 papers extracted from the search engines, 1760 articles derived from Google Scholar, 750 from Scopus, 233 from IEEE Xplore, 33 from Science Direct, 109 from WoS, and 2000 from Springer (see Figure 2).
- **Screening:** In this stage, the gathered papers were screened based on the title, abstract, and conclusion. All the 4,885 papers were screened, among which 1901 papers were excluded as duplicate entries.
- **Exclusion:** In this stage, some of the papers were excluded based on several criteria. For example, 1300 papers were ruled out for reasons, such as duplicate publications, poor quality, incomplete information, and insufficient data. Also, 529 were ejected due to irrelevancy, and 1127 owing to lacking results. Figure 3 displays the exclusion stage.
- **Inclusion:** In this stage, 28 papers were included as they were focused on either healthcare or ML, or on both, as depicted in Table I.

Based on the summary exhibited in Table I, ML applications can play a central role in the organization and management of healthcare records along with covering a variety of features, including data classification, organization, analysis, and summarization, identifying, and matching

patients, managing medications, providing clinical decision support, predicting and diagnosing diseases, engaging patients, and providing telemedicine, ML applications contribute to more valid healthcare outcomes by improving effectiveness, data accuracy, and the value of healthcare.

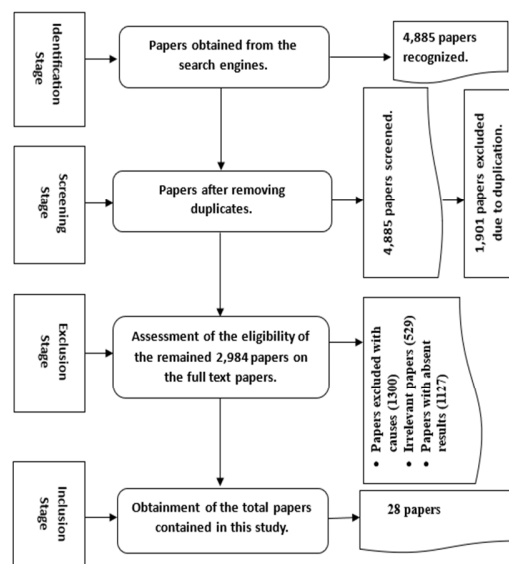


Fig. 1. The adopted methodology.

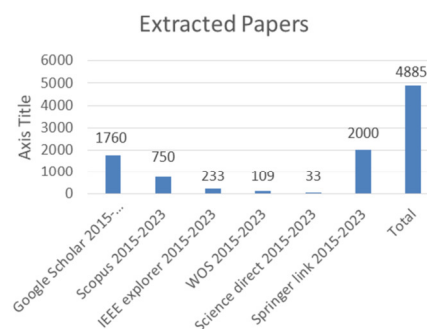


Fig. 2. Extracted papers from known search engines.

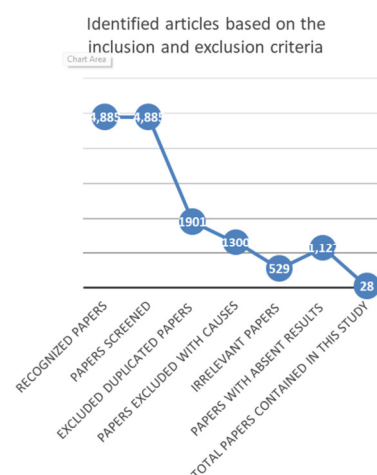


Fig. 3. Identification process based on the inclusion and exclusion criteria.

TABLE I. SELECTED PAPERS IN THIS STUDY

Ref.	Year	Focal points	ML application factors						
			Data classification	Data summarization	Patient identification	Medication management	Clinical decision support	Disease prediction	Patient engagement
[18]	2022	A framework for predicting the length of stay in lung cancer patients using ML models.	√	×	√	√	×	√	×
[19]	2015	Interactive collaborative care solutions for chronic diseases, which would empower patient-doctor relationships.	×	×	√	×	×	√	×
[20]	2016	A disease classification method using Disease Closeness, Disease Matrix, and Middle Measure to identify accurate and relevant diseases.	√	×	×	×	×	√	×
[21]	2019	A thorough review of the relationship between ML and blockchain technologies in smart healthcare. Several challenges arising when implementing ML in blockchain-based healthcare systems are discussed.	×	√	×	×	×	×	×
[22]	2020	A secure blockchain-based architecture specifically to address the requirements of e-healthcare systems.	×	×	√	√	×	×	×
[23]	2020	Combined healthcare pathway discovery and recovery time prediction models are discussed.	×	×	×	×	×	√	Q
[24]	2020	Generalizable ML approach by using data collected from a 40-year period from over 860,000 patients corresponding to over 6,700 prescription medications.	×	×	√	×	√	×	×
[25]	2020	Illustrating how ML models can be designed to have a "conversational" approach, through combining modeling decisions with domain expertise.	×	×	×	×	√	×	×
[26]	2020	A review of the importance of ML in healthcare.	×	×	×	×	×	×	×
[27]	2020	Comparison of the classification accuracy of ML methods using simulated EHR logs.	√	×	×	×	×	×	×
[28]	2021	Predicting patient survivability and determining which variables are most likely to affect it.							
[29]	2021	Presenting a decentralized workflow for trust-based federated learning without compromising privacy.	×	×	×	√	×	×	×
[30]	2021	Discussing methods already proposed for protecting privacy in cloud computing, including service-oriented architecture, and secure multiparty computation.	×	×	×	×	×	×	×
[31]	2021	Employing a variety of ML techniques to analyze healthcare records using different algorithms.	√	√	×	×	×	×	×
[32]	2022	A study on the impact of ML on the development of healthcare systems.	×	×	×	×	√	×	×
[33]	2022	Literature review regarding blockchain technology and improved health record management techniques that utilize artificial intelligence technologies.	×	×	×	×	√	×	×
[34]	2022	The possibility of generating EHRs by text-to-text translation using language models, which allows for highly flexible event imputation, is discussed.	√	×	×	×	×	×	×
[35]	2022	An optimized stacking ensemble learning model for breast cancer detection and classification using ML.	×	×	×	√	√	√	×
[36]	2022	Investigation of pneumonia patient profiles after initial RTI consultations in a large general population sample by using both conventional and ML approaches.	×	×	√	×	×	×	×
[37]	2022	Development and comparison of several ML algorithms to determine the hypertension risk among residents of Shanghai, China.	×	×	×	√	×	×	×

[38]	2022	A stochastic gradient descent method based on ML to manage medical records and optimize healthcare transactions.	√	×	×	×	×	×	×
[39]	2022	A review of the e-healthcare domain to help researchers better understand the benefits and drawbacks of using ML, blockchain technology, and other components to guarantee privacy and security in the healthcare system.	×	×	√	×	×	×	×
[40]	2022	Discussion about how ML is rooted in healthcare and some of its valuable applications in this industry.	×	×	×	×	√	×	×
[41]	2022	Discussing a variety of ML algorithms in addition to some of the ML properties of in healthcare along with their potential applications.	√	×	×	×	×	×	×
[42]	2023	Analysis of the increase in opioid prescribing in the U.S. during and after the peak of opioid prescribing.	×	×	×	×	√	×	×
[43]	2023	Applying automated intelligent practices to the early detection of heart diseases.	√	×	×	×	×	×	×
[44]	2023	Discussing various applications of quantum ML in healthcare, including quantum computing frameworks, algorithms, and concepts.	√	×	×	√	×	×	×
[45]	2024	Investigates the current state of the art in the application of ML to detect, categorize, and predict disorders stemming from cardiovascular and respiratory diseases.	√	×	×	×	×	√	×
[46]	2024	An effort to identify and overcome the limitations of traffic-centric approaches in cooperative cloud-edge networks by examining how these approaches can be utilized effectively.	√	×	×	×	√	×	×

The existing ML applications have exhibited a few limitations and challenges in their application to the management of health records, as displayed in Table II. On the other hand, several advantages have been offered by the ML application to healthcare record management as noticed in Table III.

III. FINDINGS AND DISCUSSION

Data were collected from six widely used search engines and were analyzed. At first, more than 4,000 journal articles, reports, books, book chapters, and dissertations were collected. After filtering with the above mentioned criteria, only 28 articles were finally selected to be analyzed. The results of the study reveal that ML can have a positive effect on handling and classifying healthcare records. In addition, the study found that the benefits of the ML utilization in the management and organization of healthcare records can be grouped in three categories, which are analyzed below.

A. Benefits for Healthcare Professionals

This study demonstrated the potential of ML techniques in upgrading the efficiency of healthcare systems. As healthcare professionals are expected to provide high-quality patient care while saving as much time as possible, they can benefit from the automated processing of administrative tasks, including data entry and retrieval. To improve their overall operational efficiency, minimize human error, and reduce manual processes, hospitals can utilize ML algorithms. Moreover, ML techniques have also proven to be extremely useful in ameliorating the accuracy and precision of healthcare records. If large datasets are analyzed employing ML-based algorithms, then patterns, trends, and anomalies that humans may not

observe can be detected. By automating the process of data extraction and analysis, healthcare professionals can make more informed decisions due to the use of accurate and up-to-date medical data. Keeping records in this manner not only ensures consistency and reliability, but it also minimizes the possibility of errors. Furthermore, ML techniques have been added to the healthcare record management process to improve the access, efficiency, and accuracy of healthcare records. With traditional paper-based systems, it can be difficult to access important information in a timely manner and find the information needed. Using a digital healthcare record, healthcare professionals can now access patient data, medical history, and test results from anywhere and at any time. It is easier for health professionals to collaborate, to share knowledge, to make informed decisions, and to provide timely care because of a higher degree of accessibility.

B. Benefits for Researchers

These techniques will be useful to researchers because they allow them to analyze large datasets more efficiently. The analysis of patterns, trends, and associations among different types of data can lead to the identification of risk factors and the prediction of outcomes. ML techniques can also be deployed to analyze the management of healthcare records. Manual analysis of historical data may not be able to provide researchers with the necessary insight to detect trends or patterns. ML techniques can greatly benefit healthcare research, particularly in the regulation of medical records. Incorporating these techniques into one's research can provide a valuable understanding of large datasets. By doing so, prevention, treatment, and management of diseases can be conducted more efficiently.

TABLE II. OBSERVED CHALLENGES AND LIMITATIONS IN THE APPLICATION OF ML TO HEALTHCARE RECORD MANAGEMENT

Challenge	Description
Privacy and security	It is essential that patients' privacy and security are protected when using ML algorithms. Compliance with privacy laws and robust security measures are essential to safeguarding sensitive information.
Data quality	Healthcare data must be of high quality for ML models to be accurate. For medical data to be reliable and valid, they must be cleansed, normalized, and pre-processed.
Model explainability	Transparent and understandable models should be used to make healthcare-related decisions. Healthcare professionals need to know how ML algorithms make decisions so that they can trust the information they receive.
Healthcare domain knowledge	Clinical guidelines, medical terminology, and domain knowledge are essential to the development of ML algorithms for healthcare. Developing and validating models should be the responsibility of experts in the domain.
Dataset size and diversity	Managing healthcare records involves analyzing large and diverse datasets, which makes ML implementation difficult.
Bias and fairness	ML algorithms could be trained, and data could be selected in a way that biases the results. Biased models can result in inaccurate diagnosis and treatment recommendations.
Data integration and interoperability	Healthcare data are fragmented across different systems and resources, making it difficult to integrate and extract meaningful insights from them. ML techniques require large and diverse datasets to produce accurate results. Ensuring data integration and interoperability is crucial for leveraging ML effectively in healthcare record management.
Regulatory and ethical considerations	The use of ML in healthcare raises numerous regulatory and ethical considerations. Healthcare organizations must comply with regulatory requirements related to data privacy, data security, and informed consent. Additionally, they need to consider the ethical implications of using ML technologies, such as the potential for discrimination and bias.

TABLE III. ADVANTAGES OF ML APPLICATION TO HEALTHCARE RECORD MANAGEMENT

Advantages	Description
Accuracy	ML algorithms are trained on vast amounts of data, making them capable of analyzing complex patterns and relationships. This enables them to accurately classify and extract meaningful information from healthcare records, reducing the chances of errors and inconsistencies.
Speed	ML algorithms can process large datasets quickly, giving healthcare providers the opportunity to access and retrieve information in real time. This increases productivity and allows appropriate decision-making in significant positions.
Data integration	To manage and organize healthcare records, it is often necessary to integrate data from a variety of medical systems, laboratories, and imaging devices. These data can be analyzed and integrated with ML algorithms to ensure their accessibility and normalization.
Clinical decision support	Clinical decision-making can benefit from the analysis of healthcare records by ML algorithms. The algorithms can help healthcare professionals identify patients at risk, recommend treatments, and allocate resources more effectively.
Quality assurance	Healthcare records can be analyzed using artificial intelligence algorithms to detect errors, inconsistencies, and frauds. Therefore, the integrity of the health system is improved, and personal information is safeguarded.
Enhanced efficiency	Healthcare professionals can focus more on patient care and decision-making if ML algorithms automate repetitive and time-consuming tasks.
Improved data accuracy	When using ML techniques, it is possible to validate and clean data, reducing errors and inconsistencies.
Streamlined access	Health professionals will be able to access relevant information no matter where their records are stored by ML-powered systems that integrate multiple data sources.
Enhanced security and privacy	Algorithms based on ML can be used to encrypt and protect healthcare records.
Research and innovation	ML techniques can be deployed to identify patterns, trends, and associations in healthcare records so that medical research and the development of new treatments can be facilitated.

C. Benefits for Policymakers

The findings of this study suggest healthcare policymakers should consider several important factors when designing and implementing healthcare policies in the future: increased efficiency, enhanced data security, personalized care, resource optimization, and integration with other healthcare technologies. Making informed decisions for better health outcomes can be achieved by applying ML techniques to optimize healthcare resources, identify healthcare disparities, and enhance healthcare outcomes. Moreover, ML algorithms can identify improvements that need to be made besides monitoring compliance. Among the significant impacts of this study is that policymakers can optimize the distribution of healthcare resources. Utilizing ML algorithms, policymakers can analyze huge amounts of healthcare data to identify patterns and trends, enabling them to determine the optimum use of resources, such as hospital beds, medical equipment, and healthcare professionals. When policymakers are given the ability to employ these tools in a more effective manner, resources can be allocated to maximize patient outcomes while minimizing cost. Policymakers can also use ML techniques to identify quality disparities among patients.

IV. CONCLUSION

The need to manage and organize healthcare information is becoming increasingly important as medical data grow exponentially. Through identifying the challenges, advantages, and limitations associated with ML, this study focused on how ML technology can be implemented to optimize healthcare record management. Therefore, the study contributes to the understanding of how ML can be applied in healthcare settings. The findings of this study furnish meaningful input that can help healthcare professionals, researchers, and policymakers make sound decisions regarding the employment of ML techniques for retaining and handling healthcare records. As a result of discussing the ML usage to better regulate patient data, this research may ultimately produce enhanced patient results and healthcare provision. Since ML engages different perspectives and an efficient approach, it was found advantageous in dealing with healthcare records. Future work could be concentrated on developing a comprehensive system for arranging and systematizing healthcare records utilizing ML.

ACKNOWLEDGMENT

This work was funded by the University of Jeddah, Jeddah, Saudi Arabia, under grant No. (UJ-23-DR-148). Therefore, the authors thank the University of Jeddah for its technical and financial support.

REFERENCES

- [1] K. C. Rath, A. Khang, and D. Roy, "The Role of Internet of Things (IoT) Technology in Industry 4.0 Economy," in *Advanced IoT Technologies and Applications in the Industry 4.0 Digital Economy*, 1st ed., Boca Raton, FL, USA: CRC Press, 2024, pp. 1–28.
- [2] A. E. Yahya, A. Gharbi, W. M. S. Yafooz, and A. Al-Dhaqm, "A Novel Hybrid Deep Learning Model for Detecting and Classifying Non-Functional Requirements of Mobile Apps Issues," *Electronics*, vol. 12, no. 5, Jan. 2023, Art. no. 1258, <https://doi.org/10.3390/electronics12051258>.
- [3] W. A. H. Altowayti *et al.*, "The Role of Conventional Methods and Artificial Intelligence in the Wastewater Treatment: A Comprehensive Review," *Processes*, vol. 10, no. 9, Sep. 2022, Art. no. 1832, <https://doi.org/10.3390/pr10091832>.
- [4] S. Messinis, N. Temenos, N. E. Protonotarios, I. Rallis, D. Kalogeras, and N. Doulamis, "Enhancing Internet of Medical Things security with artificial intelligence: A comprehensive review," *Computers in Biology and Medicine*, vol. 170, Mar. 2024, Art. no. 108036, <https://doi.org/10.1016/j.combiomed.2024.108036>.
- [5] M. Rasool, N. A. Ismail, A. Al-Dhaqm, W. M. S. Yafooz, and A. Alsaedi, "A Novel Approach for Classifying Brain Tumours Combining a SqueezeNet Model with SVM and Fine-Tuning," *Electronics*, vol. 12, no. 1, Jan. 2023, Art. no. 149, <https://doi.org/10.3390/electronics12010149>.
- [6] H. Askr, E. Elgeldawi, H. Aboul Ella, Y. A. M. M. Elshaier, M. M. Gomaa, and A. E. Hassanien, "Deep learning in drug discovery: an integrative review and future challenges," *Artificial Intelligence Review*, vol. 56, no. 7, pp. 5975–6037, Jul. 2023, <https://doi.org/10.1007/s10462-022-10306-1>.
- [7] K. B. Vikhyath and N. A. Prasad, "Combined Osprey-Chimp Optimization for Cluster Based Routing in Wireless Sensor Networks: Improved DeepMaxout for Node Energy Prediction," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12314–12319, Dec. 2023, <https://doi.org/10.48084/etasr.6542>.
- [8] Y. Zhang and Z. Dong, "Medical Imaging and Image Processing," *Technologies*, vol. 11, no. 2, Apr. 2023, Art. no. 54, <https://doi.org/10.3390/technologies11020054>.
- [9] Z. Ji *et al.*, "Lung Nodule Detection in Medical Images Based on Improved YOLOv5s," *IEEE Access*, vol. 11, pp. 76371–76387, 2023, <https://doi.org/10.1109/ACCESS.2023.3296530>.
- [10] C. Wang, X. Lv, M. Shao, Y. Qian, and Y. Zhang, "A novel fuzzy hierarchical fusion attention convolution neural network for medical image super-resolution reconstruction," *Information Sciences*, vol. 622, pp. 424–436, Apr. 2023, <https://doi.org/10.1016/j.ins.2022.11.140>.
- [11] B. D. Katzman, C. B. van der Pol, P. Soyer, and M. N. Patlas, "Artificial intelligence in emergency radiology: A review of applications and possibilities," *Diagnostic and Interventional Imaging*, vol. 104, no. 1, pp. 6–10, Jan. 2023, <https://doi.org/10.1016/j.diii.2022.07.005>.
- [12] P. Chakraborty, T. Chandraprasam, A. Arunachalam, and S. Rafiammal, "Artificial Intelligence-based Oral Cancer Screening System using Smartphones: Oral cancer screening system," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12054–12057, Dec. 2023, <https://doi.org/10.48084/etasr.6364>.
- [13] A. Salau, N. A. Nwojo, M. M. Boukar, and O. Usen, "Advancing Preauthorization Task in Healthcare: An Application of Deep Active Incremental Learning for Medical Text Classification," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12205–12210, Dec. 2023, <https://doi.org/10.48084/etasr.6332>.
- [14] J. F. Wolfswinkel, E. Furtmueller, and C. P. M. Wilderom, "Using grounded theory as a method for rigorously reviewing literature," *European Journal of Information Systems*, vol. 22, no. 1, pp. 45–55, Jan. 2013, <https://doi.org/10.1057/ejis.2011.51>.
- [15] I. U. Onwuegbuzie, S. A. Razak, I. F. Isnin, T. S. J. Darwish, and A. Al-dhaqm, "Optimized backoff scheme for prioritized data in wireless sensor networks: A class of service approach," *PLOS ONE*, vol. 15, no. 8, Jul. 2020, Art. no. e0237154, <https://doi.org/10.1371/journal.pone.0237154>.
- [16] H. Shamshad, F. Ullah, A. Ullah, V. R. KEBANDE, S. Ullah, and A. Al-Dhaqm, "Forecasting and Trading of the Stable Cryptocurrencies With Machine Learning and Deep Learning Algorithms for Market Conditions," *IEEE Access*, vol. 11, pp. 122205–122220, 2023, <https://doi.org/10.1109/ACCESS.2023.3327440>.
- [17] M. Q. Mohammed *et al.*, "Review of Learning-Based Robotic Manipulation in Cluttered Environments," *Sensors*, vol. 22, no. 20, Jan. 2022, Art. no. 7938, <https://doi.org/10.3390/s22207938>.
- [18] B. Alsinglawi *et al.*, "An explainable machine learning framework for lung cancer hospital length of stay prediction," *Scientific Reports*, vol. 12, no. 1, Jan. 2022, Art. no. 607, <https://doi.org/10.1038/s41598-021-04608-7>.
- [19] A. Hussain, K. Farooq, B. Luo, and W. Slack, "A Novel Ontology and Machine Learning Inspired Hybrid Cardiovascular Decision Support Framework," in *IEEE Symposium Series on Computational Intelligence*, Cape Town, South Africa, Dec. 2015, pp. 824–832, <https://doi.org/10.1109/SSCI.2015.122>.
- [20] L. S. Kumar and A. Padmapriya, "Disease Information Extraction from Healthcare Records using CTA Matrix," *Australian Journal of Basic and Applied Sciences*, vol. 10, no. 2, pp. 141–149, 2016.
- [21] N. V. Pardakhe and V. M. Deshmukh, "Machine Learning and Blockchain Techniques Used in Healthcare System," in *IEEE Pune Section International Conference*, Pune, India, Dec. 2019, pp. 1–5, <https://doi.org/10.1109/PuneCon46936.2019.9105710>.
- [22] P. Pandey and R. Litoriya, "Securing and authenticating healthcare records through blockchain technology," *Cryptologia*, vol. 44, no. 4, pp. 341–356, Jul. 2020, <https://doi.org/10.1080/01611194.2019.1706060>.
- [23] A. W. Kempa-Liehr *et al.*, "Healthcare pathway discovery and probabilistic machine learning," *International Journal of Medical Informatics*, vol. 137, May 2020, Art. no. 104087, <https://doi.org/10.1016/j.ijmedinf.2020.104087>.
- [24] A. Taylor, R. Kleiman, S. Hebbing, P. Peissig, and D. Page, "High-Throughput Approach to Modeling Healthcare Costs Using Electronic Healthcare Records," *arXiv*, Jun. 01, 2022, <https://doi.org/10.48550/arXiv.2011.09497>.
- [25] A. Datta *et al.*, "'Black Box' to 'Conversational' Machine Learning: Ondansetron Reduces Risk of Hospital-Acquired Venous Thromboembolism," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 6, pp. 2204–2214, Jun. 2021, <https://doi.org/10.1109/JBHI.2020.3033405>.
- [26] K. P. Arjun and K. S. Kumar, "Machine Learning -A Neoteric Medicine to Healthcare," *International Journal on Emerging Technologies*, vol. 11, no. 3, pp. 195–201, May 2020.
- [27] P. K. Yeng, M. Ali Fauzi, and B. Yang, "Comparative analysis of machine learning methods for analyzing security practice in electronic health records' logs," in *IEEE International Conference on Big Data*, Atlanta, GA, USA, Dec. 2020, pp. 3856–3866, <https://doi.org/10.1109/BigData50022.2020.9378353>.
- [28] O. Bardhi and B. Garcia Zapirain, "Machine Learning Techniques Applied to Electronic Healthcare Records to Predict Cancer Patient Survivability," *Computers, Materials & Continua*, vol. 68, no. 2, pp. 1595–1613, 2021, <https://doi.org/10.32604/cmc.2021.015326>.
- [29] P. Papadopoulos, W. Abramson, A. J. Hall, N. Pitropakis, and W. J. Buchanan, "Privacy and Trust Redefined in Federated Machine Learning," *Machine Learning and Knowledge Extraction*, vol. 3, no. 2, pp. 333–356, Jun. 2021, <https://doi.org/10.3390/make3020017>.
- [30] M. H. Chaitra and S. Vagdevi, "A Detailed Survey Study on Various Issues and Techniques for Security and Privacy of Healthcare Records," in *Intelligent Sustainable Systems*, J. S. Raj, R. Palanisamy, I. Perikos, and Y. Shi, Eds. New York, NY, USA: Springer, 2022, pp. 181–189.

- [31] S. Dutta and S. K. Bandyopadhyay, "Diabetes Prediction Using Machine Learning Approaches," in *Advanced Prognostic Predictive Modelling in Healthcare Data Analytics*, S. Roy, L. M. Goyal, and M. Mittal, Eds. New York, NY, USA: Springer, 2021, pp. 179–202.
- [32] D. K. Sharma, D. S. Chakravarthi, R. S. K. Boddu, A. Madduri, M. R. Ayyagari, and Md. Khaja Mohiddin, "Effectiveness of Machine Learning Technology in Detecting Patterns of Certain Diseases Within Patient Electronic Healthcare Records," in *Second International Conference in Mechanical and Energy Technology*, Greater Noida, India, Oct. 2021, pp. 73–81, https://doi.org/10.1007/978-981-19-0108-9_8.
- [33] A. Haddad, M. H. Habaebi, Md. R. Islam, N. F. Hasbullah, and S. A. Zabidi, "Systematic Review on AI-Blockchain Based E-Healthcare Records Management Systems," *IEEE Access*, vol. 10, pp. 94583–94615, 2022, <https://doi.org/10.1109/ACCESS.2022.3201878>.
- [34] Z. Wang and J. Sun, "PromptEHR: Conditional Electronic Healthcare Records Generation with Prompt Learning," arXiv, Oct. 11, 2022, <https://doi.org/10.48550/arXiv.2211.01761>.
- [35] M. Kumar, S. Singhal, S. Shekhar, B. Sharma, and G. Srivastava, "Optimized Stacking Ensemble Learning Model for Breast Cancer Detection and Classification Using Machine Learning," *Sustainability*, vol. 14, no. 21, Jan. 2022, Art. no. 13998, <https://doi.org/10.3390/su142113998>.
- [36] X. Sun, A. Douiri, and M. Gulliford, "Applying machine learning algorithms to electronic health records to predict pneumonia after respiratory tract infection," *Journal of Clinical Epidemiology*, vol. 145, pp. 154–163, May 2022, <https://doi.org/10.1016/j.jclinepi.2022.01.009>.
- [37] N. Chen *et al.*, "Evaluating the risk of hypertension in residents in primary care in Shanghai, China with machine learning algorithms," *Frontiers in Public Health*, vol. 10, Oct. 2022, Art. no. 984621, <https://doi.org/10.3389/fpubh.2022.984621>.
- [38] A. Ayub Khan *et al.*, "Healthcare Ledger Management: A Blockchain and Machine Learning-Enabled Novel and Secure Architecture for Medical Industry," *Human-centric Computing and Information Sciences*, vol. 12, Nov. 2022, Art. no. 55, <https://doi.org/10.22967/HCIS.2022.12.055>.
- [39] D. Tenepalli and N. Thandava Meganathan, "A Review on Machine Learning and Blockchain Technology in E-Healthcare," in *22nd International Conference on Intelligent Systems Design and Applications*, Dec. 2022, pp. 338–349, https://doi.org/10.1007/978-3-031-35510-3_33.
- [40] V. Kumawat, B. Umamaheswari, P. Mitra, and G. Lavania, "Machine Learning for Health Care: Challenges, Controversies, and Its Applications," in *Soft Computing: Theories and Applications*, R. Kumar, C. W. Ahn, T. K. Sharma, O. P. Verma, and A. Agarwal, Eds. New York, NY, USA: Springer, 2022, pp. 253–261.
- [41] E. S. Tumpa and K. Dey, "A Review on Applications of Machine Learning in Healthcare," in *6th International Conference on Trends in Electronics and Informatics*, Tirunelveli, India, Apr. 2022, pp. 1388–1392, <https://doi.org/10.1109/ICOEI53556.2022.9776844>.
- [42] T. J. Banks, T. D. Nguyen, J. K. Uhlmann, S. S. Nair, and J. F. Scherrer, "Predicting opioid use disorder before and after the opioid prescribing peak in the United States: A machine learning tool using electronic healthcare records," *Health Informatics Journal*, vol. 29, no. 2, Apr. 2023, Art. no. 14604582231168826, <https://doi.org/10.1177/14604582231168826>.
- [43] S. Gupta, G. F. Nama, and S. Deivasigamani, "Real-Time Monitoring of Patient Activity Using IoT and Machine Learning in Healthcare," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 7s, pp. 51–57, Jul. 2023.
- [44] S. Rani, P. Kumar Pareek, J. Kaur, M. Chauhan, and P. Bhambri, "Quantum Machine Learning in Healthcare: Developments and Challenges," in *International Conference on Integrated Circuits and Communication Systems*, Raichur, India, Feb. 2023, pp. 1–7, <https://doi.org/10.1109/ICICACS57338.2023.10100075>.
- [45] G. Parashar, A. Chaudhary, and D. Pandey, "Machine Learning for Prediction of Cardiovascular Disease and Respiratory Disease: A Review," *SN Computer Science*, vol. 5, no. 1, Jan. 2024, Art. no. 196, <https://doi.org/10.1007/s42979-023-02529-y>.
- [46] S. S. Saranya, P. Anusha, S. Chandragandhi, O. Kiran Kishore, N. Phani Kumar, and K. Srihari, "Enhanced decision-making in healthcare cloud-edge networks using deep reinforcement and lion optimization algorithm," *Biomedical Signal Processing and Control*, vol. 92, Jun. 2024, Art. no. 105963, <https://doi.org/10.1016/j.bspc.2024.105963>.

A Systematic Literature Review on Construction Management Productivity Enhancement by utilizing Business Information Modeling

Abd Alrazaq Khamees Saja

Civil Engineering Department, College of Engineering, Baghdad University, Iraq
mqsqrrt2@gmail.com (corresponding author)

Rasheed Mohammed Sawsan

Faculty of Civil Engineering, College of Engineering, Baghdad University, Iraq
dr.sawsan.r@coeng.uobaghdad.edu.iq

Received: 10 February 2024 | Revised: 27 February 2024 | Accepted: 2 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7048>

ABSTRACT

The systematic review of Business Information Modeling (BIM) plays a crucial role in understanding its significance and impact. This review allows for a comprehensive examination of the existing literature, highlighting the benefits, challenges, and success factors associated with BIM. There is a scarcity of studies dealing with this subject, and so a question about the most important advantages that will be obtained by the construction industry, especially the construction companies, as a result of the BIM application arises. Relevant previous studies were reviewed and their quality was evaluated using a systematic methodology. The current study was characterized by the use of the SWOT (Strengths, Weaknesses, Opportunities, and Threats) analysis technique. As far as is known, the current study is the first of its kind in the field of Iraqi project management. The results suggest that the BIM benefits include firm's growth, organizational performance, enhanced market value, employee motivation, and service quality.

Keywords-Business Information Modeling (BIM); construction industry; productivity improvement; systematic literature review

I. INTRODUCTION

A business information model is a systematic and visual depiction of a construction organization's business processes, entities, and information flows. Business Information Modeling (BIM) actively encourages the integration, standardization, and interoperability of information across diverse systems, creates a central warehouse with features for data entry, formatting, editing, storing, and updating throughout a project's life cycle to support businesses in implementing appropriate information management and guaranteeing that all data are available to support projects' proper implementation [1]. Aligned with the strategic goals of the construction organization, it fosters collaboration and facilitates continuous improvement. BIM is pivotal in comprehending, overseeing, and leveraging business information to enhance performance and attain desired outcomes in the construction sector. The identification, ranking, and allocation of the major risk variables affecting school construction projects in Iraq were conducted in [2] through a systematic examination of the literature. The BIM methodology addresses the design and development of data-intensive information technology and governing data by

introducing a semantic business information model as a central point of reference. Creating and capturing value can be simply described as one of the BIM functions because information systems tend to offer management tools and support all users in their activities. Their use has become commonplace and an integral component of every activity [3]. Authors in [4] explored the BIM applications in the field of engineering management. Various scientific research steps were used and multiple sources of information were investigated, to review studies and research related to BIM and its importance and assets.

II. METHODOLOGY

This study acts as an addition to the scientific engineering management literature that is related to the BIM subject in terms of its impact and the gains it provides the construction sector with. Given that literature still requires further research on identifying the most crucial benefits emerging from the BIM application in the construction sector, this study aims to provide this type of exploration. Increasing productivity is a top priority for all businesses and organizations across the globe.

However, it is difficult to give a precise definition of the former or an algorithm to gauge its rise or fall because there are many variables and metrics to consider. Work scheduling directly affects productivity [5]. Enhancing productivity concerns is essential to the success of project management in the construction sector [6]. Previous studies indicate that there is an urgent need for more analytical and exploratory research on regulating the advantages, success factors, or the improved productivity caused by the application or non-application of BIM to the construction sector. The specific uses of BIM may differ based on the industry size and type. BIM allows businesses to better manage their information assets, streamline operations, and drive innovation and competitiveness. The current systematic review is based on answering the question raised in this study as well as on identifying, evaluating, and summarizing relevant previous studies using a systematic methodology. Within the framework of the current study is to explore and determine the most important gains that construction companies will obtain from the application of BIM. The systematic review process employed in this study went through several stages, which are analyzed below.

A. Identifying Relevant Literature

This step consists of identifying the basic databases, their sources, periods, terms, and keywords used when searching for literature related to the topic (benefits of BIM in improving productivity in the construction sector). With this regard, a preliminary survey was conducted and SciSpace [7] was finally chosen as the source database.

B. Primary Research Question

There is a limited number of studies that investigate the subject of BIM employment and its benefits regarding productivity. Understanding the significance of the productivity of employees in the building industry, may result in time and

money savings [8]. The raised research question is: What is the impact of BIM in developing construction management and ameliorating productivity?

C. Topics to Include

After the basic database has been identified, the criteria that will be adopted in the inclusion of research within the literature must be determined. Among these criteria the following keywords for searching on the database can be found: Practice Management, Productivity Improvement, Business Information Modeling, Construction Information Management, Business Information Systems. The considered language was English and the publication period was confined between 2007 and 2023.

D. Topics to Exclude

A research protocol must be developed before the execution of any systematic literature review to identify exclusion topics such as, repetition in titles, no relative subject, generic BIM studies, aiming to focus on studies directly addressing BIM's influence on practice management and productivity.

III. DATA EXTRACTION AND ANALYSIS

After identifying the relevant studies, data must be extracted from them. The main purpose of the current analysis is to integrate the findings derived from different studies to answer the research question, rather than constructing a mere collection of previous researches. At this stage, earlier studies were collected from SciSpace. A total of 27 studies were initially considered. After duplicate (6) and irrelevant (10) study removal, 11 studies were reviewed. In the last stage of audit, it was found that one of them was repeated, so it was excluded. Therefore, 10 past studies were finally contemplated in total. The summary of the reviewed previous studies can be noticed in Table I.

TABLE I. REVIEW SUMMARY

Ref.	Objectives	Results
[9]	BIM for process integration in the mold-making industry is designed. To accomplish smooth information flows and business process automation, crucial business processes must be integrated and a process-oriented business information model is suggested.	The system's implementation and advantages are presented. The results may help mold manufacturing businesses to accomplish business process automation and concurrency, as well as business process integration, streamlining, and maximizing information exchange.
[10]	The study proposes an analytical framework for modeling business processes to achieve both operational and goal integrity.	The use of BIM for structuring and formalizing business needs in big data-intensive projects is proposed. The assessment of BIM is based on a sizable banking data warehouse project. The study also demonstrates how BIM can be utilized in large data landscapes and data science activities, demonstrating its suitability outside traditional setups.
[11]	A technique for modeling the dynamics and structure of intricate corporate information supply chains using a process approach is presented.	The simulation software package comprises an execution engine, a knowledge base, a process library, and a model database. The study offers a useful tool for risk analysis, planning, optimization, assessment, and management in commercial information supply chains.
[12]	BIM is explored regarding data warehouses and big data governance	BIM is suitable for data warehouse, big data, and data science projects.
[13]	The study emphasizes how crucial project management abilities are to IT initiatives, especially those involving business informatics.	To complete a project, excellent communication skills and the application of best practices and project management concepts are essential.
[14]	A BIM method for EIS is proposed.	The proposed method helps EIS development teams gain comprehensive knowledge.
[15]	The proposed Early-warning Performance Monitoring System (EPMS) offers an alternative method for objective performance monitoring and forecasting by utilizing project progress data, a theoretical model, and an index for performance monitoring and forecasting.	Changes in the project execution environment can affect performance. The EPMS was successfully applied in a Korean construction project. It offered a way to objectively monitor and forecast project performance, even in circumstances data on performance indicators were lacking.

[16]	Project management techniques are used more often in modern industry to deliver work packages in a more regulated and cost-effective manner while maximizing the utilization of human resources to satisfy client demands and gain a competitive edge. The behavioral competencies of project, program, and portfolio managers as well as the significance of information modeling and technology in project management are highlighted. The difficulties and peculiarities of the industry project management in Croatia are addressed.	Leadership is regarded as the most crucial behavioral project management competency, with teamwork and self-management following. Relationships and engagement, conflict and crisis, negotiation, and resourcefulness have less effect on the successful completion of a project.
[17]	This study examines how information systems, Human Resource Management (HRM), and decision-making impact competitiveness and sustainable corporate performance in a transitional economy.	Decision-making and HRM have a moderately beneficial impact on business performance, but information systems have a large positive influence. The study also discovered a favorable relationship between competitiveness and sustainable business performance. The study sheds light on the variables that may affect competitiveness and sustainable corporate performance in a transitional economy.
[18]	The study's objectives were to determine the advantages of methodical Construction Information Management (CIM) and how it may inspire building firms.	Company expansion, organizational performance, increased market value, employee motivation, and high-quality services are among the advantages of CIM. The use of modern technology in CIM is also emphasized as a means of improving building design, productivity, and project planning. Construction businesses should use a well-organized CIM system to increase dedication to the project, enhance professional collaboration among workers, and provide clarity on the information required at each project stage.

IV. RESULTS AND DISCUSSION

Similarities and differences are the essence of a systematic review. The goals of such a review are to summarize the results of previous relevant studies, to evaluate them using a clear logical structure, and to review these findings in an unbiased, organized, clear, and direct manner.

A. Similarities

The earlier reviewed studies bear some resemblance to our work, which will be presented in another paper. This resemblance can be detected in some of the statistical methods used.

- The considered studies used the descriptive approach due to its suitability to the nature of the problem. Often questionnaires were used as a means of collecting data.
- All studies demonstrated the positive impact of BIM and information management on the construction sector due to their ability to enhance and improve productivity.

- Some past studies attempted to discover the degree of stakeholders' commitment to professional and scientific standards for the application of BIM or information management in their companies and projects.

B. Differences

- This study represents the research community in the Iraqi building and construction sector, and is concentrated on the Ministry of Construction, Housing and Municipalities in the Republic of Iraq. The latter's tasks are to supervise the implementation of infrastructure projects.
- The research conducted was characterized by the employment of the SWOT analysis technique in assessing the reality of BIM, with identifying and diagnosing strengths, weaknesses, opportunities, and challenges. As far as is known, the current study is the first one coping with this topic in the field of Iraqi project management.

Table II illustrates a comparison between the current study and previous ones in terms of study objectives, study location, study limits, and the research gap we aim to cover.

TABLE II. COMPARISON WITH PREVIOUS STUDIES

Field	Previous studies	Our work
Place	Most earlier research was carried out in nations with a history in BIM and information management, particularly in Europe.	This study was conducted in the Republic of Iraq, which only recently began to apply the concept of BIM in the construction sector.
Subjects	Previous studies dealt with the concept of BIM and information management within a general holistic framework and focused on public projects such as IT projects.	This study focuses on an in-depth evaluation and analysis of the performance of BIM in the Iraqi construction sector, which requires specificity in the analysis, taking into account the size of projects, the scarcity of local technical resources, and the nature of infrastructure projects.
Objectives	Earlier studies focused on reviewing the benefits of information management and BIM and their relationship to the success or failure of companies or public projects. The performance evaluation was limited in IT projects.	An effort to highlight the role of BIM within construction institutions is made due to the importance of BIM in project management and enhanced productivity. The ability of BIM to achieve integration management and effective communication between the technical and administrative companies, its importance in overcoming risks, and the increasing productivity it brings are considered.

C. Benefits from Previous Works

- New ideas were generated about the method of scientific research by examining the theoretical framework of similar studies and their research methods. The literature review helped defining the research problem of this study.

Knowledge of the findings of past researches in the field of BIM was gained and was utilized in defining the research gap problem and the dimensions and fields of work. The particular study is not a repetition of previous ideas, but a renewal of knowledge, and an attempt to reach originality.

- Several sectors that can be benefited from our work were identified.
- The performance of BIM in construction companies was evaluated along with the main causes for its weak application and its role in the Iraqi construction sector by using the Root Cause Analysis Technique (RCAT).
- This work is a new addition to the field of knowledge in project management, as it addresses the basic points of BIM in a new and more in-depth manner.

V. CONCLUSION

The most important results of earlier studies related to this work were summarized and a set of conclusions were reached regarding the most crucial gains expected from the BIM application in construction companies to diagnose its impact and effectiveness on improving the competence of the construction and building sectors.

The correct application of BIM results in completing projects on time and in reducing the previously allocated costs. The former also allows continuous follow-up, reporting to senior management, and preparation of standard specifications, methodologies, and models for projects.

BIM needs human knowledge and technical capabilities to be combined and thus support the stakeholders' participation and enhance their understanding of the relationship between knowledge-sharing behaviors and of how they correspond to the functions of administrative units in companies.

This work attempts to close the research gap of identifying in practice BIM's important contribution to management and productivity improvement and to the success of the construction companies in the Iraqi construction industry.

Company expansion, organizational performance, increased market value, employee motivation, and high quality of service are all advantages generated from BIM employment.

REFERENCES

- [1] M. A. A. A.-A. Noaman and S. R. Mohammed, "Application Innovation Strategy For Digital Maintenance Management Of School Building In Iraq," *Journal of Positive School Psychology*, pp. 2023–2035, Jun. 2022.
- [2] H. R. Abed and H. A. Rashid, "Empirical Study for Capturing and Allocating Significant Risk Factors in School Construction Projects in Iraq," *Journal of Engineering*, vol. 29, no. 12, pp. 81–103, Dec. 2023, <https://doi.org/10.31026/j.eng.2023.12.06>.
- [3] M. Laaziri, S. Khouliji, K. Benmoussa, and K. M. Larbi, "Information System for the Governance of University Cooperation," *Engineering, Technology & Applied Science Research*, vol. 8, no. 5, pp. 3355–3359, Oct. 2018, <https://doi.org/10.48084/etasr.2156>.
- [4] K. R. Erzaij, W. A. Hatem, and B. H. Maula, "Applying Intelligent Portfolio Management to the Evaluation of Stalled Construction Projects," *Open Engineering*, vol. 10, no. 1, pp. 552–562, Jan. 2020, <https://doi.org/10.1515/eng-2020-0064>.
- [5] D. Pylarinos, "Investigating the Effect on Productivity of a Geospatial Ticket Management System for Power Distribution Network Studies," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11616–11621, Oct. 2023, <https://doi.org/10.48084/etasr.6202>.
- [6] M. H. Momade and M. R. Hainin, "Identifying Motivational and Demotivational Productivity Factors in Qatar Construction Projects," *Engineering, Technology & Applied Science Research*, vol. 9, no. 2, pp. 3945–3948, Apr. 2019, <https://doi.org/10.48084/etasr.2577>.
- [7] "AI Chat for scientific PDFs," *SciSpace*. <https://typeset.io>.
- [8] R. H. Fatah and J. Paslawski, "Factors Affecting Labor Productivity on Construction in Kurdistan of Iraq: Web Survey," *Journal of Engineering*, vol. 29, no. 01, pp. 14–41, Jan. 2023, <https://doi.org/10.31026/j.eng.2023.01.02>.
- [9] Q. Ni, W. F. Lu, P. K. D. V. Yarlagadda, and X. Ming, "Business information modeling for process integration in the mold making industry," *Robotics and Computer-Integrated Manufacturing*, vol. 23, no. 2, pp. 195–207, Apr. 2007, <https://doi.org/10.1016/j.rcim.2005.12.006>.
- [10] R. Khurana and V. V. Mandke, "Business process modeling with information integrity," *Business Process Management Journal*, vol. 15, no. 4, pp. 487–503, Jan. 2009, <https://doi.org/10.1108/14637150910975507>.
- [11] Y. Liu, J. Zhang, S. Sun, and H. Huang, "Business information supply chain modeling and simulation methodology," presented at the Proceedings of 2012 International Conference on Modelling, Identification and Control, Jan. 2012, pp. 339–344.
- [12] T. Priebe and S. Markus, "Business information modeling: A methodology for data-intensive projects, data science and big data governance," in *2015 IEEE International Conference on Big Data (Big Data)*, Santa Clara, CA, USA, Jul. 2015, pp. 2056–2065, <https://doi.org/10.1109/BigData.2015.7363987>.
- [13] M. H. B. Afzal and L. D. Cravens, "Utilization of Project Management and Soft Skills in Business Informatics Projects," *Journal of Advanced Management Science*, pp. 109–112, 2018, <https://doi.org/10.18178/joams.6.2.109-112>.
- [14] C. J. Montilva and A. J. Barrios, "BMM: A Business Modeling Method For Information Systems Development," *CLEI Electronic Journal*, vol. 7, no. 2, pp. 3:1-3:19, 2004, <https://doi.org/10.19153/cleiej.7.2.3>.
- [15] C.-W. Kim, W. S. Yoo, H. Lim, I. Yu, H. Cho, and K.-I. Kang, "Early-warning performance monitoring system (EPMS) using the business information of a project," *International Journal of Project Management*, vol. 36, no. 5, pp. 730–743, Jul. 2018, <https://doi.org/10.1016/j.ijproman.2018.03.010>.
- [16] R. D. Vlahov, M. Klindžić, and M. Radujković, "Information Modeling of Behavioral Project Management Competencies," *Information Technologies and Learning Tools*, vol. 69, no. 1, pp. 186–197, Feb. 2019, <https://doi.org/10.33407/itlt.v69i1.2713>.
- [17] N. Djalil, M. Nikolic, M. Bakator, and Z. Erceg, "Modeling the Influence of Information Systems on Sustainable Business Performance and Competitiveness," *Sustainability*, vol. 13, no. 17, Jan. 2021, Art. no. 9619, <https://doi.org/10.3390/su13179619>.
- [18] P. Adekunle, C. Aigbavboa, O. Akinradewo, A. Oke, and D. Aghimien, "Construction Information Management: Benefits to the Construction Industry," *Sustainability*, vol. 14, no. 18, Jan. 2022, Art. no. 11366, <https://doi.org/10.3390/su141811366>.

Two Proposed Models for Face Recognition: Achieving High Accuracy and Speed with Artificial Intelligence

Hind Moutaz Al-Dabbas

Department of Computer Science, College of Education for Pure Science (Ibn Al-Haitham), Iraq | University of Baghdad, Iraq

hind.moutaz@ihcoedu.uobaghdad.edu.iq, cs.20.04@grad.uotechnology.edu.iq (corresponding author)

Raghad Abdulaali Azeez

Information Technology Unit, College of Education Ibn-Rushd for Human Sciences, Iraq | University of Baghdad, Iraq

raghad.azeez@ircoedu.uobaghdad.edu.iq

Akbas Ezaldeen Ali

Department of Computer Science, University of Technology, Iraq

akbas.e.ali@uotechnology.edu.iq

Received: 2 February 2024 | Revised: 11 February 2024 and 16 February 2024 | Accepted: 19 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7002>

ABSTRACT

In light of the development in computer science and modern technologies, the impersonation crime rate has increased. Consequently, face recognition technology and biometric systems have been employed for security purposes in a variety of applications including human-computer interaction, surveillance systems, etc. Building an advanced sophisticated model to tackle impersonation-related crimes is essential. This study proposes classification Machine Learning (ML) and Deep Learning (DL) models, utilizing Viola-Jones, Linear Discriminant Analysis (LDA), Mutual Information (MI), and Analysis of Variance (ANOVA) techniques. The two proposed facial classification systems are J48 with LDA feature extraction method as input, and a one-dimensional Convolutional Neural Network Hybrid Model (1D-CNNHM). The MUCT database was considered for training and evaluation. The performance, in terms of classification, of the J48 model reached 96.01% accuracy whereas the DL model that merged LDA with MI and ANOVA reached 100% accuracy. Comparing the proposed models with other works reflects that they are performing very well, with high accuracy and low processing time.

Keywords-ANOVA; CNN; face recognition; LDA; MI

I. INTRODUCTION

Face recognition technology and biometric systems have been employed for security purposes in various applications including Human Computer Interaction (HCI), surveillance systems, and facial animation and expression [1, 2]. Facial recognition software can recognize people from a collection of images or a video stream based on personal physical or behavioral traits, face, fingerprints, iris, and voice [3, 4]. Deep Learning (DL)-based face recognition has shown outstanding results [5, 6]. Convolutional Neural Networks (CNNs), one of the most often used types of deep neural networks in computer vision applications, demonstrate a key benefit of autonomous visual feature extraction [7]. CNNs typically have multiple convolutional layers, each with multiple filters of different sizes, allowing them to learn increasingly complex visual

features as they process the input data [8, 9]. Several researchers have implemented face recognition systems with different classifiers like Machine Learning algorithms, such as Support Vector Machines (SVMs), K-Nearest Neighbor (KNN), Random Forest (RF), and Non-Negative Collaborative Representation-based Classifier (NCRC) using different feature extraction methods, namely Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Gabor Wavelet's (GW), and Canonical Correlation Analysis (CCA) [10-16]. More recently, DL has been applied, which is believed to be more accurate and vital in face recognition. The DL techniques, such as GoogleNet, AlexNet, residual neural network (ResNet), and Visual Geometry Group Network (VGGNet and VGG-16) consider CNN architecture as the classifier [17-23].

DL methodology has relatively wider acceptance, so it was also considered in this research. Depending on J48 ML and the

deep one dimensional Convolutional Neural Network Hybrid Model (1D-CNNHM), two facial recognition techniques are proposed in this paper. Involving LDA feature extraction, which is combined with two feature selection methods, Mutual Information (MI) and Analysis of Variance (ANOVA), is considered an innovative step of this work. The objectives of this research are to propose advanced, high-accurate, and fast Artificial Intelligent (AI) models for facing impersonation crimes.

II. RELATED WORK

Authors in [10] proposed a face recognition system that depends on appearance-based features that focus on the entire face image rather than local facial features. The Viola-Jones face detection method was used. Feature extraction and dimension reduction methods were applied, using PCA and LDA. Square Euclidean Distance (SED) was employed to measure the distance between two images. Authors in [13] proposed an automatic face recognition system based on features focusing on the whole image as well as local-based features utilizing the Local Binary Pattern Histogram (LBPH), PCA, and LDA. In addition, the system deployed the ML algorithms PART and J48. The results displayed high accuracy for detection and feature extraction. Authors in [15] suggested a facial recognition system based on Multi-Scale Local Mapped Patterns (MSLMPs), using the Genetic Algorithm (GA) to optimize parameters and weight matrices. To deal with difficult databases like MUCT, this technique was established on the average gray levels of the images in the database. The results obtained for the database are superior and have high accuracy. Authors in [16] suggested the KNN algorithm for recognizing faces on an ARM processor. To reach the best k-value and create proper face recognition with a low-power processor, PCA and LDA were used for feature extraction. Authors in [20] proposed combining RFID cards with facial biometry based on DL to strengthen the safety of an e-payment organization through the use of the CASIA Face-V5 and MUCT datasets in order to assess and validate three DL face authentication models, namely Dlip, VGG-16, and ResNet-50. Combining RESNET-50 and PCA features yielded the greatest results, achieving 99.90% accuracy and 0.08% EERs on MUCT and 99.26% accuracy and 0.75% EERs on CASIA Face V5. Authors in [22] submitted a hybrid approach based on a Modified Local Binary Patterns (MLBPs) and Layered-Recurrent Neural Network (L-RNNs) for facial recognition. Utilizing the MUCT database and various ANN performance analyses, their hybrid technique yielded a classification rate of 98%.

III. MATERIALS AND METHODS

The MUCT database [24] was selected and applied in this research [13, 23]. Sample images from the MUCT database are illustrated in Figure 1. The selected database was divided into two sets, with 70% for training and 30% for testing. Every image in the database went through some preprocessing techniques. The image's facial region was detected and cropped using the Viola-Jones method [12]. Feature extraction and feature selection were applied [25]. Finally, the output classification was created by 1D-CNNHM. A preprocessing stage is implemented as the first phase, involving the

conversion of color images to grayscale images, Histogram Equalization (HE), face detection, cropping, and resizing (Figure 2).



Fig. 1. Sample images from the MUCT database.

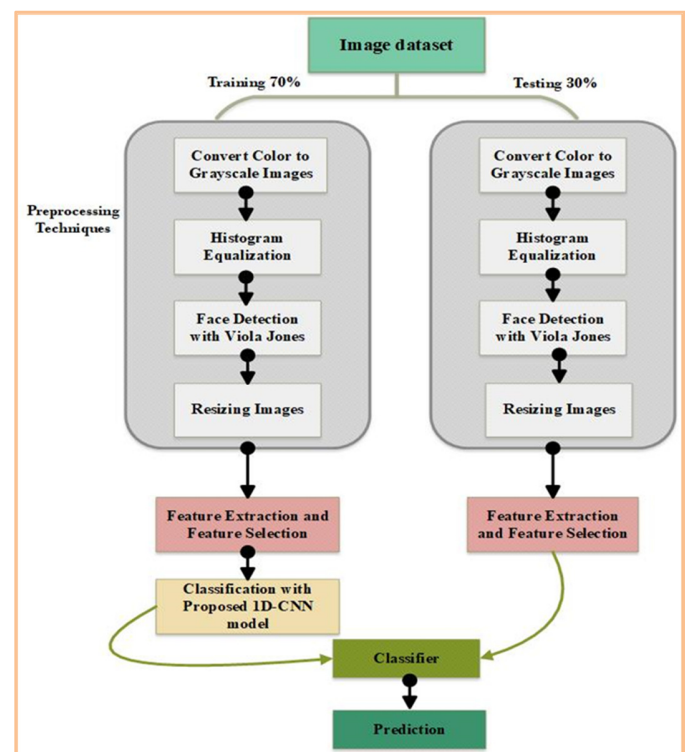


Fig. 2. The system methodology process.

A. Convert Color to Grayscale Images

The color models provide color data for each pixel in a particular image [18]. Grayscale image brightness was represented using an 8-bit value, whereas the color of a colored image's pixel is represented with a 24-bit value. The brightness in grayscale images spans from 0 to 255, with 0 intensity denoting black and 255 intensity denoting white [19, 20]. To convert the image into grayscale, the Red (R), Green (G), and Blue (B) colors should be averaged. Since each of the three hues has a unique wavelength and contributes differently to the creation of an image, the average must be calculated according to each color's contribution rather than just utilizing the average approach. This has been conducted by the luminosity method. The latter indicates that the contribution of the Red color must

be reduced, the contribution of the Green color must be increased, and the contribution of the Blue color must be placed between these two [26, 27]. Samples of the generated grayscale images are portrayed in Figure 3. The process of converting a colored to grayscale image is depicted in (1):

$$greyscale = (0.3 * R) + (0.5 * G) + (0.11 * B) \quad (1)$$



Fig. 3. Samples of grayscale images.

B. Histogram Equalization (HE)

HE is used to increase contrast in images. In order to improve low contrast images' quality and face recognition capabilities, it is common practice to distribute the most frequent intensity values evenly [27]. The image's dynamic range (contrast range) is altered as a result, making some crucial face features more noticeable [24]. Samples of images after HE are illustrated in Figure 4.



Fig. 4. Image samples after applying HE.

C. Face Detection with Viola Jones Algorithm

Viola-Jones algorithm was deployed for object detection due to its high accuracy and speed. Viola-Jones method consists of four concepts [23, 25], which are analyzed below.

1) Haar Features

The entire image is divided into $M \times M$ -sized squares or small windows. The characteristics of each window are selected separately. For face detection, three different feature types are commonly used: two-rectangle, three-rectangle, and four-rectangle features. The difference in the sums of the pixels within two rectangular sections is known as the "two-rectangle" characteristic. These rectangular areas are near to one another, either horizontally or vertically and have the same size and shape [23]. The three-rectangle feature adds the sum of the pixels in the middle rectangle to the sums of the pixels in the two outside rectangles. The difference between diagonal pairs of rectangles is computed in the four-rectangle feature, which is the final step [28]. These features are displayed in Figure 5.



Fig. 5. Haar features on the detected face.

2) Integral Image

An intermediate image representation known as an integral image was used to quickly calculate Haar-like features. Equation (2) shows the calculation formula. The integral image at position x, y contains the sum of the pixels above and to the left of (x, y) [29].

$$p(x, y) = \sum_{s=1}^x \sum_{t=1}^y I(s, t); 1 \leq x \leq M, 1 \leq y \leq M \quad (2)$$

where p denotes the integral image and I the original image.

3) Adaptive Boosting (Adaboost)

The Adaboost algorithm shrinks the size of a large set of features by removing pointless ones [30]. This method can replace thousands of features with hundreds or less. Such features are referred to as weak classifiers [26, 31].

4) Cascading

The discovered faces must proceed through a number of cascaded phases before finding the positive windows [32]. A classifier is created for each stage utilizing a few features. Every level after that adds more features, increasing the complexity of the classifier. Every stage has the option of rejecting or moving on to the next within the discovered region. Therefore, only the area that successfully completes all stages is categorized as a face [29]. After the Viola-Jones method was used, the images were in different sizes, so each image was resized employing bicubic interpolation into 100×100 size and then was cropped. The samples of the images after resizing and cropping are presented in Figure 6.



Fig. 6. Images after resizing and cropping.

Bicubic interpolation is an ideal method. It would also be a good option if quality was an issue. With bicubic interpolation, a pixel's 16 nearest neighbors are taken into account [34]. The intensity value assigned to point (x, y) is determined by:

$$v(x, y) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ij} x^i y^j \quad (3)$$

D. Feature Extraction and Feature Selection

In order to reduce the complexity of space and time required for the proposed system model, the discriminative power of ANOVA and the thorough relationship assessment of MI with LDA feature extraction were combined. LDA was

merged with different percentage cases (5%, 10%, 15%, and 20%) of two feature selection methods, MI and ANOVA. It is possible to minimize the complexity of the space and the time needed for the proposed models by identifying the most pertinent features that have the most significant impact on the goal variable. This integration may improve model performance while reducing the drawbacks of each technique [35].

1) LDA Feature Extraction

The input data are transformed into a set of features during feature extraction, and the resultant reduced representation retains the majority of the pertinent data from the original data. LDA is a method of supervised linear dimensionality reduction that seeks for the subspace that distinguishes the most effective among several classes. The objective of LDA is to select an optimal solution vector as the best projection direction based on the Fisher criterion function and consider it not sensitive to light. The main LDA steps are [33]:

- Take samples for class-1 and class-2
- Calculate the class-1 and class-2 means as μ_1 and μ_2 .
- Create class-1 and class-2 covariance matrices, denoted as C_1 and C_2 , respectively.

- Compute the within-class scatter matrix

$$S_w = C_1 + C_2 \quad (4)$$

- Compute the between-class scatter matrix depending on the classes means:

$$S_B = (\mu_1 - \mu_2) * (\mu_1 - \mu_2) \quad (5)$$

- Compute all the class means.
- The basic eigenvalue problem is then solved by the LDA scheme in (6) and (7):

$$S_w^{-1} S_B W = \lambda W \quad (6)$$

$$W = \text{eig}(S_w^{-1} S_B) \quad (7)$$

where W is the projection vector.

2) ANOVA Feature Selection

ANOVA is a group of statistical models and associated estimation methods, which are used to evaluate variances in means [36]. The f ratio of the class-to-class variance over the within-class variance is the value determined by the ANOVA equations. A class separation indicator is provided by the size of the f ratio [28]. Data are preserved as features during retention times that result in an f ratio larger than a predetermined threshold, while the rest of the data are deleted [37]. The variance between classes is calculated by:

$$\sigma_{C1}^2 = \frac{\sum (\bar{x}_i - \bar{x})^2 n_i}{(k-1)} \quad (8)$$

where \bar{x}_i is the mean of the i^{th} class, \bar{x} stands for the mean, n_i is the number of measurements in the i^{th} class, N is the total sample size of the group, and k is the group number. The within-class variance is computed by:

$$\sigma_{err}^2 = \frac{\sum \sum (x_{ij} - \bar{x})^2 - (\sum (\bar{x}_i - \bar{x})^2 n_i)}{(N-k)} \quad (9)$$

where x_{ij} is the j^{th} class's i^{th} measurement. The ratio of the two variances in (10) is used to generate the ANOVA f ratio:

$$f \text{ ratio } \frac{\sigma_{C1}^2}{\sigma_{err}^2} \quad (10)$$

3) Mutual Information (MI) Feature Selection

MI is a measure of how much knowledge one random variable has about another [39]. This concept can be used to measure how relevant a feature subset is in relation to the output vector C , which is helpful in the context of feature selection. The MI is described by [38]:

$$MI(x, y) = \sum_{i=1}^n \sum_{j=1}^n p(x(i), y(j)) \cdot \log \left(\frac{p(x(i), y(j))}{p(x(i)) \cdot p(y(j))} \right) \quad (11)$$

where MI is zero when x and y are statically independent, i.e. $p(x(i), y(j)) = p(x(i)) \cdot p(y(j))$. Considering two random variables x and y , their joint probability density function is $p(x(i), y(j))$.

E. ML Classification Model with J48 Decision Tree

ML research places a lot of emphasis on developing software that can automatically recognize complicated patterns and draw conclusions from data. This algorithm creates the rules for the target variable's prediction. The properties of J48 are accounting for missing values, Decision Tree (DT) pruning, continuous attribute value ranges, and derivation of rules. The objective is progressively generalization of a DT until it gains equilibrium of flexibility and accuracy. The output of this process is a DT classifier where each node represents a decision based on the selected attribute, and leaf nodes are labeled with class labels. Algorithm 1 displays the steps of J48 algorithm [36].

Algorithm 1: Steps of J48 Algorithm

Input: Features extracted by LDA

Output: Classifier

Processing steps:

Step 1: The leaf is labeled with a similar class if the instances belong to the same class.

Step 2: For each attribute, the potential data will be generated and the gain in the data will be derived from the test on the attribute.

Step 3: The best attribute will be chosen according to the current selection parameter

F. Architecture of 1D-CNNHM

CNNs were considered for identification and classification. Filters, kernels, or neurons that can learn their weights, parameters, and biases make up CNNs [39, 40]. Each filter receives certain inputs, performs convolution, and then, if desired, adds nonlinearity [41, 42]. The structure of a CNN consists of convolutional, pooling, rectified linear unit (ReLU),

and fully connected layers [43]. The proposed 1D-CNNHM is built in 27 layers which include: 9 convolutional layers for 1D feature extraction, 6 Max-pooling 1D layers, 8 LeakyRelu layers for its ability to speed up the training of the model by reducing the slop for the negative feature values, 3 fully connected layers for collecting, and 1 flatten layer which flattens the output of the preceding layers into a single vector that can be the input for the following stage. This hybrid model produces higher predictive performance when compared to the traditional DL models. Each layer is presented with its filters, kernel size, strides (i.e. number of steps), and padding (refers to the amount of pixels added to an image when it is being processed by the kernel of a CNN). The total number of epochs is 100 and the batch size is 64. Adam learning optimizer with learning rate equal to 0.001 was employed. The model architecture is presented in Algorithm 2. The detailed layers of the proposed 1D-CNNHM architecture are depicted in Figure 7, in which K is the kernel size, F is the filter size, P is the pool size, and S is number of strides.

Algorithm 2: The architecture of the proposed 1D-CNNHM

Input: Features of images

Output: 1D-CNNHM with optimum weights

Steps

Begin

Step 1: Add block 1

Conv1D with (F=16, K=3, S=1)

MaxPooling1D with (P=1, S=1)

LeakyReLU with (alpha=0.3)

Step 2: Add block 2

Conv1D with (F=32, K=3, S=1)

MaxPooling1D with (P=1, S=1)

LeakyReLU with (alpha=0.3)

Step 3: Add block 3

Conv1D with (F=64, K=3, S=1)

MaxPooling1D with (P=1, S=1)

LeakyReLU with (alpha=0.3)

Step 4: Add block 4

Conv1D with (F=64, K=3,

strides=1)

MaxPooling1D with (P=1, S=1)

LeakyReLU with (alpha=0.3)

Dense with (F=64, activation

= "linear")

Step 5: Add block 5

Conv1D with (F=32, K=3, S=1)

MaxPooling1D with (P=1, S=1)

LeakyReLU with (alpha=0.3)

Step 6: Add block 6

Conv1D with (F=32, K=3,

strides=1)

MaxPooling1D with (P=1, S=1)

LeakyReLU with (alpha=0.3)

Dense with (F=32, activation

= "linear")

Step 7: Add block 7

Conv1D with (F=16, K=3, S=1)

LeakyReLU with (alpha=0.3)

Step 8: Add block 8

Conv1D with (F=16, K=3, S=1)

LeakyReLU with (alpha=0.3)

Step 9: Conv1D with (F=485, K=3, S=1,

activation='linear')

Step 10: Flatten ()

Step 11: Dense with (F= 276, activation=

'softmax')

End



Fig. 7. Architecture layers of the proposed 1D-CNNHM.

G. Performance Measures

Various criteria, such as accuracy, precision, recall, and F1-measure were put into service for measuring the performance of the proposed hybrid algorithms [43, 44]. In (12)-(15), *TP* stands for True Positive, *TN* for True Negative, *FP* for False Positive, and *FN* for False Negative.

- Accuracy is the percentage of instances properly classified out of all those presented. It can be calculated by:

$$\text{Accuracy} = \frac{TN+TP}{TN+FP+FN+TP} \quad (12)$$

- Precision represents the correction ratio of predicting the positive results:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (13)$$

- Recall is the correct prediction of the actual positive results:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (14)$$

- F1-measure (score) is the harmonic mean of recall and precision:

$$F_1 = \frac{1}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (15)$$

LDA applied with J48 ML achieved results with 96.01% accuracy, 96.03% precision, 95.45% recall, and 95.37% F1-measure, whereas applying 1D-CNNHM obtained results with 99.2% accuracy, and 99.3% precision, recall, and F1-measure. This means the 1D-CNNHM is considered more accurate in identifying patterns in image data compared to LDA with J48, and better at understanding and interpreting image data, proving to be them more effective for tasks like classification.

IV. RESULTS AND DISCUSSION

A. Time Performance

The time consuming result is considered to be good and low, as the processing time by using LDA feature extraction application only reached 5400 s for the whole training procedure. One epoch reached 55 s out of the 100 total epochs and 64 patch size. The training time of merging LDA feature extraction with MI and ANOVA feature selection reached 360 s. One epoch reached 3 s of 100 total epochs and 64 patch size.

B. Statistical Performance

LDA applied with J48 ML achieved results with 96.01% accuracy, 96.03% precision, 95.45% recall, and 95.37% F1-measure. 1D-CNNHM achieved results with 99.2% accuracy, and 99.3% for precision, recall, and F-measure. This means the 1D-CNNHM is considered more accurate in identifying patterns in image data compared to LDA with J48, and better at understanding and interpreting image data, making them more effective for classification tasks.

C. Merging Feature Performance

Merging 5% of ANOVA and MI with LDA obtained an impressive accuracy of 100%, while slightly reducing precision, recall, and F-measure to 99%. Merging 10% and 15% ANOVA, the performance remained highly competitive, with accuracy of 99.66% in both cases. Precision and recall

remained consistently high, between 98.33% and 99.33%, while F1-measure ranged between 99% and 99.33%. At 20% merged features, the model was able to attain perfect performance metrics across the board, with 100% accuracy, precision, recall, and F-measure. These findings show the potential of combining features from ANOVA and MI at varying percentages to enhance predictive capability and optimize model efficiency, illustrating the efficacy of feature integration strategies in enhancing model performance and robustness. Table I summarizes the results. The outputs obtained by some previous related works using the MUCT database in several feature extraction and classification methods such as PCA, LDA, Gabor wavelet, and Haar-like features with different classifiers such as SVM, KNN, NCRC, J48, GA, and Euclidian distance, had accuracy ranging from 25.56% to 94.33% [10-16]. Other researchers used classifiers such as MLBP, L-RNN, CNN, Dlip deep, VGG-16, RESNET-50, and RESNET-101, with accuracy ranging from 91.87% to 98.55% [19-23]. The two proposed systems reached very good results with high accuracy and quick processing. The 1D-CNNHM is considered to perform better than J48. Table II illustrates the comparison between the proposed algorithms and some earlier relative works.

TABLE I. RESULTS OF PROPOSED SYSTEMS

Classifier	Accuracy	Precision	Recall	F1- score
LDA + J48 ML	96.01%	96.03%	95.45%	95.37%
LDA+1D-CNNHM	99.2%	99.3%	99.3%	99.3%
LDA+ 5% (ANOVA & MI) + 1D-CNNHM	100%	99%	99%	99%
LDA+ 10% (ANOVA & MI) + 1D-CNNHM	99.66%	98.33%	99%	99%
LDA+ 15% (ANOVA & MI) + 1D-CNNHM	99.66%	98.66%	99.33%	99.33%
LDA+ 20% (ANOVA & MI) + 1D-CNNHM	100%	100%	100%	100%

TABLE II. COMPARISON BETWEEN PROPOSED SYSTEMS AND SOME PREVIOUS WORKS

Work	Feature extraction	Classifier	Accuracy
[11]	-	SVM, KNN, NCRC	25.56%, 27.87%, 77.78%
[14]	-	Random Forest SVM	83.3% 96.3%
[10]	PCA and LDA	Euclidean distance	87.5%
[16]	PCA and LDA	KNN	91.5%
[15]	MSLMP	GA	93.49%
[12]	Gabor Wavelets	SVM	93.70%
[13]	LDA	PART, J48	91.21% 94.33%
[21]	Counterfeit feature extraction	CNN with ELA	97.6%
[22]	MLBP	L-RNN with quasi-Newton back propagation	98%
[20]	PCA	Dlip deep model, VGG-16,	96.28% 99.64%
[19]	Haar-like and LPPF	ResNet50, ResNet101	91.87% 98.55%
Proposed	LDA	J48 ML	96.01%
Proposed	LDA with ANOVA and MI	1D-CNNHM	100%

V. CONCLUSIONS

High accuracy and faster performance models were proposed in this paper. LDA feature extraction method was deployed as the input of J48 machine learning model, and the novel 1D-CNNHM model, depending on LDA feature extraction with feature selection from MI and ANOVA was presented. MUCT database was considered for model training and validation. Through preprocessing techniques, the images are detected and cropped using the Viola-Jones algorithm, owing to its high degree of accuracy detection and speed. To evaluate the classification success of the suggested systems, a number of measures designed for evaluation performance including accuracy, precision, recall, and the F1-measure were utilized. Applying the LDA feature extraction with J48, the performance measure reached 96.01% accuracy, 96.03% precision, 95.45% recall, and 95.37% F1-measure, while applying LDA with 1D-CNN HM reached 99.2% accuracy, and 99.3% precision, recall, and f1-measure. When merging LDA feature extraction with different percentages of MI and ANOVA feature selection (5%, 10%, 15%, and 20%), perfect resulting performance of 100% accuracy, precision, recall, and the F1-measure was acquired for the 20% of merging features. Consumed time reached 55 s for 1 out of the 100 total epochs and 64 patch size. The time of merging the LDA feature extraction with MI and ANOVA feature selection reached 3 s for 1 out of the 100 total epochs and 64 patch size. Such findings are considered to be excellent when compared with previous works.

REFERENCES

- [1] A. Y. Noori, S. H. Shaker, and R. A. Azeez, "3D scenes semantic segmentation using deep learning based Survey," *IOP Conference Series: Materials Science and Engineering*, vol. 928, no. 3, Aug. 2020, Art. no. 032083, <https://doi.org/10.1088/1757-899X/928/3/032083>.
- [2] A. Azeez and Raghad, "Determination Efficient Classification Algorithm for Credit Card Owners: Comparative Study," *Engineering and Technology Journal*, vol. 39, no. 1B, pp. 21–29, Mar. 2021, <https://doi.org/10.30684/etj.v39i1B.1577>.
- [3] N. H. Ali, M. E. Abdulmunem, and A. E. Ali, "Constructed model for micro-content recognition in lip reading based deep learning," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 5, pp. 2557–2565, Oct. 2021, <https://doi.org/10.11591/eei.v10i5.2927>.
- [4] W. Najah Abdullah and Y. Hussain Ali, "Face Retrieval Using Image Moments and Genetic Algorithm," *Engineering and Technology Journal*, vol. 34, no. 1B, pp. 160–171, Jan. 2016, <https://doi.org/10.30684/etj.v34i1B.18>.
- [5] H. Al-Dabbas and F. Mohammed, "The Effect of Wavelet Coefficient Reduction on Image Compression Using DWT and Daubechies Wavelet Transform," *Science International*, vol. 30, no. 5, pp. 757–762, Jan. 2018.
- [6] N. A. Taha, Z. Qasim, A. Al-Saffar, and A. A. Abdullatif, "Steganography using dual tree complex wavelet transform with LSB indicator technique," *Periodicals of Engineering and Natural Sciences*, vol. 9, no. 2, pp. 1106–1114, Jun. 2021, <https://doi.org/10.21533/pen.v9i2.2060>.
- [7] W. N. Abdullah, "Solving Job-Shop Scheduling Problem Using a Developed Particle Swarm Optimization Algorithm," *International Journal of Science and Research*, vol. 7, no. 1, pp. 1845–1848, 2015, <https://doi.org/10.21275/ART20179752>.
- [8] F. B. Ibrahim and M. E. Abdulmunim, "Real Time Face Recognition System based Hybrid Method," *International Journal of Scientific Research & Management Studies*, vol. 4, no. 4, pp. 6–11, May 2018.
- [9] A. H. Morad and H. M. Al-Dabbas, "Classification of Brain Tumor Area for MRI images," *Journal of Physics: Conference Series*, vol. 1660, no. 1, Aug. 2020, Art. no. 012059, <https://doi.org/10.1088/1742-6596/1660/1/012059>.
- [10] N. H. Barnouti, S. S. M. Al-Dabbagh, W. E. Matti, and M. A. S. Naser, "Face Detection and Recognition Using Viola-Jones with PCA-LDA and Square Euclidean Distance," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 5, pp. 371–377, 2016.
- [11] J. Zhou and B. Zhang, "Collaborative Representation Using Non-Negative Samples for Image Classification," *Sensors*, vol. 19, no. 11, Jan. 2019, Art. no. 2609, <https://doi.org/10.3390/s19112609>.
- [12] J. A. C. Moreano and N. B. L. S. Palomino, "Efficient Technique for Facial Image Recognition With Support Vector Machines in 2D Images With Cross-validation in Matlab," *Wseas Transactions on Systems and Control*, vol. 15, pp. 175–183, May 2020, <https://doi.org/10.37394/23203.2020.15.18>.
- [13] A. H. Rashed and M. H. Hamd, "Robust Detection and Recognition System based on Facial Extraction and Decision Tree," *Journal of Engineering and Sustainable Development*, vol. 25, no. 4, pp. 40–50, Jul. 2021, <https://doi.org/10.31272/jeasd.25.4.4>.
- [14] R. Szmurlo and S. Osowski, "Ensemble of classifiers based on CNN for increasing generalization ability in face image recognition," *Bulletin of the Polish Academy of Sciences. Technical Sciences*, vol. 70, no. 3, 2022, Art. no. e141004, <https://doi.org/10.24425/bpasts.2022.141004>.
- [15] E. M. Silva, M. Boaventura, I. A. G. Boaventura, and R. C. Contreras, "Face Recognition Using Local Mapped Pattern and Genetic Algorithms," in *International Conference on Pattern Recognition and Artificial Intelligence*, Union, NJ, USA, Dec. 2018, pp. 11–17, <https://doi.org/10.1145/3243250.3243262>.
- [16] E. Setiawan and A. Muttaqin, "Implementation of K-Nearest Neighbors Face Recognition on Low-power Processor," *Telkomnika*, vol. 13, no. 3, pp. 949–954, Sep. 2015, <https://doi.org/10.12928/telkomnika.v13i3.713>.
- [17] C. Szegedy et al., "Going deeper with convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, Jun. 2015, pp. 1–9, <https://doi.org/10.1109/CVPR.2015.7298594>.
- [18] I. Alhakam and N. H. Salman, "An Improved Probability Density Function (PDF) for Face Skin Detection," *Iraqi Journal of Science*, vol. 63, no. 10, pp. 4460–4473, Oct. 2022, <https://doi.org/10.24996/ijss.2022.63.10.31>.
- [19] S. O. Adeshina, H. Ibrahim, S. S. Teoh, and S. C. Hoo, "Custom Face Classification Model for Classroom Using Haar-Like and LBP Features with Their Performance Comparisons," *Electronics*, vol. 10, no. 2, Jan. 2021, Art. no. 102, <https://doi.org/10.3390/electronics10020102>.
- [20] N. K. Benamara, M. Keche, M. Wellington, and Z. Munyradzi, "Securing E-payment Systems by RFID and Deep Facial Biometry," in *1st International Conference on Artificial Intelligence and Data Analytics*, Riyadh, Saudi Arabia, Apr. 2021, pp. 151–157, <https://doi.org/10.1109/CAIDA51941.2021.9425175>.
- [21] W. Zhang and C. Zhao, "Exposing Face-Swap Images Based on Deep Learning and ELA Detection," *Proceedings*, vol. 46, no. 1, 2019, Art. no. 29, <https://doi.org/10.3390/ecea-5-06684>.
- [22] A. Al-Qaisi, M. Altarawneh, A. ElSaid, and Z. Al Qadi, "A Hybrid Method of Face Feature Extraction, Classification Based on MLBP and Layered- Recurrent Network," *Traitement du Signal*, vol. 37, pp. 555–561, Oct. 2020, <https://doi.org/10.18280/ts.370402>.
- [23] H. M. Al-Dabbas, R. A. Azeez, and A. E. Ali, "Machine Learning Approach for Facial Image Detection System," *Iraqi Journal of Science*, vol. 64, no. 10, pp. 6328–6341, Oct. 2023, <https://doi.org/10.24996/ijss.2023.64.10.44>.
- [24] "The MUCT Face Database." <http://www.milbo.org/muct/>.
- [25] F. G. Mohammed and H. M. Al-Dabbas, "Application of WDR Technique with different Wavelet Codecs for Image Compression," *Iraqi Journal of Science*, vol. 59, no. 4B, pp. 2128–2134, Nov. 2018, <https://doi.org/10.24996/ijss.2018.59.4B.18>.
- [26] A. Salim Jamil, R. Azeez, and N. Hassan, "An Image Feature Extraction to Generate a Key for Encryption in Cyber Security Medical Environments," *International Journal of Online and Biomedical*

- Engineering*, vol. 19, no. 1, pp. 93–106, Jan. 2023, <https://doi.org/10.3991/ijoe.v19i01.36901>.
- [27] J. M. AL-Tuwaijari, S. I. Mohammed, and M. A. B. Rahem, "Performance Evaluation of Face Image Recognition Based Viola-Joins with SVM," *Iraqi Journal of Information Technology*, vol. 9, no. 1, pp. 48–64, 2018.
- [28] H. Alamri, E. Alshanbari, S. Alotaibi, and M. Alghamdi, "Face Recognition and Gender Detection Using SIFT Feature Extraction, LBPH, and SVM," *Engineering, Technology & Applied Science Research*, vol. 12, no. 2, pp. 8296–8299, Apr. 2022, <https://doi.org/10.48084/etasr.4735>.
- [29] M. N. Chaudhari, M. Deshmukh, G. Ramrakhiani, and R. Parvatikar, "Face Detection Using Viola Jones Algorithm and Neural Networks," in *Fourth International Conference on Computing Communication Control and Automation*, Pune, India, Aug. 2018, pp. 1–6, <https://doi.org/10.1109/ICCCUBEA.2018.8697768>.
- [30] Y. Said, M. Barr, and H. E. Ahmed, "Design of a Face Recognition System based on Convolutional Neural Network (CNN)," *Engineering, Technology & Applied Science Research*, vol. 10, no. 3, pp. 5608–5612, Jun. 2020, <https://doi.org/10.48084/etasr.3490>.
- [31] M. K. Dabhi and B. K. Pancholi, "Face Detection System Based on Viola - Jones Algorithm," *International Journal of Science and Research*, vol. 5, no. 4, pp. 62–64, 2013.
- [32] M. E. Abdulmunem and F. B. Ibrahim, "Design of an Efficient Face Recognition Algorithm based on Hybrid Method of Eigen Faces and Gabor Filter," *Iraqi Journal of Science*, vol. 57, no. 3B, pp. 2102–2110, 2016.
- [33] H. Santoso and R. Pulungan, "A Parallel Architecture for Multiple-Face Detection Technique Using AdaBoost Algorithm and Haar Cascade," in *Information Systems International Conference*, Bali, Indonesia, Dec. 2013, pp. 592–597.
- [34] N. Singhal, V. Ganganwar, M. Yadav, A. Chauhan, M. Jakhar, and K. Sharma, "Comparative study of machine learning and deep learning algorithm for face recognition," *Jordanian Journal of Computers and Information Technology*, vol. 7, no. 3, pp. 313–325, Sep. 2021, <https://doi.org/10.5455/jicit.71-1624859356>.
- [35] Y. Aliyari Ghassabeh, F. Rudzicz, and H. A. Moghaddam, "Fast incremental LDA feature extraction," *Pattern Recognition*, vol. 48, no. 6, pp. 1999–2012, Jun. 2015, <https://doi.org/10.1016/j.patcog.2014.12.012>.
- [36] E. A. Khorsheed and Z. A. Nayef, "Face Recognition Algorithms: A Review," *Academic Journal of Nawroz University*, vol. 11, no. 3, pp. 202–207, Aug. 2022, <https://doi.org/10.25007/ajnu.v11n3a1432>.
- [37] H. M. Al-Dabbas, R. A. Azeez, and A. E. Ali, "Digital Watermarking, Methodology, Techniques, and Attacks: A Review," *Iraqi Journal of Science*, vol. 64, no. 8, pp. 4169–4186, Aug. 2023, <https://doi.org/10.24996/ijis.2023.64.8.37>.
- [38] K. J. Johnson and R. E. Synovec, "Pattern recognition of jet fuels: comprehensive GC×GC with ANOVA-based feature selection and principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 60, no. 1, pp. 225–237, Jan. 2002, [https://doi.org/10.1016/S0169-7439\(01\)00198-8](https://doi.org/10.1016/S0169-7439(01)00198-8).
- [39] S. Shakeela, N. S. Shankar, P. M. Reddy, T. K. Tulasi, and M. M. Koneru, "Optimal Ensemble Learning Based on Distinctive Feature Selection by Univariate ANOVA-F Statistics for IDS," *International Journal of Electronics and Telecommunications*, no. Vol. 67, No. 2, pp. 267–275, 2021, <https://doi.org/10.24425/ijet.2021.135975>.
- [40] Y. H. Ali and W. N. Abdullah, "A Survey of Similarity Measures in Web Image Search," *International Journal of Emerging Trends & Technology in Computer Science*, vol. 4, no. 4, pp. 191–196, 2015.
- [41] H. M. Al-Dabbas, R. A. Azeez, and A. E. Ali, "Efficient Iris Image Recognition System Based on Machine Learning Approach," *Iraqi Journal of Computers, Communications, Control & Systems Engineering*, vol. 23, no. 3, pp. 104–114, 2023, <https://doi.org/10.33103/uot.ijccce.23.3.9>.
- [42] T. Saidani, "Deep Learning Approach: YOLOv5-based Custom Object Detection," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12158–12163, Dec. 2023, <https://doi.org/10.48084/etasr.6397>.
- [43] R. Rajamohanam and B. C. Latha, "An Optimized YOLO v5 Model for Tomato Leaf Disease Classification with Field Dataset," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12033–12038, Dec. 2023, <https://doi.org/10.48084/etasr.6377>.
- [44] E. R. V. Reddy and S. Thale, "A Novel Efficient Dual-Gate Mixed Dilated Convolution Network for Multi-Scale Pedestrian Detection," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 11973–11979, Dec. 2023, <https://doi.org/10.48084/etasr.6340>.

Comparison of YOLOv5 and YOLOv6 Models for Plant Leaf Disease Detection

Ecem Iren

Izmir Kavram Vocational School, Turkiye

ecem.iren@kavram.edu.tr (corresponding author)

Received: 7 February 2024 | Revised: 7 March 2024 | Accepted: 10 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7033>

ABSTRACT

Deep learning is a concept of artificial neural networks and a subset of machine learning. It deals with algorithms that train and process datasets to make inferences for future samples, imitating the human process of learning from experiences. In this study, the YOLOv5 and YOLOv6 object detection models were compared on a plant dataset in terms of accuracy and time metrics. Each model was trained to obtain specific results in terms of mean Average Precision (mAP) and training time. There was no considerable difference in mAP between both models, as their results were close. YOLOv5, having 63.5% mAP, slightly outperformed YOLOv6, while YOLOv6, having 49.6% mAP₅₀₋₉₅, was better in detection than YOLOv5. Furthermore, YOLOv5 trained data in a shorter time than YOLOv6, since it has fewer parameters.

Keywords-deep learning; YOLOv5; YOLOv6; plant leaf disease detection; convolutional neural networks

I. INTRODUCTION

The agricultural process plays a vital role in the food supply, while the farming sector contributes to the employment [1]. Plant diseases are a significant issue since they can cause huge economic losses. Many factors lead to such infection types, such as viruses, bacteria, and fungi [2]. Farmers or experts can monitor diseases in traditional ways with some experience or training, but this method can be considered challenging, expensive, and open to mistakes. A disease can be recognized and examined at an early stage if the farmer regularly controls the crops and has sufficient information. As it is difficult to manage larger fields, a fast, healthy, and accurate disease detection system is strongly needed in cases where human evaluation is unreliable or not enough. Early detection systems can reduce large-scale crop losses by preventing the spread of diseases [2-4]. Image processing, machine learning, and deep learning methods can be used to examine infected regions of plants by predicting the class of a specific disease [5]. Machine learning algorithms try to explore hidden insights and complex patterns of a given dataset and are generally preferred for tasks related to classification, regression, and clustering. Artificial Neural Networks (ANNs) operate according to the principle of information processing in biological systems. Each processing unit is an artificial neuron that connects to others and is represented by mathematical concepts. Each connection between neurons sends signals simulating the synapses in the brain. If a specific signal ensures a certain threshold determined by an activation function, it can be processed by subsequent neurons. Typically, neurons are organized into networks with different layers. An input layer usually receives the data input, and the output layer produces the ultimate result. Deep neural networks have more than one hidden layer, usually containing complex neurons, and employ

high-level operations using multiple activation functions instead of a single neuron [6-7].

Object detection is a research area in deep learning and is more complicated than object classification. Object classification does not provide information on the location of the object in the image. Object detection can be divided into two different categories, one- and two-stage. In two-stage detection, the region proposal is generated first, and then the classification task is performed. Faster R-CNNs generally apply this mechanism to perform detection, but it cannot work in images with more than one object [8]. However, the YOLO models perform the detection process by combining two operations called classification and localization. YOLO is a one-stage detector, which means that regional proposal and classification are performed simultaneously. YOLO divides an image into a grid of cells. Each grid cell predicts the class of an object and specifies the bounding box determining the object's location. Each bounding box can be described using four descriptors: the center of a bounding box (bx, by), width (bw), height (bh), and object class (c). Non-max suppression is a common algorithm used for discarding bounding boxes when multiple boxes are predicted for the same object [9-10].

Many studies have been conducted by applying YOLO models. In [11], YOLOv7, YOLOv5, YOLO-X, and YOLO-R were compared in terms of Frames Per Second (FPS) and accuracy metrics, showing that YOLOR-p6 had the best results with 23 FPS, followed by YOLO-Xm and YOLOR6-p6 with 23 and 22 FPS, respectively, while YOLOv7 had the lowest speed of only 17 FPS. YOLOv7 showed the highest mean Average Precision (mAP) of approximately 60%, followed by YOLO-p6, YOLOv5, and YOLO-Xm with 56, 53, and 46%, respectively. In [12], a performance comparison was made between YOLOv3, YOLOv4, and YOLOv5 for poultry

recognition. The dataset was split into training and testing parts with a ratio of 80% and 20%, respectively. After the manual annotation process, different YOLO versions were trained. The experimental results showed that YOLOv5x achieved the highest accuracy of 99.5%, while the YOLOv4-tiny model had the lowest mAP and the least training speed.

In [13], YOLOv5 and YOLOv8 were compared on aerial image data taken from the Roboflow pedestrian dataset. Training and validation tasks were performed with 100 epochs using the Google Colab environment, showing that YOLOv8 had better precision and F1-score than YOLOv5, by 2.82% and 0.98%, respectively. In [14], a safety helmet detection system was developed based on YOLOv5. 6045 images were collected and annotated, and then some operations such as random flip, geometric and illumination distortion, random erase, cutout, and mix-up were used for data augmentation. Four YOLOv5 models were compared, showing that without pre-trained weights, YOLOv5x obtained the best mAP value of 93.6%, which was only 0.1% higher than YOLOv5l. The performance of all YOLOv5 models improved with pre-trained weights (transfer learning), and while YOLOv5s had the highest improvement rate with 1.3%, YOLOv5x was the leader in terms of accuracy.

In [15], the YOLOv5 and DETR models were compared in detecting sea cucumbers. The results showed that YOLOv5 outperformed DETR in terms of accuracy and computing resources. Authors in [13] aimed to find the most suitable hyperparameters for the detection of healthy and diseased tomato leaves, using private and public datasets, showing that YOLOv5 reached an accuracy rate of 93% during the evaluation of the test data. In [16], a YOLOv5 model with a special configuration of hyperparameters, such as learning rate and batch size, and transfer learning provided an mAP value of 60.9%. This study compares the detection performance of the YOLOv5 and YOLOv6 models on a leaf disease dataset.

II. MATERIALS AND METHODS

This study compares the object detection performance of the YOLOv5 and YOLOv6 models on a data set that contains different types of leaf diseases. Figure 1 illustrates each step of the detection process. At first, a dataset was obtained and split into training and validation parts, and then each detection model was trained to obtain specific metric results.

A. Dataset

The PlantDoc [17] dataset was developed at the Indian Institute of Technology and is shared on the GitHub platform. PlantDoc has 2,569 images, including 13 plant species and 30 classes as diseased and healthy for image classification and object detection purposes. The data set has 8851 annotations and was divided into training and validation parts in a 90:10 ratio. Therefore, 2328 images were used for training and 239 images were used for validation.

B. YOLOv5 Object Detection Model

YOLOv5 has four different scale variations, named S, M, L, and X that represent Small, Medium, Large, and Extra Large, respectively. These scales contain different multipliers in terms of depth and width, while the structure of the model

remains constant. Only the complexity and size of the objects are scaled in each variation [14]. YOLOv5 is an object detection framework that has been established in the CSPDarknet53 and PyTorch frameworks. CSPDarknet53 provides a backbone architecture that consists of a Focus structure and a CSP network and aggregates image features to achieve feature extraction. The Focus layer is used to decrease the Compute Unified Device Architecture (CUDA) memory, layers, and parameters and improve the backward and forward speed. Aggregated features are directed to the Neck (PANet) structure, which includes a new Feature Pyramid Network (FPN) with many bottom-up and top-down layers. Low-level features can be obtained with the help of FPN, and lower layers improve the localization accuracy of the object. In the Head block, features from the Neck are used to detect classes in terms of class, location, confidence score, and size [18-20]. Figure 2 shows the YOLOv5 architecture.

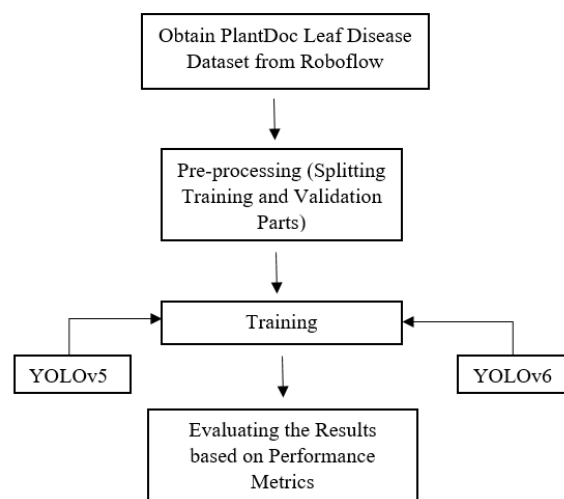


Fig. 1. Method's flow diagram.

C. YOLOv6 Object Detection Model

YOLOv6 shows better performance than YOLOv5 in terms of detection accuracy and inference speed, making it an efficient option for industrial applications. YOLOv6 has been designed with many improvements in the Backbone, Neck, and Head blocks. The Neck and Backbone structures have been replaced with Rep-PAN and EfficientRep. Since the design of the backbone network plays an important role in detection effectiveness, a RepBlock with RepVGG is preferred since it offers a re-parameterizable structure. A multi-branch topology is used during the training process, and then RepBlock is converted to RepConv stacks of 3×3 convolutional layers with ReLU activation functions during the inference stage. EfficientRep Backbone supports hardware such as GPU and CPU. Rep-Pan Neck is considered more precise and operates faster than PANet and SPP. A convolutional layer has been added between the network and the final Head to improve performance, using processing power and memory [21-22]. Figure 3 shows the architecture of YOLOv6.

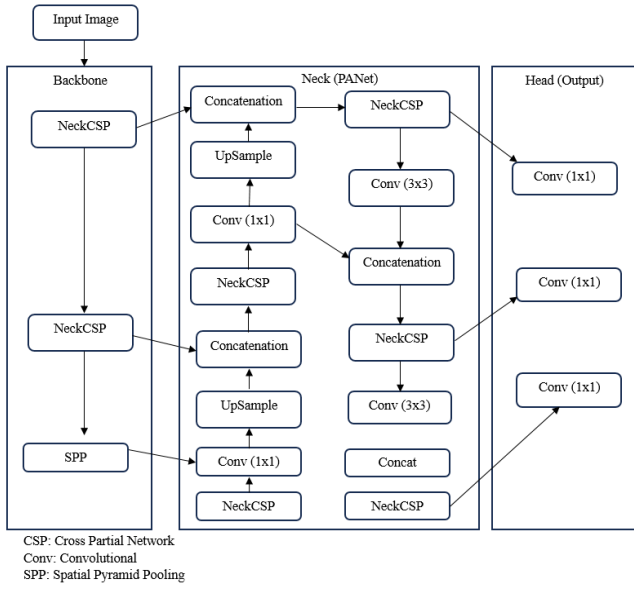


Fig. 2. YOLOv5 architecture.

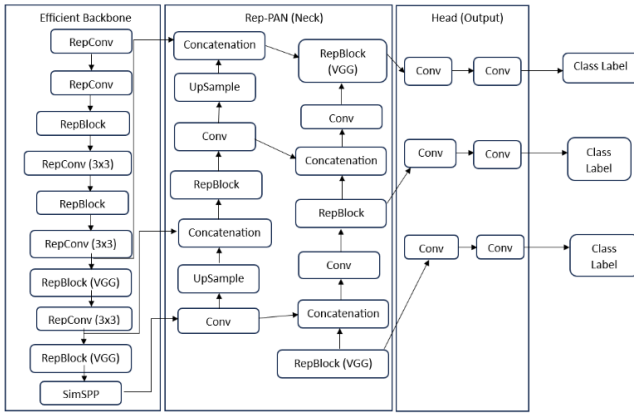


Fig. 3. YOLOv6 architecture.

D. Environment and HyperParameters

Each model was trained in the Google Colab environment, which provides a service with powerful GPUs and TPUs for machine and deep learning problems. Python and NVIDIA Tesla T4 GPU were used for the experiments. Table I shows the hyperparameter configurations of the models, which were the same for each model for a fair comparison. Epochs are the total iterations of a training task. The learning rate determines the change level of the model based on the predicted error when updating the model weights. Batch size defines the number of samples for one epoch during training, and the optimization algorithm adjusts the weights to minimize error. Lastly, momentum deals with optimizing convergence.

TABLE I. HYPERPARAMETERS OF YOLO MODELS

Parameter	Value
Epochs	100
Learning rate	0.01
Batch size	16
Optimization algorithm	SGD (Stochastic Gradient Descent)
Momentum	0.937

III. RESULTS AND DISCUSSION

True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) were used to evaluate the results of the experiments, which are generally considered indicators of detection performance. TP refers to the number of accurate detections of diseased leaves, FP refers to the number of objects that were incorrectly identified as a diseased leaf, TN represents the number of negative samples with a negative prediction, and FN is the number of diseased leaves that were missed. Precision (P) is the rate of the total number of actual positives to all predicted positives (1), and Recall (R) refers to how many samples that had to be predicted as positive were correctly estimated as positive (2). The Average Precision (AP) balances precision and recall by taking into account both FP and FN (3). The area under the precision-recall curve is calculated to find the AP for each class, while mAP is the average of the precision-recall curve of each object class (4) [13, 23], and it is a well-known metric for measuring the accuracy of detection models.

$$\text{Precision (P)} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{Recall (R)} = \frac{TP}{TP+FN} \quad (2)$$

$$\text{Average Precision (AP)} = \int_0^1 P(r) dr \quad (3)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (4)$$

The mAP50 metric represents the mean average precision in an IoU at the threshold of 0.5, while mAP50-95 is a term for the mean average precision, calculated as the average of 10 AP values for 10 IoU thresholds between 0.5 and 0.95 confidence levels [23-24]. Therefore, it is a way to examine several confidence values to assess the accuracy of the detected objects. Table II shows the accuracy, precision, recall, and training duration for each YOLO model. YOLOv5 showed a slightly better result than YOLOv6, having mAP50 of 63.5% and 62%, respectively. When examining mAP50-95, YOLOv6 had a higher performance (49.6%) than YOLOv5 (48.8%). However, there is not a huge difference between the two models when analyzing the mAP metrics, as the results were close to each other. On the other hand, YOLOv6 needed much more time to train than YOLOv5. The training process for the YOLOv6 model took nearly 2.3 hours, while the YOLOv5 training process took approximately 0.8 hours. More complex models train more parameters, leading to an increased training time. The mAP graphs can be seen in Figures 4-6.

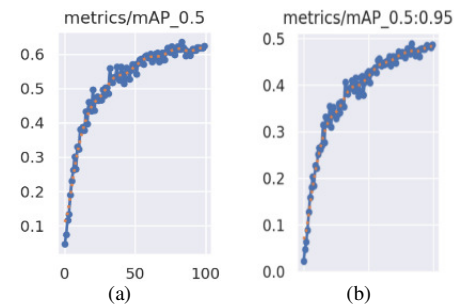


Fig. 4. YOLOv5 graphs: (a) mAP, (b) map50-95.

TABLE II. PERFORMANCE METRICS

Model	mAP50	mAP50-0.95	Precision	Recall	Duration
YOLOv5	63.5%	48.8%	64.8%	54.8%	0.772
YOLOv6	62%	49.6%	56.6%	62.4%	2.282

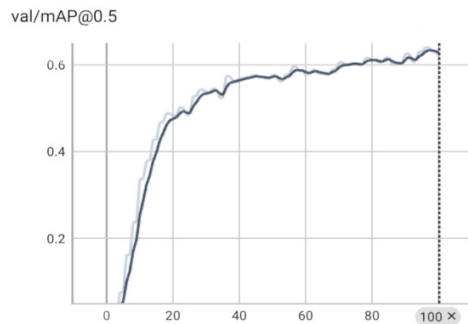


Fig. 5. mAP graph of YOLOv6.

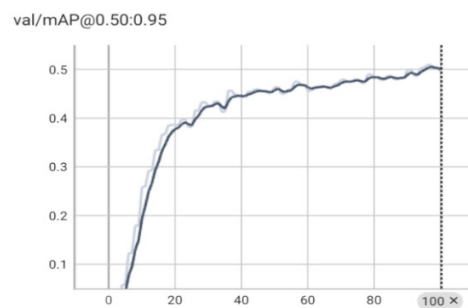


Fig. 6. mAP50-95 graph of YOLOv6.

The precision of YOLOv5 was 64.8%, which was better than YOLOv6 (56.6%). The recall of YOLOv6 (62.4%) was found to be greater than that of YOLOv5 (54.8%). Figures 7 and 8 show the precision-recall curves of each model. Figures 9-12 show the detection of some validation samples with their class categories and confidence levels for both models. In Figure 9, cherry and blueberry leaves were estimated with confidence levels of 70, 50, and 80%, respectively, using YOLOv5. Likewise, corn leaf blight disease was detected with an 80% confidence level using the same algorithm, as shown in Figure 10. Additionally, YOLOv6 detected squash powdery mildew leaf disease and apple leaf with 77% and 91% confidence levels, respectively, as shown in Figures 11 and 12.

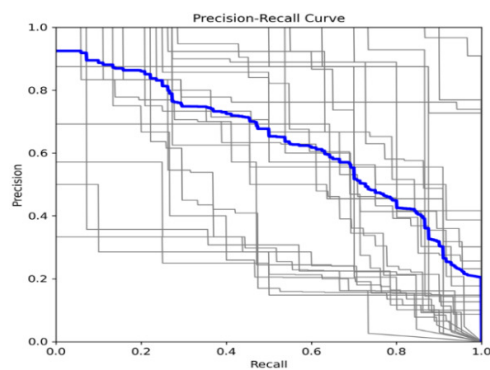


Fig. 7. Precision-recall curve of YOLOv5.

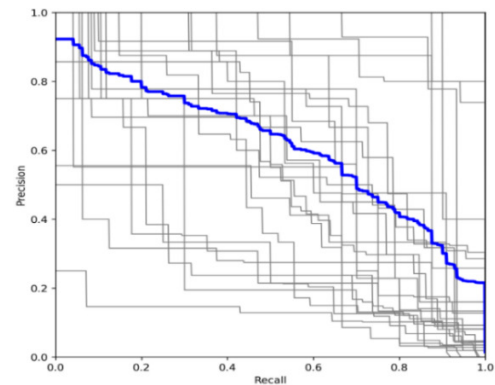


Fig. 8. Precision-recall curve of YOLOv6.

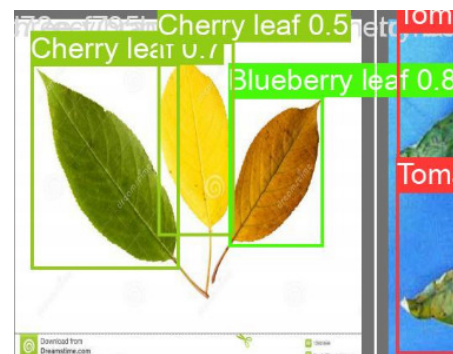


Fig. 9. YOLOv5 detection of cherry and blueberry leaves with confidence levels.

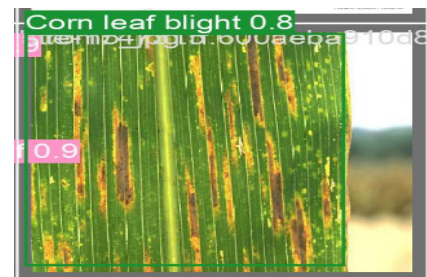


Fig. 10. YOLOv5 detection of corn leaf blight with confidence levels.

Other studies that compared YOLOv5 and YOLOv6 in different datasets showed that mAP values can vary according to different parameters such as class category distribution, total number of class categories, hyperparameters, and augmentation methods. In [25], YOLO models were trained on a dataset containing road cracks and potholes, reaching mAP50 values of 77 and 68.11% for YOLOv5-s and YOLOv6-s, respectively. In [26], YOLOv5 and YOLOv6 were compared on the COCO dataset with 30 class categories, and YOLOv5-m had a mAP of 9.28%, followed by YOLOv6-m with 8.51%. In [24], two models were compared in an oil tank dataset, showing mAP50-95 of 69.69 and 65.60% for YOLOv5-x and YOLOv6-l, respectively. In [21], the same models were compared on a lunar crater dataset, where YOLOv5 achieved 72% mAP and YOLOv6 had 62% with SGD optimization.



Fig. 11. Detection of squash powdery mildew leaf with confidence level by YOLOv6.



Fig. 12. Detection of apple leaf with confidence level by YOLOv6.

IV. CONCLUSION

This study carried out a comparative analysis of YOLOv5 and YOLOv6 object detection models on a plant leaf disease dataset to evaluate their detection and time performance. The results showed that there were no great differences between the detection performance of each model according to the mAP metric. If saving time is significant, the YOLOv5 model can be chosen since it is trained in a shorter time compared to YOLOv6. In the future, the hyperparameters of these models can be tuned by changing the learning rate, batch size, optimization method, and momentum. Furthermore, advanced YOLO versions, such as YOLOv7 and YOLOv8, can be studied with the same dataset to discover and compare their performance results.

REFERENCES

- [1] L. Christiaensen, Z. J. Rutledge, and J. E. Taylor, "The Future of Work in Agriculture: Some Reflections," The World Bank, Policy Research Working Paper 9193, Mar. 2020.
- [2] S. Nigam and R. Jain, "Plant disease identification using Deep Learning: A review," *Indian Journal of Agricultural Sciences*, vol. 90, no. 2, pp. 249–257, Mar. 2020, <https://doi.org/10.56093/ijas.v90i2.98996>.
- [3] R. Rajamohan and B. C. Latha, "An Optimized YOLO v5 Model for Tomato Leaf Disease Classification with Field Dataset," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12033–12038, Dec. 2023, <https://doi.org/10.48084/etasr.6377>.
- [4] N. Shelar, S. Shinde, S. Sawant, S. Dhumal, and K. Fakir, "Plant Disease Detection Using CNN," *ITM Web of Conferences*, vol. 44, 2022, Art. no. 03049, <https://doi.org/10.1051/itmconf/20224403049>.
- [5] V. Singh and A. K. Misra, "Detection of plant leaf diseases using image segmentation and soft computing techniques," *Information Processing in Agriculture*, vol. 4, no. 1, pp. 41–49, Mar. 2017, <https://doi.org/10.1016/j.inpa.2016.10.005>.
- [6] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," *Electronic Markets*, vol. 31, no. 3, pp. 685–695, Sep. 2021, <https://doi.org/10.1007/s12525-021-00475-2>.
- [7] J. F. Mas and J. J. Flores, "The application of artificial neural networks to the analysis of remotely sensed data," *International Journal of Remote Sensing*, vol. 29, no. 3, pp. 617–663, Feb. 2008, <https://doi.org/10.1080/01431160701352154>.
- [8] J. Kaur and W. Singh, "Tools, techniques, datasets and application areas for object detection in an image: a review," *Multimedia Tools and Applications*, vol. 81, no. 27, pp. 38297–38351, Nov. 2022, <https://doi.org/10.1007/s11042-022-13153-y>.
- [9] M. Carranza-García, J. Torres-Mateo, P. Lara-Benítez, and J. García-Gutiérrez, "On the Performance of One-Stage and Two-Stage Object Detectors in Autonomous Vehicles Using Camera Data," *Remote Sensing*, vol. 13, no. 1, Jan. 2021, Art. no. 89, <https://doi.org/10.3390/rs13010089>.
- [10] S. Norkobil Saydirasulovich, A. Abdusalomov, M. K. Jamil, R. Nasimov, D. Kozhamzharova, and Y.-I. Cho, "A YOLOv6-Based Improved Fire Detection Approach for Smart City Environments," *Sensors*, vol. 23, no. 6, Jan. 2023, Art. no. 3161, <https://doi.org/10.3390/s23063161>.
- [11] M. R. M. Ismat Saira Gillani, "Yolov5, Yolo-x, Yolo-r, Yolo-v7 Performance Comparison: A Survey," *CS & IT Conference Proceedings*, vol. 12, no. 16, Sep. 2022.
- [12] O. Kivrak and M. Z. Gürbüz, "Performance Comparison of YOLOv3, YOLOv4 and YOLOv5 algorithms: A Case Study for Poultry Recognition," *Avrupa Bilim ve Teknoloji Dergisi*, no. 38, pp. 392–397, Aug. 2022, <https://doi.org/10.31590/ejosat.1111288>.
- [13] I. P. Sary, S. Andromeda, and E. U. Armin, "Performance Comparison of YOLOv5 and YOLOv8 Architectures in Human Detection using Aerial Images," *Ultima Computing: Jurnal Sistem Komputer*, vol. 15, no. 1, pp. 8–13, Jun. 2023, <https://doi.org/10.31937/sk.v15i1.3204>.
- [14] F. Zhou, H. Zhao, and Z. Nie, "Safety Helmet Detection Based on YOLOv5," in *2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA)*, Shenyang, China, Jan. 2021, pp. 6–11, <https://doi.org/10.1109/ICPECA51329.2021.9362711>.
- [15] X. Yuan *et al.*, "Performance Comparison of Sea Cucumber Detection by the YOLOv5 and DETR Approach," *Journal of Marine Science and Engineering*, vol. 11, no. 11, Nov. 2023, Art. no. 2043, <https://doi.org/10.3390/jmse11112043>.
- [16] T. Saidani, "Deep Learning Approach: YOLOv5-based Custom Object Detection," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12158–12163, Dec. 2023, <https://doi.org/10.48084/etasr.6397>.
- [17] D. Singh, N. Jain, P. Jain, P. Kayal, S. Kumawat, and N. Batra, "PlantDoc: A Dataset for Visual Plant Disease Detection," in *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, Hyderabad, India, Jan. 2020, pp. 249–253, <https://doi.org/10.1145/3371158.3371196>.
- [18] O. E. Olorunshola, M. E. Irhebhude, and A. E. Ewwiekpaefe, "A Comparative Study of YOLOv5 and YOLOv7 Object Detection Algorithms," *Journal of Computing and Social Informatics*, vol. 2, no. 1, pp. 1–12, Feb. 2023, <https://doi.org/10.33736/jcsi.5070.2023>.
- [19] H. K. Jung and G. S. Choi, "Improved YOLOv5: Efficient Object Detection Using Drone Images under Various Conditions," *Applied Sciences*, vol. 12, no. 14, Jan. 2022, Art. no. 7255, <https://doi.org/10.3390/app12147255>.
- [20] T. Saidani, R. Ghodhban, A. Alhomoud, A. Alshammari, H. Zayani, and M. B. Ammar, "Hardware Acceleration for Object Detection using YOLOv5 Deep Learning Algorithm on Xilinx Zynq FPGA Platform," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 13066–13071, Feb. 2024, <https://doi.org/10.48084/etasr.6761>.
- [21] N. Aburaed, M. Alsaad, S. A. Mansoori, and H. Al-Ahmad, "A Study on the Autonomous Detection of Impact Craters," in *Artificial Neural*

- Networks in Pattern Recognition*, Dubai, United Arab Emirates, 2023, pp. 181–194, https://doi.org/10.1007/978-3-031-20650-4_15.
- [22] C. Li *et al.*, "YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications." arXiv, Sep. 07, 2022, <https://doi.org/10.48550/arXiv.2209.02976>.
- [23] T. Y. Lin *et al.*, "Microsoft COCO: Common Objects in Context," in *Computer Vision – ECCV 2014*, Zurich, Switzerland, 2014, pp. 740–755, https://doi.org/10.1007/978-3-319-10602-1_48.
- [24] T. Bakirman, "An Assessment of YOLO Architectures for Oil Tank Detection from SPOT Imagery," *International Journal of Environment and Geoinformatics*, vol. 10, no. 1, pp. 9–15, Mar. 2023, <https://doi.org/10.30897/ijegeo.1196817>.
- [25] N. I. M. Yusof, A. Sophian, H. F. M. Zaki, A. A. Bawono, A. H. Embong, and A. Ashraf, "Assessing the performance of YOLOv5, YOLOv6, and YOLOv7 in road defect detection and classification: a comparative study," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 1, pp. 350–360, Feb. 2024, <https://doi.org/10.11591/eei.v13i1.6317>.
- [26] M. Horvat, L. Jelečević, and G. Gledec, "Comparative Analysis of YOLOv5 and YOLOv6 Models Performance for Object Classification on Open Infrastructure: Insights and Recommendations," presented at the 34th Central European Conference on Information and Intelligent Systems, Zagreb, Croatia, 2023, pp. 317–324.

Performance Enhancement of Distributed Processing Systems using Novel Hybrid Shard Selection Algorithm

Praveen M. Dhulavvagol

School of Computer Science and Engineering, KLE Technological University, India
praveen.md@kletech.ac.in (corresponding author)

Sashikumar G. Totad

School of Computer Science and Engineering, KLE Technological University, India
totad@kletech.ac.in

Received: 22 February 2022 | Revised: 7 March 2024 | Accepted: 12 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7128>

ABSTRACT

Distributed processing systems play a crucial role in query search operations, where large-scale data are partitioned across multiple nodes using shard selection algorithms. However, the existing shard selection algorithms pose significant challenges, such as shard ranking, shard cut-off estimation, high latency, low throughput, and high processing costs. These limitations become more pronounced as the data size increases, affecting the efficiency and effectiveness of search operations. To address these challenges, the novel Hybrid Shard Selection Algorithm (HSSA) is proposed as a solution in this paper, designed specifically to enhance the effectiveness and efficiency of search operations within distributed processing systems. HSSA employs an advanced sharding approach that adeptly navigates and targets pertinent shards based on specific queries. This not only curtails search-related overhead but also enhances operational efficiency. Through rigorous testing using the Gov2 dataset, the HSSA algorithm has proven its merits. When set against well-established algorithms like CORI, Rank-S, and SHiRE, HSSA stands out, registering remarkable gains in average throughput by 21%, 16%, and 12%, while also slashing latency by 14.2%, 9.4%, and 8.2%, respectively. The insights gained from this research underscore HSSA's capability to effectively bridge the gaps inherent in traditional shard selection strategies. Furthermore, its exemplary efficacy with datasets of varied sizes amplifies its relevance for practical integration within distributed processing landscapes.

Keywords-sharding; cluster; indexing; partitioning; allocation

I. INTRODUCTION

The exponential growth of data across industries and disciplines has prompted a paradigm shift in data processing, storage, and retrieval methods. Distributed processing systems have emerged as a dominant paradigm in the era of big data, enabling the efficient processing and analysis of vast amounts of information. The cornerstone of these systems lies in their ability to divide data and tasks across multiple nodes, ensuring parallelism, scalability, and fault tolerance [1]. A critical component in optimizing the performance of such systems is sharding, which entails partitioning databases into smaller, more digestible pieces, termed as shards, which can be processed concurrently. Shard selection essentially involves the partitioning of a database or dataset into smaller, more manageable segments, known as shards. Each shard can then be processed independently and concurrently across various nodes of the distributed system. This concurrent processing not only accelerates data retrieval and computation, but also ensures a balanced load distribution across the system,

preventing potential bottlenecks and system overloads. Traditional sharding approaches have predominantly revolved around either static or dynamic methodologies. Static sharding involves dividing a database based on predetermined criteria, which, while simple and predictable, may lead to inefficiencies such as data skew or imbalance among nodes. On the other hand, dynamic sharding techniques allow the system to partition data on-the-fly based on various parameters, such as data size or workload characteristics. While more flexible, dynamic sharding can sometimes add overhead due to the continuous monitoring and adjustment requirements.

Static shard selection algorithms, like the Lexical SHiRE (Lex-S) and Connected SHiRE (Conn-S), are deterministic in nature. Their modus operandi focuses on predefined criteria set during data ingestion or index creation. For instance, Lex-S relies on the lexical attributes of data for sharding, and once formed, these shards remain unchanged irrespective of the nature of queries they confront. Similarly, Conn-S capitalizes on the intrinsic relationships or connected components within

the data to form shards, retaining their structure throughout their lifecycle [2].

In contrast, dynamic shard selection algorithms adjust and evolve based on the incoming queries, offering a more adaptive approach to shard selection. Algorithms such as Rank SHiRE (Rank-S), Collection Retrieval Inference Network (CORI), and Relevant Document Distribution Estimation (ReDDE) epitomize this category. Rank-S, for instance, dynamically ranks shards vis-à-vis their relevance to an incoming query. CORI and ReDDE, on the other hand, estimate the significance of shards based on varying parameters, like document statistics or the relevance of terms in a query, respectively [3]. While both static and dynamic sharding techniques come with their own set of merits, neither is devoid of shortcomings. Static algorithms might be compromised by unforeseen query patterns, whereas dynamic ones may incur overheads due to their adaptive nature [4]. In the wake of these challenges, there's a growing need for an algorithmic solution that seamlessly blends the predictability of static sharding with the adaptability of dynamic sharding [5].

This paper proposes the Hybrid Shard Selection Algorithm (HSSA), a novel approach designed to enhance the performance of distributed processing systems. By intelligently leveraging the strengths of both static and dynamic sharding techniques, HSSA seeks to integrate the robustness of static algorithms with the adaptability of the dynamic ones. By initially leveraging the likes of Lex-S or Conn-S for shard creation and subsequently resorting to dynamic algorithms like Rank-S, CORI, or ReDDE for shard selection during queries, HSSA aspires to strike a harmonious balance. Such an approach promises to fuse the predictability of static sharding with the nimbleness of dynamic sharding, paving the path for a more refined, efficient, and performance-enhanced distributed processing system. HSSA aims to ensure optimal data distribution, minimize processing latency, and promote balanced resource utilization across the system. The main contributions of the proposed algorithm are:

- **Hybrid Integration:** HSSA uniquely combines both static (Lex-S, Conn-S) and dynamic (Rank-S, CORI, ReDDE) shard selection methodologies, optimizing efficiency and adaptability in distributed systems.
- **Balanced Efficiency:** By leveraging static predictability and dynamic real-time relevance, HSSA ensures swift data retrieval and optimal resource utilization.
- **Scalability and Adaptability:** HSSA exhibits enhanced scalability and fault tolerance, adjusting seamlessly to diverse query patterns, and setting a benchmark for future shard selection algorithms.

The primary contribution of HSSA lies in its innovative approach of combining the benefits of both static and dynamic shard selection methodologies to presents a holistic solution. This hybrid methodology aims to ensure optimal data distribution, reduce processing latency, and guarantee efficient resource utilization across the distributed processing system. In essence, the HSSA algorithm sets a new benchmark for shard selection by harmoniously merging predictability with

flexibility, ultimately enhancing the overall performance of distributed data processing systems.

II. RELATED WORK

Sharding is a horizontal data partitioning technique that partitions data row-wise into n different shards. Authors in [6] focus on enhancing the effectiveness of data reclamation and processing in Elasticsearch through colorful shard selection ways. The study delves into the complications of Elasticsearch's armature, emphasizing the pivotal part of sharding in vertical scaling. The authors explore and compare different shard selection strategies, similar to the range-grounded, hash-grounded, and adaptive ways, assessing their effectiveness in optimizing query prosecution and cargo balancing. The optimization of data partitioning in a sharding system employing harmonious mincing is presented in [2]. The central focus lies in streamlining the process of dividing and distributing data across shards to enhance system effectiveness. The authors address the complications associated with traditional data partitioning styles and propose a simplified approach, using harmonious mincing as the underpinning mechanism. The study delves into the complications of harmonious mincing, emphasizing its part in maintaining cargo balance and easing scalable data storehouses. Authors in [7] report that, the basic CORI algorithm does not scale up with large-size data and downgrades the performance of distributed search engines. In [8], it was reported that selective search could divide the data according to topic-based shards and search could be conducted in only a few shards that contain the relevant query. The experimental results showed that the selective search is more effective than the (traditional) exhaustive search method. Dividing the data according to the topic-based motivates the cluster to search the data for the query in a few shards which contain the relevant data [9]. If the documents have similarities, then they can be grouped. In [1], a novel query-biased partitioning strategy, aligning document partitions with topics derived from query logs is presented. It specifically addresses clustering initialization and document similarity calculation.

A query-driven clustering initialization algorithm employs topics from query logs, while a query-biased similarity metric prioritizes terms important in query logs. Both methods enhance retrieval effectiveness, diminish variance, and promote a more balanced distribution of shard sizes. The SHARD Triple-Store, a cutting-edge, high performance, and massively scalable distributed system leveraging the MapReduce software framework, is presented in [10]. This innovative approach capitalizes on the inherent capabilities of MapReduce to handle vast amounts of data in a parallel and distributed manner. The triple-store architecture is specifically designed for efficient and scalable storage and retrieval of RDF (Resource Description Framework) triples, a crucial component in semantic data processing. By utilizing the MapReduce paradigm, the SHARD Triple-Store achieves exceptional scalability, allowing it to seamlessly handle large-scale datasets. MapReduce's parallel processing capabilities are harnessed to distribute the computational load across a multitude of nodes, ensuring efficient data processing and storage. This architecture excels in scenarios where traditional

relational databases may struggle to cope with the sheer volume and complexity of data. The SHARD Triple-Store represents a notable advancement in the field of distributed systems, offering a robust solution for organizations dealing with massive and dynamic datasets in the realm of semantic data processing.

Significant challenges in distributed processing are performance enhancement, fault tolerance, and scalability [11, 12]. Different approaches and heuristic techniques have been proposed to partition data and allocate shards at multiple nodes. The existing data partitioning strategies encounter high communication, low throughput, scalability, consistency, and mining overhead. The proposed approach aims to overcome the above limitations using a hybrid shard selection technique.

III. HSSA SYSTEM ARCHITECTURE

Figure 1 depicts the proposed HSSA system's architecture. The architecture offers a sophisticated solution to the challenges inherent in managing vast datasets in distributed processing systems. Central to the architecture, is a distributed framework anchored by the Master Node, a pivotal component that supervises the data's initial processing. With the aid of sharding techniques, the master node meticulously segments the massive volume of data into strategically defined shards. Each shard is intricately themed around specific topics, ensuring that every fragment or shard is a coherent aggregation of documents that share thematic resonance. As the orchestrator, the master node oversees operations across the distributed spectrum, managing queries and the critical dual processes of data partitioning and shard selection. This pivotal node's proactive role extends further; by controlling parallel processing across shards, it ensures a dramatic reduction in latency, simultaneously amplifying throughput and refining search precision.

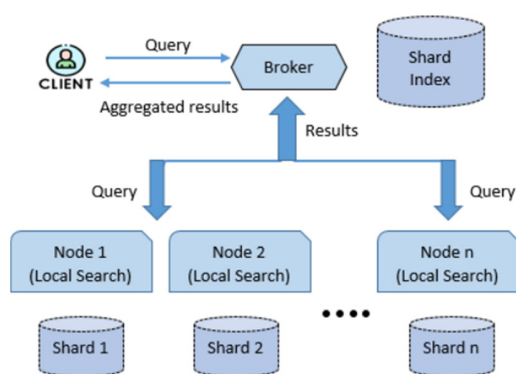


Fig. 1. HSSA system architecture.

Complementing the master node are the Data Nodes, the repositories entrusted with housing the carefully curated shards. Each data node plays steward to a shard—a self-sustaining subset of the entire dataset, a thematic archive of documents under its domain. With HSSA's architectural design, the process is seamlessly streamlined into two core phases. First, the data undergoes a rigorous partitioning process, ensuring it is dissected into topic-centric, logical shards. Following this, the

algorithm pivots to the shard selection and ranking phase, pinpointing the shards most relevant to a given query and ranking them in order of relevance. This fusion of methodological rigor and precision-driven selection gives HSSA its distinctive edge, setting new benchmarks in distributed data processing efficiency.

A. Hybrid Data Partitioning Approach in HSSA

Hybrid data partitioning, as encapsulated in HSSA, is an innovative fusion of horizontal and vertical partitioning strategies, offering a flexible and adaptive system for diverse datasets. This approach emerges from the understanding that the strengths of one partitioning type can compensate for the shortcomings of the other, thus ensuring a holistic data management system. The procedure initiates with a Preliminary Data Analysis, where the data's structure, volume, and access patterns are meticulously assessed. By probing into the unique intricacies of the dataset, HSSA can tailor its partitioning mechanisms to match the data's inherent needs.

The horizontal partitioning strategy then comes into play, slicing data at the row level. Each shard, or partition, is structured to represent a consistent portion of the overarching dataset. This approach is instrumental when systems foresee queries that span multiple columns but are concentrated on specific rows. For example, in a dataset profiling users, one might seek all data attributes (columns) for a specific user (a particular row).

Vertical partitioning focuses on column-based data segregation. By clustering columns that are frequently accessed in tandem, this approach optimizes query responses for those seeking specific data attributes across the board. For instance, in an e-commerce database, queries might predominantly target product names and prices, while seldom accessing their detailed descriptions. In such scenarios, vertical partitioning optimizes data fetch times by co-locating these commonly accessed columns.

HSSA's approach is its dynamic melding of these two strategies. The system constantly evaluates query demands and data access patterns, fine-tuning the balance between horizontal and vertical partitions. If a dataset witnesses evolving access patterns, HSSA's dynamic adaptation mechanism adjusts the partitioning mix, ensuring consistent and optimal performance. Incorporated into this is a suite of implementation strategies like range-based, list-based, and hash-based partitioning. Each serves a unique purpose: from partitioning data based on predefined categories to evenly distributing data using hash values, ensuring there are no concentration hotspots.

B. Shard Selection and Ranking in HSSA

In distributed systems, data are spread across multiple shards, making global understanding a herculean task. HSSA alleviates this through the Central Sample Index (CSI). A representative set of documents from every shard is carefully sampled and indexed in the CSI. This curated index serves as a microcosm of the entire dataset, providing a bird's eye view of the information sprawl across all shards. By relying on this subset, HSSA can rapidly ascertain the potential relevance of vast amounts of data without trawling through every individual

shard. The efficiency of the CSI comes to the forefront when a query is dispatched. Rather than wandering aimlessly across shards, the query is first mapped against the CSI. Given that the CSI encapsulates samples from all shards, it is adept at offering an initial assessment of which shards are likely to be most relevant to the query in question. This preliminary direction ensures that search operations are both swift and pinpointed.

After the identification of potential shards, HSSA takes an additional step. It arranges these shards in a hierarchy, dictated by their estimated relevance to the posed query. By directing its computational prowess towards the highest-ranking shards first, HSSA ensures optimal resource utilization, leading to faster and more relevant search results. HSSA's architectural brilliance doesn't just reside in its native capabilities but also in its adaptability. While it carries a unique approach, it pays homage to time-tested shard selection algorithms like CORI and RedDE. When scenarios demand it, HSSA can seamlessly amalgamate these methods into its operations. This synergy not only adds another layer of robustness to HSSA but also ensures its relevance across a spectrum of use-cases, making it a truly versatile tool in the domain of distributed data systems. CSI stands as a monumental pillar in the HSSA's framework, offering a streamlined approach to shard selection and relevance estimation.

C. Integrating Static and Dynamic Approaches

To combine the static and dynamic shard selection algorithms, we can use a linear combination of the algorithms. A weighted approach allows for prioritizing the strengths of each method based on their performance on the dataset.

$$HSSA(q,s) = \alpha \times (Lex(s,q) + Conn(s)) + \beta \times (CoRI(q,s) + RedDE(s,q) + Rank - s(s,q)) \quad (1)$$

where α and β are weights representing the importance of the static and dynamic approaches, respectively.

The determination of α and β is vital. If, for example, static approaches (Lex-S and Conn-S) provide consistently reliable results in a particular dataset, α might be higher than β . Alternatively, if real-time adaptability and responsiveness are more crucial, then β might weigh more. To further fine-tune the hybrid approach:

$$HSSA(q,s) = \alpha_1 \times Lex(s,q) + \alpha_2 \times Conn(s) + \beta_1 \times CoRI(q,s) + \beta_2 \times RedDE(s,q) + \beta_3 \times Rank - s(s,q) \quad (2)$$

In this expanded equation, each algorithm has its weight ($\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3$), allowing for granular control over the integration of each method into the hybrid approach. These weights can be optimized through several methods. In this study, empirical analysis was opted by running the HSSA with different weights on validation data and choosing the weights that offer the best results.

Algorithm 1: Proposed HSSA Algorithm

Require: List of shards retrieved from distributed storage systems.

Ensure: Selected shards which are relevant to the user query.

```
function hybrid_shard_selection(query):
    relevant_shards = [ ]
```

```
    static_selected_shards =
static_shard_selection(query)
    dynamic_selected_shards =
dynamic_shard_selection(query)
    combined_shards =
merge(static_selected_shards,
dynamic_selected_shards)
    relevant_shards =
rank_shards(combined_shards)
    return relevant_shards
function static_shard_selection(query):
    relevant_shards = [ ]
    for shard in shards:
        if shard.meets_static_criteria(query):
            relevant_shards.append(shard)
    return relevant_shards
function dynamic_shard_selection(query):
    relevant_shards = [ ]
    for shard in shards:
        if
shard.meets_dynamic_criteria(query):
            relevant_shards.append(shard)
    return relevant_shards
function merge(list1, list2):
// Combine two lists of shards while
removing duplicates
merged_list = list1 + list2
merged_list =
remove_duplicates(merged_list)
    return merged_list
function rank_shards(shards):
// Rank shards based on relevance score
ranked_shards =
sort_shards_by_score(shards)
    return ranked_shards
```

IV. RESULTS AND DISCUSSION

The experiment was meticulously designed to ensure the comprehensive evaluation of HSSA against prevalent shard selection algorithms. The metrics chosen offered a holistic view of the algorithm's efficiency, effectiveness, and potential applicability in real-world scenarios.

A. Performance Analysis of HSSA Considering Latency

Figure 2 shows how different shard selection algorithms behave in terms of latency (ms) as the maximum number of shards (K) increases. The x-axis represents the maximum number of shards, whereas the y-axis denotes latency. It is evident that as K increases, latency generally tends to rise for all algorithms. This is expected as a larger value of K implies that more shards are considered, leading to higher computation time. However, HSSA consistently outperforms the other algorithms across all K values, achieving significantly lower latency. At K = 80, HSSA registers a latency of 21 ms, whereas the closest competitor, Rank-S, records a latency of 28 ms, underlining a clear efficiency gain. In contrast, CORI and Redde peak with latency values of 30 ms at 80 shards, with Lex-S and Conn-S escalating by 25 ms and 26 ms respectively.

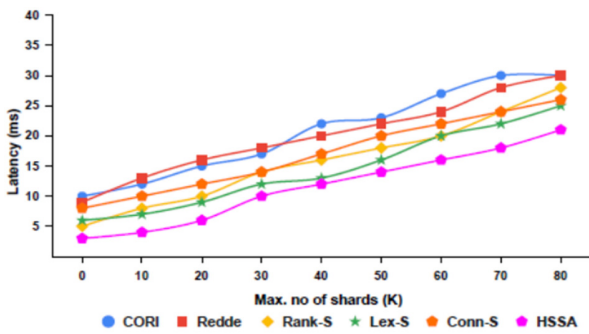


Fig. 2. Latency vs. maximum number of shards.

B. Average Document Score

Figure 3 shows the performance of various shard selection algorithms evaluated based on the average document score, a critical metric that signifies the average relevance of documents retrieved for specific queries. The proposed HSSA stands out distinctly, consistently outperforming its counterparts across all K values. For instance, at $K = 80$, HSSA achieves a score of 9.2, a testament to its efficacy. The underlying hybrid nature of HSSA, which seamlessly merges the merits of both static and dynamic shard selection methodologies, plays a pivotal role in its superior performance. Lex-S, while initiating on a modest note, exhibits a commendable growth, closely following HSSA with a score of 8.1 at $K = 80$. Rank-S, although beginning at 1.2, steadily ascends to 7.8 for $K = 80$. On the other hand, CORI and Redde display moderate growth trajectories, achieving scores of 6 and 6.5, respectively at $K = 80$. Conn-S, maintaining a consistent performance gradient, peaks at 7.5 for $K = 80$. To encapsulate, while each algorithm presents its unique advantages and growth trajectories, HSSA emerges as a clear frontrunner, underscoring its potential to set new standards in the domain of distributed processing systems.

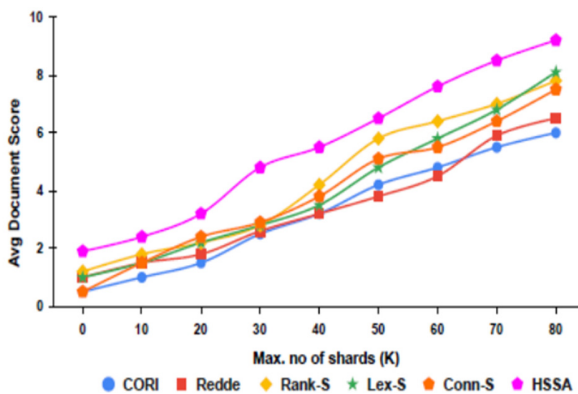


Fig. 3. Average document score vs. maximum number of shards.

C. Comparative Study Analysis

Figure 4 shows the Mean Average Precision (MAP), resource utilization cost and the latency for 60 queries on Gov2 dataset. We can observe that the MAP score of the proposed hybrid shard selection is high whereas the cost and latency are lower than those of the existing algorithms.

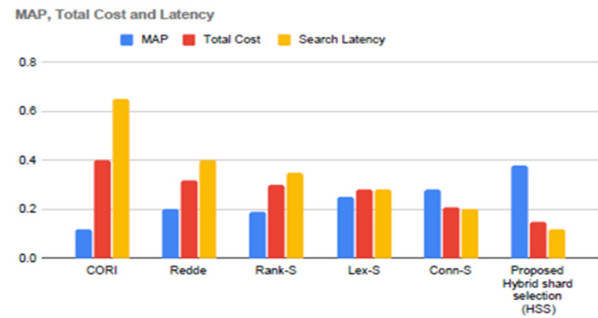


Fig. 4. MAP, total cost, and latency for the compared shard algorithms.

Table I shows the proposed algorithm's performance analysis compared to existing shard selection methods. The consistent precision at multiple cut-off points ($P@10$, $P@30$, and $P@60$) across all assessed K-values of HSSA substantiates its efficacy in discerning and retrieving relevant documents, even as the allowable shards burgeon. Its robust hybrid nature, amalgamating static and dynamic shard selection approaches, underpins its superior performance by ensuring relevance and minimized computational overhead.

TABLE I. PERFORMANCE ANALYSIS OF SHARD SELECTION ALGORITHMS CONSIDERING PRECISION, COST, AND LATENCY PARAMETERS

Shard selection algorithm	No. of shards	Search cost					
		P@30	P@60	NDCG@100	MAP	C _{Total}	C _{Latency}
Exhaustive	50	0.53	0.47	0.41	0.27	2.87	2.89
	100	0.53	0.47	0.41	0.27	2.87	0.96
	150	0.53	0.47	0.41	0.27	2.87	0.58
	200	0.53	0.47	0.41	0.27	2.87	0.51
CORI	50	0.28	0.3	0.24	0.14	0.12	0.12
	100	0.36	0.38	0.32	0.21	0.32	0.21
	150	0.42	0.42	0.37	0.25	0.41	0.24
	200	0.51	0.48	0.4	0.26	0.48	0.26
ReDDE	50	0.38	0.31	0.25	0.13	0.12	0.12
	100	0.47	0.44	0.36	0.23	0.36	0.18
	150	0.52	0.46	0.39	0.25	0.47	0.21
	200	0.56	0.51	0.43	0.28	0.52	0.28
Rank-S	50	0.34	0.3	0.2	0.16	0.14	0.12
	100	0.45	0.34	0.29	0.28	0.38	0.19
	150	0.49	0.45	0.34	0.29	0.48	0.21
	200	0.51	0.48	0.41	0.4	0.52	0.28
Modified Rank-S	50	0.34	0.3	0.2	0.16	0.14	0.16
	100	0.45	0.34	0.29	0.28	0.25	0.14
	150	0.49	0.45	0.34	0.29	0.3	0.18
	200	0.51	0.48	0.41	0.4	0.32	0.21
Proposed HSSA	50	0.3	0.31	0.3	0.28	0.16	0.15
	100	0.28	0.3	0.28	0.27	0.2	0.12
	150	0.27	0.28	0.26	0.25	0.21	0.15
	200	0.23	0.26	0.25	0.24	0.24	0.16

V. CONCLUSION

This paper presents a new approach to the distributed data processing problem: the Hybrid Shard Selection Algorithm (HSSA). The existing shard selection algorithms focus on search efficiency without significantly reducing search effectiveness. They may suffer from high query latency and load imbalance when the size of the database is enormous. The proposed hybrid shard selection approach partitions the

documents into topic-wise shards and uses cluster-skipping indexes inside each of the shards during the query processing operation. Only those shards that have relevant documents matching the user query are searched. This reduces the latency and enhances the search effectiveness by reducing the resource utilization cost. We tested the proposed algorithm on the Gov2 dataset partitioned into topic-based shards. The proposed HSSA algorithm search accuracy and effectiveness are significantly better than those of the baseline existing shard selection algorithms. At the same time, the search cost of the proposed algorithm is lower with optimized resource utilization. It reduced the search cost by 20% for the Gov2 dataset. The proposed algorithm supports query-specific minimal shard cut-off estimation, which can be used with different shard ranking algorithms to satisfy early precision or high recall goals of information retrieval applications.

REFERENCES

- [1] N. Venkateswaran and S. Changder, "Simplified data partitioning in a consistent hashing based sharding implementation," in *TENCON 2017 - 2017 IEEE Region 10 Conference*, Penang, Malaysia, Aug. 2017, pp. 895–900, <https://doi.org/10.1109/TENCON.2017.8227985>.
- [2] A. Kulkarni, A. S. Tigelaar, D. Hiemstra, and J. Callan, "Shard ranking and cutoff estimation for topically partitioned collections," in *Proceedings of the 21st ACM international conference on Information and knowledge management*, New York, NY, USA, Jul. 2012, pp. 555–564, <https://doi.org/10.1145/2396761.2396833>.
- [3] J. Kamal, M. Murshed, and R. Buyya, "Workload-aware incremental repartitioning of shared-nothing distributed databases for scalable OLTP applications," *Future Generation Computer Systems*, vol. 56, pp. 421–435, Mar. 2016, <https://doi.org/10.1016/j.future.2015.09.024>.
- [4] H. R. Mohammad, K. Xu, J. Callan, and J. S. Culpepper, "Dynamic Shard Cutoff Prediction for Selective Search," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, New York, NY, USA, Mar. 2018, pp. 85–94, <https://doi.org/10.1145/3209978.3210005>.
- [5] P. M. Dhulavvagol, V. H. Bhajantri, and S. G. Totad, "Performance Analysis of Distributed Processing System using Shard Selection Techniques on Elasticsearch," *Procedia Computer Science*, vol. 167, pp. 1626–1635, Jan. 2020, <https://doi.org/10.1016/j.procs.2020.03.373>.
- [6] Z. Dai, C. Xiong, and J. Callan, "Query-Biased Partitioning for Selective Search," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, New York, NY, USA, Jul. 2016, pp. 1119–1128, <https://doi.org/10.1145/2983323.2983706>.
- [7] A. Kulkarni and J. Callan, "Selective Search: Efficient and Effective Search of Large Textual Collections," *ACM Transactions on Information Systems*, vol. 33, no. 4, pp. 17:1–17:33, Dec. 2015, <https://doi.org/10.1145/2738035>.
- [8] P. M. Dhulavvagol, S. G. Totad, and S. Sourabh, "Performance Analysis of Job Scheduling Algorithms on Hadoop Multi-cluster Environment," in *Emerging Research in Electronics, Computer Science and Technology*, Singapore, 2019, pp. 457–470, https://doi.org/10.1007/978-981-13-5802-9_42.
- [9] N. C. Kundur, B. C. Anil, P. M. Dhulavvagol, R. Ganiger, and B. Ramadoss, "Pneumonia Detection in Chest X-Rays using Transfer Learning and TPUs," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11878–11883, Oct. 2023, <https://doi.org/10.48084/etasr.6335>.
- [10] E. Rodrigues and R. Morla, "Run Time Prediction for Big Data Iterative ML Algorithms: a KMeans case study," Oct. 2017.
- [11] M. Ali, N. Q. Soomro, H. Ali, A. Awan, and M. Kirmani, "Distributed File Sharing and Retrieval Model for Cloud Virtual Environment," *Engineering, Technology & Applied Science Research*, vol. 9, no. 2, pp. 4062–4065, Apr. 2019, <https://doi.org/10.48084/etasr.2662>.
- [12] N. Jayakumar and A. M. Kulkarni, "A Simple Measuring Model for Evaluating the Performance of Small Block Size Accesses in Lustre File System," *Engineering, Technology & Applied Science Research*, vol. 7, no. 6, pp. 2313–2318, Dec. 2017, <https://doi.org/10.48084/etasr.1557>.

Modeling of Mass Transfer and Reaction Kinetics in ZnO Nanoparticle Micro-Reactor Systems for AMX and DOX Degradation

Nidhal Becheikh

Department of Chemical and Materials Engineering, Engineering College, Northern Border University, Saudi Arabia

nidhal.becheikh@nbu.edu.sa (corresponding author)

Received: 12 January 2024 | Revised: 19 February 2024 | Accepted: 28 February 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6898>

ABSTRACT

This study aims to model the coupled phenomena of photocatalytic reaction and mass transfer in the degradation of Amoxicillin (AMX) and Doxycycline (DOX) using Zinc oxide (ZnO) nanoparticles within microreactor systems. The objective is to gain a comprehensive understanding of the dynamic interaction between the photocatalytic degradation kinetics and the mass transfer processes to optimize the conditions for efficient antibiotic removal from contaminated water. This involves characterizing the reaction kinetics via the Langmuir-Hinshelwood model, estimating the mass transfer coefficients, and analyzing the effects of axial dispersion to ensure the accurate determination of intrinsic kinetic constants and minimize mass transfer limitations. This study used a syringe pump to ensure a consistent flow of antibiotic solution into the microreactor. The results indicate that AMX reaches adsorption equilibrium more rapidly than DOX, corresponding to its faster photocatalytic degradation kinetics and higher final conversion rate (89% for AMX, 86% for DOX). The mass transfer coefficient (k_d) was estimated using the Sherwood number, derived from three different models, with the constant Sherwood model best fitting the R1 microreactor data. An analysis of the Damköhler number (Da_{II}) indicates that high flow rates minimize mass transfer limitations in the R1 microreactor, allowing the determination of near-intrinsic kinetic constants. On the contrary, at low flow rates, kinetic constants are apparent as a result of mass-transfer limitations. The study concludes that higher flow rates (≥ 10 mL/h) in the R1 microreactor are preferable to approach intrinsic kinetics and reduce mass transfer limitations during photocatalytic degradation of antibiotics. These findings underscore the potential of ZnO-based oxidation processes in treating antibiotic-contaminated water with optimized conditions, providing a pathway for efficient and sustainable wastewater treatment.

Keywords-modeling; photocatalytic degradation; micro-reactors; kinetic constants; mass transfer limitations

I. INTRODUCTION

Water pollution causes or exacerbates the degradation of its natural qualities by altering its physical, chemical, biological, or bacteriological characteristics. This contamination comes from various anthropogenic sources, including urban, agricultural, or industrial processes [1-3]. Pollution disturbs the living conditions of fauna and flora. To solve the growing problem of pollution, strict measures are implemented to regulate water quality for human consumption [4]. These regulations are constantly adapted according to technological advances in effluent treatment processes and analysis strategies [5]. Many studies have shown the presence of various antibiotics in different environmental elements, such as wastewater from treatment plants, farms, waterways, sediments, and soils [6]. In [7], the trends and factors that influence antibiotic consumption were analyzed between 2000 and 2015 in 76 countries. According to this study, global antibiotic consumption increased by 65% between 2000 and

2015, from 21.1 to 34.8 billion Defined Daily Doses (DDD). During the same period, the rate of antibiotic consumption rose by 39%, from 11.3 to 15.7 DDD per 1,000 inhabitants per day. It was also estimated that by 2030, global antibiotic consumption will have increased by 32%, reaching approximately 55.6 billion DDD [7-9].

Traditional elimination techniques, such as adsorption, coagulation, filtration, biodegradation, or reverse osmosis, display certain limitations, either in terms of high cost or in terms of incomplete removal of pollutants. Therefore, it is essential to find suitable alternative technologies to remove toxic and biologically persistent antibiotics from wastewater [10]. Advanced Oxidation Processes (AOPs) have been developed as a highly effective alternative for converting organic materials into simple mineral products. Commonly used methods include Fenton reactions, ozonation, electrochemistry, photocatalysis, and ultrasonic. Among AOPs, heterogeneous photocatalysis is the most promising solution for environmental decontamination due to its nonselective nature,

high efficiency in removing various pollutants from wastewater, including antibiotics, and its ability to effectively reduce toxicity [11-13].

AOPs rely on in situ production of highly reactive and oxidizing radical species, like peroxy organic radicals, hydroxyl radicals, molecular singlet oxygen, or superoxide anion radicals [14]. These species interact with organic pollutants and lead to their fragmentation. The hydroxyl radical is particularly reactive and relatively non-selective, has a powerful oxidation capacity, and is the second most reactive species after the fluorine atom, with a redox potential of 2.8 eV and a marked electrophilic character [15]. In this wide range of techniques, two main categories can be distinguished: homogeneous and heterogeneous phase processes. These two classes can also be subdivided into two subcategories depending on whether or not energy is used to enhance the production of oxidative radical species and thus water treatment. Heterogeneous photocatalysis has been applied to the degradation of various pollutants, namely pesticides, insecticides, and dyes, which can be completely oxidized to CO_2 and H_2O [16-17]. Photocatalytic reactions involve the participation of one or more chemical compounds, a photocatalyst, and a photon flux. Numerous physicochemical parameters, such as the concentration of dissolved oxygen, temperature, pH, crystal structure, and specific surface, have also been identified to influence photocatalytic kinetics. Extensive studies have been conducted on these physicochemical parameters to better understand their impact on the kinetics of photocatalytic reactions [18-19].

Zinc oxide (ZnO) is one of the semiconductor materials widely employed in photocatalytic applications because of its various advantages, such as non-toxicity, ecological properties, natural abundance, low cost, photochemical stability, high surface area, and straightforward preparation techniques. However, the rapid recombination of photoexcited e^-/h^+ pairs, the high bandgap (~3.3 eV), and the required excitation energy of 60 meV are inherent obstacles to achieving higher photocatalytic activity from ZnO nanostructures [20-21]. The use of ZnO for photocatalytic degradation of antibiotics has been investigated in several studies, encompassing a variety of ZnO forms, including nanoparticles, microparticles, and multiple composites [22-24]. In [25-26], it was shown that ZnO effectively degrades antibiotic pollutants, like tetracycline hydrochloride (TCH), ciprofloxacin (CIP), and sulfamethoxazole (SMX). The efficiency of ZnO in the photocatalytic degradation process is affected by factors, such as its structural form, environmental reaction conditions (including pH and flow rate), and interaction with other photocatalytic materials. The degradation mechanism is primarily facilitated by reactive species, namely superoxide radicals (O_2^-) and holes (h^+). Typically, the degradation kinetics of antibiotics via ZnO adhere to pseudo-first-order kinetics. The stability and recyclability of ZnO-based photocatalysts are commendable, underscoring their potential for sustainable use in wastewater treatment [27-28].

However, the modeling of mass transfer and photocatalytic reactions has been investigated. Several studies have proposed mathematical and numerical models to understand the kinetics

and parameters that influence the reaction [29-30]. These models consider factors, such as Langmuir-Hinshelwood (LH) kinetics, mass transfer theory, and the relative contributions of mass transport and intrinsic kinetics to the apparent reaction rate [1, 31]. Despite the significant number of proposed models that allow coupling photocatalytic reactions and mass transfer, the extrapolation of these models is generally limited by a set of considerations, involving the configuration of the reactor or microreactor, the type and concentration of the pollutant, the catalyst, and the hydrodynamic conditions. This study falls within this theme and aims to model the photocatalytic degradation of two antibiotic pollutants, Amoxicillin (AMX) and Doxycycline (DOX), by determining the kinetic constants k and adsorption constants K of the LH model.

II. EXPERIMENTAL METHOD

A. Pollutants: AMX and DOX

The pollutants tested in the study belong to two distinct classes of antibiotics, which are differentiated by their unique molecular structures. These classes are beta-lactams and tetracyclines. The specific pollutants examined in this case are AMX and DOX, and Table I depicts their characteristics.

TABLE I. PHYSICOCHEMICAL PROPERTIES OF PHARMACEUTICAL PRODUCTS.

Antibiotic	Class	Formula	Molecular weight (g/mol)	λ_{max} (nm)
Amoxicillin	Beta-lactams	$\text{C}_{16}\text{H}_{19}\text{N}_3\text{O}_5\text{S}$	365.4	230 and 270
Doxycycline	Tetracycline	$\text{C}_{22}\text{H}_{24}\text{N}_2\text{O}_8$	444.4	274 and 346

A Thermo Scientific Evolution 300 model spectrophotometer, which offers a spectral range of 190 to 1100 nm, was put into service. Its precision is ± 1 nm, its reproducibility is 0.1 nm, and its spectrum scanning speed is 240 nm/min. A 10 mm quartz cuvette was used for the measurement. Two lamps were employed: a mercury lamp to serve as the main reference standard to validate and calibrate the instrument, and a xenon lamp, which was activated only during measurements. Blanks were performed utilizing ultrapure water in a quartz cuvette identical to the one used for the samples. Residual concentrations were determined by interpolation deploying calibration curves or spectra $A = f(\lambda)$. AMX and DOX antibiotics were utilized for photocatalytic degradation, which absorb in UV-visible at wavelengths of 230 and 270 nm and 275 and 347 nm, as observed in Figures 1 and 2, respectively. These wavelengths were implemented to monitor the degradation kinetics by UV-visible spectroscopy.

B. Photocatalytic Degradation

The experimental setup consists of a microreactor where the photocatalytic reaction occurs, a UV lamp that irradiates the microchannel, and a syringe pump that ensures the flow of an aqueous solution containing the pollutant (AMX or DOX) at a known concentration ($C = 10$ mg/L) into the microreactor. The microreactors had ZnO nanoparticles dispersed in a water/ethanol solution (30% ethanol). The surface of the microchannel was covered by the water/ethanol solution containing the ZnO particles. The microreactor was then placed in an oven at 60°C for one hour to ensure drying. Figure 3

exhibits the experimental assembly applied. The microreactors were manufactured in polymer with different geometries (width, depth, and length) using a 3D printer to study the photocatalytic activity as a function of the microchannel characteristics. Table II lists the different characteristic dimensions of the microreactors.

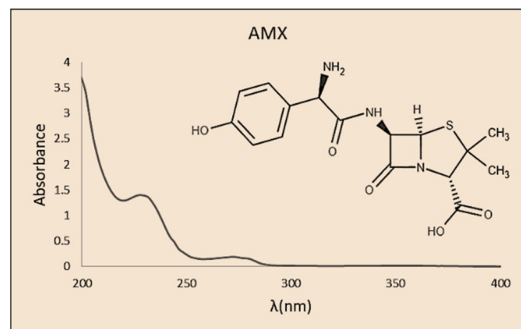


Fig. 1. UV absorption spectrum of AMX at 10 mg/L.

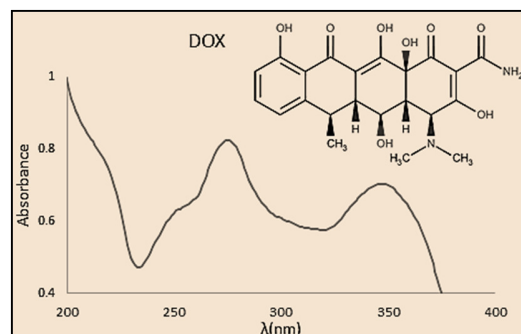


Fig. 2. UV absorption spectrum of DOX at 10 mg/L.

TABLE II. DIFFERENT DIMENSIONS OF MICROREACTORS

	Width w (mm)	Depth d (mm)	Length L (mm)	Volume V (mm ³)
R1	1	2	120	240
R2	3	2	120	720
R3	2	3	120	720

III. RESULTS AND DISCUSSION

A. Results of Photocatalytic Degradation

The syringe pump, demonstrated in Figure 3, was employed to inject a solution containing an initial antibiotic concentration (C_e) of 10 mg/L into the microreactor. Once the flow rate was set, the syringe pump ensured regular injection of the solution. At regular time intervals, samples were taken from the treated solution and their concentration was measured utilizing UV-visible spectroscopy. These measurements allow the determination of the exit concentration (C_{so}) of the antibiotic at a given time, which corresponds to the experimental time required to obtain a representative sample. The irradiation of the ZnO photocatalyst was initiated after the establishment of the adsorption-desorption equilibrium, which occurs at 52 minutes. Figure 4 portrays the degradation of AMX and DOX antibiotics on the ZnO photocatalyst before and after the initiation of irradiation.

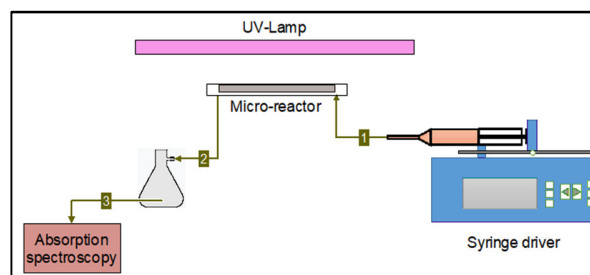


Fig. 3. Experimental setup for photocatalytic degradation in microreactor.

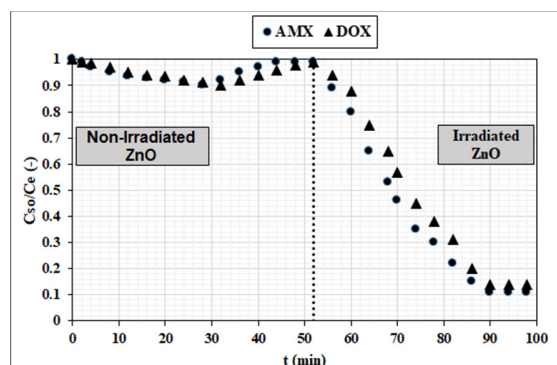


Fig. 4. Evolution of C_{so}/C_e in R1 at a flow rate of 2.5 mL/h, in the absence and with irradiation. $C_e = 10$ mg/L.

At a fixed flow rate and identical initial concentrations, the times required to reach the adsorption equilibrium states differ between the various antibiotics. In this case, the adsorption equilibrium states of AMX are established after 40 minutes, while the adsorption-desorption equilibrium of DOX, as well as the mixture of antibiotics, requires 52 minutes. This suggests that the adsorption processes of each antibiotic have different rates, leading to different equilibrium times. Irradiation instantly causes a drop in AMX and DOX concentrations, achieving final conversion rates of 89% and 86%, respectively. It is important to note that the kinetics of photocatalytic degradation for AMX and DOX are different. Moreover, the maximum conversion of AMX is notably significant. This behavior is similar to the concentration fluctuations observed during the adsorption-desorption processes of antibiotics. The processes of adsorption, desorption, and photocatalytic reaction proceed concurrently. Because of the faster adsorption kinetics of AMX, the molecules adsorbed on the catalyst surface rapidly oxidize and decompose into byproducts.

B. Modeling of the Photocatalytic Reaction

Using this axial dispersion plug flow reactor model, it is possible to account for the effects of axial dispersion in the reactor, which allows for a more accurate representation of reality compared to the simplified model of the ideal plug flow reactor. This is particularly important when significant concentration variations occur in the reactor due to the effects of diffusion, convection, or non-ideality. The balance equation in a plug flow reactor with axial dispersion is written as follows:

$$D_A \frac{d^2 C}{dx^2} = U \frac{dC}{dx} + r \quad (1)$$

Considering the concentration variation between the reaction medium and the catalytic surface, the LH model can be applied to describe the apparent rate of the photocatalytic reaction.

$$r_a = r_D = K_D a (C_b - C_s) = k_r \frac{K C_s}{1 + K C_s} \quad (2)$$

According to this model, this rate is in balance with the mass transfer rate, denoted as r_D . This equation expresses the specific relationship and allows for the calculation of the surface concentration C_s , the conversion rate, and the mass transfer coefficient. In this equation, r_a is the apparent rate of the photocatalytic reaction (mmol/L/s), r_D is the mass transfer rate (mmol/L/s), K_D is the mass transfer coefficient (m/s), a is the specific surface area of the catalytic surface (mm²/mm³), C_s is the surface concentration on the catalytic surface (mmol/L), C_b is the concentration in the reaction phase (mmol/L), and k_r is the rate constant of the chemical reaction (mmol/L/s). Through this equation, it is possible to determine C_s from the concentration in the reaction phase C and other parameters. Furthermore, the conversion rate can be calculated based on C_s and the initial concentration of the reactive species.

$$X = 1 - \exp\left(\frac{L}{2} \left(\frac{U}{D_A} - \sqrt{\left(\frac{U}{D_A}\right)^2 + \frac{4}{D_A} \frac{k_r K k_d}{k_d(1 + K C_s) + k_r K}} \right)\right) \quad (3)$$

The mass transfer coefficient k_d can be estimated from the Sherwood number using:

$$Sh = \frac{k_d d_h}{D_m} \quad (4)$$

Table III illustrates common models utilized to estimate the Sherwood number in microreactors.

TABLE III. MODELS USED TO ESTIMATE THE SHERWOOD NUMBER IN MICRO-REACTORS

Model	Equation
Constant Sherwood Model	$Sh = 3.66$
Graetz-L��v��que (GL)	$Sh = 1.86 \left(\frac{d_h}{L}\right)^{0.33} R_e^{0.33} S_c^{0.33}$
Ranz-L��venspiel (RL)	$Sh = 2 + 1.8 R_e^{0.33} S_c^{0.33}$

In this table, Sh denotes the Sherwood number (dimensionless), d_h is the hydraulic diameter (m), which is equal to $4(w \times d)/(2(w + d))$, L is the characteristic length (m), R_e is the Reynolds number (dimensionless), and S_c is the Schmidt number (dimensionless). Equation 3 was solved by fitting the experimental results engaging an Excel solver. This technique allowed the estimation of the kinetic and adsorption constants. Moreover, the determination of the mass transfer constant was carried out by testing the three models proposed in Table III. Kinetic constant determination was performed by fitting the described models to the experimental results for the micro-reactor R1 as detected in Figure 5. The findings of the conversion rates in the microreactor R1 are accurately modeled by the constant Sherwood model. However, this conclusion cannot be generalized to other microreactors with different hydraulic diameters. To verify the effect of changes in hydraulic diameter on the mass transfer coefficient, a comparison of the experimental conversion rate results in microreactors R2 and R3 and the mass transfer models was carried out, as displayed in Figures 6 and 7.

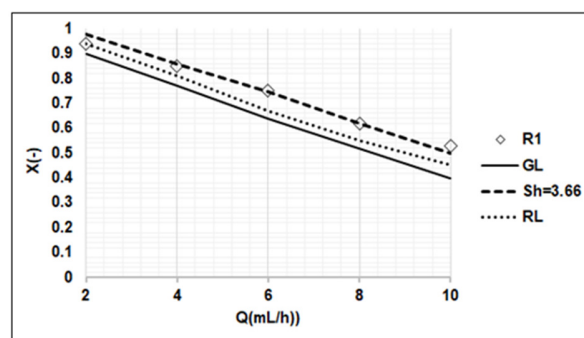


Fig. 5. Experimental and theoretical conversion rates calculated by the models of $Sh=3.66$, GL and RL as a function of flow rate in R1, $C_e = 10$ mg/L, $I = 1.5$ mW/cm², $K = 75 \cdot 10^{-4}$ L/mmol and $k_r=50$ mmol/L.s.

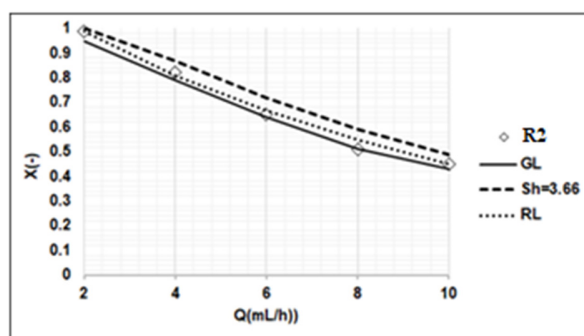


Fig. 6. Experimental and theoretical conversion rates calculated by the models of $Sh=3.66$, GL, and RL as a function of flow rate in R2, $C_e = 10$ mg/L, $I = 1.5$ mW/cm², $K = 75 \cdot 10^{-4}$ L/mmol, and $k_r=50$ mmol/L.s.

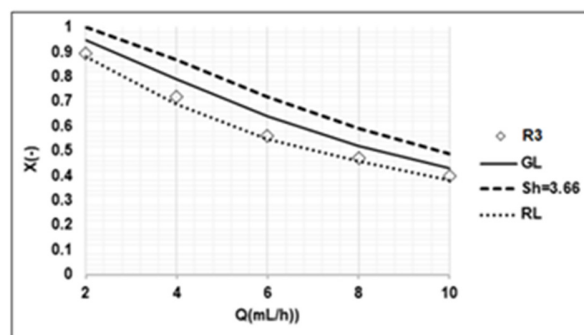


Fig. 7. Experimental and theoretical conversion rates calculated by the models of $Sh=3.66$, GL and RL as a function of flow rate in R3, $C_e = 10$ mg/L, $I = 1.5$ mW/cm², $K = 75 \cdot 10^{-4}$ L/mmol, and $k_r=50$ mmol/L.s.

The conversion rate outcomes in the microreactors R2 and R3 were accurately modeled as a function of the flow rates by the RL model. It is interesting to note that the model allowed the linking of mass transfer to the flow regime via the Reynolds number. Indeed, at different flow rates, it was observed that the model with $Sh = 3.66$ does not accurately reproduce the experimental conversion rates.

C. Influence of Mass Transfer

To ensure that the kinetic data measured for the degradation of AMX and DOX accurately reflect the intrinsic kinetics, it is necessary to analyze the mass transfer limitations. In the context of microchannel reactors, if the heterogeneous

Damköhler number (Da_{II}) is below 0.1, the error in the measured rate constants is kept below 3%. This condition allows for the assumption that these constants are a valid representation of the intrinsic kinetics of the photocatalytic degradation of AMX and DOX.

$$Da_{II} = \frac{k_r}{a.k_d./K+a.k_d./C_b} \quad (8)$$

Figure 8 demonstrates the values of the Damköhler number Da_{II} versus the flow rate for different microreactors. Analysis of the variation of the Damköhler number as a function of flow rate indicates that the photocatalytic degradation in microreactor R1 exhibits a slight limitation by mass transfer and that the experimentally determined kinetic constants at high volumetric flows are close to the intrinsic constants. On the other hand, for low flow rates, kinetic constants are considered apparent. Similarly, for microreactors R2 and R3, the photocatalytic degradation process is limited by mass transfer. In conclusion, using the R1 at high flow will allow overcoming the limitation of mass transfer.

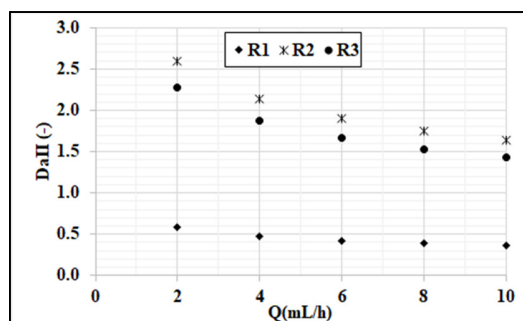


Fig. 8. Damköhler number Da_{II} for different micro-reactors vs flow rate.

IV. CONCLUSION

This study models the mass transfer and reaction kinetics of ZnO nanoparticle microreactor systems, focusing on the degradation of AMX and DOX antibiotics. The former provides a deep understanding of the dynamic interaction between photocatalytic degradation kinetics and mass transfer processes, which is vital for optimizing antibiotic removal from contaminated water. Three microreactors were manufactured in polymer with different geometries using a 3D printer to study the photocatalytic activity as a function of the microchannel characteristics. The results reveal distinct adsorption and degradation behavior for AMX and DOX, with AMX reaching adsorption equilibrium more swiftly and achieving a higher conversion rate of 89%, compared to the conversion rate of 86% for DOX. Employing this axial dispersion plug flow reactor model, it is possible to account for the effects of axial dispersion in the reactor, which allows for a more accurate representation of reality compared to the simplified model of the ideal plug flow reactor. This is particularly important when significant concentration variations occur in the reactor due to the effects of diffusion, convection, or non-ideality. The determination of the conversion rate was solved by fitting the experimental results deploying an Excel solver. This technique allowed for the estimation of the kinetic and adsorption constants. Moreover, the determination of the mass transfer

constant was carried out by testing three different Sherwood models. The conversion rate results in the microreactor R1 are accurately modeled by the constant Sherwood model. However, the R2 and R3 microreactors were accurately modeled using the RL model. The analysis of the variation of the Damköhler number discloses that the photocatalytic degradation in the R1 microreactor exhibits a minor limitation by mass transfer and that the experimentally determined kinetic constants at high volumetric flows are close to the intrinsic constants. On the other hand, for low flow rates, the determined kinetic constants are considered apparent. For microreactors R2 and R3, the photocatalytic degradation process is limited by mass transfer. In conclusion, utilizing the R1 microreactor at high flow rates (≥ 10 mL/h) enables overcoming the mass transfer limitation.

V. ACKNOWLEDGMENTS

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through project number NBU-FFR-2024- 2933-01.

REFERENCES

- [1] Md. G. Uddin, S. Nash, and A. I. Olbert, "A review of water quality index models and their use for assessing surface water quality," *Ecological Indicators*, vol. 122, Mar. 2021, Art. no. 107218, <https://doi.org/10.1016/j.ecolind.2020.107218>.
- [2] N. Akhtar, M. I. Syakir Ishak, S. A. Bhawani, and K. Umar, "Various Natural and Anthropogenic Factors Responsible for Water Quality Degradation: A Review," *Water*, vol. 13, no. 19, Jan. 2021, Art. no. 2660, <https://doi.org/10.3390/w13192660>.
- [3] F. Zenati, A. Djellali, and D. Sarker, "Wastewater Assessment and Biochemical Oxygen Demand Value Prediction from Mining Operations: A Case Study," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10754–10758, Jun. 2023, <https://doi.org/10.48084/etasr.5721>.
- [4] D. S. Malik, A. K. Sharma, A. K. Sharma, R. Thakur, and M. Sharma, "A review on impact of water pollution on freshwater fish species and their aquatic environment," in *Advances in Environmental Pollution Management: Wastewater Impacts and Treatment Technologies*, Haridwar, India: Agro Environ Media, 2020, pp. 10–28.
- [5] R. Alduina, "Antibiotics and Environment," *Antibiotics*, vol. 9, no. 4, Apr. 2020, Art. no. 202, <https://doi.org/10.3390/antibiotics9040202>.
- [6] M. Bilal, S. Mehmood, T. Rasheed, and H. M. N. Iqbal, "Antibiotics traces in the aquatic environment: persistence and adverse environmental impact," *Current Opinion in Environmental Science & Health*, vol. 13, pp. 68–74, Feb. 2020, <https://doi.org/10.1016/j.coesh.2019.11.005>.
- [7] E. Y. Klein *et al.*, "Global increase and geographic convergence in antibiotic consumption between 2000 and 2015," *Proceedings of the National Academy of Sciences*, vol. 115, no. 15, pp. E3463–E3470, Apr. 2018, <https://doi.org/10.1073/pnas.1717295115>.
- [8] A. Z. Al Meslamani, "Antibiotic resistance in low- and middle-income countries: current practices and its global implications," *Expert Review of Anti-infective Therapy*, vol. 21, no. 12, pp. 1281–1286, Dec. 2023, <https://doi.org/10.1080/14787210.2023.2268835>.
- [9] P. Kovalakova, L. Cizmas, T. J. McDonald, B. Marsalek, M. Feng, and V. K. Sharma, "Occurrence and toxicity of antibiotics in the aquatic environment: A review," *Chemosphere*, vol. 251, Jul. 2020, Art. no. 126351, <https://doi.org/10.1016/j.chemosphere.2020.126351>.
- [10] R. N. Abbas and A. S. Abbas, "The Taguchi Approach in Studying and Optimizing the Electro-Fenton Oxidation to Reduce Organic Contaminants in Refinery Wastewater Using Novel Electrodes," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 8928–8935, Aug. 2022, <https://doi.org/10.48084/etasr.5091>.

- [11] J. Wang and S. Wang, "Reactive species in advanced oxidation processes: Formation, identification and reaction mechanism," *Chemical Engineering Journal*, vol. 401, Dec. 2020, Art. no. 126158, <https://doi.org/10.1016/j.cej.2020.126158>.
- [12] J. Wang and R. Zhuan, "Degradation of antibiotics by advanced oxidation processes: An overview," *Science of The Total Environment*, vol. 701, Jan. 2020, Art. no. 135023, <https://doi.org/10.1016/j.scitotenv.2019.135023>.
- [13] S. Li *et al.*, "Antibiotics degradation by advanced oxidation process (AOPs): Recent advances in ecotoxicity and antibiotic-resistance genes induction of degradation products," *Chemosphere*, vol. 311, Jan. 2023, Art. no. 136977, <https://doi.org/10.1016/j.chemosphere.2022.136977>.
- [14] P. Krystynik, "Advanced Oxidation Processes (AOPs) – Utilization of Hydroxyl Radical and Singlet Oxygen," in *Reactive Oxygen Species*, IntechOpen, 2021.
- [15] C. Amor, J. R. Fernandes, M. S. Lucas, and J. A. Peres, "Hydroxyl and sulfate radical advanced oxidation processes: Application to an agro-industrial wastewater," *Environmental Technology & Innovation*, vol. 21, Feb. 2021, Art. no. 101183, <https://doi.org/10.1016/j.eti.2020.101183>.
- [16] A. G. Akerdi and S. H. Bahrami, "Application of heterogeneous nano-semiconductors for photocatalytic advanced oxidation of organic compounds: A review," *Journal of Environmental Chemical Engineering*, vol. 7, no. 5, Oct. 2019, Art. no. 103283, <https://doi.org/10.1016/j.jece.2019.103283>.
- [17] K. Bano, S. Kaushal, and P. P. Singh, "A review on photocatalytic degradation of hazardous pesticides using heterojunctions," *Polyhedron*, vol. 209, Nov. 2021, Art. no. 115465, <https://doi.org/10.1016/j.poly.2021.115465>.
- [18] S. Bagheri, A. TermehYousefi, and T.-O. Do, "Photocatalytic pathway toward degradation of environmental pharmaceutical pollutants: structure, kinetics and mechanism approach," *Catalysis Science & Technology*, vol. 7, no. 20, pp. 4548–4569, Oct. 2017, <https://doi.org/10.1039/C7CY00468K>.
- [19] I. Khan, K. Saeed, N. Ali, I. Khan, B. Zhang, and M. Sadiq, "Heterogeneous photodegradation of industrial dyes: An insight to different mechanisms and rate affecting parameters," *Journal of Environmental Chemical Engineering*, vol. 8, no. 5, Oct. 2020, Art. no. 104364, <https://doi.org/10.1016/j.jece.2020.104364>.
- [20] A. Mir, N. Becheikh, L. Khezami, M. Bououdina, and A. Ouderni, "Synthesis, Characterization, and Study of the Photocatalytic Activity upon Polymeric-Surface Modification of ZnO Nanoparticles," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12047–12053, Dec. 2023, <https://doi.org/10.48084/etasr.6373>.
- [21] F. H. Abdullah, N. H. H. A. Bakar, and M. A. Bakar, "Current advancements on the fabrication, modification, and industrial application of zinc oxide as photocatalyst in the removal of organic and inorganic contaminants in aquatic systems," *Journal of Hazardous Materials*, vol. 424, Feb. 2022, Art. no. 127416, <https://doi.org/10.1016/j.jhazmat.2021.127416>.
- [22] T. H. Nguyen, T. A. N. Cong, and A.-T. Vu, "Synthesis of Carnation-Like ZnO for Photocatalytic Degradation of Antibiotics, Including Tetracycline Hydrochloride," *Environmental Engineering Science*, vol. 40, no. 8, pp. 329–339, Aug. 2023, <https://doi.org/10.1089/ees.2023.0034>.
- [23] S. Zeinali Heris, M. Etemadi, S. B. Mousavi, M. Mohammadpourfard, and B. Ramavandi, "Preparation and characterizations of TiO₂/ZnO nanohybrid and its application in photocatalytic degradation of tetracycline in wastewater," *Journal of Photochemistry and Photobiology A: Chemistry*, vol. 443, Sep. 2023, Art. no. 114893, <https://doi.org/10.1016/j.jphotochem.2023.114893>.
- [24] M. R. da Silva, D. L. Cunha, and E. M. Saggioro, "Insight into antibiotic degradation through photocatalysis processes: Photocatalysts semiconductors, transformation products and influencing factors," *Research Square*, Jun. 06, 2023, <https://doi.org/10.21203/rs.3.rs-2865739/v1>.
- [25] X. Yu *et al.*, "Preparation of ZnO/Cu₂O Composite Particles and Its Degradation of Ciprofloxacin: Analysis of Degradation Performance and Active Species," *Integrated Ferroelectrics*, vol. 236, no. 1, pp. 96–108, Jul. 2023, <https://doi.org/10.1080/10584587.2023.2194831>.
- [26] N. Roy, K. Kannabiran, and A. Mukherjee, "Integrated adsorption and photocatalytic degradation based removal of ciprofloxacin and sulfamethoxazole antibiotics using Fc@rGO-ZnO nanocomposite in aqueous systems," *Chemosphere*, vol. 333, Aug. 2023, Art. no. 138912, <https://doi.org/10.1016/j.chemosphere.2023.138912>.
- [27] R. K. Dharman, A. Mariappan, and T. H. Oh, "Accelerated photocatalytic degradation of sulfonamide antibiotic pollutant using oxygen vacancy in metal-organic framework ZIF-8/Ag₃PO₄ heterostructure," *Surfaces and Interfaces*, vol. 39, Jul. 2023, Art. no. 102998, <https://doi.org/10.1016/j.surfin.2023.102998>.
- [28] M. Sharma *et al.*, "Photocatalytic degradation of four emerging antibiotic contaminants and toxicity assessment in wastewater: A comprehensive study," *Environmental Research*, vol. 231, Aug. 2023, Art. no. 116132, <https://doi.org/10.1016/j.envres.2023.116132>.
- [29] J. C. G. da Silva, J. L. F. Alves, G. D. Mumbach, and M. Di Domenico, "Photocatalytic degradation of ethylene in tubular microreactor coated with thin-film of TiO₂: Mathematical modeling with experimental validation and geometry analysis using computational fluid dynamics simulations," *Chemical Engineering Research and Design*, vol. 196, pp. 101–117, Aug. 2023, <https://doi.org/10.1016/j.cherd.2023.06.036>.
- [30] R. Mondal, "Micro-segregated two-dimensional fluid and mass transport modelling in unsteady rotating annular photocatalytic reactor," *Results in Engineering*, vol. 16, Dec. 2022, Art. no. 100752, <https://doi.org/10.1016/j.rineng.2022.100752>.
- [31] N. A. B. Timmerhuis, J. A. Wood, and R. G. H. Lammertink, "Connecting experimental degradation kinetics to theoretical models for photocatalytic reactors: The influence of mass transport limitations," *Chemical Engineering Science*, vol. 245, Dec. 2021, Art. no. 116835, <https://doi.org/10.1016/j.ces.2021.116835>.

Digital Forensics Readiness Framework (DFRF) to Secure Database Systems

Ahmed Albugmi

Computer and Information Technology Department, The Applied College, King Abdulaziz University, Saudi Arabia

analbugome@kau.edu.sa (corresponding author)

Received: 21 February 2024 | Revised: 2 March 2024 | Accepted: 10 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7116>

ABSTRACT

Database systems play a significant role in structuring, organizing, and managing data of organizations. In this regard, the key challenge is how to protect the confidentiality, integrity, and availability of database systems against attacks launched from within and outside an organization. To resolve this challenge, different database security techniques and mechanisms, which generally involve access control, database monitoring, data encryption, database backups, and strong passwords have been proposed. These techniques and mechanisms have been developed for certain purposes but fall short of many industrial expectations. This study used the design science research method to recommend a new Digital Forensic Readiness Framework, named DFRF, to secure database systems. DFRF involves risk assessments, data classification, database firewalls, data encryption, strong password policies, database monitoring and logging, data backups and recovery, incident response plans, forensic readiness, as well as education and awareness. The proposed framework not only identifies threats and responds to them more effectively than existing models, but also helps organizations stay fully compliant with regulatory requirements and improve their security. The design of the suggested framework was compared with existing models, confirming its superiority.

Keywords-database systems; digital forensics; forensic readiness; design science method

I. INTRODUCTION

In general, Database Management Systems (DBMS) help create, modify, share, and manage database transactions among users and applications [1-2]. These systems have several advantages as they help separate user applications from the underlying physical databases and facilitate the use and management of the actual database. Despite their numerous benefits, these systems have also some disadvantages. Several attacks can compromise the integrity, confidentiality, and availability of databases [3-4]. The security measures of databases are continuously improved despite the increasing complexity of attacks. For this reason, many traditional security measures were suggested, including using secure and strong passwords, controlling access to data, encrypting data, performing database backups, and monitoring databases. Although these measures are effective under some conditions, some limitations, such as the use of weak passwords, inadequate encryption of the data, and risky behavior of users, still exist. Organizations must use digital evidence as effectively as possible to meaningfully reduce investigation costs [5]. Companies need to expand their forensic capabilities, establish efficient processes for collecting and preserving data, make effective collaborations with external experts, and employ high-standard investigative methods to have more rapid and efficient forensic investigations. This study applies the Design Science Research Methodology (DSRM) to design a

digital forensic readiness framework, named DFRF, to ensure database security. This framework is expected to help organizations identify and competently deal with the security challenges that can arise in digital databases. A key objective of DFRF is the identification and mitigation of database vulnerabilities. This framework performs periodic assessments and audits to unveil any weaknesses that may exist in a system's configuration, access control, and data encryption measures. Successfully managing these problems, minimizing the risk of data breaches, securing databases, and preventing data breaches could result in a significant decrease in database security risks.

II. RELATED WORKS

Several studies have argued that database security models might fail when applied [6-11]. DBMSs differ greatly in terms of functionality, which may explain this failure. In addition, database forensics is focused on one dimension (file system), which primarily involves identifying, gathering, handling, storing, responding to incidents, and training. However, in some cases, it is possible to trace database incidents when digital investigators cooperate to analyze the database [8]. Owing to the multidimensionality and diversity of DBMSs, it is difficult to develop a standardized approach to database forensics. Current digital forensic models do not cover the full range of database concepts [12]. According to [13-14], most of the studies on database forensics focused on resolving database

contents and metadata, which is consistent with documents rather than database incidents. In [13], it was shown that investigation processes can be utilized to collect data related to operations executed using Oracle database concepts by performing certain tasks. Also, four steps were proposed to resolve the problem: reconstructing databases, canceling database operations, collecting data, and repairing integrity issues. In [15], an audit reconstruction tool was presented to extract information from logs when auditing features are disabled. The Oracle database has been the subject of several forensic investigation models. For example, in [15], it was displayed how an Oracle log file can be used to detect attack events by examining the binary format of redo logs. Furthermore, it was demonstrated how attackers conceal their tracks after a failed attack, as well as the way to detect them. According to [16], evidence that has already been deleted could be recovered in the case of Oracle objects, and data files extracted from a compromised server can be indirectly recovered by investigators using this procedure. Additionally, malicious entities can drop objects. The listener's log file and audit trail can be put into service to capture evidence of attacks against the authentication mechanism. An instance name and the IP address of the server are also recorded in this log file, along with the Service Identifier (SID) and the IP address of the connection. On the contrary, the audit trail can indicate whether a log-in or log-out was successful or unsuccessful. The listener log file and audit trails can be used by investigators to collect evidence against the authentication mechanism, but first, the respective database must be configured with an audit trail enabled.

In [17], a forensic model in which the database servers are disconnected from the network to capture volatile data was proposed. The recommended evidence collection and identification processes can be employed to retrieve fragile data from the database server. Forensic techniques are applied to move the captured data once the server is disconnected from the network and the forensic environment. A compromised database server is obtained for evidence collection in the evidence collection process. The recovery and careful storage of volatile data is necessary for forensic research. A human-readable form makes it easier for forensic inspectors to examine non-volatile data, as opposed to stored binary forms. In [18], a detection investigation model was presented to help the examiner find evidence of data theft. A DBA or incident responder can use the model to determine whether such a breach has occurred in a situation where no audit trail exists, but the assumption is that unauthorized access has been obtained to the data on the server. In [19], a forensic analysis method was proposed for MSSQL servers, consisting of four phases: preparation of the investigation, verification of the incident, collection of artifacts, and analysis. In [20], another model was proposed to detect and investigate database servers. It involved three phases: server detection, data collection, and data analysis, but this model cannot handle volatile artifacts. In [21], a database model was introduced to identify and name inconsistencies in the MySQL database system. In [22], a reconstruction model was developed to restore already deleted or updated values from redo logs to reconstruct basic SQL statements. The basic DDL statement was overlooked in this

proposed model, as it was based on the DML statements. In [23], a practical forensic method was discussed to reconstruct basic SQL DDL statements and improve the previous method. In [14], a framework was showcased to identify, collect, analyze, validate, and document digital evidence that has been altered. This framework collected, analyzed, and reconstructed volatile and non-volatile data.

In addition to the different forensic tamper detection and analysis algorithms proposed for DBMSs, several forensic tamper detection models have also been suggested. In [24], methods and scenarios were recommended to detect covert databases. In [25], a model was proposed to efficiently collect digital evidence. This study stated that a database business environment could gather evidence against authorized and unauthorized events using triggers, replication, and log file backup. In [26], a forensic tamper detection model was introduced deploying a one-way hash function that could detect compromised audit logs. However, this model could not detect when tampering occurred, which data were altered, and the identity of the attacker, as it was not able to analyze intruder activities. In [7], a model was introduced to investigate compromised databases by involving two examination processes: identification and collection. In [27], a method was presented to collect, preserve, and analyze database metadata to prevent attacks on databases using four investigation processes: collecting and preserving evidence, analyzing anti-forensic attacks, analyzing database attacks, and preserving evidence reports. In [28], it was attempted to reconstruct database events to uncover intruders' actions by collecting and reconstructing evidence. In [12], forensic investigation frameworks were developed for NoSQL DBMSs based on their unique features. This process involved five parallel phases: preparation, acquisition and preservation, identification of distributed evidence, examination and analysis, and finally reporting and presentation. The framework did not include database schemes' evaluation or analysis of database characteristics, such as gathering logs for assessing operations. In [29], MongoDB was studied, which is among the most widely employed NoSQL DBMSs, and a forensic tool was proposed to explore the internal structure and format of the data files. In [30], a comprehensive review of database forensic investigation processes was presented to help domain researchers gain a deeper understanding of database forensics from various perspectives. This study also discussed the issues and drawbacks that emerged and proposed solutions. In [31], database forensic models proposed from 2009 to 2015 were evaluated but did not address limitations, challenges, issues, directions, or proposed solutions for database forensics. In [32], another review was conducted for the period 2015-2017 in database forensics. For the database forensics field, five stages of forensic analysis were proposed: determining, examining, presenting, documenting, and reporting the event.

III. METHODOLOGY

This study aimed to design a digital forensics readiness framework to secure database systems using the DSRM, which is suitable for designing and validating artifices in a digital forensics context. DSRM was utilized as the analysis method for several reasons: 1) It is a solution-oriented analysis method

that is deployed to produce logical, testable, and communicable products, 2) the formal process it follows facilitates the research procedure, and 3) ensures the smooth, cohesive link between the design and development of the model and its evaluation and demonstration. Figure 1 displays the adapted DSRM and the operational framework. DSRM consists of the following four stages.

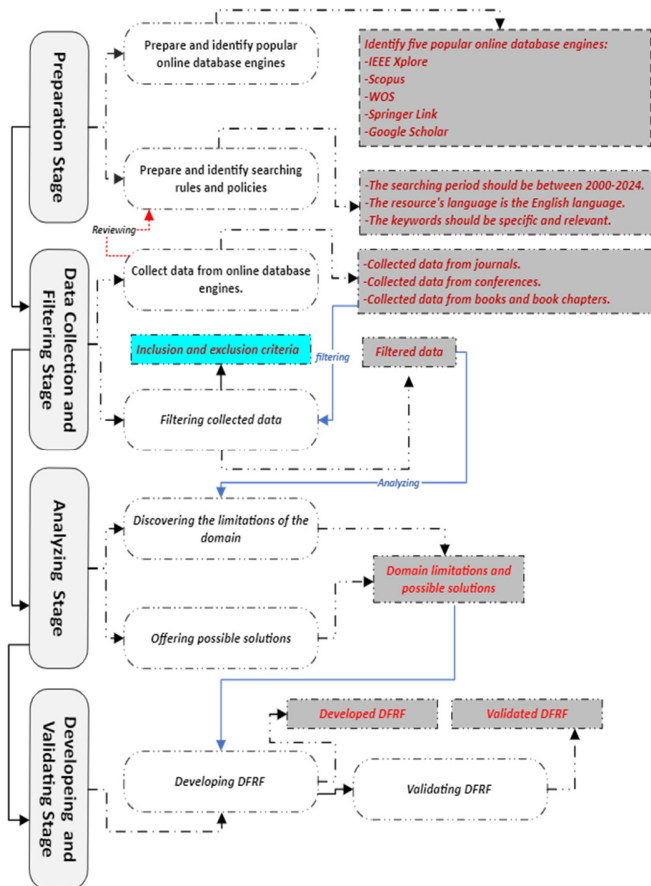


Fig. 1. Adapted DSRM and operational framework.

A. Preparation

This stage prepares the search guidelines that will govern and control the research process and consists of two parts: identifying popular online databases and determining the rules for the search procedure. For the first part, five online databases were identified: IEEE Xplore, Scopus, Web of Science (WOS), Springer Link, and Google Scholar. For the second part, three rules were set: a) the search period was set to be from 2000 to 2024, b) the language of the resource should be English, and c) the keywords were: "digital forensic", "forensics readiness", "database forensic", and "database systems".

B. Data Collection and Filtering

This stage aimed to collect and filter the data gathered in the previous stage. Figures 2-5 show the data collected from the above databases based on the defined keywords.

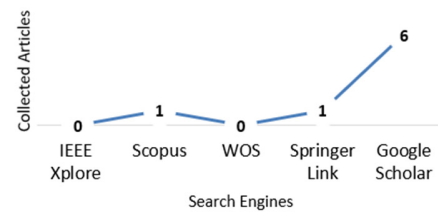


Fig. 2. Collected data based on all four keywords.

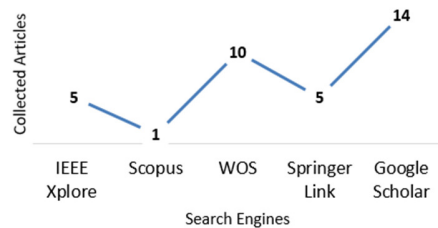


Fig. 3. Collected data based on three keywords.

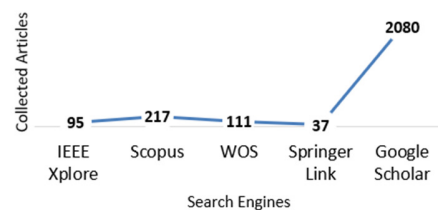


Fig. 4. Collected data based on two keywords.

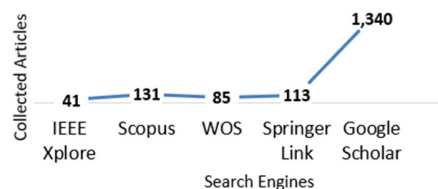


Fig. 5. Collected data based on one keyword.

These results helped to determine the direction of the research and the attention and motivation of the researchers. Therefore, this study did not pay more attention to database systems from a digital forensic perspective but focused instead on identifying the reasons behind this limitation in studies on database systems and database forensics. Table I presents the digital forensic models that concentrate on database systems and the way they could be implemented. As a result, a total of 57 digital forensic models, frameworks, approaches, techniques, tools, and algorithms, which were purely related to database systems, were identified. There was only one model that focused on forensic readiness for database systems [33].

TABLE I. DIGITAL FORENSIC MODELS FOCUSED ON DATABASE SYSTEMS

ID	Year	Ref	Forensic Readiness	Investigation Processes			
				Identification	Collection and Preservation	Analysis & Examination	Presentation & Documentation
	2004	[13]	x	√	√	√	x
	2004	[34]	x	x	x	√	x
	2006	[35]	x	√	x	x	x
	2007	[2]	x	√	x	√	x
	2007	[36]	x	√	x	x	x
	2007	[24]	x	√	x	x	x
	2007	[15]	x	x	x	√	x
	2007	[16]	x	x	x	√	x
	2007	[30]	x	x	x	√	x
	2007	[31]	x	x	x	√	x
	2007	[32]	x	x	x	√	x
	2008	[19]	x	√	√	√	√
	2008	[37]	x	x	x	√	x
	2009	[38]	x	x	x	√	x
	2009	[39]	x	x	x	√	x
	2010	[21]	x	√	x	x	x
	2011	[40]	x	√	x	x	x
	2011	[41]	x	x	√	x	x
	2011	[20]	x	√	√	√	x
	2011	[42]	x	x	x	√	x
	2012	[43]	x	√	x	x	x
	2012	[14]	x	√	√	√	√
	2012	[6]	x	√	√	√	x
	2012	[44]	x	x	x	√	x
	2012	[45]	x	x	x	√	x
	2012	[22]	x	x	√	√	x
	2012	[46]	x	x	x	√	x
	2013	[47]	x	x	x	√	x
	2013	[48]	x	x	√	√	x
	2013	[49]	x	x	x	√	x
	2013	[50]	x	x	x	√	x
	2013	[51]	x	x	x	√	x
	2013	[52]	x	x	x	√	x
	2013	[53]	x	x	x	√	x
	2013	[54]	x	x	x	√	x
	2013	[23]	x	x	x	√	x
	2014	[55]	x	x	x	√	x
	2014	[27]	x	x	x	x	x
	2014	[56]	x	x	x	x	x
	2014	[7]	x	√	√	x	x
	2014	[28]	x	x	√	√	x
	2014	[57]	x	x	x	√	x
	2015	[58]	x	x	x	√	x
	2015	[59]	x	x	x	√	x
	2015	[60]	x	x	x	√	x
	2016	[61]	x	x	x	√	x
	2016	[62]	x	√	√	√	√
	2016	[8]	x	√	√	√	√
	2017	[63]	x	x	x	√	x
	2017	[64]	x	√	√	√	√
	2017	[65]	x	√	√	√	√
	2018	[66]	x	√	√	√	√
	2019	[67]	x	√	√	√	√
	2020	[33]	√	√	√	√	√
	2020	[68]	x	√	√	√	√
	2021	[69]	x	x	x	√	x

C. DFRF Development

The development and validation of DFRF require the identification and selection of models from Table I based on the development criteria. This table summarizes several digital forensic models for database systems. Coverage factors were selected based on [70-71]. It is essential to cover a wide range of investigation processes to meet the development objective of the DFRF. A model is said to be well covered (i.e., it has a high coverage value) if it can incorporate all the four investigation processes mentioned in Table I. The coverage value of the model will be lower if it merely describes three or two investigation processes. As a result, this study identified two categories of models for developing and validating the proposed framework. Models related to the four different investigation processes are included in the first group. It is also possible to find models that cover three or two investigation processes in the second group. Consequently, and based on this categorization, this study found 11 models that met the first group, while eight models represented the second group. Tables II and III portray the selected development and validation models, respectively, which have redundant and overlapped investigation processes and concepts for investigation.

The common investigation procedures and concepts of digital forensics were collected from the 11 selected models on the basis of the criteria adapted from [68, 70]. In general, the investigation processes and concepts were extracted from the diagram or the main textual models. To understand the purpose and meaning of an investigation process and concept, a definition, activity, or task is also necessary. Investigation processes and concepts that were not relevant to the investigation were excluded. However, as part of the model, implicit and explicit investigation processes and concepts were included. The extracted investigations and concepts represent the input and output of the DFRF to secure database systems. The relationships between the extracted components were examined to ensure their applicability to the development process of the DFRF. Figure 6 illustrates the proposed DFRF system for securing database systems. It consists of 10 primary components: risk assessment, data classification, data encryption, database firewalls, strong password policies, database monitoring and logging, data backups and recovery, incident response plans, forensic readiness, education, and awareness.

1) Risk Assessment

Risk assessment refers to the identification of potential risks and weaknesses in a system and the measurement of the potential impacts of each risk. By taking this step into action, organizations would be able to properly distribute the available resources and also prioritize their efforts when mitigating the risks. An example of assessing risks in a database system is the identification and evaluation of the potential impacts of unauthorized access to or misuse of sensitive customer data. Through the assessment of the risks that may arise because of this vulnerability, companies can have a deep insight into the best security measures to take, for example, using strong password policies and/or applying suitable techniques for encrypting the data.

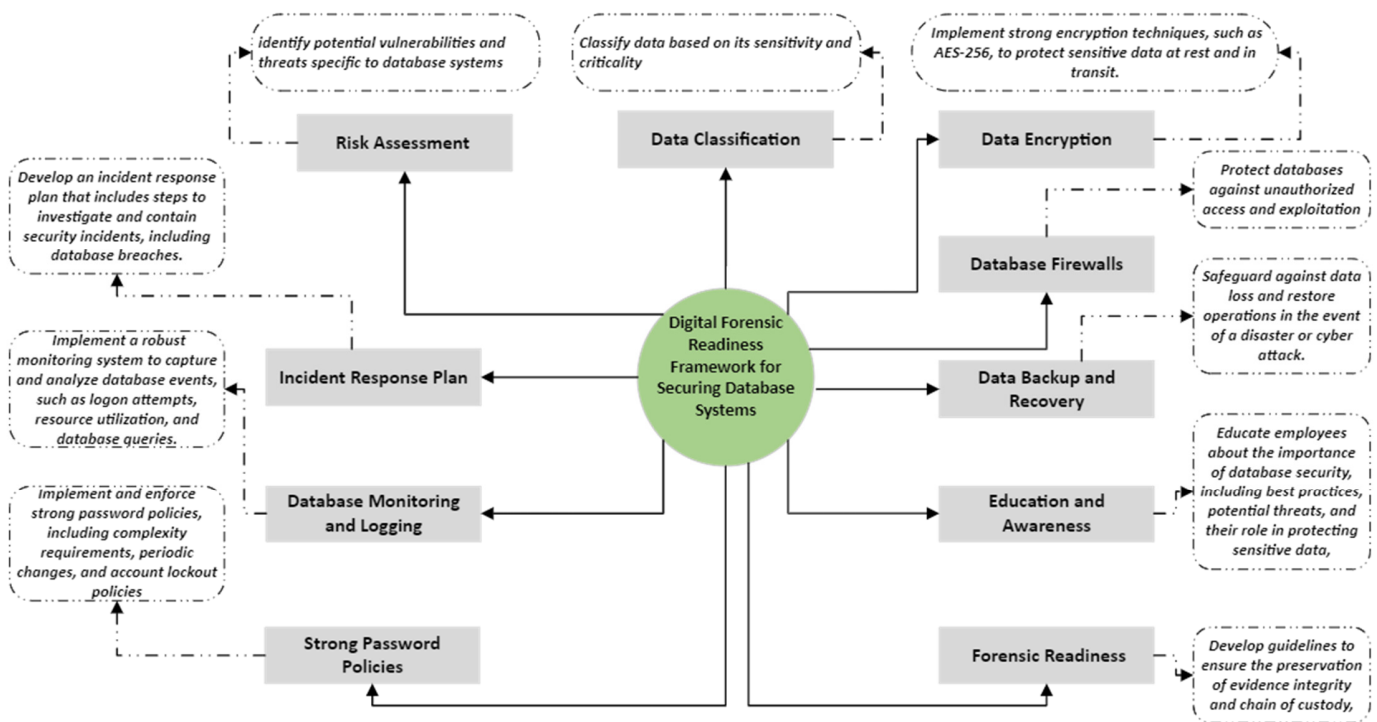


Fig. 6. DFRF design for securing database systems.

2) Data Classification

This step involves categorizing the data according to their importance and sensitivity. Data classification helps companies identify and protect data of the highest criticality and sensitivity, allowing users to manage less sensitive data with less rigorous security measures. Actual instances of data classification in a database system are the identification and labeling of customer data, financial records, or intellectual property. Data classification helps companies ensure that proper security measures are performed for the protection of the high-sensitive data while allowing more flexible treatment with data of lower criticality.

3) Data Encryption

This process converts data into unreadable codes, known as ciphertext, with the help of a key, and only authorized parties can decipher it. Data encryption protects the data against unauthorized access, even in the case of compromise of the underlying database system. An example of data encryption in a database system is to encrypt sensitive personal data, such as financial information or social security numbers. This helps companies prevent unauthorized parties from gaining access or misusing the data, even if the underlying database system is compromised.

4) Database Firewalls

Firewalls add another layer of protection to database systems. Not only do they build a barrier between the databases themselves and an external network, but also monitor and analyze network traffic to prevent unauthorized access to data or malevolent activities. It is well exemplified using a web application firewall. This type of firewall is responsible for

monitoring incoming HTTP requests and then filtering any suspicious activities. It could protect the underlying database against any potential attacks.

5) Strong Password Policies

Such policies prevent unauthorized parties from having access to database systems. Companies need to ask their users to set passwords of high complexity and uniqueness, change them regularly, and protect them from others' access. An example is to ask users to choose passwords of at least eight characters that include numbers, both uppercase and lowercase letters, and special characters. In addition, organizations could ask their users to regularly change their passwords so that any unauthorized access could be obstructed.

6) Database Monitoring and Logging

This activity results in acquiring critical awareness of potential security breaches or attempts made to gain unauthorized access. Through regular monitoring and logging of database activities, companies would be capable of detecting and responding to potential threats in real time. An instance of this activity in a database system is to use event logging through which any suspicious activities are logged in a secure and centralized location. This process helps the company in investigating and analyzing the logs to determine the potential security breaches and review the attempts made by unauthorized parties to gain access to the data.

7) Data Backups and Recovery

Data loss or system failure can be hindered by taking backups regularly. Every organization needs to establish an inclusive backup strategy to ensure that critical data is consistently backed up. For example, this can be done by

implementing automated backup routines through which data are backed up on a daily or weekly basis in separate locations. It helps to restore the data when they get lost or corrupted, which causes business operations to be subject to minimum potential disruptions. In addition, backup and recovery can be considered a proper strategy against system failures and security breaches. Companies are required to define well-established processes to recover their data and systems so that downtime can be minimized and normal business operations can be restored. Data backup and recovery of a database system can be performed using automated recovery scripts or tools. This helps users quickly restore data and systems from backups and have minimum downtime and disruption rates for their business operations.

8) Incident Response Plans

These plans help manage and give suitable responses to security breaches or incidents that may occur in a database system. Companies need well-defined incident response procedures, such as communication plans, roles and responsibilities, and mitigation measures. An example of these plans in a database system is to activate a defined incident response team that can investigate and respond to security breaches or incidents. The team will be also capable of coordinating with IT and security teams, making effective communication with affected stakeholders, and taking appropriate mitigation measures into action.

9) Forensic Readiness

This refers to the ability to collect and preserve evidence concerning security breaches or incidents in database systems. It involves taking appropriate measures to preserve forensic artifacts, namely system configurations, log files, and network traffic, and being able to analyze and interpret this evidence, which is collected throughout forensic investigations. Forensic readiness in a database system is exemplified by using forensically sound software/hardware to acquire and preserve forensic evidence. Companies must have well-trained personnel in forensic analysis techniques that can competently assist them in the analysis and interpretation of forensic evidence.

10) Education and Awareness

For sensitive data to be well protected, employees need to be well-trained in database security and understand its importance. To reduce the risks of data breach, companies should take the most appropriate practices, find and mitigate potential risks, and enable their employees to resist them. This helps sensitive information to be kept with high confidentiality and integrity.

The DFRF's securing data storage system consists of 10 components working together to build a holistic view of the way sensitive data are protected. Organizations can create a strong security posture through the following practices: risk assessments, data classification, data encryption, implementation of database firewall, enforcement of strong password policies, monitoring, logging, backup administration, data recovery, and creation of a response plan when an incident occurs.

D. DFRF Validation

During this step, the developed framework was validated against the validation models identified in the previous step. Table IV provides a summary of the validation process. The results obtained disclosed that the developed DFRF model is comprehensive and covers most of the security components used in the database security and investigation domain.

TABLE II. VALIDATION PROCESS

Components involved in DFRF	Validation models							
	[6]	[7]	[13]	[20]	[22]	[28]	[36]	[48]
Risk assessment	√	√	×	√	×	×	×	×
Data classification	×	√	×	×	×	×	×	×
Data encryption	√	×	√	×	×	√	√	√
Database firewalls	×	×	×	×	×	√	×	×
Strong password policies	×	×	×	×	×	×	×	×
Database monitoring and logging	√	√	√	√	×	×	√	×
Data backups and recovery	×	×	√	×	×	×	×	×
Incident response plans	×	×	√	×	×	×	×	×
Forensic readiness	×	×	×	×	×	×	×	×
Education and awareness	×	×	×	×	×	×	×	×

IV. RESULTS AND DISCUSSION

A total of 57 models were collected and filtered. The review revealed that only one study [33] had covered the perspective of forensic readiness for database systems. The model proposed in [33] consists of three phases: pre-incident, during incident, and post-incident. However, it lacks some major forensic readiness components, such as risk assessment, strong password policies, and education and awareness. In addition, other models represented different investigation processes, as shown in Figure 7. Forensic readiness was covered only once, while the identification process was covered by 21 models and 19 models focused on collection and preservation. The analysis and examination process were covered by 47 models, while presentation and documentation were covered by 10 models. Therefore, most of the studies covered the analysis and examination process, followed by collection and preservation. It can be concluded that database systems still do not receive sufficient attention from researchers working in the field of digital forensics. This could be due to the heterogeneity and complexity of the architectures of database systems.

During this study, DSRM was utilized to develop DFRF, and the models were divided into two main categories: one for design and development and the other for validation. Figure 8 categorizes the development and validation models. Classifying the models into development and validation categories allows for a systematic approach to model development and validation. The development category encompasses 11 models that were initially identified, selected, and adapted for development. These models undergo a detailed development process that includes requirement gathering, combining, and harmonizing the investigation processes. The primary objective of the development category is to ensure that the model meets the desired specifications and delivers the functionality requested.

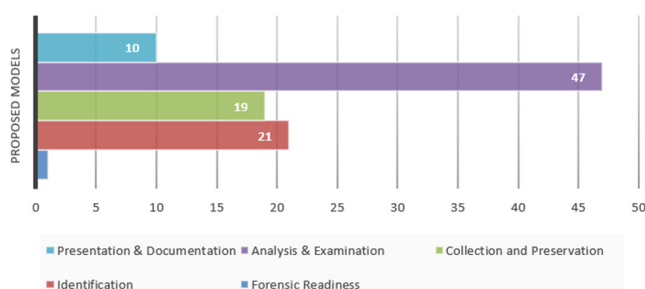


Fig. 7. Digital forensic models covering different investigation processes.

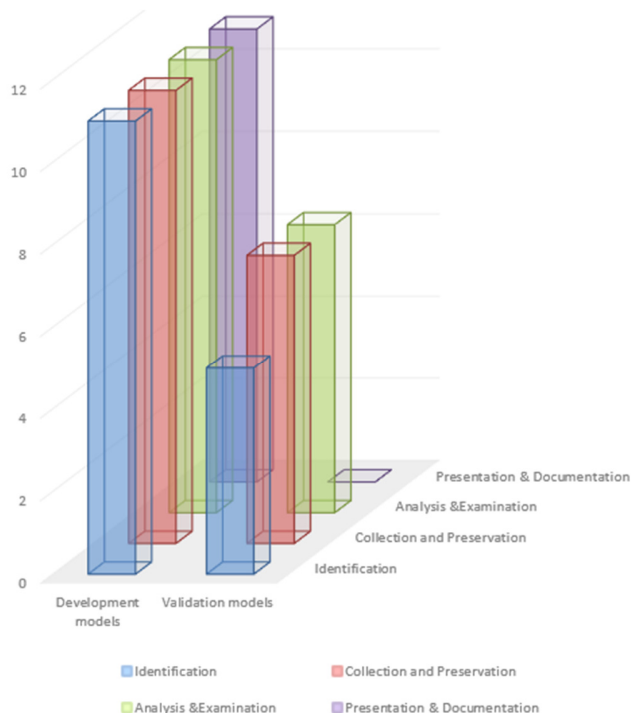


Fig. 8. Categorization of the development and validation models.

However, the validation category focuses on the assessment of the accuracy and reliability of the models developed. In this study, eight models were selected to validate the robustness, accuracy, and suitability of the proposed DFRF. Validation involves statistical analysis, empirical testing, or other validation techniques that can be used to ensure the effectiveness of a model's performance. This study proposed DFRF that combines all proactive components used in other previously proposed models to secure database systems. A key objective of the DFRF is to address the challenges that can be faced in the procedure of securing digital databases. It also helps to identify and respond proactively to potential security threats and attacks launched against data. Furthermore, DFRF contributes to the data security domain by helping organizations find and mitigate weaknesses in database systems. DFRF conducts routine audits and evaluations, which could result in the identification of possible weaknesses in system configuration, encryption measures, and access controls. Organizations can significantly reduce the risk of data breaches and improve the security of their databases by effectively dealing with these weaknesses.

V. CONCLUSION

Database systems perform different tasks, including structuring, organizing, and managing data. The integrity, confidentiality, and accessibility of an organization's database system could be compromised by many threats launched within and outside the organization. Parallel to the increase in the sophistication of attackers, security measures for database protection are continuously improved. In this sense, researchers and practitioners have proposed many traditional safety measures, such as data encryption, access control, database backups, database monitoring, and strong passwords. However, there are many challenges to the effectiveness of such security measures, including a lack of user awareness, insufficient data encryption, weak passwords, and limited capacity to detect and respond to database security threats. This study designed a digital forensic readiness framework, called DFRF, to secure database systems. This framework consists of 10 components: risk assessment, data classification, data encryption, database firewalls, strong password policies, database monitoring and logging, data backup and recovery incident response plans, and forensic readiness education and awareness. DFRF could help organizations improve their security postures, detect and provide appropriate responses to incidents, and fully comply with regulatory requirements. DFRF, compared to existing databases, offers a comprehensive approach to digital forensics. Future research should evaluate the performance quality of the developed DFRF in real-world scenarios.

REFERENCES

- [1] M. Alam and K. A. Shakil, "Cloud Database Management System Architecture," *UACEE International Journal of Computer Science and its Applications*, vol. 3, no. 1, pp. 27–31.
- [2] A. Alshammari, "A Novel Security Framework to Mitigate and Avoid Unexpected Security Threats in Saudi Arabia," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11445–11450, Aug. 2023, <https://doi.org/10.48084/etasr.6091>.
- [3] M. Ngadi, R. Al-Dhaqm, and A. Mohammed, "Detection and prevention of malicious activities on RDBMS relational database management systems," *International Journal of Scientific & Engineering Research*, vol. 3, no. 9, Sep. 2012.
- [4] F. Alotaibi, A. Al-Dhaqm, and Y. D. Al-Otaibi, "A Conceptual Digital Forensic Investigation Model Applicable to the Drone Forensics Field," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11608–11615, Oct. 2023, <https://doi.org/10.48084/etasr.6195>.
- [5] A. A. Alhussan, A. Al-Dhaqm, W. M. S. Yafouz, S. B. A. Razak, A.-H. M. Emara, and D. S. Khafaga, "Towards Development of a High Abstract Model for Drone Forensic Domain," *Electronics*, vol. 11, no. 8, Jan. 2022, Art. no. 1168, <https://doi.org/10.3390/electronics11081168>.
- [6] R. Susaimanickam, "A workflow to support forensic database analysis - Murdoch University," MSc Thesis, Murdoch University, Australia, 2012.
- [7] H. Q. Beyers, "Database forensics: Investigating compromised database management systems," MSc Thesis, University of Pretoria, South Africa, 2013.
- [8] A. Al-Dhaqm, S. Abd Razak, S. H. Othman, A. Nagdi, and A. Ali, "A Generic Database Forensic Investigation Process Model," *Jurnal Teknologi*, vol. 78, no. 6–11, Jun. 2016, <https://doi.org/10.11113/jt.v78.9190>.
- [9] O. M. Fasan and M. Olivier, "Reconstruction in Database Forensics," in *Advances in Digital Forensics VIII*, Pretoria, South Africa, 2012, pp. 273–287, https://doi.org/10.1007/978-3-642-33962-2_19.
- [10] O. M. Fasan and M. S. Olivier, "On Dimensions of Reconstruction in Database Forensics," in *Proceedings of the Seventh International*

- Workshop on Digital Forensics and Incident Analysis (WDFIA 2012), 2012.
- [11] I. S. Alansari, "A Detection and Investigation Model for the Capture and Analysis of Network Crimes," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11871–11877, Oct. 2023, <https://doi.org/10.48084/etasr.6316>.
 - [12] J. Yoon, D. Jeong, C. Kang, and S. Lee, "Forensic investigation framework for the document store NoSQL DBMS: MongoDB as a case study," *Digital Investigation*, vol. 17, pp. 53–65, Jun. 2016, <https://doi.org/10.1016/j.diin.2016.03.003>.
 - [13] D. Wong and K. Edwards, "System and method for investigating a data operation performed on a database," US20050289187A1, Dec. 29, 2005.
 - [14] H. K. Khanuja and D. S. Adane, "A framework for database forensic analysis," *Computer Science & Engineering: An International Journal*, vol. 2, no. 3, pp. 27–41, 2012.
 - [15] D. Litchfield, "Oracle Forensics - Part 1: Dissecting the Redo Logs," NGSSoftware Insight Security Research (NISR), Mar. 2007.
 - [16] D. Litchfield, "Oracle Forensics Part 2: Locating Dropped Objects," NGSSoftware Insight Security Research (NISR), Mar. 2007.
 - [17] D. Litchfield, "Oracle Forensics - Part 3: Isolating evidence of attacks against the authentication mechanism," NGSSoftware Insight Security Research (NISR), Mar. 2007.
 - [18] D. Litchfield, "Oracle Forensics Part 4: Live Response," NGSSoftware Insight Security Research (NISR), Apr. 2007.
 - [19] K. Fowler, *SQL Server Forensic Analysis*. Pearson Education, 2008.
 - [20] N. Son, K. Lee, S. Jeon, H. Chung, S. Lee, and C. Lee, "The Method of Database Server Detection and Investigation in the Enterprise Environment," in *Secure and Trust Computing, Data Management and Applications*, Loutraki, Greece, 2011, pp. 164–171, https://doi.org/10.1007/978-3-642-22339-6_20.
 - [21] P. Frühwirth, M. Huber, M. Mulazzani, and E. R. Weippl, "InnoDB Database Forensics," in *2010 24th IEEE International Conference on Advanced Information Networking and Applications*, Perth, WA, Australia, Apr. 2010, pp. 1028–1036, <https://doi.org/10.1109/AINA.2010.152>.
 - [22] P. Frühwirth, P. Kieseberg, S. Schrittwieser, M. Huber, and E. Weippl, "InnoDB Database Forensics: Reconstructing Data Manipulation Queries from Redo Logs," in *2012 Seventh International Conference on Availability, Reliability and Security*, Prague, Czech Republic, Aug. 2012, pp. 625–633, <https://doi.org/10.1109/ARES.2012.50>.
 - [23] P. Frühwirth, P. Kieseberg, S. Schrittwieser, M. Huber, and E. Weippl, "InnoDB database forensics: Enhanced reconstruction of data manipulation queries from redo logs," *Information Security Technical Report*, vol. 17, no. 4, pp. 227–238, May 2013, <https://doi.org/10.1016/j.istr.2013.02.003>.
 - [24] G. T. Lee, S. Lee, E. Tsoenko, and S. Lee, "Discovering Methodology and Scenario to Detect Covert Database System," in *Future Generation Communication and Networking (FGCN 2007)*, Jeju, Korea (South), Sep. 2007, vol. 2, pp. 130–135, <https://doi.org/10.1109/FGCN.2007.106>.
 - [25] J. Azemovi, "Efficient Model for Detection Data and Data Scheme Tempering with Purpose of Valid Forensic Analysis," presented at the International Conference on Computer Engineering and Applications, Singapore, 2011.
 - [26] R. T. Snodgrass, S. S. Yao, and C. Collberg, "Tamper detection in audit logs," in *Proceedings of the Thirtieth international conference on Very large data bases*, Vol. 30, 2004, pp. 504–515.
 - [27] H. Khanuja and S. S. Suratar, "Role of metadata in forensic analysis of database attacks," in *2014 IEEE International Advance Computing Conference (IACC)*, Gurgaon, India, Feb. 2014, pp. 457–462, <https://doi.org/10.1109/IAdCC.2014.6779367>.
 - [28] P. Frühwirth, P. Kieseberg, K. Krombholz, and E. Weippl, "Towards a forensic-aware database solution: Using a secured database replication protocol and transaction management for digital investigations," *Digital Investigation*, vol. 11, no. 4, pp. 336–348, Dec. 2014, <https://doi.org/10.1016/j.diin.2014.09.003>.
 - [29] J. Yoon and S. Lee, "A method and tool to recover data deleted from a MongoDB," *Digital Investigation*, vol. 24, pp. 106–120, Mar. 2018, <https://doi.org/10.1016/j.diin.2017.11.001>.
 - [30] D. Litchfield, "Oracle Forensics Part 5: Finding Evidence of Data Theft in the Absence of Auditing," NGSSoftware Insight Security Research (NISR), Aug. 2007.
 - [31] D. Litchfield, "Oracle Forensics Part 6: Examining Undo Segments, Flashback and the Oracle Recycle Bin," NGSSoftware Insight Security Research (NISR), Aug. 2007.
 - [32] D. Litchfield, "Oracle Forensics Part 7: Using the Oracle System Change Number in Forensic Investigations," NGSSoftware Insight Security Research (NISR), Nov. 2008.
 - [33] A. Al-Dhaqm, S. A. Razak, K. Siddique, R. A. Ikuesan, and V. R. KEBANDE, "Towards the Development of an Integrated Incident Response Model for Database Forensic Investigation Field," *IEEE Access*, vol. 8, pp. 145018–145032, 2020, <https://doi.org/10.1109/ACCESS.2020.3008696>.
 - [34] P. M. Wright, "Oracle Database Forensics using LogMiner," SANS Institute, Jun. 2004.
 - [35] A. Basu, "Forensic Tamper Detection in SQL Server," <http://amitfrombangalore.blogspot.com/2015/08/forensic-tamper-detection-in-sql-server.html>.
 - [36] M. J. Malmgren, "An Infrastructure for Database Tamper Detection and Forensic Analysis," BSc Thesis, University of Arizona, 2007.
 - [37] K. E. Pavlou and R. T. Snodgrass, "Forensic analysis of database tampering," *ACM Transactions on Database Systems*, vol. 33, no. 4, Sep. 2008, <https://doi.org/10.1145/1412331.1412342>.
 - [38] M. S. Olivier, "On metadata context in Database Forensics," *Digital Investigation*, vol. 5, no. 3, pp. 115–123, Mar. 2009, <https://doi.org/10.1016/j.diin.2008.10.001>.
 - [39] D. Lee, J. Choi, and S. Lee, "Database forensic investigation based on table relationship analysis techniques: 2009 2nd International Conference on Computer Science and Its Applications, CSA 2009," in *Proceedings of the 2009 2nd International Conference on Computer Science and Its Applications*, 2009, <https://doi.org/10.1109/CSA.2009.5404235>.
 - [40] F. Fatima, "Detecting database attacks using computer forensics tools," Texas A&M University-Corpus Christi, 2011.
 - [41] H. Beyers, M. Olivier, and G. Hancke, "Assembling Metadata for Database Forensics," in *Advances in Digital Forensics VII*, Orlando, FL, USA, 2011, pp. 89–99, https://doi.org/10.1007/978-3-642-24212-0_7.
 - [42] H. Beyers and M. Olivier, "An Approach to Examine the Metadata and Data of a Database Management System by making use of a Forensic Comparison Tool," 2011.
 - [43] S. Tripathi and B. B. Meshram, "Digital Evidence for Database Tamper Detection," vol. 2012, Apr. 2012, <https://doi.org/10.4236/jis.2012.32014>.
 - [44] S. Jeon, J. Bang, K. Byun, and S. Lee, "A recovery method of deleted record for SQLite database," *Personal and Ubiquitous Computing*, vol. 16, no. 6, pp. 707–715, Aug. 2012, <https://doi.org/10.1007/s00779-011-0428-7>.
 - [45] P. D. Abhonkar and A. Kanthe, "Enriching forensic analysis process for tampered data in database," *International Journal of Computer Science and Information Technologies*, vol. 3, no. 5, pp. 5078–5085, 2012.
 - [46] H. Q. Beyers, M. S. Olivier, and G. P. Hancke, "Arguments and Methods for Database Data Model Forensics," in *Proceedings of the Seventh International Workshop on Digital Forensics and Incident Analysis (WDFIA 2012)*, 2012.
 - [47] H. K. Khanuja and Dr. D. S. Adane, "Forensic Analysis of Databases by Combining Multiple Evidences," *International Journal of Computers and Technology*, vol. 7, no. 3, pp. 654–663, Jun. 2013, <https://doi.org/10.24297/ijct.v7i3.3446>.
 - [48] K. E. Pavlou and R. T. Snodgrass, "Generalizing database forensics," *ACM Transactions on Database Systems*, vol. 38, no. 2, Apr. 2013, <https://doi.org/10.1145/2487259.2487264>.
 - [49] O. M. Adedayo and M. S. Olivier, "On the Completeness of Reconstructed Data for Database Forensics," in *Digital Forensics and*

- Cyber Crime*, Lafayette, IN, USA, 2013, pp. 220–238, https://doi.org/10.1007/978-3-642-39891-9_14.
- [50] P. P. Gawali, "Forensic Analysis Algorithm: By using the Tiled Bitmap with Audit Log Mechanism," *International Journal of Computer Applications*, vol. 63, no. 11, pp. 36–42, Feb. 2013.
- [51] B. Wu, M. Xu, H. Zhang, J. Xu, Y. Ren, and N. Zheng, "A Recovery Approach for SQLite History Recorders from YAFFS2," in *Information and Communication Technology*, Yogyakarta, Indonesia, 2013, pp. 295–299, https://doi.org/10.1007/978-3-642-36818-9_30.
- [52] J. H. Choi, D. W. Jeong, and S. Lee, "The method of recovery for deleted record in Oracle Database," *Journal of the Korea Institute of Information Security & Cryptology*, vol. 23, no. 5, pp. 947–955, 2013, <https://doi.org/10.13089/JKIISC.2013.23.5.947>.
- [53] M. Xu *et al.*, "A metadata-based method for recovering files and file traces from YAFFS2," *Digital Investigation*, vol. 10, no. 1, pp. 62–72, Jun. 2013, <https://doi.org/10.1016/j.diin.2013.02.006>.
- [54] P. P. Gawali, "Database tampering and detection of data fraud by using the forensic scrutiny technique," *International Journal of Emerging Technology and Advanced Engineering*, vol. 3, no. 2, pp. 439–446, Feb. 2013.
- [55] M. Xu *et al.*, "A Reconstructing Android User Behavior Approach based on YAFFS2 and SQLite," *Journal of Computers*, vol. 9, no. 10, pp. 2294–2302, 2014.
- [56] W. K. Hauger and M. S. Olivier, "The role of triggers in database forensics," in *2014 Information Security for South Africa*, Johannesburg, South Africa, Dec. 2014, pp. 1–7, <https://doi.org/10.1109/ISSA.2014.6950506>.
- [57] H. K. Khanuja and D. S. Adane, "Forensic Analysis for Monitoring Database Transactions," in *Security in Computing and Communications*, Delhi, India, 2014, pp. 201–210, https://doi.org/10.1007/978-3-662-44966-0_19.
- [58] O. M. Adedayo, "Reconstruction in Database Forensics," Ph.D. dissertation, University of Pretoria, South Africa, 2015.
- [59] J. Wagner, A. Rasin, and J. Grier, "Database forensic analysis through internal structure carving," *Digital Investigation*, vol. 14, pp. S106–S115, Aug. 2015, <https://doi.org/10.1016/j.diin.2015.05.013>.
- [60] O. M. Adedayo and M. S. Olivier, "Ideal log setting for database forensics reconstruction," *Digital Investigation*, vol. 12, pp. 27–40, Mar. 2015, <https://doi.org/10.1016/j.diin.2014.12.002>.
- [61] J. O. Ogutu, "A Methodology To Test The Richness Of Forensic Evidence Of Database Storage Engine: Analysis Of MySQL Update Operation In InnoDB And MyISAM Storage Engines," MSc Thesis, University of Nairobi, Kenya, 2016.
- [62] A. Aldhaqm, S. A. Razak, S. H. Othman, A. Ali, and A. Ngadi, "Conceptual Investigation Process Model for Managing Database Forensic Investigation Knowledge," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 12, no. 4, pp. 386–394, Feb. 2016, <https://doi.org/10.19026/rjaset.12.2377>.
- [63] J. Wagner, A. Rasin, T. Malik, K. Heart, H. Jehle, and J. Grier, "Database Forensic Analysis with DBCarver," in *CIDR 2017, 8th Biennial Conference on Innovative Data Systems Research*, Jan. 2017.
- [64] A. Al-Dhaqm, S. Razak, S. H. Othman, A. Ngadi, M. N. Ahmed, and A. A. Mohammed, "Development and validation of a Database Forensic Metamodel (DBFM)," *PLOS ONE*, vol. 12, no. 2, 2017, Art. no. e0170793, <https://doi.org/10.1371/journal.pone.0170793>.
- [65] M. Alam and K. A. Shakil, "Cloud Database Management System Architecture," *UACEE International Journal of Computer Science and its Applications*, vol. 3, no. 1, pp. 27–31, <https://doi.org/10.1109/ACCESS.2017.2762693>.
- [66] A. Al-Dhaqm, S. Razak, and S. H. Othman, "Model Derivation System to Manage Database Forensic Investigation Domain Knowledge," in *2018 IEEE Conference on Application, Information and Network Security (AINS)*, Langkawi, Malaysia, Nov. 2018, pp. 75–80, <https://doi.org/10.1109/AINS.2018.8631468>.
- [67] R. Bria, A. Retnowardhani, and D. N. Utama, "Five Stages of Database Forensic Analysis: A Systematic Literature Review," in *2018 International Conference on Information Management and Technology (ICIMTech)*, Jakarta, Indonesia, Sep. 2018, pp. 246–250, <https://doi.org/10.1109/ICIMTech.2018.8528177>.
- [68] A. Al-Dhaqm *et al.*, "Categorization and Organization of Database Forensic Investigation Processes," *IEEE Access*, vol. 8, pp. 112846–112858, 2020, <https://doi.org/10.1109/ACCESS.2020.3000747>.
- [69] H. Choi, S. Lee, and D. Jeong, "Forensic Recovery of SQL Server Database: Practical Approach," *IEEE Access*, vol. 9, pp. 14564–14575, 2021, <https://doi.org/10.1109/ACCESS.2021.3052505>.
- [70] M. F. Caro, D. P. Josyula, M. T. Cox, and J. A. Jiménez, "Design and validation of a metamodel for metacognition support in artificial intelligent systems," *Biologically Inspired Cognitive Architectures*, vol. 9, pp. 82–104, Jul. 2014, <https://doi.org/10.1016/j.bica.2014.07.002>.
- [71] S. Kelly and R. Pohjonen, "Worst Practices for Domain-Specific Modeling," *IEEE Software*, vol. 26, no. 4, pp. 22–29, Jun. 2009, <https://doi.org/10.1109/MS.2009.109>.

Evaluation and Improvement of the Accuracy of Reanalysis and Analysis Datasets for Wind Resource Assessment in Sudan

Youssef Kassem

Department of Mechanical Engineering, Engineering Faculty, Near East University, Cyprus | Energy, Environment, and Water Research Center, Near East University, Cyprus | Science, Technology, Engineering Education Application and Research Center, Near East University, Cyprus
yousseuf.kassem@neu.edu.tr (corresponding author)

Huseyin Camur

Department of Mechanical Engineering, Engineering Faculty, Near East University, Cyprus | Science, Technology, Engineering Education Application and Research Center, Near East University, Cyprus
huseyin.camur@neu.edu.tr

Mohamedalmojtba Hamid Ali Abdalla

Department of Mechanical Engineering, Engineering Faculty, Near East University, Cyprus
taabo129@gmail.com

Received: 8 February 2024 | Revised: 5 March 2024 | Accepted: 7 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7035>

ABSTRACT

Wind speed datasets are used to evaluate wind resources and energy production of wind farms. In locations where measured data are not available, reanalysis and analysis datasets can be used as an alternative to assess wind resources. This study evaluated the accuracy of wind speed data collected from reanalysis and analysis datasets against mast-measured data between 1975 and 1985 in Sudan, using monthly statistical analyses. Three bias correction methods, based on Measure-Correlate-Predict (MCP) and Linear Adaptation (LA1 and LA2), were applied to determine the original wind speed. The results indicate that LA1 outperformed MCP and LA2. Furthermore, the Weibull distribution function was employed to analyze the wind speed characteristics. In addition, wind power density was calculated using data from different sources. The findings show that although the wind power potential of the chosen locations is not suitable for large wind turbines, wind power can still be exploited with small wind turbines. Consequently, this study introduces a wind energy roadmap to attract investors in clean energy for sustainable development in Sudan, address energy problems, and meet domestic demands. The study also identifies the most important grid datasets for assessing the country's wind potential, enhancing the accuracy of assessments for investors and policymakers.

Keywords-Sudan; bias correction methods; mast measurement; reanalysis and analysis datasets; wind energy potential

I. INTRODUCTION

Energy plays a crucial role in driving economic growth, whereas its scarcity can negatively affect developing countries such as Sudan [1-2]. Increasing energy demands, consumption of fossil fuels, global warming, and climate change necessitate the transition from conventional to renewable energy sources [3]. Not only does the use of renewable energy improve energy security, but it also fosters sustainable development [4]. Renewable energy, a vital component of the energy mix, offers cost-effective and environmentally friendly solutions, reducing dependence on fossil fuels [5]. This shift towards clean energy

addresses environmental concerns and supports sustainable development for future generations [6]. Wind energy is the second largest source of renewable energy worldwide and plays a crucial role in sustainable and economic energy generation [7-8]. Wind speed data are significant to assess the potential of wind energy and optimize the design and layout of wind farms for maximum energy production [9]. Therefore, accurate and consistent wind speed data is essential for informed decision-making, ensuring the economic viability and efficiency of wind energy projects.

A. Literature Review on Reanalysis and Analysis of Datasets

In developing countries such as Sudan, the scarcity of wind speed data requires alternative sources for conducting preliminary wind resource evaluations. One of the primary issues in determining wind potential at a site or region is the long-term availability of accurate and uniform datasets [10-11]. This is not necessarily true for meteorological measurement stations, which may show discontinuities in data records due to equipment inefficiencies and performance, particularly in extreme climate conditions. Therefore, satellite data, including reanalysis or analysis data, become vital for the assessment of the wind energy potential [10]. Despite their value, it is important to recognize that satellite data may contain errors in estimations [10-11]. Meteorological prediction departments around the world offer satellite datasets as an alternative form of meteorological data. An assimilation scheme and a numerical weather prediction model are part of the data integration system that processes satellite data. Numerous sources, including satellites, buoys, aircraft, and ground surface stations, have provided historical weather data that this system integrates.

Reanalysis and analysis of datasets on a broad scale have gained attention as a competitive substitute for measured data obtained at meteorological stations in the field of wind resource evaluation, due to its ability to cover huge regions at many timescales. Numerous studies have used satellite databases to evaluate the wind energy potential in different places [10-30]. For example, in [10], the wind energy potential in Libya's coastal agricultural regions was assessed using TerraClimate, ERA5, ERA5-Land, MERRA-2, and CFSR. In [11], measured data from 2009 to 2018 were utilized to examine the reliability of the ERA5 dataset at three windy locations in the southwest of the Algerian Sahara. In [12], the accuracies of ERA-Interim, JRA-55, CFS, and MERRA-2 were assessed in the northern hemisphere employing measured wind speed. In [13], a NASA dataset was put into service to assess the wind energy potential of three coastal cities in Cameroon. As a result, the review of previous studies highlights the importance of carefully considering the accuracy of the data when deploying satellite-based wind speed datasets for wind energy evaluations.

B. Literature Review on Wind Resource Assessment in Sudan

Sudan is one of the largest countries in Africa. Fossil fuels currently supply 92% of Sudan's energy production, while hydropower provides the remaining 8% of the country's power needs. Sudan has six dams that use the waters of the Nile to produce energy. In addition, there are three thermal power plants, Umm Dabakir in the south and Garria and Bahri in the north. In general, the total power generation capacity of the whole country is estimated to be 4,000 MW. However, with a capacity of approximately 1,820 MW, the power generation ability falls short of this potential. The mean amount of electricity consumed in Sudan is 4,500MW, which meets all the needs of industrial and residential customers. Additionally, population expansion and the ongoing need for a variety of new appliances contribute to increasing electricity usage, according to the International Energy Agency (IEA). Sudan has been suffering from an electricity crisis for many years. According to [2], about 60% of people do not have access to

electricity. The country is increasingly requiring more power and the available supply has difficulty keeping up with this demand [31]. The country's outdated electrical infrastructure is deteriorating, there is a lack of investment in the electricity industry, and climate change negatively affects hydropower generation, exacerbating the problem [31]. The country's dependence on the expensive maintenance and operation of diesel generators is a direct effect of the power crisis [32]. Many Sudanese homes and businesses cannot purchase electricity due to its high cost.

The Sudanese government has recognized the value of renewable energy in addressing the country's electrical issues [33]. Wind energy is an alternative energy source that has gained popularity in recent years. Sudan has significant wind resources that could be used to generate electricity, especially in the northeast and eastern parts of the country [34]. The Global Wind Atlas published by the World Bank indicates that the typical wind power density at a 50-meter altitude ranges between 75-830 W/m². In Sudan, there have been recent attempts to increase wind energy usage. According to a government goal, by 2023, the country's electrical mix should contain 20% more renewable energy [35]. Furthermore, wind power systems can act as a substitute to supply electrical power needs and reduce CO₂ emissions [36]. In [37], it was stated that more than half of Sudan's land offers favorable conditions for the installation of wind farms. According to [38], most of Sudan is not suitable for small-scale wind-solar hybrid systems. In [39], it was concluded that wind energy could be an alternative option to reduce Sudan's energy production from fossil fuels. In [40], the mean wind speed in more than 50% of Sudan was reported to vary between 4.5 and 6 m/s, which is an appropriate speed range for wind power generation. Furthermore, several studies have proposed that wind energy can help solve the electricity crisis in the country [41-50]. In [41], it was shown that a small-scale wind power system would be more financially beneficial for remote rural areas.

C. Importance of the Study

Sudan suffers from an electricity crisis characterized by chronic electricity shortages and a weakening power sector, which has adverse effects on businesses, healthcare facilities, and households, affecting daily life and hindering development efforts. To address the electricity crisis, Sudan needs significant investments to promote renewable energy sources, such as wind energy, and implement energy efficiency measures to ensure a reliable and sustainable electricity supply. Most of the previous studies on wind energy evaluation in Sudan used data from the NASA database. However, only the studies in [42-45] and [50] employed measured data to determine the country's wind energy potential. This study aims to improve and evaluate the accuracy and reliability of four satellite-based wind energy sources by comparing them with measured data collected from eight locations in Sudan. Bias correction methods are adopted to reduce the bias of satellite-derived wind energy products. This evaluation is critical before deploying these data sources for wind potential evaluations and wind farm development. The measured data were collected from [42]. Additionally, statistical analysis of observed data and satellite-based wind energy outputs is performed implementing the two-parameter

Weibull distribution function, both before and after applying bias correction approaches. Moreover, an assessment of the Wind Power Density (WPD) is performed to evaluate the wind power potential of the country. Engaging data from several sources, this study offers a novel approach to decision-making in Sudan's first stages of wind farm development. This all-inclusive strategy closes a significant research gap and offers insightful information to potential wind farm investors in the country.

II. MATERIALS

A. Study Area

This study focuses on the eight sites listed in Table I, which were chosen due to the availability of measured data for the period 1975–1985.

B. Mast-Measured Data

The mean monthly wind speed data used in this analysis were taken from [42]. These wind data were obtained from the Meteorological Department of Khartoum, Sudan. These wind speed data were gathered between 1975 and 1985 at a height of 10 m.

TABLE I. INFORMATION OF THE SELECTED LOCATIONS

Location number	Location name	Latitude [°N]	Longitude [°E]	Elevation [m]
L#1	Atbara	17.70	34.01	358
L#2	Shendi	16.67	33.45	364
L#3	Aroma	15.81	36.14	443
L#4	Khartoum	15.50	32.56	387
L#5	Kassla	15.46	36.40	505
L#6	El Showak	14.40	35.87	516
L#7	Wad Madani	14.39	33.54	412
L#8	E1 Gadarif	14.02	35.37	598

C. Reanalysis and Analysis of Dataset

Reanalysis and analysis of datasets include large-scale meteorological data produced by sophisticated data assimilation methods [51]. These datasets are produced by combining many observational sources utilizing advanced mathematical models, including observations from weather stations, satellite data, and other relevant inputs [2, 8, 10]. This study chose four datasets, ERA5, AgERA5, TerraClimate, and NASA, based on their high spatial resolution, coverage domain, and periods of availability, as observed in Table II.

TABLE II. MAIN CHARACTERISTICS OF THE SELECTED SATELLITE DATABASE USED IN THE STUDY

Sources	Description/Full name of the dataset	Resolution	Period	Spatial extent
ERA5	Fifth-generation reanalysis product of the European Centre for Medium-Range Weather Forecasts	0.05°/1 d	1979–present	Global
AgERA5	Daily surface meteorological data for 1979–present is input for agriculture and agro-ecological studies. It is based on the hourly ECMWF ERA5 data at the surface level.	0.1° × 0.1°	1979–present	Global
TerraClimate	Global gridded dataset of meteorological and water balance for global terrestrial surfaces	0.500°×0.625°	1958– present	Global
NASA	NASA/Forecasting of World Energy Resources (NASA/POWER)	0.05° × 0.05°	1981–present	Global

III. METHODS

A. Bias Correction Methods

The estimated wind speed values from datasets may contain errors due to inadequate parameterization techniques, leading to either overestimation or underestimation depending on data assimilation. This discrepancy, known as bias, often arises from locally dominant atmospheric components. To improve accuracy, bias correction methods can be applied deploying on-site measured wind speed data. These methods help mitigate bias impact, enhancing the overall reliability of wind speed estimates. Based on [51–52], Measure-Correlate-Predict (MCP) and Linear Adaptation (LA) are commonly employed to reduce the bias of model-derived data.

1) Measure-Correlate-Predict (MCP)

The bias in the estimated v , using measured v , can be reduced by finding a correction or scaling factor adopting the MCP method. The method utilized for bias correction, known as ratio bias correction [24], is mathematically represented in (1). The scaling factor is calculated by dividing the mean of measured v (\bar{v}_m) by the mean of estimated v (\bar{v}_e), and the scaling factor is multiplied by the estimated v data to get corrected v data (v_{ec}).

$$v_{ec,i} = \frac{\bar{v}_m}{\bar{v}_e} \times v_{e,i} \quad (1)$$

2) Linear-Adaptation (LA)

The bias correction by LA is another method to correct the estimated value of wind speed. The line of best fit between measured and estimated v is developed using (2) and then subtracted from the $v_e = v_m$ line to get (3). It should be noted that v_{en} represents the new estimated v with negligible bias.

$$v_{e,i} = m_1 v_{m,i} + c_1 \quad (2)$$

$$v_{en,i} = V_{e,i} - [(m_1 - 1)v_{m,i} + c_1] \quad (3)$$

Furthermore, the correction performed based on mast measurements can be extended to non-overlapping years by developing a linear fit between new estimated v data (v_{en}) and estimated v (v_e) as in (5) to get the slope and intercept. Equation (6) is employed to find a time series of corrected v data. This bias correction method is referred to as LA1.

$$v_{en,i} = m_2 v_{e,i} + c_2 \quad (4)$$

$$v_{ec,i} = m_2 v_{e,i} + c_2 \quad (5)$$

Moreover, bias correction based on on-site mast measurements can be accomplished by establishing a linear relationship between the measured and estimated wind speed, represented in (6). The slope and gradient of the best-fit line from this relationship are then deployed to derive the corrected time series using (7). This specific approach of bias correction utilizing linear adaptation is denoted as LA2.

$$v_{m,i} = mv_{e,i} + c \quad (6)$$

$$v_{ec,i} = mv_{e,i} + c \quad (7)$$

B. Statistical Indices

The performance of these datasets is assessed using statistical metrics like Coefficient of Determination (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), as outlined in (8-10).

$$R^2 = 1 - \frac{\sum_{i=1}^n (a_{a,i} - a_{p,i})^2}{\sum_{i=1}^n (a_{p,i} - a_{a,ave})^2} \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (a_{a,i} - a_{p,i})^2} \quad (9)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |a_{a,i} - a_{p,i}| \quad (10)$$

where n is the number of data, $a_{p,i}$ is the predicted value, $a_{a,i}$ is the actual value, $a_{a,ave}$ is the average actual value and i is the number of input variables.

C. Weibull Distribution Function and Wind Power Density

The Weibull distribution is the most common among the probability distributions employed in wind energy. This statistical model represents the pattern of wind speed data well and is adaptable and flexible. The probability density function $f(v)$ and cumulative distribution function $F(v)$, which are given in (11) and (12), respectively, are characteristics of the Weibull distribution [53]. The curve is defined by two parameters: the scale parameter c affects the mean wind speed, which indicates the typical wind speed at the site, and the form parameter k determines the curve's peak or spread.

$$f(v) = \frac{k}{c} \left(\frac{v}{c}\right)^{k-1} \exp\left[-\left(\frac{v}{c}\right)^k\right], v > 0, k > 0, c > 0 \quad (11)$$

$$F(v) = 1 - \exp\left[-\left(\frac{v}{c}\right)^k\right], v > 0, k > 0, c > 0 \quad (12)$$

In this study, the Maximum Likelihood Method (MLM) was selected due to its widely acknowledged efficiency. Using MLM, the parameters k and c of the Weibull distribution are estimated as follows [54]:

$$k = \left[\frac{\sum_{i=1}^n v_i^k \ln(v_i)}{\sum_{i=1}^n v_i^k} - \frac{1}{n} \sum_{i=1}^n \ln v_i \right] \quad (13)$$

$$c = \left(\frac{1}{n} \sum_{i=1}^n v_i^k \right) \quad (14)$$

Equation (12) involves numerical solving for the parameter k , and in this study, the bisection method was employed for solving (14).

D. Wind Power Density (WPD)

WPD is a crucial metric to assess the potential of wind resources and estimate the amount of wind energy that is accessible at particular locations [55]. This parameter is essential for evaluating wind turbine power and identifying the best kind of turbine for a particular location [56]. Equation (15) is used to calculate the WPD in W/m^2 based on wind speed data. Equation (16) can also be utilized to calculate WPD based on the Weibull distribution [55, 57].

$$\frac{P}{A} = \frac{1}{2} \rho \bar{v}^3 \quad (15)$$

$$\frac{P}{A} = \frac{1}{2} \rho c^3 \Gamma\left(\frac{k+3}{k}\right) \quad (16)$$

where v_i is the measured/observed wind speed (m/s), n is the sample size of the wind speed data, and ρ denotes the air density (kg/m^3). The standard air density is taken to be $\rho = 1.225 \text{ kg/m}^3$.

IV. RESULTS AND DISCUSSION

A. Descriptive Statistics of Mast Measurement and Estimated Data before using Bias Correction Methods

Figure 1 presents the descriptive statistics of wind speed data, including the mean, standard deviation, minimum, and maximum for all regions.

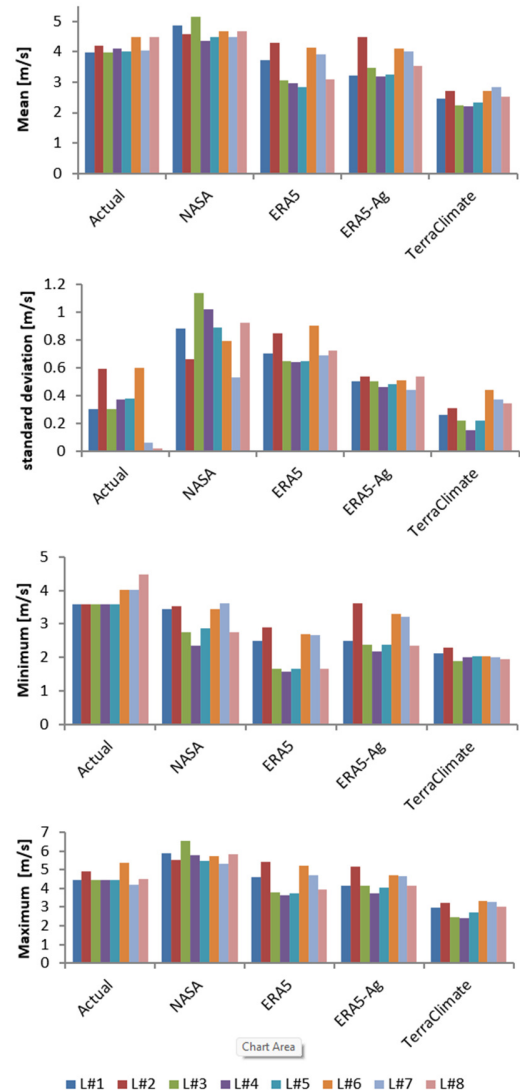


Fig. 1. Descriptive statistics of wind speed data for all locations before using bias correction methods.

Focusing on the mast-measured data, the mean wind speed ranged from 3.99 to 4.48 m/s. The maximum value of the mean monthly wind speed was recorded in Wad Madani, while the minimum value was found in Aroma and El Gadarif. Moreover, CV is a statistical measure that represents the relative variability or dispersion of a dataset compared to its mean. CV was found to be moderately low, ranging from 0.39 to 13.09%, which means that the wind speed data for all regions demonstrate relatively low variability compared to their respective mean values. On the other hand, when considering the satellite database, the mean wind speed ranges between 2.47 and 4.78 m/s for Aroma, 2.72 and 4.59 m/s for Atbara, 2.25 and 5.16 m/s for El Gadarif, 2.22 and 4.36 m/s for El Showak, 2.33 and 4.49 m/s for Kassala, 2.71 and 4.66 m/s for Khartoum, 2.83 and 4.50 m/s for Shendi, and 2.53 and 4.66 m/s for Wad Madani. Notably, the maximum and minimum wind speed values were obtained from NASA and TerraClimate datasets, respectively. The mean values collected from ERA5, NASA, and ERA5-Ag were found to be close to the mast-measured data.

B. Statistical Analysis before using Bias Correction Methods

Figure 2 depicts the R^2 , MAE, and RMSE statistical parameters for the wind data, by comparing estimated datasets with the measured data for all locations.

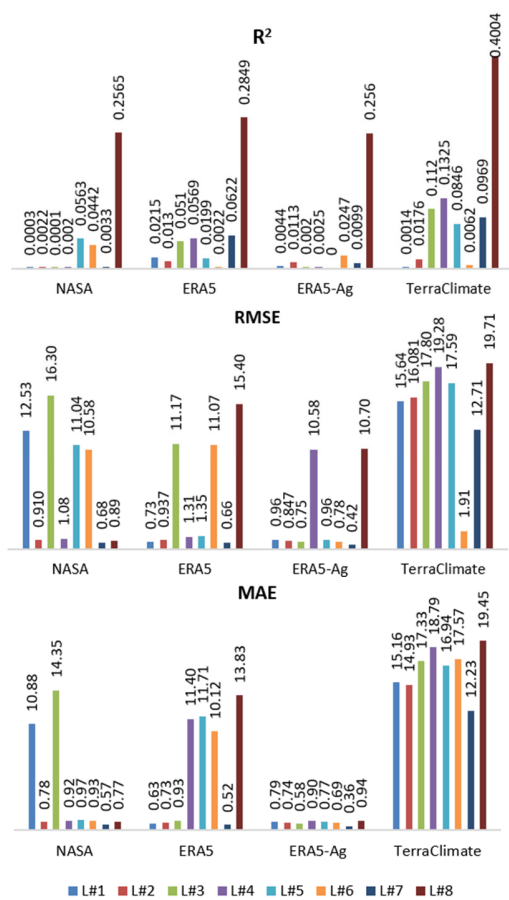


Fig. 2. Statistical parameters for wind speed using monthly mean wind speed.

R^2 ranges from 0 to 1, and as reported in [58], an acceptable R^2 value is typically considered to be higher than 0.5. Figure 2 exhibits that R^2 is within the range of 0.00-0.400, suggesting that the dataset does not fit the data well and that there is a significant amount of unexplained variability or error in the estimation of wind speed data. Furthermore, RMSE is employed to measure the level of agreement between observed and estimated data. An RMSE value of 0 indicates a perfect match between the observed and estimated values, and increasing RMSE values indicate a progressively poorer match. The analysis reveals that the RMSE varied between 0.4242 and 1.9707 m/s. The minimum RMSE value was observed when using the ERA5-Ag dataset at Shendi. The maximum RMSE value was obtained when utilizing the TerraClimate dataset at Wad Madani. MSE is a statistical measure that quantifies the average squared difference between the actual and the estimated data. A lower MAE indicates better accuracy, as it represents smaller differences between the observed and estimated values. MAE ranged between 0.3593 and 1.9449m/s. and the lowest and highest values were recorded at Shendi and Wad Madani, respectively.

C. Determination of Weibull Parameters and Wind Power Density before using Bias Correction Methods

Figure 3 presents the Weibull distribution parameters for the chosen locations.

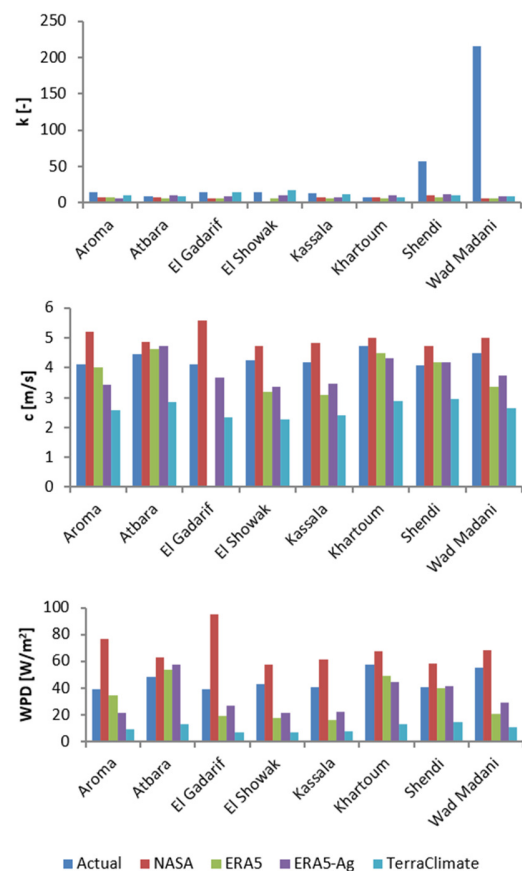


Fig. 3. Weibull parameters and wind power density for all locations.

These parameters were determined implementing the maximum likelihood approach based on the mean monthly wind speed data collected from different datasets. The k values were within the range of 7.9-216.00, 5.56-10.34, 5.73-7.81, 6.86-12.18, and 7.32-17.36 for the mast-measured, NASA, ERA5, ERA5-Ag, and TerraClimate data, respectively. The c value was found to range from 4.08 to 4.74 m/s for the mast-measured dataset. In contrast, c ranged from 2.29 to 5.59 m/s for all datasets. The maximum value of 5.59 m/s was obtained in El Gadarif, while the minimum value was recorded in El Showak. k varied between 7.9 and 216.00 for the mast-measured data, 5.56 and 10.34 for NASA data, 5.73 and 7.81 for ERA5 data, 6.86 and 12.18 for ERA5-Ag data, and 7.32 and 17.36 for TerraClimate data. WPD was calculated to evaluate the wind potential at the selected locations. In the measured dataset, WPD ranged from 39.31 to 57.83 W/m² with an average value of 45.56 W/m². Based on wind power classification at 10 m height [10], the selected locations can be considered fair (WPD < 100 W/m²). As a result, in the designated zones, small-scale wind turbines are appropriate for utilizing the wind energy potential. Moreover, it can be deduced that wind energy potential in the designated sites can be collected by high-capacity wind turbines (MWs) that are 90 m or higher. Concentrating on NASA, ERA5, ERA5-Ag, and TerraClimate datasets, it was found that WPD varied between 6.77 W/m² (El Showak, TerraClimate dataset) and 94.77 W/m² (El Gadarif, NASA dataset). In this case, the wind energy potential in the selected locations is also categorized as fair. Therefore, small wind turbines are suitable for use in areas selected to exploit the potential of available wind energy.

D. Descriptive Statistics of Mast Measurement and Estimated Data after using Bias Correction Methods

The reduction in error in the reanalysis and analysis datasets was checked employing the bias correction methods MCP, LA1, and LA2. Using a variety of statistical indices, including R^2 , RMSE, and MAE, the performance of the corrective reanalysis and analysis datasets was compared with the measured data to determine the optimal bias correction methods. The R^2 values were within the range of 0.00-0.400 deploying the MCP and LA2 methods, indicating that the data do not fit well and there is a significant amount of unexplained variability or error in the estimation of the wind speed data. When utilizing the LA1 method, the R^2 values were within the range of 0.0033-0.8792. Based on these findings, some of the R^2 values acquired from the LA1 method could fall below the acceptable range. Therefore, it can be concluded that the corrected data gathered from the TerraClimate dataset is the best option compared to the other datasets, especially for the regions of Aroma, Atbara, El Gadarif, El Showak, Kassala, and Khartoum. The TerraClimate dataset yielded higher R^2 values, indicating a better fit of the model to the observed wind speed data in these regions compared to the other datasets and methods applied. The LA1 method produced the lowest RMSE and MAE values compared to other methods. RMSE and MAE varied from 0.0129 to 1.512 m/s and from 0.0112 to 1.3829 m/s, accordingly. In this case, the ranges of RMSE and MAE values suggest that some datasets have a better fit, with smaller differences between the observed and predicted wind speed values, while others may exhibit larger discrepancies, resulting

in higher RMSE values. Overall, the corrected data from TerraClimate is the best option compared to the other datasets.

Figure 4 portrays the R^2 , RMSE, and MAE values for the optimum bias correction method, LA1. Figure 5 illustrates the descriptive statistics for wind speed data in all regions. Based on the mast-measured data, the mean wind speed ranged between 3.99 and 4.48 m/s. The maximum mean monthly wind speed was recorded in Wad Madani, while the minimum was found in Aroma and El Gadarif. Regarding the corrected satellite database, it was found that the mean wind speed was within the range of 3.99 and 4.48 m/s, indicating that the mean value of the corrected satellite database is close to the measured value for all selected locations. Additionally, the findings demonstrated that the standard deviation, minimum, and maximum values of the corrected wind speed satellite database are close to the measured value for all the selected locations. This agreement between the corrected satellite data and the measured values demonstrates that the correction process was effective in producing reliable and accurate wind speed estimates for all the selected locations, boosting confidence in using them for further analysis and decision-making related to wind energy projects and other applications.

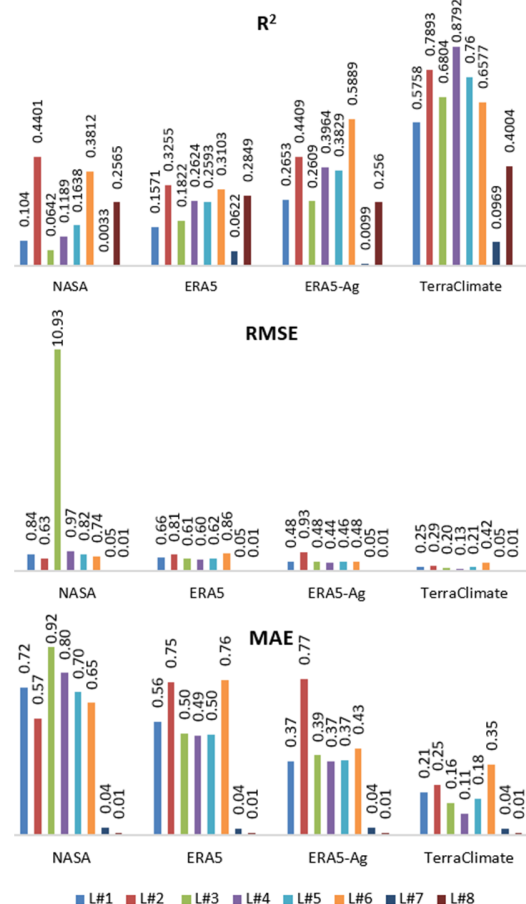


Fig. 4. Statistical parameters for wind speed using monthly mean wind speed.

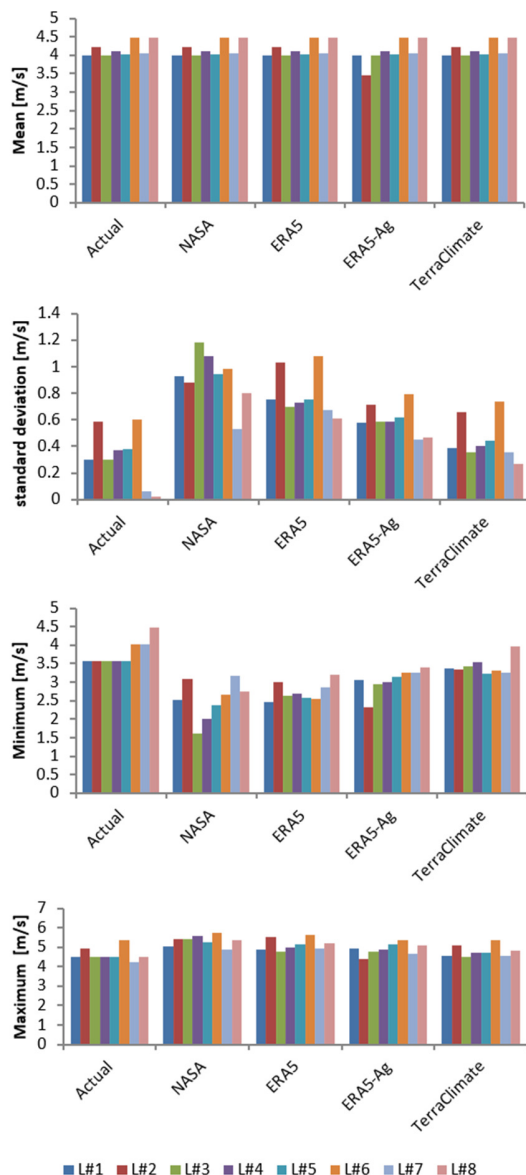


Fig. 5. Descriptive statistics of wind speed data for all locations after using bias correction methods.

E. Determination of Weibull Parameters and Wind Power Density after Using Bias Correction Methods

Figure 6 displays the Weibull distribution parameters for all locations, calculated implementing the maximum likelihood approach. WPD values were determined based on these parameters. The k values were found to fall within the following ranges: 7.99-216.00 for mast-measured data, 4.31-9.34 for NASA data, 4.85-10.07 for ERA5 data, 5.76-13.77 for ERA5-Ag data, and 4.14-4.77 for TerraClimate data. Additionally, for the mast-measured dataset, c ranged from 4.08 to 4.74 m/s. However, for all datasets utilized, c ranged from 3.72 to 4.86 m/s. The maximum c value of 4.86 m/s was obtained in Khartoum, while the minimum value was recorded in Atbara. In the mast-measured dataset, WPD ranged from 39.31 to 57.83 W/m², with an average value of 45.56 W/m².

The wind energy generation potential of the sites is classified according to the average WPD values. Based on the classification of wind power at 10 m height, the selected locations can be considered poor (WPD < 100 W/m²). Therefore, small-scale wind turbines are suitable to be used in selected regions to exploit the available wind energy potential. Furthermore, it can be concluded that high-capacity wind turbines with a height of 90 m and higher can be suitable for gathering the wind energy potential in the selected locations. This is investigated employing the power-law model, i.e., the collected data at 10 m height is synthesized to the 90 m height, which is the height of most 1 MW or above wind turbines. Based on NASA, ERA5, ERA5-Ag, and TerraClimate datasets, WPD was found to vary between 6.77 W/m² (El Showak, TerraClimate dataset) and 94.77 W/m² (El Gadarif, NASA dataset). In this case, the wind energy potential in the selected locations is also classified as fair. Therefore, small wind turbines are suitable for use in selected areas to exploit the potential of available wind energy.

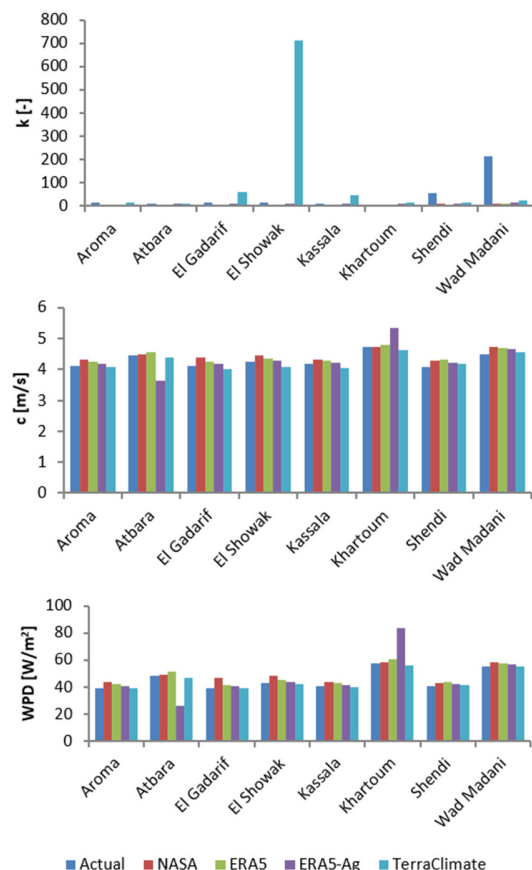


Fig. 6. Weibull parameters and wind power density for all locations.

V. CONCLUSIONS

Due to the unavailability of historically measured wind data in the country, assessing and forecasting next-generation wind resources becomes more important when using data from satellite observations and different reanalyses. These alternate data sources are essential resources for assessing wind energy

potential and financial viability in certain areas. However, it is important to validate the suitability of these datasets for specific regions by comparing them with the available mast-measured data. Therefore, the reanalysis and analysis estimates of wind speed data were evaluated and compared to the mean monthly measured data collected from eight locations in Sudan. In general, ERA5-Ag performs better than other datasets. The original wind speed was modified employing three bias correction techniques, MCP, LA1, and LA2. According to the statistical criteria applied, LA1 performed better than MCP and LA2.

Especially in developing countries, such as Sudan, corrected wind speed data from reanalysis and analysis datasets can be utilized in the early assessment of wind resources. Wind speed characteristics were analyzed deploying the two-parameter Weibull distribution, whose parameters were used to determine the annual WPD values at the chosen places. The results revealed that the selected locations can be categorized as fair based on wind power at a height of 10 m. Therefore, small-scale wind turbines are suitable for harnessing the wind energy potential at the selected locations. Furthermore, it can be concluded that high-capacity wind turbines of 90 m or more height can exploit the wind energy potential. This study presents a wind energy road map with the aim of attracting investors interested in clean energy technology to mitigate the impact of global warming and achieve sustainable technological development in the country. In the future, onshore wind farms should be the focus of technical and economic analysis based on a variety of data sources, focusing on the investment payback period, the selection of the most suitable turbines, and the present value of the electricity production cost.

REFERENCES

- [1] Y. Kassem, H. Gokcekus, I. A. A. Albakoush, and K. S. B. Abdullah, "Solar-Powered Solutions for the Water and Energy Shortage Problem: The Case Study of Nahr El Bared, Lebanon," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10861–10869, Jun. 2023, <https://doi.org/10.48084/etasr.5858>.
- [2] Y. Kassem, H. Gökçekuş, and M. H. A. Abdalla, "Wind energy resource assessment based on the use of multiple satellite data for sustainable energy production in Sudan," *Environment, Development and Sustainability*, Nov. 2023, <https://doi.org/10.1007/s10668-023-04221-w>.
- [3] Y. Kassem, H. Gokcekus, H. Camur, and A. H. A. Abdelnaby, "Wind Power Generation Scenarios in Lebanon," *Engineering, Technology & Applied Science Research*, vol. 12, no. 6, pp. 9551–9559, Dec. 2022, <https://doi.org/10.48084/etasr.5258>.
- [4] Y. Kassem and M. H. A. Abdalla, "Modeling predictive suitability to identify the potential of wind and solar energy as a driver of sustainable development in the Red Sea state, Sudan," *Environmental Science and Pollution Research*, vol. 29, no. 29, pp. 44233–44254, Jun. 2022, <https://doi.org/10.1007/s11356-022-19062-9>.
- [5] W. S. Ebhota and T. C. Jen, "Fossil Fuels Environmental Challenges and the Role of Solar Photovoltaic Technology Advances in Fast Tracking Hybrid Renewable Energy System," *International Journal of Precision Engineering and Manufacturing-Green Technology*, vol. 7, no. 1, pp. 97–117, Jan. 2020, <https://doi.org/10.1007/s40684-019-00101-9>.
- [6] Y. Kassem, H. Gokcekus, and F. A. R. Agila, "Techno-Economic Feasibility Assessment for the promotion of Grid-Connected Rooftop PV Systems in Botswana: A Case Study," *Engineering, Technology & Applied Science Research*, vol. 13, no. 2, pp. 10328–10337, Apr. 2023, <https://doi.org/10.48084/etasr.5668>.
- [7] H. Patidar, V. Shende, P. Baredar, and A. Soni, "Comparative study of offshore wind energy potential assessment using different Weibull parameters estimation methods," *Environmental Science and Pollution Research*, vol. 29, no. 30, pp. 46341–46356, Jun. 2022, <https://doi.org/10.1007/s11356-022-19109-x>.
- [8] A. Kanwal *et al.*, "Evaluation of Reanalysis and Analysis Datasets against Measured Wind Data for Wind Resource Assessment," *Energy & Environment*, vol. 34, no. 5, pp. 1258–1284, Aug. 2023, <https://doi.org/10.1177/0958305X221084078>.
- [9] Z. Wang and W. Liu, "Wind energy potential assessment based on wind speed, its direction and power data," *Scientific Reports*, vol. 11, no. 1, Aug. 2021, Art. no. 16879, <https://doi.org/10.1038/s41598-021-96376-7>.
- [10] H. S. A. Lagili, A. Kiraz, Y. Kassem, and H. Gökçekuş, "Wind and Solar Energy for Sustainable Energy Production for Family Farms in Coastal Agricultural Regions of Libya Using Measured and Multiple Satellite Datasets," *Energies*, vol. 16, no. 18, Jan. 2023, Art. no. 6725, <https://doi.org/10.3390/en16186725>.
- [11] A. Fekih, M. Abdelouahab, and Y. Marif, "Evaluation of wind resource and mapping during 2009–2018 based on ERA5 reanalysis data: a case study over Algeria," *International Journal of Energy and Environmental Engineering*, vol. 14, no. 1, pp. 15–34, Mar. 2023, <https://doi.org/10.1007/s40095-022-00500-w>.
- [12] H. Miao, D. Dong, G. Huang, K. Hu, Q. Tian, and Y. Gong, "Evaluation of Northern Hemisphere surface wind speed and wind power density in multiple reanalysis datasets," *Energy*, vol. 200, Jun. 2020, Art. no. 117382, <https://doi.org/10.1016/j.energy.2020.117382>.
- [13] N. A. Arreyndip, E. Joseph, and A. David, "Wind energy potential assessment of Cameroon's coastal regions for the installation of an onshore wind farm," *Heliyon*, vol. 2, no. 11, Nov. 2016, Art. no. e00187, <https://doi.org/10.1016/j.heliyon.2016.e00187>.
- [14] J. Sanz Rodrigo, J.-M. Buchlin, J. van Beeck, J. T. M. Lenaerts, and M. R. van den Broeke, "Evaluation of the antarctic surface wind climate from ERA reanalyses and RACMO2/ANT simulations based on automatic weather stations," *Climate Dynamics*, vol. 40, no. 1, pp. 353–376, Jan. 2013, <https://doi.org/10.1007/s00382-012-1396-y>.
- [15] D. Carvalho, A. Rocha, M. Gómez-Gesteira, and C. Silva Santos, "Offshore wind energy resource simulation forced by different reanalyses: Comparison with observed data in the Iberian Peninsula," *Applied Energy*, vol. 134, pp. 57–64, Dec. 2014, <https://doi.org/10.1016/j.apenergy.2014.08.018>.
- [16] R. McKenna, S. Hollnaicher, P. Ostman v. d. Leye, and W. Fichtner, "Cost-potentials for large onshore wind turbines in Europe," *Energy*, vol. 83, pp. 217–229, Apr. 2015, <https://doi.org/10.1016/j.energy.2015.02.016>.
- [17] D. Mentis, S. Hermann, M. Howells, M. Welsch, and S. H. Siyal, "Assessing the technical wind energy potential in Africa a GIS-based approach," *Renewable Energy*, vol. 83, pp. 110–125, Nov. 2015, <https://doi.org/10.1016/j.renene.2015.03.072>.
- [18] D. Mentis, S. H. Siyal, A. Korkovelos, and M. Howells, "A geospatial assessment of the techno-economic wind power potential in India using geographical restrictions," *Renewable Energy*, vol. 97, pp. 77–88, Nov. 2016, <https://doi.org/10.1016/j.renene.2016.05.057>.
- [19] N. A. Arreyndip and E. Joseph, "Small 500 kW onshore wind farm project in Kribi, Cameroon: Sizing and checkers layout optimization model," *Energy Reports*, vol. 4, pp. 528–535, Nov. 2018, <https://doi.org/10.1016/j.egy.2018.08.003>.
- [20] F. Onea and E. Rusu, "Sustainability of the Reanalysis Databases in Predicting the Wind and Wave Power along the European Coasts," *Sustainability*, vol. 10, no. 1, Jan. 2018, Art. no. 193, <https://doi.org/10.3390/su10010193>.
- [21] J. Olauson, "ERA5: The new champion of wind power modelling?," *Renewable Energy*, vol. 126, pp. 322–331, Oct. 2018, <https://doi.org/10.1016/j.renene.2018.03.056>.
- [22] H. Gökçekuş, Y. Kassem, and M. A. Hassan, "Evaluation of Wind Potential at Eight Selected Locations in Northern Lebanon Using Open Source Data," *International Journal of Applied Engineering Research*, vol. 14, no. 11, pp. 2789–2794, 2019.

- [23] O. Assowe Dabar, M. O. Awaleh, D. Kirk-Davidoff, J. Olauson, L. Söder, and S. I. Awaleh, "Wind resource assessment and economic analysis for electricity generation in three locations of the Republic of Djibouti," *Energy*, vol. 185, pp. 884–894, Oct. 2019, <https://doi.org/10.1016/j.energy.2019.07.107>.
- [24] S. M. Boudia and J. A. Santos, "Assessment of large-scale wind resource features in Algeria," *Energy*, vol. 189, Dec. 2019, Art. no. 116299, <https://doi.org/10.1016/j.energy.2019.116299>.
- [25] A. J. A. Gudo, J. Deng, M. Belete, and G. A. Abubakar, "Estimation of Small Onshore Wind Power Development for Poverty Reduction in Jubek State, South Sudan, Africa," *Sustainability*, vol. 12, no. 4, Jan. 2020, Art. no. 1483, <https://doi.org/10.3390/su12041483>.
- [26] Y. Kassem, H. Çamur, and R. A. F. Aateg, "Exploring Solar and Wind Energy as a Power Generation Source for Solving the Electricity Crisis in Libya," *Energies*, vol. 13, no. 14, Jan. 2020, Art. no. 3708, <https://doi.org/10.3390/en13143708>.
- [27] S. A. Gil Ruiz, J. E. C. Barriga, and J. A. Martínez, "Wind power assessment in the Caribbean region of Colombia, using ten-minute wind observations and ERA5 data," *Renewable Energy*, vol. 172, pp. 158–176, Jul. 2021, <https://doi.org/10.1016/j.renene.2021.03.033>.
- [28] G. Gualtieri, "Reliability of ERA5 Reanalysis Data for Wind Resource Assessment: A Comparison against Tall Towers," *Energies*, vol. 14, no. 14, Jan. 2021, Art. no. 4169, <https://doi.org/10.3390/en14144169>.
- [29] B. Hiçsönmez, Y. Kassem, and H. Gökçekuş, "Techno-economic assessment of wind potential at five locations in northern cyprus using open source wind data," *International Journal of Scientific & Technology Research*, vol. 10, no. 2, pp. 294–299, 2021.
- [30] C. Jung and D. Schindler, "On the influence of wind speed model resolution on the global technical wind energy potential," *Renewable and Sustainable Energy Reviews*, vol. 156, Mar. 2022, Art. no. 112001, <https://doi.org/10.1016/j.rser.2021.112001>.
- [31] R. H. Basheir and M. S. Abdelrahman, "The electricity crisis in Sudan I Transnational Institute," Feb. 21, 2024, <https://www.tni.org/en/article/the-electricity-crisis-in-sudan>.
- [32] S. O. Fadlallah and D. E. Benhadji Serradj, "Determination of the optimal solar photovoltaic (PV) system for Sudan," *Solar Energy*, vol. 208, pp. 800–813, Sep. 2020, <https://doi.org/10.1016/j.solener.2020.08.041>.
- [33] A. M. Omer, "Renewable energy resources for electricity generation in Sudan," *Renewable and Sustainable Energy Reviews*, vol. 11, no. 7, pp. 1481–1497, Sep. 2007, <https://doi.org/10.1016/j.rser.2005.12.001>.
- [34] R. A. Hamza and A. O. Abdelatif, "Foundations, Design, and Dynamic Performance of Wind Turbines: Overview and Challenges in Sudan," *FES Journal of Engineering Sciences*, vol. 9, no. 1, pp. 96–103, 2020, <https://doi.org/10.52981/fjes.v9i1.665>.
- [35] M. Abdalla and T. Qarmout, "An analysis of Sudan's energy sector and its renewable energy potential in a comparative African perspective," *International Journal of Environmental Studies*, vol. 80, no. 4, pp. 1169–1187, Jul. 2023, <https://doi.org/10.1080/00207233.2023.2177417>.
- [36] S. O. Fadlallah, D. E. Benhadji Serradj, and D. M. Sedzro, "Is this the right time for Sudan to replace diesel-powered generator systems with wind turbines?," *Renewable Energy*, vol. 180, pp. 40–54, Dec. 2021, <https://doi.org/10.1016/j.renene.2021.08.018>.
- [37] A. S. Zalhaf *et al.*, "A High-Resolution Wind Farms Suitability Mapping Using GIS and Fuzzy AHP Approach: A National-Level Case Study in Sudan," *Sustainability*, vol. 14, no. 1, Jan. 2022, Art. no. 358, <https://doi.org/10.3390/su14010358>.
- [38] Z. A. Khan, M. Imran, A. Altamimi, O. E. Diemuodeke, and A. O. Abdelatif, "Assessment of Wind and Solar Hybrid Energy for Agricultural Applications in Sudan," *Energies*, vol. 15, no. 1, Jan. 2022, Art. no. 5, <https://doi.org/10.3390/en15010005>.
- [39] A. A. Mohammed, "Evaluation of potentials for renewable energy in the power system in Sudan," Ph.D. dissertation, Sudan University of Science and Technology, 2018.
- [40] T. M. Saeed, E. B. M. Tayeb, and G. Osman, "Sustainable Energy Potential in Sudan," *SUST Journal of Engineering and Computer Sciences*, vol. 2, no. 3, 2019.
- [41] Y. Kassem and M. H. A. Abdalla, "Modeling predictive suitability to identify the potential of wind and solar energy as a driver of sustainable development in the Red Sea state, Sudan," *Environmental Science and Pollution Research*, vol. 29, no. 29, pp. 44233–44254, Jun. 2022, <https://doi.org/10.1007/s11356-022-19062-9>.
- [42] A. M. Omer, "Compilation and evaluation of solar and wind energy resources in Sudan," *Renewable Energy*, vol. 12, no. 1, pp. 39–69, Sep. 1997, [https://doi.org/10.1016/S0960-1481\(97\)00009-8](https://doi.org/10.1016/S0960-1481(97)00009-8).
- [43] A. M. Omer, "Wind energy in Sudan," *Renewable Energy*, vol. 19, no. 3, pp. 399–411, Mar. 2000, [https://doi.org/10.1016/S0960-1481\(99\)00054-3](https://doi.org/10.1016/S0960-1481(99)00054-3).
- [44] A. M. Omer, "On the wind energy resources of Sudan," *Renewable and Sustainable Energy Reviews*, vol. 12, no. 8, pp. 2117–2139, Oct. 2008, <https://doi.org/10.1016/j.rser.2006.10.010>.
- [45] A. M. Omer and E. Braima, "Renewable Energy Potential in Sudan," *International Energy Journal*, vol. 2, no. 1, 2001.
- [46] T. Salih, Y. Wang, and M. A. A. Adam, "Renewable Micro Hybrid System of Solar Panel and Wind Turbine for Telecommunication Equipment in Remote Areas in Sudan," *Energy Procedia*, vol. 61, pp. 80–83, Jan. 2014, <https://doi.org/10.1016/j.egypro.2014.11.911>.
- [47] A. R. S. Salih and A. E. Aydrus, "Compilation and Evaluation of wind energy resources in Jebel Awlia area, south Khartoum," *International Journal of Life Science and Engineering*, vol. 1, no. 3, pp. 97–100, 2015.
- [48] A. M. A. Khadam, E. A. Hamouda, and K. R. Doud, "Wind Power Harnessing in Sudan, Opportunities and Challenges," *University of Khartoum Engineering Journal*, vol. 6, no. 2, Oct. 2022, <https://doi.org/10.53332/kuej.v6i2.998>.
- [49] M. R. Elkadeem, S. Wang, S. W. Sharshir, and E. G. Atia, "Feasibility analysis and techno-economic design of grid-isolated hybrid renewable energy system for electrification of agriculture and irrigation area: A case study in Dongola, Sudan," *Energy Conversion and Management*, vol. 196, pp. 1453–1478, Sep. 2019, <https://doi.org/10.1016/j.enconman.2019.06.085>.
- [50] A. Younis *et al.*, "Wind Speed Forecast for Sudan Using the Two-Parameter Weibull Distribution: The Case of Khartoum City," *Wind*, vol. 3, no. 2, pp. 213–231, Jun. 2023, <https://doi.org/10.3390/wind3020013>.
- [51] Z. ul R. Tahir, M. Asim, M. Azhar, G. Moenuddin, and M. Farooq, "Correcting solar radiation from reanalysis and analysis datasets with systematic and seasonal variations," *Case Studies in Thermal Engineering*, vol. 25, Jun. 2021, Art. no. 100933, <https://doi.org/10.1016/j.csite.2021.100933>.
- [52] Y. Kassem, H. Gökçekuş, and N. Alijl, "Gridded Precipitation Datasets and Gauge Precipitation Products for Driving Hydrological Models in the Dead Sea Region, Jordan," *Sustainability*, vol. 15, no. 15, Jan. 2023, Art. no. 11965, <https://doi.org/10.3390/su151511965>.
- [53] Y. Kassem, M. M. Mizran, and S. M. Alsayas, "Evaluation of the Wind Energy Potential in Lebanon's Coastal Regions using Weibull Distribution Function," *International Journal of Engineering Research and Technology*, vol. 12, no. 6, 2019.
- [54] Y. Kassem, H. Çamur, S. Abughinda, and A. Sefik, "Wind Energy Potential Assessment in Selected Regions in Northern Cyprus Based on Weibull Distribution Function," *Journal of Engineering and Applied Sciences*, vol. 15, pp. 128–140, Oct. 2019, <https://doi.org/10.36478/jeasci.2020.128.140>.
- [55] M. M. Alayat, Y. Kassem, and H. Çamur, "Assessment of Wind Energy Potential as a Power Generation Source: A Case Study of Eight Selected Locations in Northern Cyprus," *Energies*, vol. 11, no. 10, Oct. 2018, Art. no. 2697, <https://doi.org/10.3390/en11102697>.
- [56] Y. Kassem, H. Çamur, and A. H. A. Abdelnaby, "Wind Power Prediction in Mediterranean Coastal Cities Using Multi-layer Perceptron Neural Network," in *Proceedings of Data Analytics and Management*, Singapore, 2023, pp. 253–264, https://doi.org/10.1007/978-981-99-6553-3_20.
- [57] Y. Kassem, H. Çamur, M. A. H. A. Abdalla, B. D. Erdem, and A. M. R. Al-ani, "Evaluation of wind energy potential for different regions in Lebanon based on NASA wind speed database," *IOP Conference Series*:

Earth and Environmental Science, vol. 926, no. 1, Aug. 2021, Art. no. 012093, <https://doi.org/10.1088/1755-1315/926/1/012093>.

- [58] Y. Kassem, H. Gökçekuş, and A. A. S. Mosbah, "Prediction of monthly precipitation using various artificial models and comparison with mathematical models," *Environmental Science and Pollution Research*, vol. 30, no. 14, pp. 41209–41235, Mar. 2023, <https://doi.org/10.1007/s11356-022-24912-7>.

Magnesium Oxide (MgO) as a Sustainable Catalyst for Biodiesel Production from Waste Cooking Oil: A Comparative Study with KOH

Aboulbaba Eladeb

Department of Chemical and Materials Engineering, Northern Border University, Saudi Arabia
eladebboulbaba@gmail.com (corresponding author)

Received: 13 February 2024 | Revised: 26 February 2024 | Accepted: 2 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7055>

ABSTRACT

The present study investigates the efficiency of magnesium oxide (MgO) as a heterogeneous catalyst in the production of biodiesel from waste cooking oil (WCO), putting an emphasis on its environmental benefits, cost-effectiveness and operational efficacy. Through a series of experiments, we optimized the reaction conditions, including catalyst concentration, reaction temperature, and ethanol to WCO molar ratio, to achieve a high biodiesel yield. The results indicate that an optimal MgO concentration of 3 wt%, a reaction temperature of 65 °C, and a molar ratio of 9:1 result into the highest biodiesel production efficiency. Additionally, MgO demonstrated significant reusability without a decrease in performance, underscoring its economic and environmental advantages. Comparative analysis revealed that MgO outperforms conventional KOH catalysts in terms of yield, purity, and sustainability. Our study suggests future research directions, including the optimization of MgO preparation methods and the exploration of co-catalyst systems to further enhance biodiesel production from WCO. This research contributes to the development of sustainable biodiesel production methods, aligning with global energy and environmental goals.

Keywords-biodiesel; WCO; MgO; transesterification

I. INTRODUCTION

The need for sustainable energy sources becomes more imperative day by day due to population growth, urbanization, and industrialization. Fossil fuels, sometimes known as traditional energy sources, have powered our societies for ages. Yet, their finite nature poses a significant challenge, as they cannot be restored once depleted. Moreover, the consumption of fossil fuels is associated with a significant environmental impact. In response to the increasing global energy demand and the pressing need for sustainable alternatives, biodiesel emerges as a pivotal renewable energy source. Derived from Waste Cooking Oil (WCO), biodiesel not only mitigates waste disposal issues but also reduces reliance on non-renewable energy sources. Its production and utilization underscore a significant stride towards achieving energy sustainability and reducing the environmental impact, positioning biodiesel as a crucial component in the renewable energy mix. This study explores in detail the efficiency of magnesium oxide (MgO) as a sustainable catalyst in biodiesel production, aiming to enhance its viability and contribute to the renewable energy landscape. Biodiesel can be synthesized utilizing WCO as the primary feedstock. Biodiesel derived from WCO is renewable, biodegradable, and non-toxic [1-5]. Various catalysts have been explored to optimize the transesterification process. Among these, magnesium oxide (MgO) has garnered attention due to its unique properties, such as a high surface area,

alkalinity, and thermal stability, which contribute to its efficiency as a heterogeneous catalyst. This section presents a comprehensive analysis of existing research on MgO, comparing its performance and applicability with other commonly used catalysts, including homogeneous catalysts like sodium hydroxide (NaOH) and potassium hydroxide (KOH), as well as heterogeneous catalysts such as calcium oxide (CaO) and zinc oxide (ZnO) [2-4].

As the relevant literature suggests, MgO offers several advantages, including higher biodiesel yield, operational stability and reusability compared to traditional homogeneous catalysts that tend to be prone to saponification and call for extensive purification. Furthermore, compared to heterogeneous catalysts, MgO demonstrates superior catalytic activity that can be attributed to its higher basicity and surface area. Various studies have also highlighted MgO's environmental benefits by lowering energy consumption and minimizing waste during biodiesel production [6]. However, challenges such as the need for catalyst activation and the optimization of reaction conditions to maximize MgO's effectiveness are areas that continue to be explored. Future research directions include improving the synthesis and modification of MgO to enhance its catalytic properties as well as investigating its performance in biodiesel production from diverse feedstocks. This study extends our previous

investigation on the ethanolysis of WCOs using KOH catalysts [6].

Magnesium oxide (MgO) has emerged as a highly efficient and promising heterogeneous catalyst for biodiesel production from WCOs due to its distinctive properties and advantages. MgO catalysts have been shown to enhance the activity and stability of transesterification reactions, leading to high biodiesel yields [7]. The use of MgO catalysts in biodiesel synthesis offers several benefits, including the reduction of pretreatment costs and the elimination of the negative dependence on water content of biodiesel yield [8]. Additionally, MgO catalysts have been found to exhibit excellent performance in terms of conversion efficiency and yield which some studies bring up to 92.63% [9]. The use of MgO catalysts also allows for the reusability of the catalyst, further enhancing its cost-effectiveness and sustainability [10]. MgO exhibits excellent catalytic activity due to its high surface area and inherent basicity, which are crucial for facilitating the transesterification reaction [11]. The stability of MgO under the reaction conditions is noteworthy, as it maintains its structural integrity and catalytic performance, making it a strong choice both for continuous and batch processes [12-14]. Mg-doped CaO-MgO catalysts prepared by co-precipitation showed improved activity and stability in transesterification for biodiesel production. Mg₃Al_{1-x}Ce_xO composite oxide catalysts prepared by co-precipitation exhibited enhanced catalytic activity in the transesterification of ethylene glycol to ethylene carbonate. A highly active heterogeneous catalyst, MgP, prepared by co-precipitation in the presence of pectin, showed efficient performance in biodiesel production and depolymerization of PolyEthylene Terephthalate (PET). High-Entropy Oxide (HEO) nanoparticles synthesized by photonic flash synthesis demonstrated prolonged stability and superior catalytic activity in the Oxygen Evolution Reaction (OER) compared to commercial IrO₂ catalyst. Various MgO catalysts prepared by different methods showed different catalytic activities and stability in Dimethyl Carbonate (DMC) synthesis via transesterification. MgO is a reusable catalyst that can be regenerated and reused in multiple cycles without a substantial decrease in its catalytic performance [15]. This reusability feature enhances the economic sustainability of the biodiesel production process by lowering operational costs [16].

MgO is recognized for its minimal environmental impact, primarily due to its non-toxicity and the absence of hazardous by-products during its use [17, 18]. Its application in biodiesel production aligns with the growing emphasis on green and sustainable chemical processes, reducing the dependency on conventional acid or base homogeneous catalysts that pose disposal and neutralization challenges [19]. MgO nanoparticles synthesized using green methods are non-toxic, eco-friendly, and have high stability for a wide range of biological, medical, and catalytic applications [20]. The use of waste coconut and fish oil as feedstock for biodiesel production, with MgO as a solid nano-catalyst, offers a low-cost and sustainable approach [10]. Mg-doped CaO-MgO catalysts prepared by co-precipitation, show improved activity and stability in transesterification reactions, making them suitable for biodiesel production. The utilization of MgO as a catalyst in the ethanolysis of WCO represents a significant advancement in

the field of biodiesel production. MgO has been found to address several limitations associated with traditional catalysts. It allows for lower catalyst concentrations, eliminates the need for complex post-reaction treatments, and reduces environmental concerns related to catalyst disposal [21]. Additionally, MgO has been shown to improve the activity and stability of catalysts in transesterification reactions, leading to higher biodiesel yields [22]. This advancement is important in making the biodiesel production process more sustainable and efficient.

Overall, MgO catalysts demonstrate great potential for the transesterification process in biodiesel production from WCO, offering a greener and more efficient alternative to traditional catalysts. This research contributes to the development of sustainable biodiesel production methods by utilizing MgO as a catalyst in the ethanolysis of WCO. The process can be carried out more effectively, contributing to the development of a more environmentally friendly and economically viable biodiesel industry.

II. METHODOLOGY

The preparation of the MgO catalyst is a critical step that significantly influences its catalytic performance in the biodiesel production process. For this study, the MgO catalyst was synthesized through a calcination method, which involves the thermal decomposition of magnesium carbonate or magnesium hydroxide at high temperatures. The precursor material (magnesium carbonate or magnesium hydroxide) was subjected to calcination in a muffle furnace at various temperatures ranging from 400°C to 800°C for 4 hours. This temperature range was chosen to investigate the impact of calcination temperature on the MgO's properties and its subsequent catalytic activity. The calcination temperature plays a key role in determining the surface area, pore size, and crystallinity of the MgO catalyst. A higher calcination temperature typically leads to an increase in crystallinity and a decrease in surface area due to sintering effects.

To evaluate the physicochemical properties of the synthesized MgO catalysts, several characterization techniques were employed. The BET (Brunauer-Emmett-Teller) method was used to measure the surface area and porosity of the MgO. A higher surface area offers increased availability of active sites, enhancing, therefore, the catalytic activity. Pore size distribution was determined using mercury intrusion porosimetry. This was necessary so as to investigate the implications the pore structure has for reactant accessibility and diffusion. The acidity of MgO, which influences its ability to catalyze the transesterification reaction, was assessed using the Temperature-Programmed Desorption (TPD) of ammonia.

The WCO used in the biodiesel production process was collected from different restaurants located in Arar city, Saudi Arabia [6]. Transesterification involves the reaction of alcohol and esters (triglycerides) to produce ethyl/methyl esters (biodiesel) and glycerol. This method is widely used for producing biodiesel from sources such as edible oils, waste cooking oils, animal fats, and yellow grease [3-5, 23]. The transesterification of WCO into biodiesel using the MgO catalyst was conducted under a set of optimized reaction

conditions to maximize the conversion efficiency. The process was carried out in a three-neck flask equipped with a reflux condenser and a magnetic stirrer. Key reaction parameters were a reaction temperature of 65 °C, atmospheric pressure, and a reaction time of 3 hours. These conditions were selected based on preliminary experiments aimed at optimizing MgO's performance.

A. Catalyst Loading and Ethanol to Oil Ratio

An optimal catalyst loading of 5 wt% relative to WCO was determined, along with an ethanol to oil molar ratio of 9:1, to ensure complete conversion of the triglycerides to biodiesel. The choice of a lower ethanol to oil ratio compared to conventional KOH-catalyzed processes reflects the high efficiency of MgO in facilitating the transesterification reaction.

B. Product Recovery and Analysis

Following the reaction, the mixture was allowed to settle, leading to the separation of biodiesel from glycerol. The biodiesel layer was washed off with warm water to remove any residual catalyst along with any impurities. It was, then, dried over anhydrous sodium sulfate before being characterized for its fuel properties according to ASTM standards.

III. RESULTS

A. Effect of Catalyst Concentration

The investigation into the effect of varying MgO concentrations on the yield of biodiesel from waste cooking oil (WCO) reveals a distinct trend. At lower MgO concentrations (1–3 wt%), a significant increase in biodiesel yield was observed, reaching an optimal value at 3 wt%. This suggests that MgO's higher surface area and catalytic efficiency allow for lower catalyst amounts compared to conventional KOH catalysis. Beyond this concentration, however, the yield reached a plateau while, above 5 wt%, it, slightly, decreased. A likely cause could be the mass transfer limitations and the increased viscosity of the reaction mixture, which hinders effective mixing whilst minimizing contact between reactants (Figure 1).

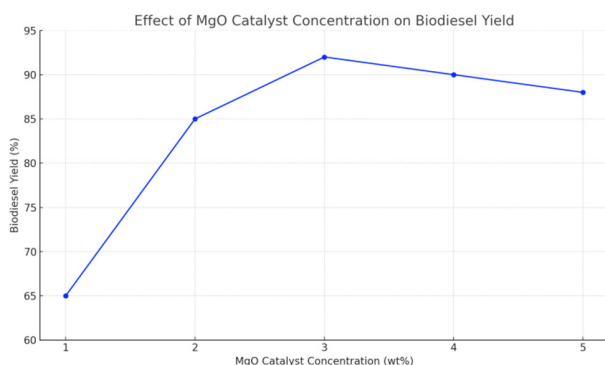


Fig. 1. Effect of MgO catalyst concentration on biodiesel yield.

B. Reaction Conditions Optimization

Optimizing reaction conditions for MgO-catalyzed transesterification showed a marked improvement in biodiesel

yield (Figure 2). The optimal reaction temperature was found to be 65°C, significantly lower than the typical temperatures required for KOH catalysis. This reduction in reaction temperature is attributed to MgO's operational stability and effective catalytic action at lower temperatures. Furthermore, the optimal ethanol to WCO molar ratio was determined to be 9:1, lower than the ratios commonly used with KOH, underscoring MgO's efficiency in catalyzing the transesterification process without the need for excess alcohol. These optimized conditions not only enhance the conversion efficiency but also contribute to energy savings and cost reduction.

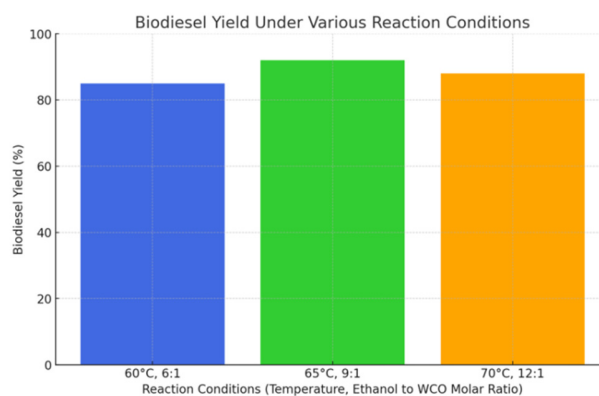


Fig. 2. Biodiesel yield under various transesterification conditions.

C. Catalyst Reusability and Environmental Impact

MgO demonstrated excellent reusability, with negligible loss in catalytic activity over multiple reaction cycles. This reusability factor significantly contributes to the process's economic viability and sustainability, reducing the need for frequent catalyst replacement with its associated costs. Additionally, the use of MgO minimizes environmental concerns typically associated with biodiesel production, such as soap formation and the release of hazardous chemicals. The lower energy requirements for the process further underscore its environmental benefits, making MgO an appealing catalyst for green biodiesel production (Figure 3). To further elucidate the sustainable aspect of utilizing WCO and MgO for biodiesel production, we have extended our analysis to include a comparative environmental impact assessment. This encompasses a preliminary Life Cycle Analysis (LCA) to evaluate the environmental footprints of biodiesel production processes catalyzed by MgO in comparison to those catalyzed by KOH. LCA focuses on key environmental indicators such as greenhouse gas emissions, energy use, and potential environmental toxicities associated with each catalyst's production, use, and disposal. Preliminary findings suggest that the MgO-catalyzed process is more environmentally friendly, primarily due to the reduced energy requirements and the avoidance of hazardous waste production. These findings highlight the environmental advantages of MgO as a catalyst, reinforcing its potential for sustainable biodiesel production. After rigorous testing, we observed that MgO maintains its catalytic efficiency over multiple cycles without significant loss in performance, which is critical for its long-term stability and

economic viability in the biodiesel production process. Unlike KOH, which is not reusable, MgO can be regenerated and reused over several cycles, demonstrating negligible degradation in its catalytic activity.

This reusability is a testament to MgO's robustness and contributes to lowering the overall operational costs of biodiesel production. In our study, MgO showcased consistent catalytic activity for up to 5 cycles of reuse, with only a minimal decrease in biodiesel yield, thereby significantly outperforming KOH in terms of sustainability and operational efficiency. This finding is critical for industrial applications where long-term stability and reusability are paramount for cost-effective biodiesel production.

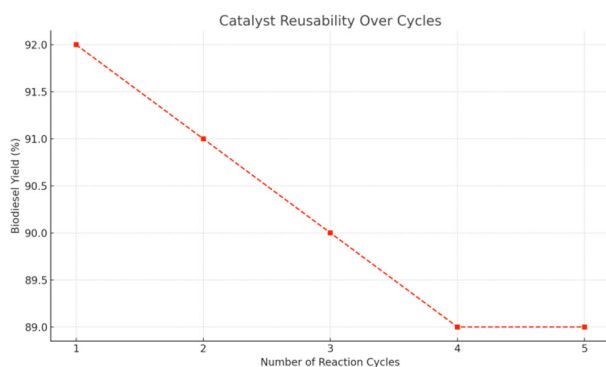


Fig. 3. Efficiency of MgO catalyst in biodiesel production over multiple cycles.

D. Comparison with KOH Catalysis

The comparative analysis of biodiesel properties produced using MgO and KOH catalysts, as presented in Table I, provides several key insights into the effectiveness and efficiency of these catalysts relative to ASTM and EU standards for biodiesel quality. The biodiesel produced using MgO as a catalyst exhibits a slightly higher cetane number than that produced with KOH. This indicates, potentially, a better combustion quality and engine performance. This is significant as it surpasses the minimum requirements set by both ASTM and EU standards, suggesting that MgO-catalyzed biodiesel could offer enhanced engine compatibility and lower emissions. Density and viscosity values for the biodiesel produced with either catalysts fall within the ASTM and EU specified ranges, ensuring good fuel atomization and pumpability at various temperatures. However, the slightly lower viscosity and density of the MgO-catalyzed biodiesel might favor better cold weather performance and fuel efficiency. The flash, cloud, and pour points are crucial for operational safety and performance in varying climatic conditions.

The biodiesel produced using MgO shows marginally better low-temperature properties, which could reduce the risk of gelation in colder environments. The acid number as well as the water and sediment content are within acceptable limits for both catalysts, indicating high-quality fuel that would minimize engine corrosion and wear. Nevertheless, MgO-catalyzed biodiesel has a lower acid number, which could further reduce

the risk of corrosion and extend engine life. The sulfur content of either biodiesel samples is well below the set thresholds, highlighting the environmental benefits of using biodiesel in reducing sulfur oxide emissions. Finally, the heating value of biodiesel produced with MgO is slightly higher than that of KOH-catalyzed one, which may translate into a marginal improvement in energy content and fuel economy. To conclude, the biodiesel produced using MgO as a catalyst meets the quality standards set by ASTM and EU, compared to KOH-catalyzed biodiesel. This suggests that MgO could be a more effective and environmentally friendly catalyst for biodiesel production, offering potential advantages in terms of fuel quality, engine performance, and environmental footprint. Further research and optimization of the MgO-catalyzed process could enhance these benefits, making it a viable alternative to traditional biodiesel production methods.

TABLE I. COMPARATIVE ANALYSIS OF BIODIESEL PROPERTIES PRODUCED USING MgO AND KOH CATALYSTS

Property	Biodiesel			
	MgO results	KOH results [6]	ASTM standard	EU standard
Cetane number	54	52	48-65	>51
Density (kg/m ³)	880	884	860-900	860-900
Viscosity (mm ² /s)	4.1	5.7	1.9-6.0	3.5-5.0
Flash point (°C)	62	60	>52	>55
Cloud point (°C)	-3	7	Variable	Variable
Pour point (°C)	-6	1	Variable	Variable
Acid number (mg KOH/g)	0.2	0.8	<0.5	<0.8
Water and sediment content (%)	0.01	-	<0.05	<0.02
Sulfur content (%)	0.001	0.0015	<0.002	<0.001
Heating value (MJ/kg)	37	36.5	Variable	Variable

The comparison between MgO and KOH catalyzed processes highlights the advantages of using MgO (Table II). The yield and purity of biodiesel produced with MgO were comparable or superior to those obtained with KOH. They were achieved under milder reaction conditions and with lower catalyst concentrations. The operational stability of MgO, coupled with its reusability and lower environmental impact, not only presents significant improvements over traditional KOH catalysis but also underlines MgO's economic viability. To address the economic implications of utilizing MgO as a catalyst, we have conducted a detailed analysis, focusing on the cost of catalyst preparation, operational expenses, and the scalability of the process for industrial applications.

Our analysis reveals that MgO offers a cost-effective alternative to KOH due to its lower consumption rate and higher activity, leading to a reduction in the overall amount of catalyst required. This, combined with MgO's durability and reusability, contributes to minimal operational costs over time. Specifically, the cost-effectiveness of MgO is supported by its efficient performance at lower concentrations and temperatures, which translates into reduced energy requirements and process costs.

Furthermore, the potential for scaling up the biodiesel production process using MgO is promising, given its availability and the feasibility of large-scale synthesis. These

factors collectively position MgO as a highly efficient and sustainable catalyst for biodiesel production from waste cooking oil, offering not only environmental but also economic benefits that are critical for the commercial viability of biodiesel production.

TABLE II. COMPARATIVE ANALYSIS OF BIODIESEL PRODUCTION USING MGO AND KOH CATALYSTS

Process Results	Catalyst	
	MgO	KOH [6]
Optimal concentration (wt%)	3	5
Optimal temperature (°C)	65	70
Optimal molar ratio (ethanol:oil)	9:1	12:1
Biodiesel yield (%)	92	88
Number of reusable cycles	5	Not reusable

IV. DISCUSSION

The exploration of magnesium oxide (MgO) as a catalyst in biodiesel production from WCO contributes to a sustainable and cost-effective biofuel production. MgO's non-toxic nature and its ability to catalyze reactions under milder conditions, reduce energy demands and carbon footprints, enhancing the environmental sustainability of the biodiesel production process.

MgO contributes positively to environmental sustainability by avoiding the production of harmful by-products during the transesterification process. Its use supports green chemistry principles, reducing waste and circumventing the production of toxic substances. The operational stability and effective catalytic action of MgO at lower temperatures lead to significant reductions in energy requirements. This, coupled with MgO's high stability and reusability, decreases the carbon footprint of biodiesel manufacturing. MgO's excellent reusability, with negligible loss in activity over multiple cycles, brings out its economic and environmental benefits. Synthesized from the abundant magnesium carbonate (MgCO_3), MgO requires less energy for production, resulting in lower CO_2 emissions.

The LCA comparing MgO and KOH catalysis methods, highlights MgO's lower environmental impact, particularly in terms of energy use and production of hazardous waste. MgO's reusability enhances the economic viability of the biodiesel production process not only by reducing the need for frequent catalyst replacement but also by leveraging on magnesium's abundance. This contributes to MgO's cost-efficiency, making it an attractive option for large-scale biodiesel production. The solid nature of MgO makes separation from the product easy, reducing the need for complex purification processes, further minimizing operational costs.

MgO's catalytic efficiency allows for the transesterification process to occur at lower temperatures with reduced catalyst concentrations, offering both operational and energy savings. The scalability of using MgO from WCO presents a promising candidate for industrial applications, making the MgO-catalyzed process an attractive option for meeting the growing demand for renewable energy sources.

Summarizing, MgO significantly advances sustainable biodiesel production, offering a multifaceted approach to environmental sustainability, economic viability, and operational efficiency. Its potential is pivotal for the transition towards more sustainable industrial practices. Therefore, the continued research and development in this area is regarded as imperative. Future research should focus on refining MgO's preparation methods to enhance its surface area and catalytic activity. Detailed lifecycle analyses are suggested to assess the sustainability of biodiesel production using MgO comprehensively.

V. CONCLUSION

In conclusion, MgO emerges as a highly advantageous catalyst for the production of biodiesel from Waste Cooking Oil (WCO), having a plethora of environmental, economic, and operational advantages. The future of biodiesel production using WCO, appears promising with a, potentially, significant contribution to the renewable energy landscape. Continued research and development efforts are essential in order to optimize the use of MgO and explore innovative ways that further enhance the sustainability and efficiency of biodiesel production.

ACKNOWLEDGEMENT

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number NBU-FFR-2024-2505-02.

REFERENCES

- [1] A. A. Khaskheli, G. D. Walasai, A. S. Jamali, Q. B. Jamali, Z. A. Siyal, and A. Mengal, "Performance Evaluation of Locally-Produced Waste Cooking Oil Biodiesel with Conventional Diesel Fuel," *Engineering, Technology & Applied Science Research*, vol. 8, no. 6, pp. 3521–3524, Dec. 2018, <https://doi.org/10.48084/etasr.2333>.
- [2] A. H. Ulukardesler, "Biodiesel Production from Waste Cooking Oil Using Different Types of Catalysts," *Processes*, vol. 11, no. 7, Jul. 2023, Art. no. 2035, <https://doi.org/10.3390/pr11072035>.
- [3] D. H. Park, F. I. Nana, and H. M. Cho, "A Review of the Emission, Performance, Combustion, and Optimization Parameters in the Production of Biodiesel from Waste Cooking Oil," *Automotive Experiences*, vol. 5, no. 3, pp. 371–388, Jun. 2022, <https://doi.org/10.31603/ae.7005>.
- [4] Monika, S. Banga, and V. V. Pathak, "Biodiesel production from waste cooking oil: A comprehensive review on the application of heterogeneous catalysts," *Energy Nexus*, vol. 10, Jun. 2023, Art. no. 100209, <https://doi.org/10.1016/j.nexus.2023.100209>.
- [5] M. M. Tunio, M. R. Luhur, Z. M. Ali, and U. Daher, "Performance and Emission Analysis of a Diesel Engine Using Linseed Biodiesel Blends," *Engineering, Technology & Applied Science Research*, vol. 8, no. 3, pp. 2958–2962, Jun. 2018, <https://doi.org/10.48084/etasr.2028>.
- [6] A. Eladeb, A. Aydi, and I. Alenezi, "Ethanolsis of Waste Cooking Oils Using KOH Catalyst," *Oriental Journal of Chemistry*, vol. 37, no. 6, pp. 1344–1349, Dec. 2021, <http://doi.org/10.13005/ojc/370611>.
- [7] W. Widayat *et al.*, "Preparation of MgO-CaO/SiO₂ catalyst from dolomite and geothermal solid waste for biodiesel production," *International Journal of Renewable Energy Development*, vol. 12, no. 3, pp. 541–549, May 2023, <https://doi.org/10.14710/ijred.2023.51573>.
- [8] K. A. V. Miyuranga, U. S. P. R. Arachchige, T. M. M. Marso, and G. Samarakoon, "Biodiesel Production through the Transesterification of Waste Cooking Oil over Typical Heterogeneous Base or Acid Catalysts," *Catalysts*, vol. 13, no. 3, Mar. 2023, Art. no. 546, <https://doi.org/10.3390/catal13030546>.

- [9] A. Ahmed, A. Ali, M. Mubashir, H. R. Lim, K. S. Khoo, and P. L. Show, "Process optimization and simulation of biodiesel synthesis from waste cooking oil through supercritical transesterification reaction without catalyst," *Journal of Physics: Energy*, vol. 5, no. 2, Feb. 2023, Art. no. 024003, <https://doi.org/10.1088/2515-7655/acb6b3>.
- [10] M. Hu, J. Pu, E. W. Qian, and H. Wang, "Biodiesel Production Using MgO–CaO Catalysts via Transesterification of Soybean Oil: Effect of MgO Addition and Insights of Catalyst Deactivation," *BioEnergy Research*, vol. 16, no. 4, pp. 2398–2410, Dec. 2023, <https://doi.org/10.1007/s12155-023-10580-z>.
- [11] F. Guo, L. Wang, Y. Cao, P. He, and H. Li, "Efficient synthesis of ethylene carbonate via transesterification of ethylene glycol with dimethyl carbonate over Mg₃Al₁–xCe_xO composite oxide," *Applied Catalysis A: General*, vol. 662, Jul. 2023, Art. no. 119273, <https://doi.org/10.1016/j.apcata.2023.119273>.
- [12] P. A. Gangotena, S. Ponce, Á. Gallo-Córdova, D. A. Streitwieser, and J. R. Mora, "Highly Active MgP Catalyst for Biodiesel Production and Polyethylene Terephthalate Depolymerization," *ChemistrySelect*, vol. 7, no. 15, 2022, Art. no. e202103765, <https://doi.org/10.1002/slct.202103765>.
- [13] A. Abdelhafiz *et al.*, "Pulsed Light Synthesis of High Entropy Nanocatalysts with Enhanced Catalytic Activity and Prolonged Stability for Oxygen Evolution Reaction," *Advanced Science*, vol. 10, no. 18, 2023, Art. no. 2300426, <https://doi.org/10.1002/adv.202300426>.
- [14] J. Liu *et al.*, "Excess soluble alkalis to prepare highly efficient MgO with relative low surface oxygen content applied in DMC synthesis," *Scientific Reports*, vol. 11, no. 1, Oct. 2021, Art. no. 20931, <https://doi.org/10.1038/s41598-021-00323-5>.
- [15] S. Xia *et al.*, "Sustainable biodiesel production via transesterification of vegetable oils and waste frying oil over reusable magnetic Ca₂Fe₂O₅/CaO@MgFe₂O₄-Fe₂O₃ catalyst," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 45, no. 3, pp. 8047–8061, Aug. 2023, <https://doi.org/10.1080/15567036.2023.2225448>.
- [16] Y. Xie *et al.*, "CaO-MgFe₂O₄@K₂CO₃ as a novel and retrievable nanocatalyst for two-step transesterification of used frying oils to biodiesel," *Process Safety and Environmental Protection*, vol. 172, pp. 195–210, Apr. 2023, <https://doi.org/10.1016/j.psep.2023.02.005>.
- [17] U. N. Khan *et al.*, "Green synthesis of magnesium oxide nanosheets by using Citrullus colocynthis fruit extract and its use in biofuel production," *Biomass and Bioenergy*, vol. 167, Art. no. 106640, Dec. 2022, <https://doi.org/10.1016/j.biombioe.2022.106640>.
- [18] N. Balaba *et al.*, "Polysaccharides as Green Fuels for the Synthesis of MgO: Characterization and Evaluation of Antimicrobial Activities," *Molecules*, vol. 28, no. 1, Jan. 2023, Art. no. 142, <https://doi.org/10.3390/molecules28010142>.
- [19] M. Ramezani Farani, M. Farsadrooh, I. Zare, A. Gholami, and O. Akhavan, "Green Synthesis of Magnesium Oxide Nanoparticles and Nanocomposites for Photocatalytic Antimicrobial, Antibiofilm and Antifungal Applications," *Catalysts*, vol. 13, no. 4, Apr. 2023, Art. no. 642, <https://doi.org/10.3390/catal13040642>.
- [20] I. Y. Dharmegowda, L. M. Muniyappa, P. Siddalingaiah, A. B. Suresh, M. P. Gowdru Chandrashekarappa, and C. Prakash, "MgO Nano-Catalyzed Biodiesel Production from Waste Coconut Oil and Fish Oil Using Response Surface Methodology and Grasshopper Optimization," *Sustainability*, vol. 14, no. 18, Jan. 2022, Art. no. 11132, <https://doi.org/10.3390/su141811132>.
- [21] M. Cerón Ferrusca, R. Romero, S. L. Martínez, A. Ramírez-Serrano, and R. Natividad, "Biodiesel Production from Waste Cooking Oil: A Perspective on Catalytic Processes," *Processes*, vol. 11, no. 7, Jul. 2023, Art. no. 1952, <https://doi.org/10.3390/pr11071952>.
- [22] N. C. Joshi, P. Gururani, P. Bhatnagar, V. Kumar, and M. S. Vlaskin, "Advances in Metal Oxide-based Nanocatalysts for Biodiesel Production: A Review," *ChemBioEng Reviews*, vol. 10, no. 3, pp. 258–271, 2023, <https://doi.org/10.1002/cben.202200019>.
- [23] A. A. Khaskheli, H. J. Arain, I. A. Memon, U. A. Rajput, and M. J. Ahsan, "Emission and Noise Characteristics of a Diesel Engine Fuelled with Diesel-Chicken Oil Biodiesel Blends," *Engineering, Technology &*

Applied Science Research, vol. 10, no. 2, pp. 5387–5391, Apr. 2020, <https://doi.org/10.48084/etasr.3348>.

Deep Learning, Ensemble and Supervised Machine Learning for Arabic Speech Emotion Recognition

Wahiba Ismaiel

Department of Science and Technology, University College of Ranyah, Taif University, Saudi Arabia
w.wahiba@tu.edu.sa (corresponding author)

Abdalilah Alhalangy

Department of Computer Engineering, College of Computer, Qassim University, Saudi Arabia
a.alhalangy@qu.edu.sa

Adil. O. Y. Mohamed

Department of Computer Science, College of Computer, Qassim University, Saudi Arabia
adi.mohamed@qu.edu.sa

Abdalla Ibrahim Abdalla Musa

Department of Computer Science, College of Computer, Qassim University, Saudi Arabia
ab.musa@qu.edu.sa

Received: 24 February 2024 | Revised: 10 March 2024 | Accepted: 12 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7134>

ABSTRACT

Today, automatic emotion recognition in speech is one of the most important areas of research in signal processing. Identifying emotional content in Arabic speech is regarded as a very challenging and intricate task due to several obstacles, such as the wide range of cultures and dialects, the influence of cultural factors on emotional expression, and the scarcity of available datasets. This study used a variety of artificial intelligence models, including Xgboost, Adaboost, KNN, DT, and SOM, and a deep-learning model named SERDNN. ANAD was employed as a training dataset, which contains three emotions, "angry", "happy", and "surprised", with 844 features. This study aimed to present a more efficient and accurate technique for recognizing emotions in Arabic speech. Precision, accuracy, recall, and F1-score metrics were utilized to evaluate the effectiveness of the proposed techniques. The results showed that the Xgboost, SOM, and KNN classifiers achieved superior performance in recognizing emotions in Arabic speech. The SERDNN deep learning model outperformed the other techniques, achieving the highest accuracy of 97.40% with a loss rate of 0.1457. Therefore, it can be relied upon and deployed to recognize emotions in Arabic speech.

Keywords-Arabic speech emotion recognition; ANAD; SERDNN; SOM; Xgboost; Adaboost; DT; KNN

I. INTRODUCTION

Speech is a common and natural mode of human communication and is considered the simplest, quickest, and most natural type of human connection but does not apply to machines [1-3]. Although human-machine-speech connections are developing, machines cannot grasp human emotions and are incapable of engaging in a genuine and natural debate [4]. Emotion recognition is the first step in improving human-computer interaction with speech-based computing systems. Recognizing emotions in Arabic speech is considered challenging due to several issues, such as cultural disparities, the wide range of Arabic dialects, and how they are circulated

among Arab people. Accurate identification of emotions in speech remains challenging due to the absence of a specific definition of emotion and the extensive and complicated influence of creating and expressing emotions. Machine Learning (ML) is a breakthrough technology in which significant amounts of data are provided to computer systems [5]. This involves making computers identify human speech and capable of detecting emotions. Computers use data to enhance their algorithm and better process other data in the future [2]. Researchers are looking for techniques to extract the emotional state of a speaker from their words and to successfully understand human emotions [6]. Emotions may be

represented in speech and then exploited to extract useful connotations from spoken words, therefore improving speech recognition implementations [4, 7].

Today, artificial emotional intelligence, often known as emotional AI, is a popular topic of study. People communicate their emotions through nonverbal indications, such as external facial expressions, gestures, and nonverbal communication, using body language and tone of voice. Emotional AI seeks to identify emotions in the same way that humans do, through many channels. Emotional AI is frequently used in intelligent security, intelligent contact centers, customer assistance, medical applications, forensics, banking, and stress and anxiety treatment [8]. In [9], facial expression-based surveillance was employed to detect neurodegenerative diseases. Utilizing the LSTM model, emotions are identified from both speech and text [10]. In [11], a smart scarf was presented to help people who struggle to communicate their feelings. In [12], Speech Emotion Recognition (SER) was compared in Arabic and English, suggesting that the introduction of particular acoustic elements ameliorates emotion recognition for Arabic words.

Automatic emotion identification from voice data without language clues is an important and developing study field. In [13-14], emotion identification in spoken Arabic data was investigated using five ensemble models on a voice emotion identification system. The SMO achieved the highest accuracy (95.52%) among single classifiers in recognizing the "happy", "angry", and "surprised" emotions in Arabic real speech. Due to the lack of Arabic speech emotion datasets, few studies have built Arabic SER systems. In [15], emotion identification in Egyptian Arabic speech was investigated deploying prosodic, spectral, and wavelet data, implementing an Egyptian television series to generate a semi-natural Egyptian Arabic speech emotion dataset. Identifying and classifying the emotional content of a speech signal is a key problem for researchers. In [3], the KS Arabic spoken emotion corpus was engaged along with classification approaches, such as K-Nearest Neighbor (KNN) and Support Vector Machine (SVM). Convolutional Neural Networks (CNNs) can emerge as the prevailing method within this domain. In [16], two neural architectures were applied to identify emotions in Arabic voice, and the results showed that the proposed strategy may lead to considerable improvement (2.2%) over a powerful deep CNN baseline system, due to the use of sophisticated layers of deep learning paradigms, namely Hubert, wav2vec2.0, wav2vecU, and WavBERT, which allow better representation learning and greater information capture. In [17], a deep learning model of emotional recognition was introduced for spoken Arabic dialogue.

In the field of emotion detection, when a person becomes emotional, their voice alters to reflect their emotional state. In [18], a deep learning-based emotional recognition model was presented for spoken Arabic dialogues. In [19], a multi-head attention multi-learning model (ABMD) was investigated, using Dilated Convolution (DC) and Residual Dilated Causal Convolution (RDCC) blocks for emotion recognition from speech patterns. Based on three datasets, EMOB, RAVDESS, and SAVEE, the ABMD model achieved accuracy rates of 95.93, 85.89, and 93.75%, respectively. In [20], a feature

vector, which consisted of 28 measures, was submitted for emotion recognition. This study adopted a Moroccan Arabic Dialect Emotional Database (MADED) with a KNN classifier to detect four emotions: "joy", "anger", "neutral", and "sadness". This KNN obtained an accuracy rate of 64.65% for all emotions. In [21], a recognition and classification model, named TLBOML-ERC, was suggested to detect emotional expressions and sentiments expressed in Arabic tweets. This study employed a denoising autoencoder to classify emotional expressions in an Arabic tweet dataset into four emotion classes, "anger", "sadness", "joy", and "fear". Three different ratios were put into service for the training and testing sets, 90% train and 10% test, 80% train and 20% test, and 70% train and 30% test set, and TLBOML-ERC achieved 97.86, 98.84, and 98.24% accuracy, respectively. In [22], a CNN model was proposed for Arabic speech emotion recognition based on the BASE-DB training dataset with four emotions (anger, neutral, sadness, and joy) and five features (Mel-scaled spectrogram, MFCC, tonal centroid, chromatogram, and spectral contrast), achieving an accuracy rate of 98.18%.

In [23], a 1DCNN was integrated with LSTM and an attention model to recognize emotions in speech. This model utilized RMSE, MFC, and ZCR features and was evaluated on four datasets: BAVED, ANAD, SAVEE, and EMO-DB. The accuracy of the proposed model varied across datasets, with results of 88.39, 96.72, 97.13, and 96.72%, accordingly. In comparison, the accuracy rates obtained by the 2DCNN model were 83.39, 90.16%, 51.30, and 50%, correspondingly. In [24], GWO-KNN, which employed the Grey Wolf Optimizer (GWO) to extract SER features in both Arabic and English languages, was presented. This study adopted the Emirati dialect Arabic speech database, the RAVDESS, and the SAVEE datasets. In [25], two sets of features were generated by combining spectral and prosodic characteristics from the Egyptian Arabic semi-natural emotion speech dataset, which consists of 579 utterances. Five machine learning classifiers, namely Random Forest (RF), Multilayer Perceptron (MLP), Ensemble learning, Logistic Regression (LR), and Support Vector Machine (SVM), were implemented to predict four emotions: "happy", "sad", "angry", and "neutral". The results showed that ensemble learning obtained the highest rate of 87.59%, whereas it achieved 64% for predicting all multi-emotions. In [26], SRE was deployed for the Algerian dialect, based on an Algerian speech emotion dataset with four emotions: "happy", "sad", "neutral", and "angry". This study applied various machine learning methods, and LSTM-CNN had the highest accuracy rate of 93.34%. In [27], an Arabic SER system was built based on supervised machine learning, using ANAD as a training set with three emotions: "angry", "happy", and "surprised". The proposed model was named C-SVM and its highest score was achieved for "angry" with 98.1, 98.7, and 98.3% precision, recall, and F1-score, respectively. In [28], SER was performed on Arabic speech utilizing the ANAD dataset with three emotions and implementing 35 classifiers.

Many problems remain unsolved for effective AI systems, including poor accuracy rates, high computing complexity of hybrid classifier models, and a paucity of natural datasets. Multiple studies have suggested different ways to recognize

vocal emotions. This study aims to fill the existing gap employing the Arabic Natural Audio Dataset (ANAD) for SER, relying on four supervised machine learning classifiers: extreme boosting (Xgboost), Adaptive boosting (Adaboost), KNN, and DT. The purpose of this research is to assist apps and institutions that rely on speech to make correct and ideal judgments, which are free from prejudice. Table I depicts the contributions of related works and their suggestions for identifying emotions in Arabic speech. This study also presents and evaluates a simple deep-learning model to precisely recognize and classify emotions from Arabic speech. The following is a list of the contributions of this study:

- Present the most recent advances in Arabic SER.
- Propose a simple and effective deep learning model.
- Implement five machine learning techniques.
- Examine the accuracy, recall, and F1-score of the models.
- Evaluate the accuracy of the proposed models against the results of previous studies.

TABLE I. CONTRIBUTION OF RELATED WORKS

Reference	Year	Models	Datasets	Language	Emotion types
[20]	2021	KNN	MADED	Arabic/Moroccan	KNN
[21]	2023	TLBOML-ERC	Arabic tweets	Arabic	Joy, angry, fear, sadness
[22]	2022	CNN	BASE-DB	Arabic	Joy, angry, neutral, sadness
[23]	2023	IDCNN integrated with LSTM and attention, 2DCNN	BAVED, ANAD, SAVEE, EMO-DB	Arabic	high, low, neutral, sad, fear, bored, disgust, happy, angry, surprised
[24]	2023	GWO-KNN	SAVEE, RAVDESS, Emirati Dialect Arabic Speech Database	Arabic/English	Surprised, neutral, angry, happy, fear, disgust, sad, calm
[25]	2023	MLP, Ensemble learning, RF, MLP&SVM	Egyptian Arabic Semi-Natural Emotion speech	Arabic/Egyptian	happy, sad, angry, neutral
[26]	2021	LSTM, BLSTM, CNN	Algerian speech emotion	Arabic/Algerian	happy, sad, neutral, angry
[27]	2022	C-SVM	ANAD	Arabic	angry, happy, surprised
[28]	2018	SOM	ANAD	Arabic	angry, happy, surprised

II. METHODOLOGY

Four supervised machine learning methods (Xgboost, Adaboost, KNN, and DT) and one unsupervised model (SOM) were used to detect and recognize emotions in Arabic speech. Figure 1 portrays the methodology of this study.

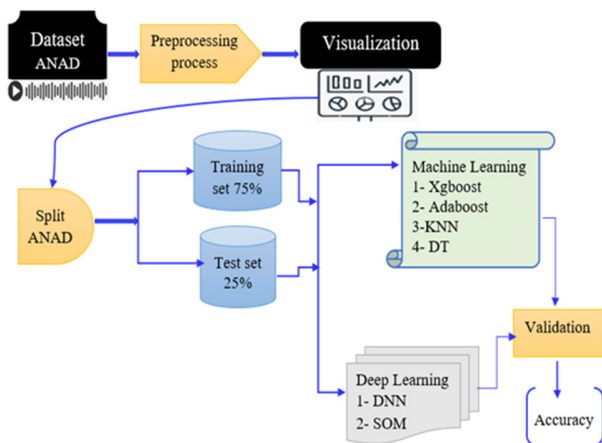


Fig. 1. Methodology.

A. Data Collection and Description

The ANAD and Arabic Speech Emotion Recognition (ASER) datasets were used [29]. The ANAD dataset consists of 846 attributes. The first attribute contains 1383 speech files, the second one includes the target field of emotion types, and the rest 844 attributes are speech features. This dataset provides 25 ANAD features that are listed in Table II. Eighteen (18) listeners participated in reporting their feelings, listening to

eight different videos and choosing from surprised, happy, and angry. The dataset contained 137 surprised, 505 happy, and 741 angry rows, and Figure 2 displays the emotion categories.

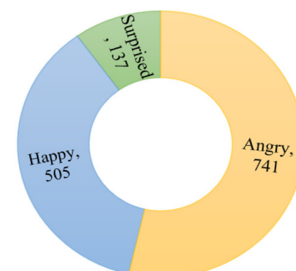


Fig. 2. Emotion categories.

B. Data Preprocessing and Visualization

This process is considered one of the most basic steps in dealing with machine and deep learning algorithms to achieve prediction accuracy based on data free from bias and noisy values. After examination, it was determined that ANAD has no null values but its data consist of different formats: float, int, and categorical (object type). The categorical format refers to two attributes: speech name and label. The label attribute represents three emotions: "angry", "happy", and "surprised". This format is not compatible with the model, so it was converted into numerical values of 0 or 1, split into three columns named 0, 1, and 2, corresponding to the emotions "angry", "happy", and "surprised", respectively. The audio files were processed to exclude any noise, quiet, and laughter. All audio files had a length of 1 second and a signal rate of 22050. Figures 3-5 exhibit the waveform visualization of each emotion's sampling, highlighting the distinction between the three emotions. All 844 characteristics were extracted from the

dataset and divided into a training set, including 75% of the data, and a test set, entailing the remaining portion. The training set consisted of 1037 records, while the test set consisted of 346 records. Following the process of data splitting, the subsequent step involved adjusting the feature values to fall within a specific range using the normalization method provided by the scikit-learn module in Python. The normalization process depends on (1), which calculates the new value (V) based on the minimum feature value (f_{min}), the maximum feature value (f_{max}), and the current value of the feature (v):

$$V = \frac{v - f_{min}}{f_{max} - f_{min}} \quad (1)$$

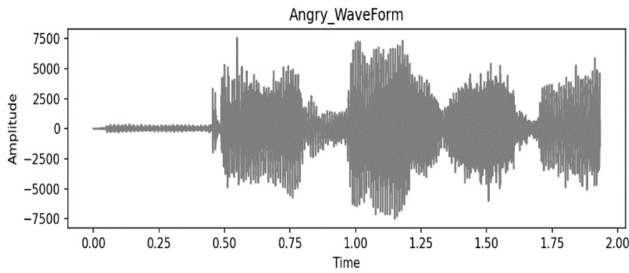


Fig. 3. Angry sampling waveform.

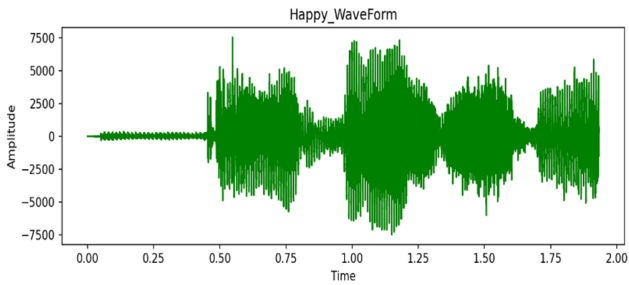


Fig. 4. Happy sampling waveform.

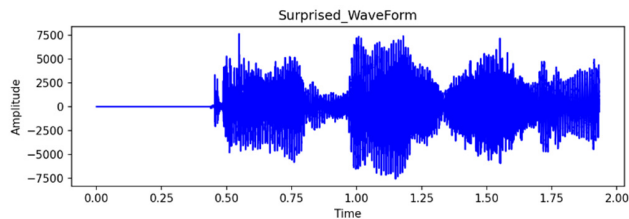


Fig. 5. Surprised sampling waveform.

C. Proposed Speech Emotion Recognition Deep Neural Network (SERDNN)

SERDNN was built predicated on the Keras module in Python language and dense class. SERDNN had 844 features as the input stage and three hidden layers that implemented a ReLU activation function. The last layer had three neurons to output the emotion with a sigmoid function. The model was compiled using the Adam optimizer function and categorical cross entropy for the loss function. Figure 6 illustrates the configuration of the SERDNN model. Table III demonstrates the configuration details of SERDNN and the activation size

calculation. The train parameters for each layer were estimated according to the previous layer neuron P_n and the current layer neuron C_n (2). The SERDNN model was trained for 200 epochs with a fixed batch size of 64, to obtain the prediction of speech emotion results.

$$No\ of\ Parameters = P_n * C_n + 1 * C_n \quad (2)$$

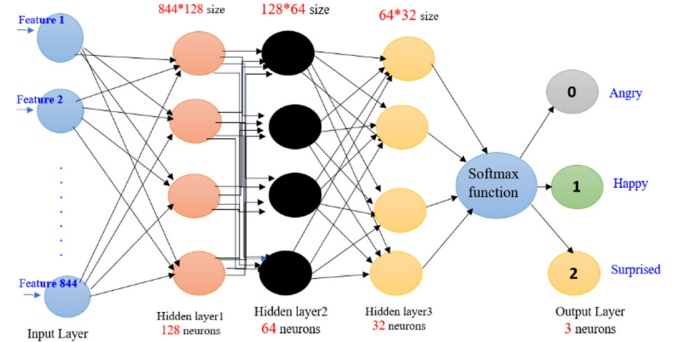


Fig. 6. SERDNN model configuration.

TABLE II. CONFIGURATION DETAILS OF SERDNN

Layer	output shape	Activation size	#Parameters
dense	(none, 128)	$844 \times 128 = 108032$	$844 \times 128 + 128 = 108160$
dense_1	(none, 128)	$128 \times 128 = 16384$	$128 \times 128 + 128 = 16512$
dense_2	(none, 64)	$128 \times 64 = 8192$	$128 \times 64 + 64 = 8256$
dense_3	(none, 32)	$64 \times 32 = 2048$	$64 \times 32 + 32 = 2080$
dense_4	(none, 3)	$32 \times 3 = 96$	$32 \times 3 + 3 = 99$

D. Proposed Artificial Neural Network (SOM)

Self-Organizing Map (SOM) is an unsupervised learning technique that accepts higher dimensional training data grouped according to their similarity to decrease their dimensions. It contains two layers, one for the input and a second for the output, which entail 2D neurons [30]. SOM is trained on the basis of the following steps:

- Initialize neurons by assigning a small random value to weight 00 and initialize the learning rate a
- Find the Euclidean distance value using:

$$distance(j) = \sum (weight_{ij} - x_i)^2 \quad (3)$$

where $j = 1 \dots m$ and $i = 1 \dots n$

- Assign the winning index with the smallest value of distance weight ($D(W_j)$)
- Determine the neighborhood of j for all i utilizing:

$$weight_{ij}(current\ value) = weight_{ij}(old\ -\ value) + a[x_i - weight_{ij}(old\ -\ value)] \quad (4)$$

- Update the learning rate at time t employing:

$$a(t+1) = 0.5 \times t \quad (5)$$

- Stop training at the iteration end.

This study deployed the Minisom library in Python to build an SOM of 12×12 neurons, with 844 features as input values.

The SOM model implemented the Gaussian function to calculate the distance between the neighborhood, the learning rate was 0.5 and the sigma value was assigned to 1.25. The sigma is the radius of a different neighborhood in the neural network. This model was trained in 3000 iterations. Figure 7 presents the proposed SOM architecture.

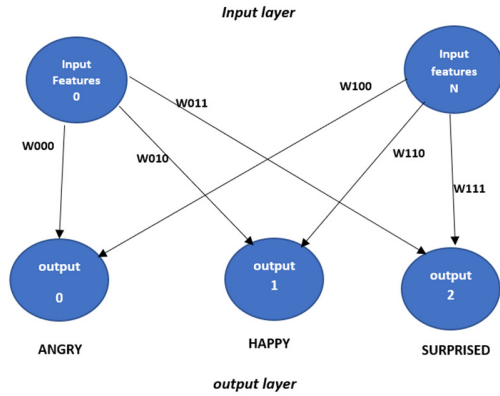


Fig. 7. SOM architecture.

E. Proposed Supervised Machine Learning Classifiers

Four supervised classifiers Xgboost, Adaboost, KNN, and DT were also utilized to classify and detect speech emotions. Xgboost and Adaboost are ensemble classifiers.

1) K-Nearest Neighbor (KNN)

KNN can perform regression and classification processes and K represents the number of nearest neighbors. KNN focuses on finding the smallest distance between neighbors, so it predicts the target based on the Euclidean distance (6). Figure 8 depicts the KNN classifier. In this study, K was assigned to 3.

$$Distance = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (6)$$

where (x_1, y_1) and (x_2, y_2) are the coordinates of the two points. The goal is to find the smaller distance between the two points, supposing that these points are two features.

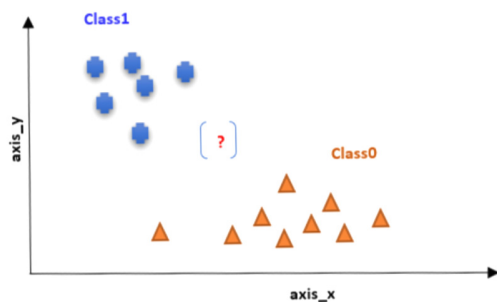


Fig. 8. KNN classifier.

2) Decision Tree Classifier (DT)

DT is commonly employed to solve classification and regression problems. The root of the tree is a point to the

particular dataset where the DT classifier starts the prediction process based on the compared results of the root value with the rows of a dataset, taking the branch direction for the next node.

3) Extreme Gradient Boosting (Xgboost)

Xgboost is an ensemble classifier that can train a large dataset of different domains [31].

4) Adaptive Boosting (Adaboost)

Adaboost can also be considered an ensemble learning technique, such as Xgboost, for classification and regression problems. This technique merges multiple weak classifiers to provide a strong one and depends on the weight used to solve the problem of incorrect prediction [32]. Equation (7) illustrates the Adaboost method, where N corresponds to the weight, and f_N explains the N weak classifiers.

$$F = \sin\left(\sum_{N=1}^N \theta_N f_N(X)\right) \quad (7)$$

III. EXPERIMENTAL RESULTS

All the proposed models were run on an MSI laptop with a Core™ i7-11th Intel® @ 3.00GHz CPU, and 32GB of RAM, using Python and Jupyter Notebook on Windows 10 Pro 64-bit. Three experiments were carried out to obtain the results of the proposed models on the ANAD dataset. The first experiment involved building the SERDNN deep learning model, the second was utilized for the supervised learning algorithms (Xgboost, Adaboost, KNN, and DT), and the final experiment employed the SOM model. Four metrics were applied: accuracy, precision, recall, and F1-score. Furthermore, a confusion matrix was also put into service to measure the performance of the proposed models in Arabic SER. The confusion matrix displays the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) results. TP represents the cases where the model predicts a positive result and an actual positive result. TN signifies cases where the model predicts a negative result for an actual negative result. FP denotes cases where the model forecasts a positive result but the actual result is negative. FN represents cases where the model predicts a negative result but the actual result is positive. The following equations demonstrate how these metrics are calculated.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

$$F1 - score = \frac{2 * (Precision) * Recall}{Precision + Recall} \quad (11)$$

Figure 9 exhibits the train and test accuracy results of the SERDNN model. Figure 10 demonstrates the train and test loss of the SERDNN model.

Figure 11 illustrates the train and test accuracies of Xgboost, Adaboost, KNN, DT, and SERDNN. Figure 12-14 provide the precision, F1-score, and recall scores of all models tested to detect the "angry", "happy", and "surprised" emotions. Figure 15 presents the confusion matrices.

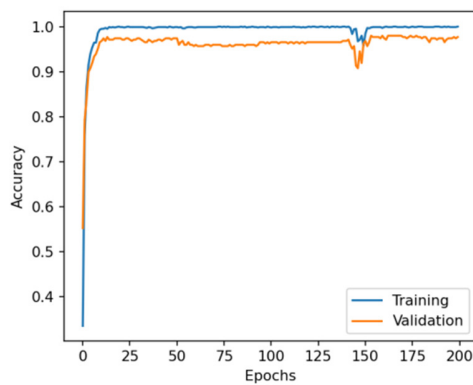


Fig. 9. Accuracy of the SERDNN model.

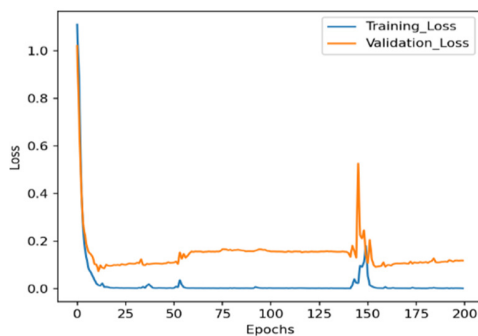


Fig. 10. Loss of the SERDNN model.

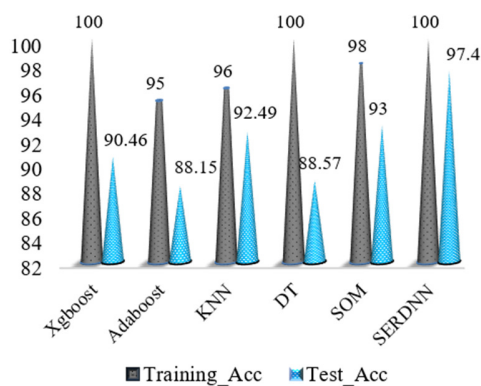


Fig. 11. Train and test accuracies of proposed models.

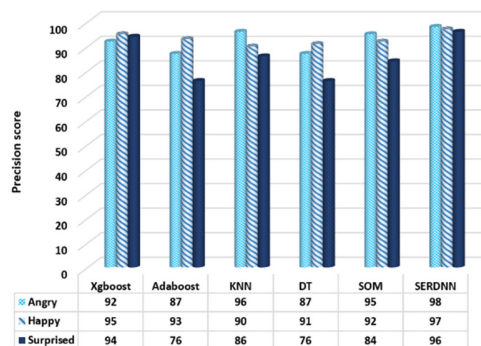


Fig. 12. Precision on surprised, happy, and angry emotions.

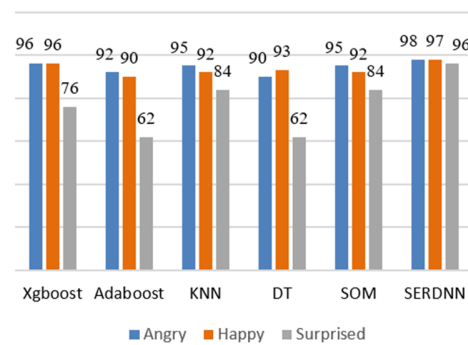


Fig. 13. F1-score of surprised, happy, and angry emotions.

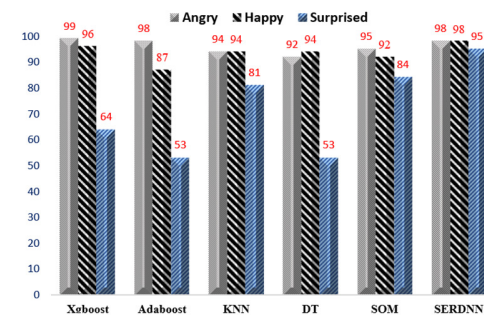


Fig. 14. Recall on surprised, happy, and angry emotions.

IV. DISCUSSION

The highest training accuracy result was obtained by SERDNN, DT, and Xgboost with a 100% rate, while the highest prediction accuracy was acquired by SERDNN with 97.40%, followed by SOM with 93%, and KNN with 92.49%, while Adaboost had the lowest accuracy at 88.15%. SERDNN achieved the highest precision for all emotions (98, 97, and 96% for "angry", "happy", and "surprised", respectively). The lowest precision score for "angry" was obtained by Adaboost and DT (87%), the lowest precision for "happy" was achieved by KNN (90%), and Adaboost had the same precision (76%) for the "surprised" emotion. SERDNN achieved the highest F1-scores (98, 97, and 95% for "angry", "happy", and "surprised", respectively). The lowest F1-score for "angry" was accomplished by DT (90%), the lowest F1-score for "happy" was acquired by Adaboost (90%), and DT and Adaboost both rated 62% for the "surprised" emotion. The highest recall score was obtained by Xgboost, reaching 99% for "angry", followed by Adaboost and SERDNN (both at 98%). SERDNN had the highest recall rates for "happy" and "surprised" (98 and 95%, respectively). DT had the lowest recall score for "angry" (92%), Adaboost had 87% recall for "happy", and Adaboost and DT obtained the lowest recall for "surprised" (53%). The results disclose that all models have a satisfactory ability to recognize these three emotions. SERDNN, Xgboost, and KNN were the best at detecting the "angry" emotion. For the "happy" emotion, SERDNN and Xgboost were the best models. The emotion prediction rates of "surprised" were low compared to other emotions. Table III displays a comparison of the results between the models in this study and previous studies that used the same dataset.

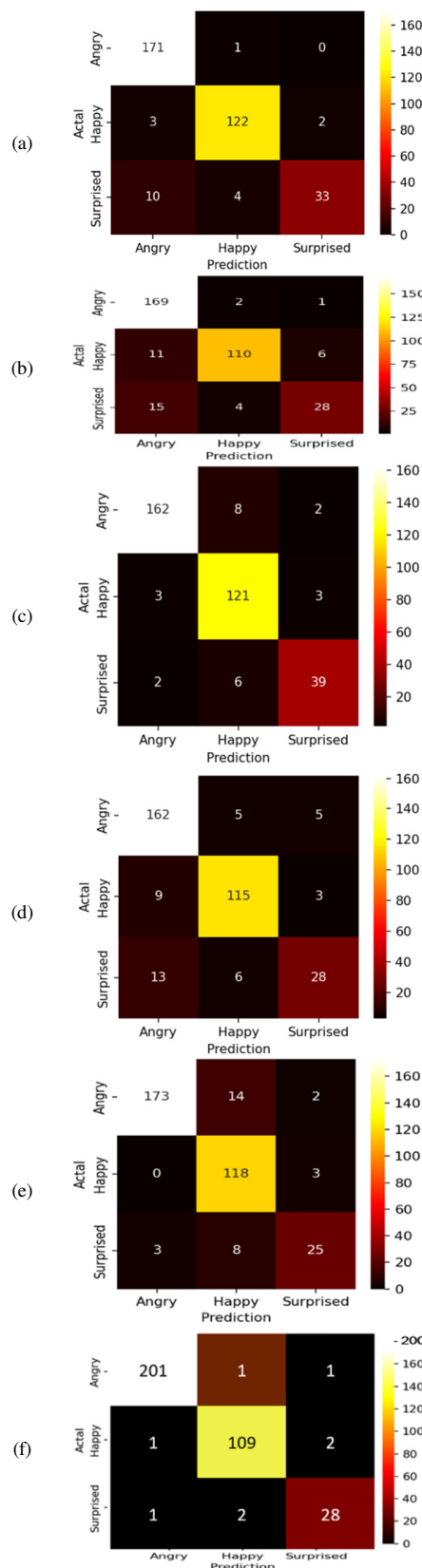


Fig. 15. Confusion matrices of (a) Xgboost, (b) Adaboost, (c) KNN, (d) DT, (e) SOM, (f) SERDNN.

TABLE III. PREVIOUS STUDIES WITH PROPOSED MODELS BASED ON ANAD DATASET

Reference	Models	Highest accuracy
[23]	1DCNN with LSTM & attention	96.72%,
	2DCNN	90.16%
[27]	C-SVM	Precision 98.1%
		Recall 98.7%
		F1-score 98.3%
[28]	SOM	95.52%
	SERDNN	97.40%
This study	SOM	95%
	DT	88.57%
	KNN	92.49%
	Xgboost	90.46%
	Adaboost	88.15%

V. CONCLUSION

This study developed and evaluated models that can recognize emotions in Arabic speech. A variety of artificial intelligence techniques, including Xgboost, Adaboost, KNN, DT, and SOM, were adopted. In addition, this study developed a simple deep-learning model, called SERDNN, to fill the gap in this domain. These models were trained and tested on the ANAD dataset to assess their ability to recognize three different emotions: "angry", "happy", and "surprised". The results manifested that all models were able to recognize the "angry" and "happy" emotions with high accuracy, while the precision, F1-score, and recall were high for the "surprised" emotion by SERDNN, SOM, Xgboost, and KNN compared to the DT and Adaboost models. The highest accuracy for all emotions was achieved by the proposed deep learning model SERDNN, at 97.40%. In the future, the SERDNN model will be improved and evaluated on other datasets, while other machine learning models will also be investigated to extract emotional features.

REFERENCES

- [1] L. Chen, W. Su, Y. Feng, M. Wu, J. She, and K. Hirota, "Two-layer fuzzy multiple random forest for speech emotion recognition in human-robot interaction," *Information Sciences*, vol. 509, pp. 150–163, Jan. 2020, <https://doi.org/10.1016/j.ins.2019.09.005>.
- [2] A. S. Alluhaidan, O. Saidani, R. Jahangir, M. A. Nauman, and O. S. Neffati, "Speech Emotion Recognition through Hybrid Features and Convolutional Neural Network," *Applied Sciences*, vol. 13, no. 8, Jan. 2023, Art. no. 4750, <https://doi.org/10.3390/app13084750>.
- [3] A. H. Meftah, M. Qamhan, Uks.-A. 22nd I. C. on C. M. and S. Alotaibi, and Y. A. Zakariah, "Arabic Speech Emotion Recognition Using KNN and KSUEmotions Corpus," presented at the UKSim-AMSS 22nd International Conference on Computer Modelling and Simulation, Mar. 2020, <https://doi.org/10.5013/IJSSST.a.21.02.21>.
- [4] R. H. Aljuhani, A. Alshutayri, and S. Alahdal, "Arabic Speech Emotion Recognition From Saudi Dialect Corpus," *IEEE Access*, vol. 9, pp. 127081–127085, 2021, <https://doi.org/10.1109/ACCESS.2021.3110992>.
- [5] K. Mountzouris, I. Perikos, and I. Hatzilygeroudis, "Speech Emotion Recognition Using Convolutional Neural Networks with Attention Mechanism," *Electronics*, vol. 12, no. 20, Jan. 2023, Art. no. 4376, <https://doi.org/10.3390/electronics12204376>.
- [6] S. Akinpelu and S. Viriri, "Speech emotion classification using attention based network and regularized feature selection," *Scientific Reports*, vol. 13, no. 1, Jul. 2023, Art. no. 11990, <https://doi.org/10.1038/s41598-023-38868-2>.
- [7] Muljono, M. R. Prasetya, A. Harjoko, and C. Supriyanto, "Speech Emotion Recognition of Indonesian Movie Audio Tracks based on MFCC and SVM," in *2019 International Conference on contemporary*

- Computing and Informatics (IC3I)*, Singapore, Dec. 2019, pp. 22–25, <https://doi.org/10.1109/IC3I46837.2019.9055509>.
- [8] S. Hamsa, I. Shahin, Y. Iraqi, and N. Werghi, "Emotion Recognition From Speech Using Wavelet Packet Transform Cochlear Filter Bank and Random Forest Classifier," *IEEE Access*, vol. 8, pp. 96994–97006, 2020, <https://doi.org/10.1109/ACCESS.2020.2991811>.
 - [9] S. Xefteris, N. Doulamis, V. Andronikou, T. Varvarigou, and G. Cambourakis, "Behavioral Biometrics in Assisted Living: A Methodology for Emotion Recognition," *Engineering, Technology & Applied Science Research*, vol. 6, no. 4, pp. 1035–1044, Aug. 2016, <https://doi.org/10.48084/etasr.634>.
 - [10] S. C. Venkateswarlu, S. R. Jeevakala, N. U. Kumar, P. Munaswamy, and D. Pendyala, "Emotion Recognition From Speech and Text using Long Short-Term Memory," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11166–11169, Aug. 2023, <https://doi.org/10.48084/etasr.6004>.
 - [11] W. Almukadi, "Smart Scarf: An IOT-based Solution for Emotion Recognition," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10870–10874, Jun. 2023, <https://doi.org/10.48084/etasr.5952>.
 - [12] A. Meftah, Y. Alotaibi, and S.-A. Selouani, "Emotional speech recognition: A multilingual perspective," in *2016 International Conference on Bio-engineering for Smart Technologies (BioSMART)*, Dubai, United Arab Emirates, Sep. 2016, <https://doi.org/10.1109/BIOSMART.2016.7835600>.
 - [13] S. Klaylat, Z. Osman, L. Hamandi, and R. Zantout, "Emotion recognition in Arabic speech," *Analog Integrated Circuits and Signal Processing*, vol. 96, no. 2, pp. 337–351, Aug. 2018, <https://doi.org/10.1007/s10470-018-1142-4>.
 - [14] R. Zantout, S. Klaylat, L. Hamandi, and Z. Osman, "Ensemble Models for Enhancement of an Arabic Speech Emotion Recognition System," in *Advances in Information and Communication*, 2020, pp. 174–187, https://doi.org/10.1007/978-3-030-12385-7_15.
 - [15] L. Abdel-Hamid, "Egyptian Arabic speech emotion recognition using prosodic, spectral and wavelet features," *Speech Communication*, vol. 122, pp. 19–30, Sep. 2020, <https://doi.org/10.1016/j.specom.2020.04.005>.
 - [16] A. Ali and Y. Hifny, "Efficient Arabic emotion recognition using deep neural networks," arXiv, Oct. 31, 2020, <https://doi.org/10.48550/arXiv.2011.00346>.
 - [17] O. Mohamed and S. A. Aly, "Arabic Speech Emotion Recognition Employing Wav2vec2.0 and HuBERT Based on BAVED Dataset," arXiv, Oct. 08, 2021, <https://doi.org/10.48550/arXiv.2110.04425>.
 - [18] O. A. Mohammad and M. Elhadeif, "Arabic Speech Emotion Recognition Method Based On LPC And PPSD," in *2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, Jan. 2021, pp. 31–36, <https://doi.org/10.1109/ICCAKM50778.2021.9357769>.
 - [19] S. Kakuba, A. Poulouse, and D. S. Han, "Attention-Based Multi-Learning Approach for Speech Emotion Recognition With Dilated Convolution," *IEEE Access*, vol. 10, pp. 122302–122313, 2022, <https://doi.org/10.1109/ACCESS.2022.3223705>.
 - [20] A. Agrima, I. Mounir, A. Farchi, L. Elmaazouzi, and B. Mounir, "Emotion recognition from syllabic units using k-nearest-neighbor classification and energy distribution," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 6, pp. 5438–5449, Dec. 2021, <https://doi.org/10.11591/ijece.v11i6.pp5438-5449>.
 - [21] I. Alwayle et al., "Parameter Tuned Machine Learning Based Emotion Recognition on Arabic Twitter Data," *Computer Systems Science and Engineering*, vol. 46, no. 3, pp. 3423–3438, 2023, <https://doi.org/10.32604/csse.2023.033834>.
 - [22] M. Tajalsir, S. M. Hernandez, and F. A. Mohammed, "ASERS-CNN: Arabic Speech Emotion Recognition System based on CNN Model," *Signal & Image Processing: An International Journal*, vol. 13, no. 1, pp. 45–53, Feb. 2022, <https://doi.org/10.5121/sipij.2022.13104>.
 - [23] W. Alsabhan, "Human-Computer Interaction with a Real-Time Speech Emotion Recognition with Ensembling Techniques 1D Convolution Neural Network and Attention," *Sensors*, vol. 23, no. 3, Jan. 2023, Art. no. 1386, <https://doi.org/10.3390/s23031386>.
 - [24] I. Shahin, O. A. Alomari, A. B. Nassif, I. Afyouni, I. A. Hashem, and A. Elnagar, "An efficient feature selection method for arabic and english speech emotion recognition using Grey Wolf Optimizer," *Applied Acoustics*, vol. 205, Mar. 2023, Art. no. 109279, <https://doi.org/10.1016/j.apacoust.2023.109279>.
 - [25] M. El Seknedy and S. A. Fawzi, "Emotion Recognition System for Arabic Speech: Case Study Egyptian Accent," in *Model and Data Engineering*, Cairo, Egypt, 2023, pp. 102–115, https://doi.org/10.1007/978-3-031-21595-7_8.
 - [26] R. Y. Cherif, A. Moussaoui, N. Frahta, and M. Berrimi, "Effective speech emotion recognition using deep learning approaches for Algerian dialect," in *2021 International Conference of Women in Data Science at Taif University (WiDSTaif)*, Taif, Saudi Arabia, Mar. 2021, <https://doi.org/10.1109/WiDSTaif52235.2021.9430224>.
 - [27] W. G. S. Al Fadhli, R. K. S. Al Hinai, P. C. Sherimon, V. Sherimon, and R. K. Remya, "An Automated Emotion Recognition from Arabic Speech Using Machine Learning Technique," *International Journal of Creative Research Thoughts*, vol. 10, no. 10, pp. a435–a438, Oct. 2022.
 - [28] S. Klaylat, Z. Osman, L. Hamandi, and R. Zantout, "Enhancement of an Arabic Speech Emotion Recognition System," *International Journal of Applied Engineering Research*, vol. 13, no. 5, pp. 2380–2389, 2018.
 - [29] "Arabic Natural Audio Dataset." [Online]. Available: <https://www.kaggle.com/datasets/susol72/arabic-natural-audio-dataset>.
 - [30] E. de Bodd, M. Cottrell, P. Letremy, and M. Verleysen, "On the use of self-organizing maps to accelerate vector quantization," *Neurocomputing*, vol. 56, pp. 187–203, Jan. 2004, <https://doi.org/10.1016/j.neucom.2003.09.009>.
 - [31] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, May 2016, pp. 785–794, <https://doi.org/10.1145/2939672.2939785>.
 - [32] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, Aug. 1997, <https://doi.org/10.1006/jcss.1997.1504>.

The Effect of Waste Marble Dust and Corncob Ash on the Engineering and Micro-Structural Properties of Expansive Soil for Use in Road Subgrades

Leonardo Z. Wongbae

Department of Civil Engineering, Pan African University Institute for Basic Sciences, Technology and Innovation, Kenya

z.leonardo@students.jkuat.ac.ke (corresponding author)

Charles Kabubo

School of Civil, Environmental, and Geomatics Engineering, Jomo Kenyatta University of Agriculture and Technology, Kenya

kabcha@jkuat.ac.ke

Alphonse Owayo

School of Engineering, Moi University, Kenya

aowayo@mu.ac.ke

Received: 8 February 2024 | Revised: 27 February 2024 and 6 March 2024 | Accepted: 12 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7034>

ABSTRACT

This research investigated the effect of Waste Marble Dust (WMD) and Corncob Ash (CCA) on expansive soil's engineering and microstructural properties. Various laboratory experiments were performed on the natural soil to ascertain its characteristics. The corncobs underwent pre-water treatment for fourteen days to remove excess potassium and increase their silica content, resulting in a rise in the silica level from 0% to 50%. At first, only WMD was added to the soil in increments of 5% to 30% using compaction and California bearing tests. The optimum dosage of 15% WMD addition yielded the best result. CCA was then incorporated by the weight of the soil from 2% to 10% in increments of 2% to the first optimum (15% WMD) to obtain the overall optimum for the study (15% WMD and 8% CCA). Stabilization of the natural soil using both materials led to the modification and solidification of the soil mass, evident by the rise in California bearing ratio values from 1.68% to 15.53% and unconfined compressive strength from 41.33 kN/m² to 174.68 kN/m². There was also a decrease in the soil's free swell from 120% to 15% as well as reductions in the liquid limits from 56.23% to 36.01% and in the plasticity index from 29.74% to 8.72%, respectively. The microstructural images showed the formation of cementitious compounds in the form of calcium silicate hydrate and calcium aluminate hydrate gels. The findings indicate that using WMD and CCA as a unit has great potential in enhancing engineering properties, like strength parameters and the swell potential of expansive soils.

Keywords-corncob ash; expansive soil; free swell index; microstructure; waste marble dust

I. INTRODUCTION

Maintaining and replacing existing pavement hugely increases the budget of transportation agencies [1]. The subsoil-supporting road infrastructure must have a sufficient load-bearing capacity to withstand vehicular traffic [2]. Expansive soils, which exhibit features of shrinking and swelling, are the major causative agents for structural and geotechnical damage to infrastructure worldwide [2, 3]. The structures built on these soils develop defects due to their swell and shrink activities [4-

6]. The majority of pavement failures are due to poor subgrade conditions in expansive soils. As a result, the subgrade of road pavements would require some kind of alteration and redesigning to strengthen their ability to support the load [7]. There are huge economic losses due to construction on expansive soils in Africa, Europe, America, and Asia [3, 7-9]. As a way of mitigating these challenges, engineers have adopted various soil improvement techniques. Removing expansive soils and replacing them with coarser soil has become a common practice. However, this process, described

as a cost-effective and environmentally friendly solution [10], is expensive and time-consuming, causing unnecessary delays in the project schedule. Thus, it may not always be economical [11, 12]. This is because there are increasing costs associated with the construction and transfer of appropriate soil from distant locations [13], affecting the investment cost of road projects [14].

Geotechnical engineers and transportation engineering professionals have made efforts to develop solutions for expansive soil, with a renewed focus on ecologically friendly mitigation methods, leading to the enhancement of soil properties either through stabilization and modification, or through both [15]. Lime and cement are two frequently used stabilizers for modifying the characteristics of soils. However, environmental pollution is greatly increased by the continuous emissions of carbon dioxide during the manufacture of cement [16]. Additionally, the production of cement requires the massive acquisition of non-renewable raw materials, like limestone, and the activities leading to the acquisition of these resources from the natural surroundings affect the environment. Furthermore, the processing of these raw materials, particularly in factories, releases pollutants into the environment [17]. Stabilizing the subgrade of expansive soils using cement has proved to be expensive and unsustainable [18]. Industrial and agricultural wastes like fly ash and rice husk ash have been utilized to stabilize expansive soils just as efficiently as cement and lime [19, 21]. Waste Marble Dust (WMD), a by-product produced in large amounts during the cutting and processing of marble stones, has been utilized in soil stabilization [20]. It is a mechanical stabilizer, often referred to as stone dust, which primarily works by enhancing gradation, plasticity, and compaction of the soil [11]. This material has an abundance of calcium oxide (CaO) or quicklime [22, 25], a primary ingredient of Portland cement. Its use as a stabilizer for weak soil leads to cation exchange, flocculation, and pozzolanic reactions [14]. Corncob Ash (CCA), an agricultural by-product obtained from the production of corn or maize, on the other hand, has been extensively employed in soil stabilization. It is similar to bagasse ash and rice husk ash with high silica, alumina, and iron oxide content. It can be used as a Supplementary Cementitious Material (SCM) in cement and concrete [12, 21-27].

To the authors' best knowledge, no study has been published regarding the usage of both WMD and CCA on the engineering and microstructural properties of expansive soils. It should be noted that the independent utilization of WMD and CCA did not lead to significant increase in the bearing capacity and reduction in the swell potential of the soils. The extent of chemical bonding between soil particles after stabilizing with these materials through microstructural analysis has yet to be studied. Hence, the current paper focuses on the effect of the combined employment of these materials on the engineering and microstructural properties of expansive soils. The results from this study can ensure the proper utilization of locally available agricultural and mining wastes in road construction, thus contributing to eco-friendly construction.

II. MATERIALS AND METHODS

A. Materials

Expansive soil, WMD, and CCA were the materials employed to conduct this study. Materials from various places across Kenya were sourced. The expansive soil was collected from the Jomo Kenyatta University of Agriculture and Technology (JKUAT) campus in Kiambu County, Kenya. The soil sample was obtained by excavating at a depth of 2 m below the ground surface. The marble dust was collected from the industrial area in Nairobi, Kenya, and was taken to the Ministry of Mining and Petroleum for X-Ray Fluorescence (XRF) to determine its chemical composition. The corn cobs were gathered from local marketers in Juja and Eldoret and transported to the Structural Engineering Laboratory of Jomo Kenyatta University of Agriculture and Technology for pre-water treatment and incineration. The materials are displayed in Figure 1.



Fig. 1. (a) Expansive soil, (b) CCA, (c) WMD, (d) soil mixed with WMD, (e) soil mixed with CCA, (f) Soil mixed with WMD and CCA.

B. Methods

To remove the excess potassium content in the corn cobs and increase the silica content, which was initially low, the cobs were subjected to water treatment. Figure 2 illustrates the pre-water treatment process for the corn cobs. Water was placed into the tanks containing the corn cobs and was renewed three times a week to facilitate the elimination of the excess potassium. After being removed from the tanks, the cobs were left to dry in the sun for several days in preparation for incineration. The material was subsequently incinerated in open air [28] and filtered through a 0.42 mm sieve [16, 29].

To determine the optimum dosage of the two materials, waste marble dust was initially added to the soil at 5, 10, 15, 20, 25, and 30% by weight of dry soil [24, 30]. Several researchers have adopted this approach to stabilize expansive soils [31, 32]. Two separate tests, compaction and CBR tests [33], were carried out to obtain the optimum WMD content. Thereafter, CCA at 2% intervals, up to the maximum of 10% by weight of the dry soil was added to the soil with the optimum (15%) WMD dosage. The optimum dosage obtained (15% WMD and 8% CCA) from the two materials was used on various experiments, based on prescribed guidelines. The soil

particle size analysis was conducted as per [34] procedures. The Atterberg limits (liquid limit, plasticity index, and plastic limit) of the parent materials and their combinations were done as per BS 1377-2:1990 [35, 36]. The specific gravity and free swell index of the samples were determined as per BS 1377-2:1990 [35] and IS 2720 (Part 3) 1980 [37], respectively. Maximum Dry Density (MDD) and Optimum Moisture Content (OMC) values of all the samples were determined based on BS1377- 4:1990 [38].



Fig. 2. Treatment and preparation of CCA: (a) Soaking of cobs, (b) drying of cobs, (c) incineration, (d) sieving, (e) ash.

California Bearing Ratio test was carried out as per BS1377-4:1990 [38]. Each Unconfined Compressive Strength (UCS) test was performed on all combinations [39]. For this test, the samples were prepared by being statically compacted at their respective MDD and OMC, and tested after seven days of curing. The microstructural characteristics were determined with the help of Scanning Electron Microscopy (SEM) and X-Ray Diffraction (XRD). The samples for the test were prepared at their moisture contents of the optimum mix of 15% WMD and 8% CCA using the UCS test. Initially, SEM and XRD analyses of the untreated samples were determined after seven (7) days of curing. Then, the analysis of the soil samples blended with the optimum percentage of WMD and CCA, was determined after 7, 28, 56, and 90 days of curing. The samples for SEM analysis were mounted onto aluminum SEM stubs using carbon tape and were then coated with carbon to make them conductive. The images were taken on the Tescan MIRA SEM. The operating conditions are observed in the data bar at the bottom of the images.

III. RESULTS AND DISCUSSION

A. Characterization of Expansive Soil, WMD, and CCA

1) Particle Size Distribution

As indicated in Figure 3 and Table I, the soil was found to have 62% clay, 28% silt, 8% and 2% gravel. This finding describes the soil as gravelly sandy and silty clay. The soil was classified as A-7-6 according to the AASTHO classification system [40], being placed under the category of clay soil. The abundance of clay in the soil suggests that the latter was weak and poor to be utilized as a subgrade material, and it was therefore necessary for it to be stabilized to become befitting for use as a road subgrade material [41].

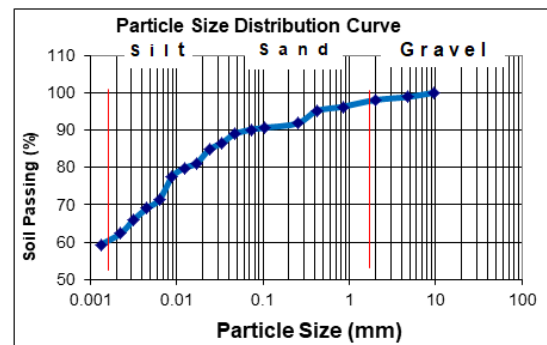


Fig. 3. Particle size distribution curve.

TABLE I. PHYSICO-MECHANICAL PROPERTIES OF EXPANSIVE SOIL

Property	Value
Specific gravity, %	2.47
Moisture content, %	6.41
Gravel, %	2
Silt content, %	28
Sand content, %	8
Clay content, %	62
Liquid limits, %	56.23
Plastic limit, %	26.50
Plasticity index, %	29.74
FSI, %	120
OMC, %	32.7
MDD, g/cm ³	1.33
UCS, kN/m ²	41.33
AASHTO classification	A-7-6
CBR (%)	1.68

2) Chemical Composition of Materials

As shown in Table II, the chemical compositions of the research materials were determined using the XRF apparatus. The most abundant oxides in the expansive soil were silicon oxide (SiO₂) at 76.136%, aluminum oxide (Al₂O₃) at 11.1833%, and ferrous oxide (Fe₂O₃) at 8.887%. Calcium oxide (CaO) was discovered to be the most abundant oxide in WMD, accounting for 51.65%. These results indicate that WMD has a sufficient amount of CaO, similar to cement. The main oxides detected in CCA were silica oxide (SiO₂) (50.009%) magnesium oxides (MgO) (12.599%), ferrous oxide (Fe₂O₃) at 4.639%, and aluminum oxide (Al₂O₃) at 4.530%. The cementitious properties of a CaO content greater than 20% or the pozzolanic and cementitious properties of a CaO content between 10% and 20%, as discussed by [43-44] were satisfied. The ratio of (CaO + MgO)/SiO₂ to exceed 1 for cementitious materials [44] was not satisfied. However, a conclusion that CCA could possess pozzolanic properties was reached since the BSI's requirement of a SiO₂ content of at least 25% was satisfied. The obtained results indicate that the CCA contains a SiO₂ content of 50.009%, which exceeds the recommended minimum requirement of 25% according to BS EN 197-1 (2011) [44] and 39.90% according to ASTM C618 [45]. The total SiO₂+Fe₂O₃+Al₂O₃ content is 59.178% and falls short of the minimum requirement of 70% of ASTM C618 for classification as Class N and Class F pozzolana. However, it meets the minimum requirement of 50% as per ASTM C618 and can be consequently classified as a class C pozzolana.

TABLE II. CHEMICAL COMPOSITION OF MATERIALS

Analysis (%)	Expansive soil	WMD	CCA
CaO	1.478	51.65	13.89
SiO ₂	76.14	3.42	50.01
Al ₂ O ₃	11.18	2.05	4.530
Fe ₂ O ₃	8.887	1.26	4.639
MgO	-	-	12.6
K ₂ O	0.511	0.01	5.784
SO ₃	-	-	-
Na ₂ O	0.002	0.001	0.01
LOI	-	39.59	2.65

B. Influence of WMD and CCA Atterberg Limits

The combination of the two materials (15% WMD and 8% CCA) exhibited a reduction of the liquid limit from 56.23% to 36.01%, as shown in Figure 4. Based on the reduction in the liquid limit, the soil behaves as a highly expansive soil [46]. As demonstrated in Figure 5, the plastic limit which defines the moisture content at which the soil changes from semisolid to plastic increased with the addition of stabilizers. This could be due to the introduction of fine particles into the soil, which altered its mineralogy and cation exchange capacity.

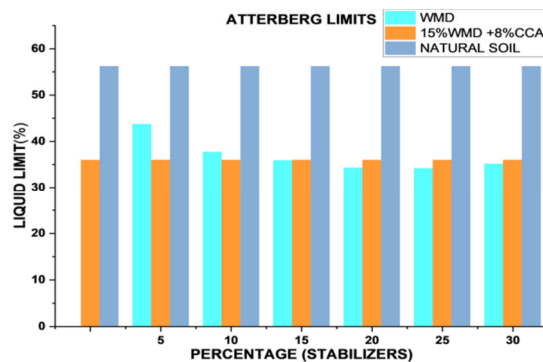


Fig. 4. Liquid limit.

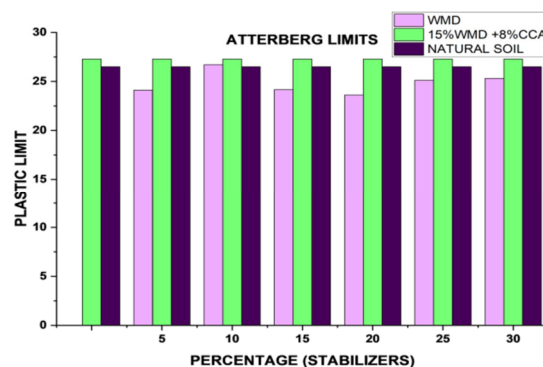


Fig. 5. Plastic limit.

Figure 6 presents the Plasticity Index (PI), which represents the difference between the liquid limit and the plastic limit. The combination of 15% WMD and 8% CCA resulted in a reduction in PI from 29.74% to 8.72%, which is a 70.6% decrease. This result satisfies the 15% minimum PI value, suggested by the Pavement Guidelines for Low-Volume Sealed Roads in Kenya [47]. The process of conducting the test is depicted in Figure 11.

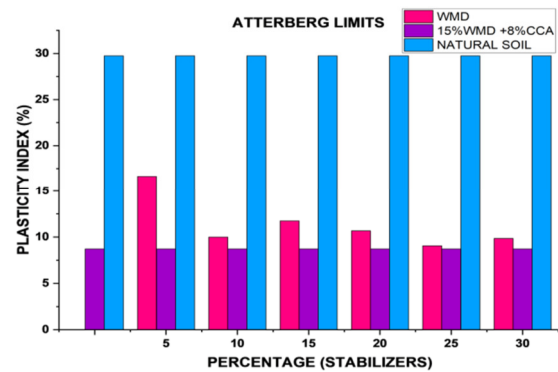


Fig. 6. Plasticity index.

C. Influence WMD and CCA Compaction Parameters

The combination of 15% WMD obtained through utilizing waste dust as a stand-alone material and varying percentages of CCA during compaction experienced a rise in MDD at only 2% addition of CCA. This could be due to the replacement of light soil particles by the larger WMD particles [25, 31]. Beyond 2%, the MDD kept decreasing, as spotted in Figure 7. This could be an outcome of the soil becoming lighter when the content of stabilizer increased. On the other hand, OMC kept reducing gradually with up to 6% CCA addition to 15% WMD.

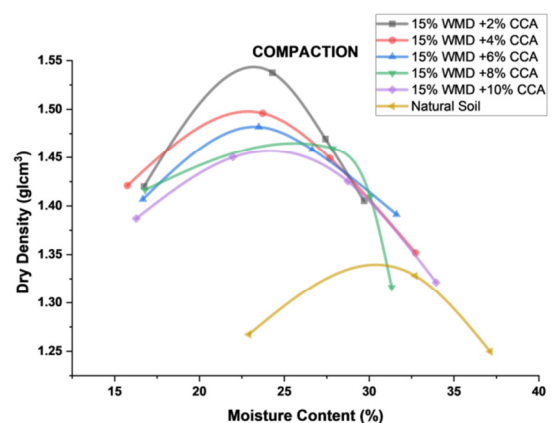
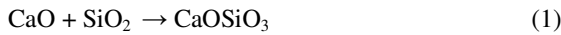


Fig. 7. Compaction with the optimum mix.

D. Influence of WMD and CCA on CBR Expansive Soil

The CBR test, an essential parameter for examining the strength of the subgrade of roads, was determined. Before testing, samples for the natural and blended soils were prepared at their optimum moisture contents and were soaked for four days [30, 48]. Adding different percentages of WMD increased the CBR from 1.68% for the natural soil to 4.50%, with the highest CBR value achieved with 15% of the material. Using the optimum of 15% WMD with varying proportions of CCA, the CBR increased from 4.50% to 15.53% for 8% CCA addition. The 15% WMD and 8% CCA were thus used as the optimum addition percentage for all the other experiments. The rise in CBR is attributed to the pozzolanic activities emanating from the reaction between CaO and SiO₂ in WMD and CCA, respectively. Equation (1) highlights the combined reaction. The process of CBR testing is indicated in Figure 11.



The formation of cementitious compounds like Calcium Silicate Hydrate (CSH) could also occur, leading to an augmentation in the bearing capacity. Equation (2) highlights the process of the CSH reaction. The result of the CBR test is outlined in Figure 8.

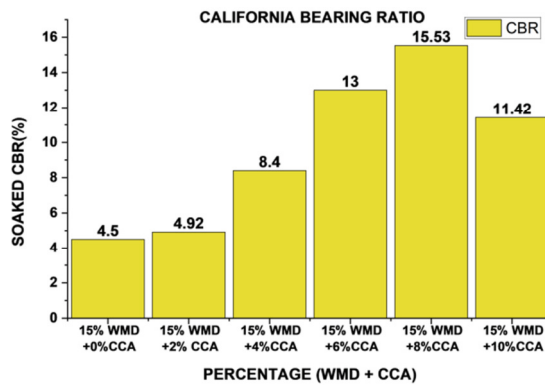
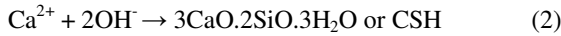


Fig. 8. CBR with the optimum mix.

E. Influence of WMD and CCA on Unconfined Compressive Strength

There was an increase in the Unconfined Compressive Strength (UCS) value from 41.33 kN/m² for the natural soil to 174.68 kN/m², using the optimum mixture with 15% WMD and 8% CCA. The results were obtained after 7 days of curing of the samples. There were further increments in the UCS value after 14, 28, and 56 days of curing. These increments in the UCS values could be caused by the pozzolanic reaction [51,53] from the use of the two stabilizers with curing time. These results were higher than the ones acquired when the stabilizers were separately utilized. The findings from the UCS test are displayed in Figure 9.

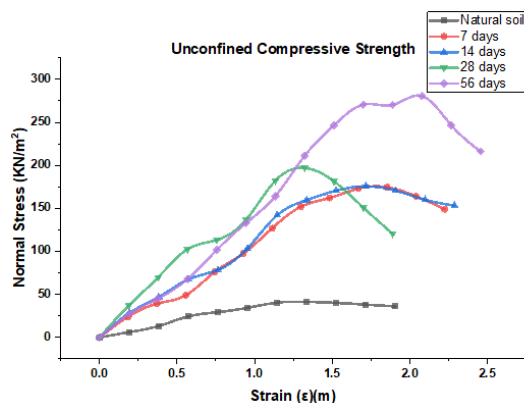
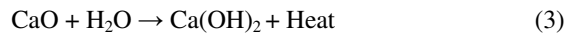


Fig. 9. Unconfined compressive strength.

F. Influence of WMD and CCA on Free Swell Index

The findings obtained from the free swell index test are detected in Figure 10. The results show a reduction in swelling

with increasing content of WMD. This pattern aligns with the findings of [31]. The presence of CaO could be responsible for this modification and reduction in swell, due to the cation ion exchange that emerges from the reaction of CaO or quicklime with water to form calcium hydroxide (Ca(OH)₂) or slaked lime. The reduction in the free swell could be an outcome of the decrease in the specific surface as well as the placement of the soil with materials that have non-swelling and pozzolanic characteristics [31]. As reported in [52], further dissolution of Ca(OH)₂ in water breaks into OH⁻ and Ca²⁺, thus reducing the double diffused layer thickness of the soil particles. The reaction is indicated in (4) and (5).



The optimum combination of WMD and CCA leads to a drop in the swell from 120% to 15%. This result changes the soil type from expansion to low expansion [47, 55], exhibiting a significant reduction in the swell potential.

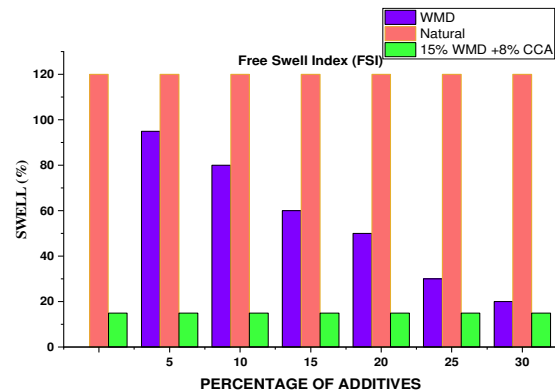


Fig. 10. Free swell index with the optimum mix.



Fig. 11. Tests conducted during the study.

G. Microstructural Analysis of Expansive Soil Blended with WMD and CCA

1) Scanning Electron Microscopy

From the microstructural analysis using SEM, large voids having honeycomb-like structures were observed in the natural soil. Figure 12 displays significant pores in the expansive soil,

indicating shrinkage fractures caused by its interaction with moisture. Authors in [54] observed a similar pattern. The SEM image of the soil clearly showed the existence of several larger void spaces and a less cohesive soil structure. However, the introduction of the WMD-CCA mix resulted, after 7 days of curing, in the filling of the voids. The soil particles underwent flocculation, leading to the development of calcium silicate hydrate gel. After 7 days of curing, there was a creation of bigger particles and a reduction in the size of larger pores, as portrayed in Figure 12. As a result, the attraction forces between particles reduce the material's ability to change shape and ultimately increase its ability to withstand compression. Additionally, the soil matrix experienced enhancement after a 14-day curing period. The WMD-CCA mixture and the natural soil reaction with moisture caused an alteration in the distribution of pore sizes. This reaction led to an increase in cementing agents and the aggregation of particles [55]. There was an overall increment in the compressive strength of the soil matrix, after 56 days of curing [49], due to the presence of gritty and rough surfaces that were detected in the SEM images. At 90 days of curing, there was a full densification of the soil particles, leading to the formation of cementitious compounds.

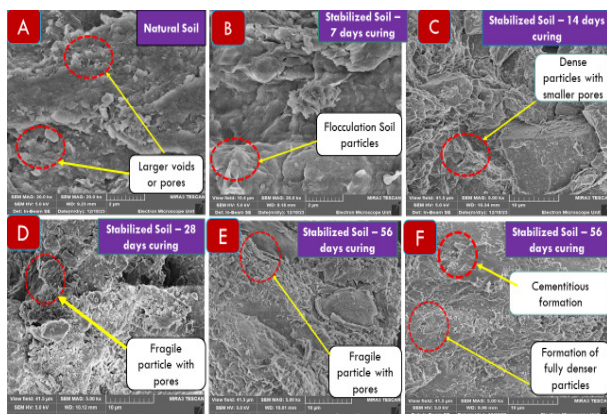


Fig. 12. SEM images for 7 - 90 days of curing.

2) X-Ray Diffraction (XRD) Analysis

XRD analysis determines the crystallographic arrangement of a substance. It is advantageous for determining the mineralogical composition of crystalline substances, such as minerals, rocks, soils, and other solid materials. Figures 13 and 14 present the XRD images of the untreated expansive soil and the soil with stabilizing materials (15% WMD and 8% CCA) after a 28-day curing period. These images provide an insight into the mineralogical modifications in the soil treated with stabilizers.

Dolomite, a kind of CaO and magnesium oxide (MgO) exhibited a greater peak in the soil-WMD-CCA combination. The presence of dolomite in WMD could be responsible for the production of a densely packed soil matrix, leading to an increase in the soil strength. Quartz and orthoclase were also identified as the dominant components in the soil and CCA. The presence of quartz in the soil and the CCA contributes to durability and aids the formation of calcium silicate hydrate

and calcium aluminate hydrate gel in the presence of water [49]. These minerals contributed to the increase of the CBR and UCS of the soil. The mineralogical compositions of the materials can be pinpointed in Table III.

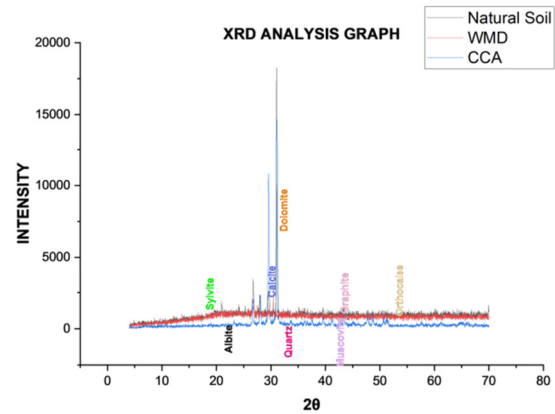


Fig. 13. XRD analysis of the materials.

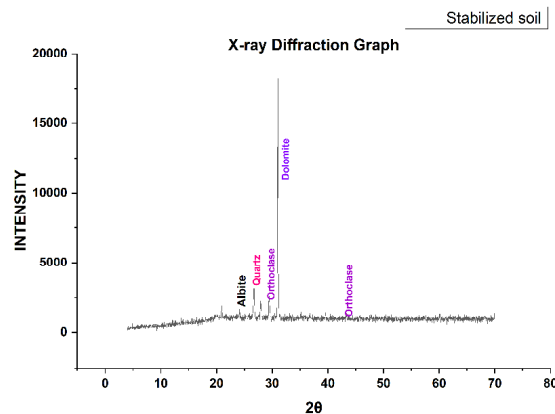


Fig. 14. XRD analysis of stabilized soil.

TABLE III. MINERALOGICAL COMPOSITION OF THE MATERIALS THROUGH XRD ANALYSIS

Materials	Phases identified from XRD
Natural soil	Quartz, Muscovite, Albite, Orthoclase
WMD	Dolomite, Calcite, Orthoclase, Quartz
CCA	Sylvite, Albite, Graphite, Orthoclase, Quartz
Stabilized Soil	Dolomite, Albite, Calcite, Orthoclase, Quartz

IV. CONCLUSION

In this study, the effects of the addition of Waste Marble Dust (WMD) and Corncob Ash (CCA) on the engineering and microstructural properties of the soil were assessed and the following conclusions were drawn:

- Integrating WMD and CCA led to enhancements in Atterberg limits, by specifically decreasing the Plasticity Index (PI) and the Liquid Limit (LL), and slightly increasing the Plastic Limit (PL), compared to the use of these materials separately. The improvement of the compaction parameters, witnessed through an increase in

the Maximum Dry Density (MDD) and a decrease in Optimum Moisture Content (OMC), leads to higher soil stability and appropriateness for road subgrades.

- The optimum addition mixture was defined as 15% WMD and 8% CCA.
- The California Bearing Ratio (CBR) and Unconfined Compressive Strength (UCS) values showed satisfactory improvements with the combined optimum mixture. The high strength gain indicates that WMD and CCA, when used as a unit have far greater potential to increase the load-bearing capacity and overall strength of expansive soils.
- The Scanning Electron Microscopy (SEM) analysis on the stabilized soil samples demonstrated noticeable alterations in the soil mass with increasing curing time. The denseness in the soil mass and the filling of large voids are traceable to the formation of CSH and CAH gel leading to the development of cementitious compounds and pozzolanic activities. The X-Ray Diffraction (XRD) examination on the stabilized soil matrix revealed the presence of newly created stable compounds, which correlates with the enhanced strength.

FUNDING STATEMENT

The authors express their thanks and appreciation to the African Union Commission for supporting financially this research and publication.

ACKNOWLEDGEMENT

The authors acknowledge the Pan African University Institute for Basic Sciences, Technology and Innovation (PAUSTI) and Jomo Kenyatta University of Agriculture and Technology (JKUAT), for allowing the use of their laboratory facilities while conducting this research.

REFERENCES

- [1] K. Jamaluddin and R. P. Munirwan, "Improvement of geotechnical properties of clayey soil with saw dust ash stabilization," *E3S Web of Conferences*, vol. 340, 2022, Art. no. 01009, <https://doi.org/10.1051/e3sconf/202234001009>.
- [2] M. K. Atahu, "The Effect of Coffee Husk Ash on Geotechnical Properties of Expansive Soil," Ph.D. dissertation, University of Rostock, Rostock, Germany, 2020.
- [3] S. Y. Amakye, S. J. Abbey, C. A. Booth, and A.-M. Mahamadu, "Enhancing the Engineering Properties of Subgrade Materials Using Processed Waste: A Review," *Geotechnics*, vol. 1, no. 2, pp. 307–329, Dec. 2021, <https://doi.org/10.3390/geotechnics1020015>.
- [4] S. J. Abbey, E. U. Eyo, and S. Ng'ambi, "Swell and microstructural characteristics of high-plasticity clay blended with cement," *Bulletin of Engineering Geology and the Environment*, vol. 79, no. 4, pp. 2119–2130, May 2020, <https://doi.org/10.1007/s10064-019-01621-z>.
- [5] X. Yang, "Mitigating Dry Shrinkage Pavement Cracking by Geocell," University of Oklahoma, Norman, OK, USA, Technical Report SPTC15.1-06, Nov. 2019.
- [6] H. Sellaf and B. Balegh, "An Experimental Study on the Effect of Plastic Waste Powder on the Strength Parameters of Tuff and Bentonite Soils Treated with Cement," *Engineering, Technology & Applied Science Research*, vol. 13, no. 2, pp. 10322–10327, Apr. 2023, <https://doi.org/10.48084/etasr.5580>.
- [7] S. Y. Amakye and S. J. Abbey, "Understanding the performance of expansive subgrade materials treated with non-traditional stabilisers: A review," *Cleaner Engineering and Technology*, vol. 4, Oct. 2021, Art. no. 100159, <https://doi.org/10.1016/j.clet.2021.100159>.
- [8] F. E. Jalal, Y. Xu, B. Jamhiri, and S. A. Memon, "On the Recent Trends in Expansive Soil Stabilization Using Calcium-Based Stabilizer Materials (CSMs): A Comprehensive Review," *Advances in Materials Science and Engineering*, vol. 2020, Mar. 2020, Art. no. e1510969, <https://doi.org/10.1155/2020/1510969>.
- [9] M. Zumrawi, A. Abdelmarouf, and A. Gameil, "Damages of Buildings on Expansive Soils: Diagnosis and Avoidance," *International Journal of Multidisciplinary and Scientific Emerging Research*, vol. 6, pp. 108–116, May 2017.
- [10] O. F. Fadmo, S. S. Kar, D. Tiwari, and A. Singh, "Environmental and Economic Impact of Mixed Cow Dung and Husk Ashes in Subgrade Soil Stabilization," *International Journal of Pavement Research and Technology*, vol. 15, no. 4, pp. 835–846, Jul. 2022, <https://doi.org/10.1007/s42947-021-00056-8>.
- [11] S. Mishra, S. N. Sachdeva, and R. Manocha, "Subgrade Soil Stabilization Using Stone Dust and Coarse Aggregate: A Cost Effective Approach," *International Journal of Geosynthetics and Ground Engineering*, vol. 5, no. 3, Jul. 2019, Art. no. 20, <https://doi.org/10.1007/s40891-019-0171-0>.
- [12] W. Yifru, N. Getu, D. Kifile, A. Mesfin, A. Sewunet, and M. Tamene, "Effects of Corn Cob Ash as Partial Replacement of Cement for Stabilization of an Expansive Clay," *Advances in Civil Engineering*, vol. 2022, Nov. 2022, Art. no. e6788120, <https://doi.org/10.1155/2022/6788120>.
- [13] A. Soltani, A. Taheri, M. Khatibi, and A. R. Estabragh, "Swelling Potential of a Stabilized Expansive Soil: A Comparative Experimental Study," *Geotechnical and Geological Engineering*, vol. 35, no. 4, pp. 1717–1744, Aug. 2017, <https://doi.org/10.1007/s10706-017-0204-1>.
- [14] J. Nelson and D. J. Miller, *Expansive Soils: Problems and Practice in Foundation and Pavement Engineering*. New York, NY, USA: John Wiley & Sons, 1997.
- [15] R. P. Munirwan, D. Sundary, Munirwansyah, and Bunyamin, "Study of coffee husk ash addition for clay soil stabilization," *IOP Conference Series: Materials Science and Engineering*, vol. 1087, no. 1, Oct. 2021, Art. no. 012016, <https://doi.org/10.1088/1757-899X/1087/1/012016>.
- [16] O. Adebisi, A. M. Taiwo, O. B. Julius, and A. E. Olusola, "Partial Replacement of Cement with Corn Cob Ash -A Review," *Global Scientific Journals*, vol. 7, no. 11, pp. 82–87, 2019.
- [17] G. Tariq *et al.*, "Influence of green technology, green energy consumption, energy efficiency, trade, economic development and FDI on climate change in South Asia," *Scientific Reports*, vol. 12, no. 1, Sep. 2022, Art. no. 16376, <https://doi.org/10.1038/s41598-022-20432-z>.
- [18] N. S. Ikhlef, M. S. Ghembaza, and M. Dadouch, "Effect of Cement and Compaction on the Physicochemical Behavior of a Material in the Region of Sidi Bel Abbes," *Engineering, Technology & Applied Science Research*, vol. 4, no. 4, pp. 677–680, Aug. 2014, <https://doi.org/10.48084/etasr.467>.
- [19] M. R. Hakro *et al.*, "Compaction Characteristics and Permeability of Expansive Shale Stabilized with Locally Produced Waste Materials," *Materials*, vol. 15, no. 6, Jan. 2022, Art. no. 2138, <https://doi.org/10.3390/ma15062138>.
- [20] N. Gupta and T. Sharma, "Experimental Study on The Effect of Marble Dust and Paddy Straw on The Strength Characteristics of Clayey Soil," *International Journal of Advanced Science and Technology*, vol. 29, pp. 5267–5281, Jun. 2020.
- [21] A. K. Jain, A. K. Jha, and Shivanshi, "Improvement in Subgrade Soils with Marble Dust for Highway Construction: A Comparative Study," *Indian Geotechnical Journal*, vol. 50, no. 2, pp. 307–317, Apr. 2020, <https://doi.org/10.1007/s40098-020-00423-5>.
- [22] H. A. El-Sayed, A. B. Farag, A. M. Kandeel, A. A. Younes, and M. M. Yousef, "Characteristics of the marble processing powder waste at Shaq El-Thoaban industrial area, Egypt, and its suitability for cement manufacture," *HBRC Journal*, vol. 14, no. 2, pp. 171–179, Aug. 2018, <https://doi.org/10.1016/j.hbrj.2016.06.002>.
- [23] C. A. Hwayyiz, A. M. Hasan, R. A. Hummadi, and H. H. Ibrahim, "Impact of Marble Powder on the Geotechnical Behavior of Expansive

- Soil," *Zanco Journal of Pure and Applied Sciences*, vol. 35, no. 1, pp. 30–37, 2023, <https://doi.org/10.21271/ZJPAS.35.1.4>.
- [24] H. A. M. Abdelkader, M. M. A. Hussein, and H. Ye, "Influence of Waste Marble Dust on the Improvement of Expansive Clay Soils," *Advances in Civil Engineering*, vol. 2021, Sep. 2021, Art. no. e3192122, <https://doi.org/10.1155/2021/3192122>.
- [25] S. Amena and W. F. Kabeta, "Mechanical Behavior of Plastic Strips-Reinforced Expansive Soils Stabilized with Waste Marble Dust," *Advances in Civil Engineering*, vol. 2022, May 2022, Art. no. e9807449, <https://doi.org/10.1155/2022/9807449>.
- [26] P. Murthi, K. Poongodi, P. O. Awoyera, R. Gobinath, and R. Saravanan, "Enhancing the Strength Properties of High-Performance Concrete Using Ternary Blended Cement: OPC, Nano-Silica, Bagasse Ash," *Silicon*, vol. 12, no. 8, pp. 1949–1956, Aug. 2020, <https://doi.org/10.1007/s12633-019-00324-0>.
- [27] S. Mohan and P. Chandrasekaran, "Effect of Artificial Fibers and Corn Cob Ash on Mechanical Behavior of High Performance Concrete," *Polish Journal of Environmental Studies*, vol. 31, no. 4, pp. 3713–3721, Jul. 2022, <https://doi.org/10.15244/pjoes/146991>.
- [28] O. D. Afolayan, O. M. Olofinade, and I. I. Akinwumi, "Use of some agricultural wastes to modify the engineering properties of subgrade soils: A review," *Journal of Physics: Conference Series*, vol. 1378, no. 2, Sep. 2019, Art. no. 022050, <https://doi.org/10.1088/1742-6596/1378/2/022050>.
- [29] N. Bheel *et al.*, "Utilization of Corn Cob Ash as Fine Aggregate and Ground Granulated Blast Furnace Slag as Cementitious Material in Concrete," *Buildings*, vol. 11, no. 9, Sep. 2021, Art. no. 422, <https://doi.org/10.3390/buildings11090422>.
- [30] H. A. M. Abdelkader, A. S. A. Ahmed, M. M. A. Hussein, H. Ye, and J. Zhang, "An Experimental Study on Geotechnical Properties and Micro-Structure of Expansive Soil Stabilized with Waste Granite Dust," *Sustainability*, vol. 14, no. 10, Jan. 2022, Art. no. 6218, <https://doi.org/10.3390/su14106218>.
- [31] P. Singh, A. Boora, and A. K. Gupta, "Sub-grade characteristics of clayey soil incorporating municipal solid waste incineration ash and marble dust," *Journal of Engineering, Design and Technology*, vol. 20, no. 3, pp. 712–726, Dec. 2021, <https://doi.org/10.1108/JEDT-08-2020-0347>.
- [32] D. Alemshet, B. Fayissa, A. Geremew, and G. Chala, "Amelioration Effect of Fly Ash and Powdered Ground Steel Slag for Improving Expansive Subgrade Soil," *Journal of Engineering*, vol. 2023, Feb. 2023, Art. no. e1652373, <https://doi.org/10.1155/2023/1652373>.
- [33] *BS 1924-2:1990 Stabilized materials for civil engineering purposes - Methods of test for cement-stabilized and lime-stabilized materials*. London, UK: BSI, 1990.
- [34] *International Standard ISO11277. Soil quality — Determination of particle size distribution in mineral soil material — Method by sieving and sedimentation*. ISO, 2020.
- [35] *BS 1377-2(1990), Methods of test for soils for civil engineering purposes. Classification tests*. London, UK: British Standards Institution, 1990.
- [36] M. Otieno, Z. Gariy, and C. Kabubo, "An Evaluation of the Performance of Lateritic Soil Stabilized with Cement and Biochars to be Used in Road Bases of Low-Volume Sealed Roads," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11366–11374, Aug. 2023, <https://doi.org/10.48084/etasr.6040>.
- [37] *IS 2720. Indian Standard Methods of Test for Soils. Part XI. Determination of Free Swell Index of Soils*. Bureau of Indian Standards, 1978.
- [38] *BS 1377-2:2022. Methods of test for soils for civil engineering purposes - Classification tests and determination of geotechnical properties*. London, UK: BS, 2022.
- [39] *ASTM D2166-06(2006), Standard Test Method for Unconfined Compressive Strength of Cohesive Soil*. West Conshohocken, PA, USA: ASTM International, 2006.
- [40] H. Jamal, "AASHTO Soil Classification System - AASHTO Chart," *About Civil*. <https://www.aboutcivil.org/aashto-soil-classification-system>.
- [41] *Road Design Manual. Part III. Material and Pavement Design for New Roads*. Nairobi, Kenya: Ministry of Transport & Communications, 1987.
- [42] J. Kamau, A. Ahmed, P. Hirst, and J. Kangwa, "Suitability of Corn Cob Ash as a Supplementary Cementitious Material," *International Journal of Materials Science and Engineering*, vol. 4, no. 4, pp. 215–228, 2016, <https://doi.org/10.17706/ijmse.2016.4.4.215-228>.
- [43] J. Kamau and A. Ahmed, "Suitability of Maize Cob Ash as a Partial Cement Replacement," *Juniper Online Journal Material Science*, vol. 2, no. 5, Sep. 2017, Art. no. 555599, <https://doi.org/10.19080/JOJMS.2017.02.555599>.
- [44] *BS EN 197-1(2011), Cement - Composition, specifications and conformity criteria for common cements*. London, UK: British Standards Institution, 2011.
- [45] H. Poernomo, "Preliminary Study of the Utilization of the Fly Ash from Coal-Fired Power Plant for Immobilization of Radioactive Waste," *Indonesian Journal of Chemistry*, vol. 11, no. 3, pp. 258–266, Dec. 2011, <https://doi.org/10.22146/ijc.21390>.
- [46] S. Asuri and P. Keshavamurthy, "Expansive Soil Characterisation: an Appraisal," *INAE Letters*, vol. 1, no. 1, pp. 29–33, Jun. 2016, <https://doi.org/10.1007/s41403-016-0001-9>.
- [47] *Pavement Design Guidelines*. Nairobi, Kenya: Ministry of Transport, Infrastructure, Housing, and Urban Development, 2011.
- [48] R. Malaoui, E. H. Harkati, M. R. Soltani, A. Djellali, A. Soukeur, and R. Kechiched, "Geotechnical Characterization of Phosphate Mining Waste Materials for Use in Pavement Construction," *Engineering, Technology & Applied Science Research*, vol. 13, no. 1, pp. 10005–10013, Feb. 2023, <https://doi.org/10.48084/etasr.5493>.
- [49] M. Otieno, C. Kabubo, and Z. Gariy, "Mechanical and Structural Correlation of Lateritic Soil Road Base Stabilized with Cement and Selected Biochars," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11070–11077, Aug. 2023, <https://doi.org/10.48084/etasr.5973>.
- [50] S. Singh, "Experimental investigation of corn cob ash on silty clay stabilized with calcium carbide," *Materials Today: Proceedings*, vol. 37, pp. 3658–3660, Jan. 2021, <https://doi.org/10.1016/j.matpr.2020.10.146>.
- [51] O. S. Olafusi, W. K. Kupolati, E. R. Sadiku, J. Snyman, and J. M. Ndambuki, "Characterization of corncob ash (CCA) as a pozzolanic material," *International Journal of Civil Engineering and Technology*, vol. 9, no. 12, pp. 1016–1024, 2018.
- [52] D. Barman and S. Dash, "Stabilization of expansive soils using chemical additives: A review," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 14, pp. 1319–1342, Apr. 2022, <https://doi.org/10.1016/j.jrmge.2022.02.011>.
- [53] K. Prakash and A. Sridharan, "Free Swell Ratio and Clay Mineralogy of Fine-Grained Soils," *Geotechnical Testing Journal*, vol. 27, no. 2, pp. 220–225, Mar. 2004, <https://doi.org/10.1520/GTJ10860>.
- [54] A. K. Jain, A. K. Jha, and Shivanshi, "Geotechnical behaviour and micro-analyses of expansive soil amended with marble dust," *Soils and Foundations*, vol. 60, no. 4, pp. 737–751, Aug. 2020, <https://doi.org/10.1016/j.sandf.2020.02.013>.
- [55] F. E. Jalal, S. Mulk, S. A. Memon, B. Jamhiri, and A. Naseem, "Strength, Hydraulic, and Microstructural Characteristics of Expansive Soils Incorporating Marble Dust and Rice Husk Ash," *Advances in Civil Engineering*, vol. 2021, Nov. 2021, Art. no. e9918757, <https://doi.org/10.1155/2021/9918757>.

Assessing the Influence of Various Work Breakdown Structures on Project Completion Time

Babatunde Omoniyi Odedairo

Department of Industrial and Production Engineering, University of Ibadan, Ibadan, 200284, Nigeria |
Department of Industrial Engineering, College of Engineering in Al-Kharj, Prince Sattam Abdulaziz
University, Al-kharj, 11942, Saudi Arabia
b.odedairo@psau.edu.sa, bo.odedairo@ui.edu.ng (corresponding author)

Received: 6 February 2024 | Revised: 8 March 2024 | Accepted: 16 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7023>

ABSTRACT

In project management, a clear definition of the objective is required for the success of a project. Scope management is a performance indicator used to ascertain compliance with predefined project boundaries. The Work Breakdown Structure (WBS) is an essential part of the scope management process and a tool in project planning. Although there is much research on WBS, there is a lack of information regarding the relationship between the selection of WBS orientation and project completion time. In this paper, the influence of alternative WBS orientations on project completion time is assessed. The Project Life Cycle (PLC) and technology (T) WBS were applied across two projects—the construction of a Liquefied Petroleum Gas (LPG) facility and the Renovation of an Office Complex (ROC)—using a top-down decomposition methodology. The PLC-WBS and T-WBS were created utilizing Figma software. The project duration was determined using the critical path method, which was implemented in the Python programming language. Based on WBS selection, differences were discovered in the definition of the project deliverables, network construction, and aggregation of work packages. These discrepancies had an impact on the technological relationships between activities by reducing opportunities for parallel processing. The LPG project was completed in 86 days using the PLC-WBS and in 80 days using the T-WBS orientation. For ROC, the project can be accomplished within 128 and 126 days, using the PLC-WBS and T-WBS orientation, respectively. This outcome suggested that there might be an association between the WBS and the project objective. Therefore, an assessment of different WBSs in project scope management demonstrated their potential influence on decision-making in activity planning and scheduling, network construction, and project objectives.

Keywords-*wbs-orientation; project completion time; project lifecycle-WBS; technology-WBS*

I. INTRODUCTION

Scope Management (SM) is a performance indicator and a significant component of the project knowledge area [1-2]. Performance metrics are essential for evaluating the effectiveness of decisions taken to achieve organisational goals [3]. SM is essential for obtaining a comprehensive understanding of the tasks, limitations, and anticipated results/deliverables of a project [4]. A lack of clarity in defining and analyzing the SM process will have a direct influence on the project's cost, duration, and quality [5]. Furthermore, a deficient description of the scope could result in multiple challenges occurring during the next phases of project development [6, 7]. This sort of situation is often referred to as scope creep. The occurrence of scope creep arising from unplanned and uncontrollable changes is evident in project completion time and cost overrun [8, 9]. Hence, to reduce the frequency of project failures due to scope creep and other factors, it is necessary to encourage transparent and consistent

communication between stakeholders and project team members [10, 11].

The SM process comprises the following steps: (i) planning to define, validate, and control the scope, (ii) planning for requirements and documentation, (iii) determining the scope and stakeholders' needs, (iv) developing a Work Breakdown Structure (WBS), (v) requesting approval from the project owner, and (vi) assessing the performance of the scope [1-2]. The intention of developing a WBS is to generate a formal statement of the project's deliverables. The WBS presents project deliverables in a structured manner, including planning information such as the actual work needed, the impact on resources and costs, and schedule details [12, 13]. A WBS is a hierarchical breakdown of the work that the project team must do to achieve the project objective and serves as a project taxonomy. While the WBS is derived from the project scope, different approaches (or orientations) used to construct the WBS can greatly impact the objective of the project [14].

A. WBS: Design, Orientation, and Decomposition

WBS is an essential document that establishes a direct connection between project objectives and their implementation through a continuous iterative process. In Project Management (PM), WBS is a multilevel system indicating logical connections among the work activities that constitute a project. The WBS breaks down the project into a hierarchical framework, enabling monitoring and control [15]. A comprehensive WBS supports the clear identification and description of project deliverables [16]. Therefore, the attainment of deliverables in a project can be achieved through the implementation of a strong WBS [17]. In the design of a WBS, measures such as man-hours, dollar value, and completion time can be introduced to determine the performance of the work packages [2, 18]. Work Packages (WPs) are deliverables that exist at the lowest feasible level of a WBS (i.e. work to be done to complete each task). An inadequately planned WBS can result to unclear job allocations, scope modifications, unsatisfactory deliverables, prolonged project timelines, etc.

Before deciding on the direction of the decomposition process, it is necessary to determine the WBS orientation or pattern. The design of a WBS orientation can be established by considering the following factors: (i) the product or deliverable, (ii) the stages of the Project Life Cycle (PLC), (iii) the technology involved, (iv) the geographical aspects, and (v) the underlying processes [19]. A Deliverable WBS (D-WBS) is a systematic grouping of components required for the manufacturing of a product or the successful completion of a project [2]. A PLC-WBS facilitates the procedure of converting a concept into a tangible outcome. The PLC-WBS can be planned or change-driven. If change-driven, the degree of change can be predictive, iterative, or incremental [20]. Furthermore, the utilization of the PLC strategy in designing the work breakdown structure facilitates the efficient tracking of significant project milestones. Nevertheless, its utilization is uncommon [21]. The Technology WBS (T-WBS) is preferred for projects with a high level of specialization, functional hierarchy, and discrete technology. Additionally, the T-WBS facilitates centralized project control in the presence of diverse technologies. For a Geography-based WBS (G-WBS), the project is divided into homogeneous subprojects with similar activities. For example, a project involving the construction of similar structures in three regions may require three project managers. Therefore, the culture, language, and legal system associated with the project's execution may have an impact on a G-WBS. The Process-WBS (P-WBS) involves breaking down a project into discrete steps, work activity phases, and functions, with each level having specific deliverables [17]. A decentralized system supports the use of a P-WBS.

Decomposition, also known as disintegration, is a planning technique in which the scope of a project is divided into manageable segments or elements [22]. This involves the division and further subdivision of the scope into smaller entities. Decomposition facilitates the analysis of a project at both the work package and organisational levels. A guideline in the decomposition process is the 100% rule. According to this rule, the subsequent breakdown of a WBS element, also known

as a child, must encompass the entire previous task, referred to as the parent [23]. In the literature, there are two decomposition approaches: bottom-up and top-down [2]. The top-down strategy is recommended for centralized decision-making and the division of a project into subprojects. In the bottom-up approach, the project framework is constructed starting from the work package level. A bottom-up strategy is preferred for projects with a high level of complexity [24]. Hence, the selection of a WBS technique is subjective and typically relies on factors such as the project manager's expertise, the project team's composition, the management approach, consideration of alternative methods, and the nature of the project.

B. Related Literature

The utilization of the WBS, a fundamental tool in project scope management, has been associated with the level of project success [19, 25]. Specifically, scheduling, network planning, resource allocation, and cost allocation require the WBS as a point of reference [26]. In [27], it was demonstrated that it is possible to use several WBS approaches to break project activities into tasks and work packages. In addition, several organisational structures can be utilized during project conceptualization and implementation. The design and integration of WBS have received much attention in the PM literature. Using neural networks, authors in [12] designed a WBS as a decision support system. The authors in [29] created a comprehensive decision support system that utilizes building information management and WBS. Authors in [17] integrated the cost breakdown structure and WBS to enhance decision-making in construction projects. For corporate events, a WBS was created in [29]. Authors in [30] examined the relationships between critical success factors and PM techniques in the construction sector of Malaysia. Their analysis revealed the limited use of PM tools and methods. Authors in [31] interviewed 40 professional staff members at a university planning department. They observed that the WBS is a statistically significant tool for determining the successful completion of a project based on its scope. Nevertheless, the literature on methods for comparing different work breakdown structures and the expected impact on project completion time is sparse.

C. Research Objective

In traditional and modern projects, the choice of a WBS orientation is different. For example, the type of PLC and organisational structure is different in construction and IT projects [32]. Invariably, a discrepancy between the WBS and organisational structure may delay or hinder the success of a project [33]. The WBS has not been given sufficient attention in the literature [27, 34], although, it is considered a fundamental and reliable instrument for planning and specifying work packages. Furthermore, information is sparse regarding the relationship between the selection of different WBS orientations and project completion time. The current research aimed to assess the influence of the PLC-WBS and T-WBS on project completion time.

II. METHODOLOGY

This section focuses on the assumptions made during project selection, the framework that connects the work

breakdown structure to the project objective, the process of selecting WBS, and the estimation of the project completion time.

A. Assumptions

The following assumptions guide the selection of projects for this research:

- The project scope remains constant throughout its entire duration.
- The cost of materials and resources remains constant throughout the project duration.

- The equipment will be fully functional during the project's implementation.
- The project's quality is assured.
- Organisation breakdown structure is assured.

B. Framework for Determining the Relationship between WBS Orientations and Project Completion Time

In Figure 1, the framework for assessing the influence of the selected WBS orientations on the project completion time is presented.

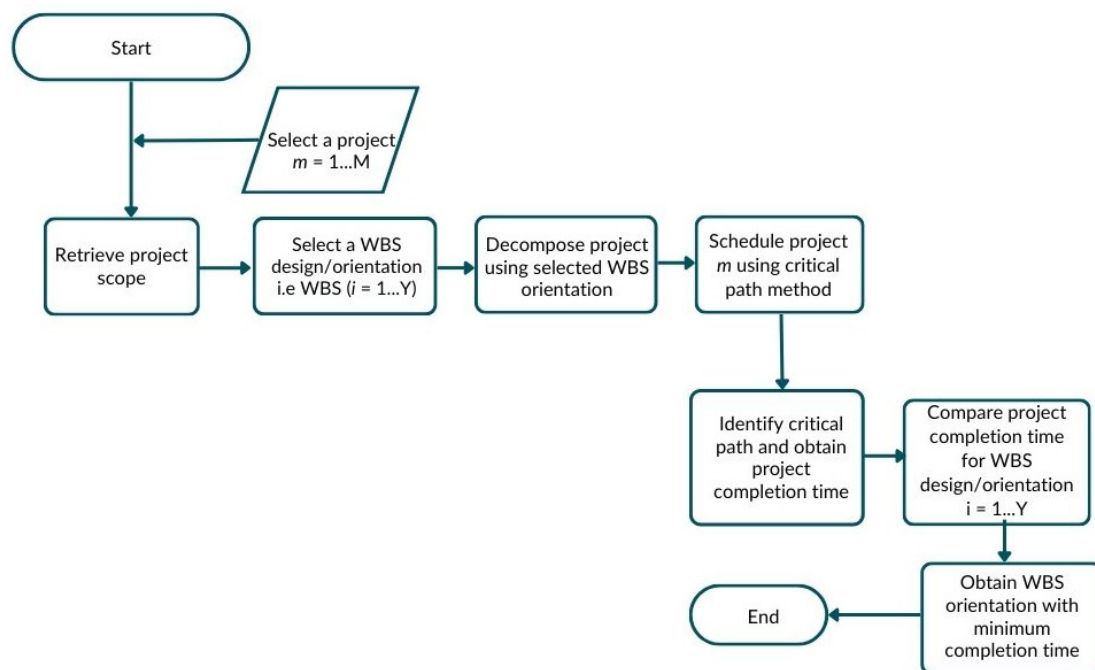


Fig. 1. Framework for assessing the influence of alternative WBSs on project completion time.

C. Project Selection

Data were gathered from two projects. The projects involve building a Liquefied Petroleum Gas (LPG) facility and the Reconstruction of an Office Complex (ROC). The LPG and ROC are denoted as $m = 1$ and $m = 2$, respectively. The aim is to minimize the time required to complete these projects.

D. Selection of WBS design and Orientation

The design and orientation of the WBS were based on the PLC and T pattern using the top-down decomposition approach. In the design of the WBS, the following frameworks were considered: (i) the 100% rule, (ii) work package naming and coding, and (iii) sizing. The relevant tasks in the project were identified and classified at the parent and child levels. Subsequently, PLC-WBS and T-WBS were designed. The design of the WBS was achieved using Figma software [35].

E. Determination of Project Completion Time

In a project, three fundamental objectives are common: (i) maximizing the use of limited resources, (ii) using allocated

resources within the specified time frame, and (iii) delivering a product or service within the agreed quality or standard requirements. Minimizing project completion time (or makespan) is the most researched objective in the literature. The development of the Critical Path Method (CPM) and Project Evaluation and Review Technique (PERT) advanced the use of network diagrams in project scheduling. The use of CPM assumes unlimited resource availability with time restrictions (i.e. a project has an imposed date to be completed). In this work, the project completion time was determined using the CPM methodology implemented in Python v3.10.9.

III. RESULTS

A. Project Decomposition

In Figure 2, the PLC-WBS is presented for the LPG project. The parent elements are start-up/design (1.1), procurement (1.2), execution (1.3), quality control (1.4), and finishing (1.5). In Figure 3, the PLC-WBS is presented for the ROC project. The parent elements are design (1.1), procurement (1.2),

equipment/procurement (1.3), start-up operations (1.4), and training and retraining (1.5). In Figure 4, the T-WBS is presented for the LPG project. The parent elements are plan/schedule (1.1), site preparation (1.2), construction (1.3), procurement (1.4), operations and maintenance (1.5), and

finishing (1.6). In Figure 5, the T-WBS is presented for the ROC project. The parent elements are planning and scheduling (1.1), construction (1.2), procurement (1.3), maintenance (1.4), operations (1.5), and human resources (1.6).

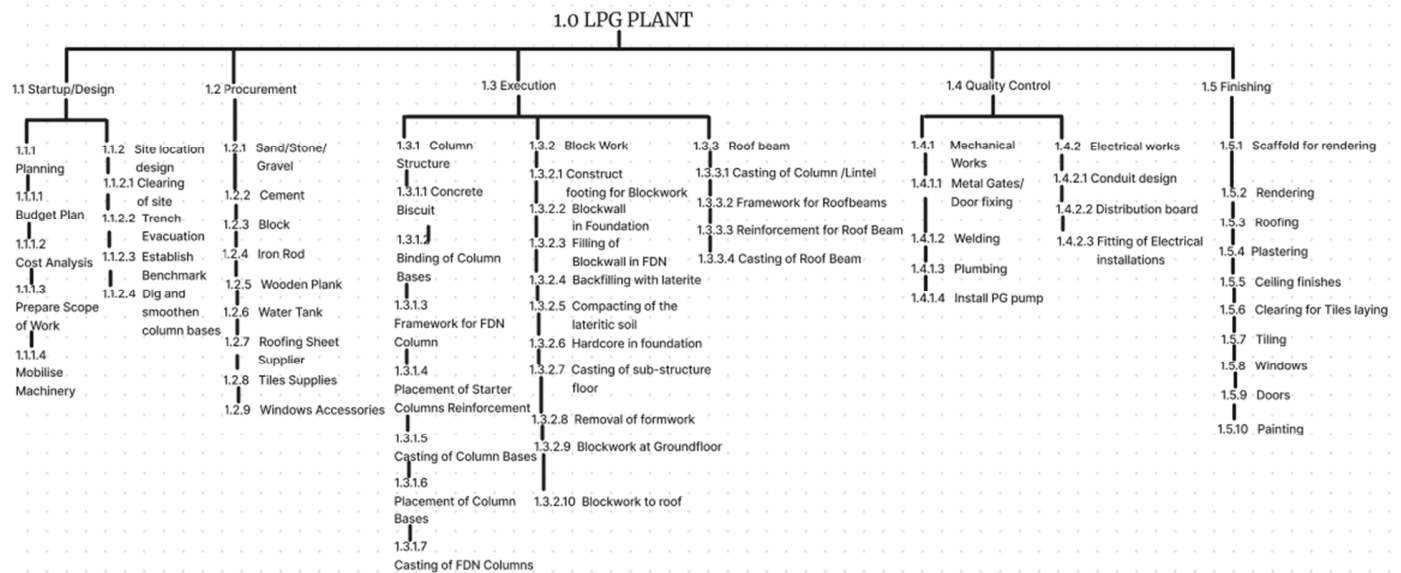


Fig. 2. PLC-WBS for LPG project.

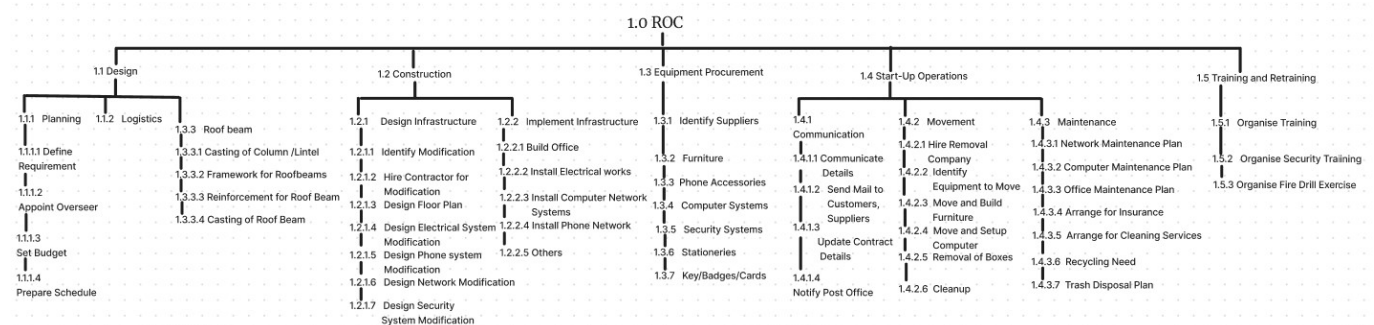


Fig. 3. PLC-WBS for the ROC project.

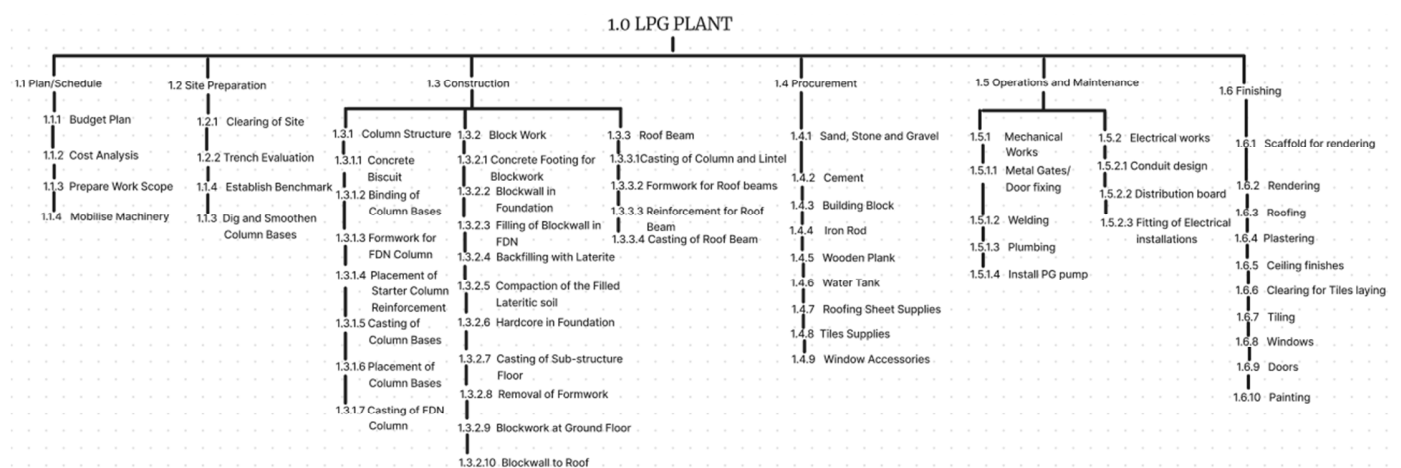


Fig. 4. T-WBS for LPG project.

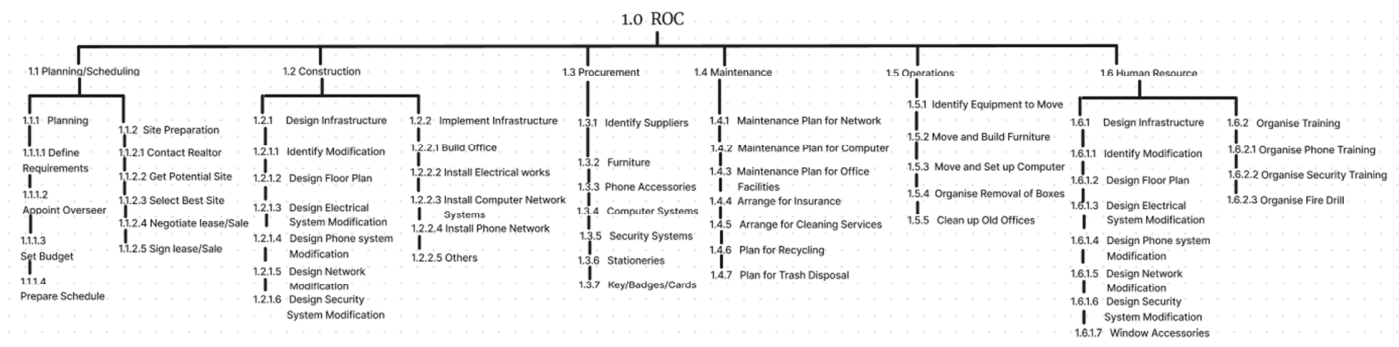


Fig. 5. T-WBS for the ROC project.

B. Project Completion Time

The impact of the two WBS orientations on project completion time (in days) is presented in Table I.

TABLE I. QUANTITIES IN DPS

m	Project Name	WBS orientation	
		PLC	T
1	LPG	87	80
2	ROC	128	126

The LPG project ($m = 1$) can be completed in 86 and 80 days using the PLC-WBS and T-WBS orientations, respectively. For the ROC ($m = 2$), the project can be completed in 128 and 126 days using the PLC-WBS and T-WBS orientations, respectively. The use of the T-WBS resulted in a 6-day and 2-day reduction in the project completion time for the LPG and ROC projects, respectively. The CPM implementation in Thonny 4.1.4 with Python version 3.10.11 is presented in Figure 6. The scheduling procedure utilized to derive the outcome presented in Table I was assumed to be deterministic, meaning that it involved planned activity duration. Clearly, the project completion time will vary when the actual duration of activities is utilized to calculate the project completion time. In practical scenarios, the total float acquired can be utilized to minimize time deviation.

```

Thonny - C:\Users\ababati\Documents\HP Pavilion DocPART A\Thonny\CPM LPG FINAL WORKING CODE 27022024.py @ 16:50
File Edit View Run Tools Help

CPM LPG FINAL WORKING CODE 27022024.py 18 CPM FINAL WORKING CODE 27022024.py 18
69 activity_41, activity_42,
70 activity_43, activity_44, activity_45, activity_46,
71 activity_47, activity_48,
72 activity_49, activity_50, activity_51, activity_52, activity_53,
73 activity_54, activity_55]
74
75 reverse_1 = activity_data[0:-1]
76
77 for activity in activity_data:
78     predecessors = activity["predecessors"]
79     if predecessors == []:
80         es = int(input("Enter the project start: "))
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
```

decision-making procedure illustrated in Figure 1 to evaluate the impact of WBS on project success by analyzing the time taken for completion.

Nevertheless, previous research claimed a relationship between the two factors, but the approach to support this assertion was sparse. In further work, the impact of multiple WBS orientations on the completion time of modern projects (e.g. research, new products, and software development) can be researched.

REFERENCES

- [1] *A Guide to the Project Management Body of Knowledge*, 7th ed. Newtown Square, PA, USA: Project Management Institute, 2021.
- [2] R. Mulcahy, *CAPM Exam Prep, Fourth Edition*, 4th ed. Minnetonka, MN, USA: RMC Publications, Inc., 2018.
- [3] B. O. Odedairo and N. Nwabukei, "Framework for Operational Performance Measurements in Small and Medium Scale Industries Using Discrete Event Simulation Approach," *Engineering, Technology & Applied Science Research*, vol. 8, no. 4, pp. 3103–3107, Aug. 2018, <https://doi.org/10.48084/etasr.2106>.
- [4] A. A. Fashina, F. F. Fakunle, and M. A. Omar, "A Study on the Effects of Construction Project Delays in Somaliland Construction Industry," *Journal of Management, Economics, and Industrial Organization*, vol. 4, no. 3, pp. 89–102, Sep. 2020, <https://doi.org/10.31039/jomeino.2020.4.3.6>.
- [5] P. R. Dumont, G. E. Gibson, and J. R. Fish, "Scope Management Using Project Definition Rating Index," *Journal of Management in Engineering*, vol. 13, no. 5, pp. 54–60, Sep. 1997, [https://doi.org/10.1061/\(ASCE\)0742-597X\(1997\)13:5\(54\)](https://doi.org/10.1061/(ASCE)0742-597X(1997)13:5(54)).
- [6] M. K. Fageha and A. A. Aibinu, "Managing Project Scope Definition to Improve Stakeholders' Participation and Enhance Project Outcome," *Procedia - Social and Behavioral Sciences*, vol. 74, pp. 154–164, Mar. 2013, <https://doi.org/10.1016/j.sbspro.2013.03.038>.
- [7] "What is project scope? Defining and outlining project success," *CIO*, <https://www.cio.com/article/193441/what-is-project-scope-defining-and-outlining-project-success.html>.
- [8] Q. H. S. Al-Rubaieci, F. A. A. Nifa, and S. Musa, "Project scope management through multiple perspectives: A critical review of concepts," *AIP Conference Proceedings*, vol. 2016, no. 1, Sep. 2018, Art. no. 020025, <https://doi.org/10.1063/1.5055427>.
- [9] "What Are Project Management Process Groups?," in *Effective Project Management: Traditional, Agile, Extreme, Hybrid, Eighth Edition*, New York, NY, USA: John Wiley & Sons, 2019, pp. 115–149.
- [10] B. O. Odedairo and V. Oladokun, "Relevance and Applicability of Multi-objective Resource Constrained Project Scheduling Problem: Review Article," *Engineering, Technology & Applied Science Research*, vol. 1, no. 6, pp. 144–150, Dec. 2011, <https://doi.org/10.48084/etasr.53>.
- [11] B. O. Odedairo and A. Alarjani, "A System Dynamics Approach to Feedback Processes in Project Scheduling," *Engineering, Technology & Applied Science Research*, vol. 14, no. 2, pp. 13201–13207, Apr. 2024, <https://doi.org/10.48084/etasr.6666>.
- [12] S. A. Hashemi Golpayegani and B. Emamizadeh, "Designing work breakdown structures using modular neural networks," *Decision Support Systems*, vol. 44, no. 1, pp. 202–222, Nov. 2007, <https://doi.org/10.1016/j.dss.2007.03.013>.
- [13] S. Kashyap, "Work Breakdown Structure: Key To Project Management Success," *ProofHub*, Mar. 13, 2024, <https://www.proofhub.com/articles/work-breakdown-structure>.
- [14] H. H. Bodicha, "How to Measure the Effect of Project Risk Management Process on the Success of Construction Projects: A Critical Literature Review," *The International Journal of Business & Management*, vol. 3, no. 12, pp. 99–112, Dec. 2015.
- [15] P. Landau, "16 Project Management Tools & Techniques for Project Managers," *ProjectManager*, Dec. 16, 2022, <https://www.projectmanager.com/blog/project-management-techniques-for-every-pm>.
- [16] D. M. Burghate, "Work Breakdown Structure: Simplifying Project Management," *International Journal of Commerce and Management Studies*, vol. 3, no. 2, pp. 1–5, 2018.
- [17] A. Cerezo-Narváez, A. Pastor-Fernández, M. Otero-Mateo, and P. Ballesteros-Pérez, "Integration of Cost and Work Breakdown Structures in the Management of Construction Projects," *Applied Sciences*, vol. 10, no. 4, Jan. 2020, Art. no. 1386, <https://doi.org/10.3390/app10041386>.
- [18] A. Cerezo-Narváez, A. Pastor-Fernández, M. Otero-Mateo, and P. Ballesteros-Pérez, "Integration of Cost and Work Breakdown Structures in the Management of Construction Projects," *Applied Sciences*, vol. 10, no. 4, Jan. 2020, Art. no. 1386, <https://doi.org/10.3390/app10041386>.
- [19] W. A. Z. Wan Abd Rahman, N. I. Mohd Zaki, and M. K. Abu Husain, "Work breakdown structure application for man-hours calculation in hull construction shipbuilding in Malaysia," *Cogent Engineering*, vol. 6, no. 1, Jan. 2019, Art. no. 1599524, <https://doi.org/10.1080/23311916.2019.1599524>.
- [20] R. Ahmed and M. Jawad, "Avoiding or disregarding: Exploring the relationship between scope creep, project complexity, and the success of construction projects," *Project Leadership and Society*, vol. 3, Dec. 2022, Art. no. 100064, <https://doi.org/10.1016/j.plas.2022.100064>.
- [21] B. O. Odedairo and I. O. Raji, "Quantitative approach to organisational design in project management office," in *Advancing Industrial Engineering in Nigeria through Teaching, Research and Innovation: A Festschrift in honour of Professor*, Ibadan, Nigeria: University of Ibadan, 2020, pp. 1–17.
- [22] B. Golany and A. Shtub, "Work Breakdown Structure," in *Handbook of Industrial Engineering*, New York, NY, USA: John Wiley & Sons, 2001, pp. 1263–1280.
- [23] W. A. Z. Wan Abd Rahman, N. I. Mohd Zaki, and M. K. Abu Husain, "A review of work breakdown structure and man-hours estimation method used in shipbuilding production," *International Journal of Mechanical Engineering and Technology*, vol. 10, no. 1, pp. 1141–1158, 2019.
- [24] "Work Breakdown Structure (WBS) in Project Management," *BigPicture*, <https://bigpicture.one/blog/create-work-breakdown-structure-project-management/>.
- [25] J.-H. Nam, J. H. Lee, and J. H. Woo, "Construction of standardised data structure for simulation of mid-term scheduling of shipbuilding process," *International Journal of Computer Integrated Manufacturing*, vol. 29, no. 4, pp. 424–437, Apr. 2016, <https://doi.org/10.1080/0951192X.2015.1066031>.
- [26] W. Halli, "Scope management through a WBS: key to success for the Logan Expansion Project," *PM Network*, vol. 7, no. 5, pp. 12–18, 1993.
- [27] A. Lester, *Project Management, Planning and Control*. Oxford, UK: Butterworth-Heinemann, 2013.
- [28] S. Globerson, "Impact of various work-breakdown structures on project conceptualization," *International Journal of Project Management*, vol. 12, no. 3, pp. 165–171, Aug. 1994, [https://doi.org/10.1016/0263-7863\(94\)90032-9](https://doi.org/10.1016/0263-7863(94)90032-9).
- [29] M. Al-Kasasbeh, O. Abudayyeh, and H. Liu, "An integrated decision support system for building asset management based on BIM and Work Breakdown Structure," *Journal of Building Engineering*, vol. 34, Feb. 2021, Art. no. 101959, <https://doi.org/10.1016/j.jobte.2020.101959>.
- [30] L.-E. Dounavi, E. Dermizakis, G. Chatzistelios, and K. Kiriopoulou, "Project Management for Corporate Events: A Set of Tools to Manage Risk and Increase Quality Outcomes," *Sustainability*, vol. 14, no. 4, Jan. 2022, Art. no. 2009, <https://doi.org/10.3390/su14042009>.
- [31] N. A. Haron, P. Devi, S. Hassim, A. H. Alias, M. M. Tahir, and A. N. Harun, "Project management practice and its effects on project success in Malaysian construction industry," *IOP Conference Series: Materials Science and Engineering*, vol. 291, no. 1, Sep. 2017, Art. no. 012008, <https://doi.org/10.1088/1757-899X/291/1/012008>.
- [32] R. O. Pelumi, O. I. Olateju, I. I. Oshin, and W. G. Odukaiye, "Work Breakdown Structure and Timely Delivery of Project. (A Study of Lagos State University).," *LASU Journal Of Transport & Logistics*, vol. 4, no. 1, pp. 1–10, 2022.
- [33] N. G. Hall, "Project management: Recent developments and research opportunities," *Journal of Systems Science and Systems Engineering*,

- vol. 21, no. 2, pp. 129–143, Jun. 2012, <https://doi.org/10.1007/s11518-012-5190-5>.
- [34] S. A. Brotherton, R. T. Fried, and E. S. Norman, "Applying Work Breakdown Structure Project Lifecycle," presented at the PMI® Global Congress 2008—North America, Denver, CO, USA.
- [35] T. A. Susiawan, Y. Latief, and L. S. Riantini, "Development of WBS (Work Breakdown Structure) risk based standard for safety planning at seaport project," *Journal of Physics: Conference Series*, vol. 1360, no. 1, Jul. 2019, Art. no. 012007, <https://doi.org/10.1088/1742-6596/1360/1/012007>.
- [36] "Learn Design & Design Basics," *Figma*. <https://www.figma.com/resource-library/design-basics/>.

An Approach to Determine and Categorize Mental Health Condition using Machine Learning and Deep Learning Models

B. H. Bhavani

JSS Academy of Technical Education, Bengaluru, Karnataka, India
bhavanibh@jssateb.ac.in (corresponding author)

N. C. Naveen

JSS Academy of Technical Education, Bengaluru, Karnataka, India
naveennc@jssateb.ac.in

Received: 29 February 2024 | Revised: 11 March 2024 | Accepted: 12 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7162>

ABSTRACT

The mental health of the human population, particularly in India during and after the COVID-19 pandemic is a major concern. All age groups have undergone mental stress during and after COVID-19, especially college students in urban areas and individuals belonging to the age group from 16 to 25. Early detection of mental stress among urban students will help in the resolution of major related issues that may hurt one's career. Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) have enabled the prediction of mental health status. Numerous studies have been conducted using various approaches, but there is still no agreement on how to predict mental symptoms across age groups. In the current study, proposed DL, Long Short-Term Memory (LSTM), and ML models, namely Support Vector Machine (SVM), ADA Boost, Random Forest (RF), K-Nearest Neighbor (K-NN), Logistic Regression (LR), and Multi-Layer Perceptron (MLP) are trained and tested on a real-world dataset. The DL LSTM model outperformed the conventional ML models with an accuracy of 100%.

Keywords-mental health; machine learning; health status detection; mental health data

I. INTRODUCTION

The COVID-19 pandemic had a negative impact on humanity in a variety of ways, including social, economic, physical, emotional, and psychological aspects. Specifically, elements related to changes in academic structures, tests, and a battle with limited resources can be directly associated with anxiety, tension, frustration, and depressive disorders during the COVID-19 lockdown period [1]. An ongoing influence on the mental health of students aged between 16 and 25 was observed. Students experienced significant psychological distress as a result of increased screen usage and frequent exposure to social media and COVID-19 news [2]. COVID-19 significantly influenced students' physical health, whereas the social distancing norms may have affected their psychological well-being [3]. Individuals, particularly students, express their thoughts and opinions on various social media platforms. The imposition of the lockdown to halt the pandemic's rapid spread caused psychological issues, such as despair and anxiety, in many students and also affected their life quality [4]. Younger people, particularly college students have considerably suffered in terms of their mental health as a consequence of numerous factors, involving their transitioning to new modes of instruction, loss of friend circles, cancellation of classes and

even entire semesters, financial hardships, shrinking job markets, failing relationships, etc. [5]. Many students experienced psychological issues as a result of the COVID-19 pandemic's spread, which impacted their academics as well as their overall personality [5]. Multiple factors, such as possible issues between students and teachers, lack of parental attention, bad eating habits, and lack of sleep can cause student stress, which is rather common among college students [6]. In July 2020, an online survey conducted by the Indian Psychiatry Society utilized the Depression Anxiety Stress Scale (DASS) 21 to gauge the mental health status of the Indian population amid the COVID-19 pandemic. The survey, distributed via WhatsApp through Survey Monkey, aimed to assess depression, anxiety, stress, and well-being among the public during the pandemic lockdown. Additionally, it included other psychological measures, namely the Patient Health Questionnaire-9, Generalized Anxiety Disorder-7, and Warwick-Edinburgh scale to comprehensively evaluate mental health [7, 8].

Authors in [9] investigated the effects of the COVID-19 pandemic and lockdown on the mental health of children and youth. The Short Self-Rating Questionnaire (SSRQ) was employed to evaluate the stress levels among individuals aged from 9 to 18 during this period. Utilizing an observational

approach with a descriptive cross-sectional design, the study conducted an online survey involving 369 schoolchildren. Stress levels were categorized as Low, Moderate, or Severe based on a scoring system and subsequent analysis. Authors in [10] compared the mental health symptoms and quality of life of frontline clinicians treating and not treating COVID-19 following the initial outbreak. They employed the Patient Health Questionnaire-9 (PHQ-9), Generalized Anxiety Disorder Scale-7 (GAD-7), Insomnia Severity Index (ISI), and World Health Organization Quality of Life Questionnaire-brief version (WHOQOL-BREF) to assess depression, anxiety, insomnia, and overall Quality of Life (QoL). Using a cross-sectional design with convenient sampling, the study was conducted between October 13 and 22, 2020, five months after China's first COVID-19 outbreak. Comparable groups were identified implementing Propensity Score Matching (PSM), whereas differences were analyzed deploying a Generalized Linear Model (GLM).

There is an upward trend in mental illness detection in the Natural Language Processing (NLP) research. Deep Learning (DL) methods receive more attention and outperform traditional Machine Learning (ML) methods [11]. Authors in [12] examined fear, depression, and anxiety symptoms among 324 college students in India during the COVID-19 pandemic. Utilizing a Google Forms questionnaire, the cross-sectional web-based survey included sociodemographic questions and psychometric scales. Results showed that 68.8% of students had a high fear of COVID-19, 28.7% experienced moderate to severe depression, and 51.5% had mild to severe anxiety. Additionally, the COVID-19 Fear Scale was moderately correlated with anxiety and depression scales. In [13], the COVID-19 impact on students, including changes in online learning, sleep patterns, and fitness routines, affecting weight, social life, and mental health, was highlighted. The findings emphasized the need for public officials to address COVID-19's adverse effects on learning experience.

Authors in [14] explored the psychosocial issues arising from academic stress in children and teenagers, and potential short and long-term psychological effects. A cross-sectional online survey of 4,342 Shanghai primary and secondary students from March 13 to 23, 2020, assessed psychological distress, life satisfaction, and the impact of home quarantine. Anxiety (24.9%), depression (19.7%), and stress (15.2%) were prevalent. Children experienced stress, missed growth opportunities, and lacked access to school meals [15]. A study in eastern India evaluated the impact of COVID-19 lockdowns on the mental well-being and behavior of children receiving psychiatric care. Conducted via telephone interviews from June 1 to July 8, 2020, the study involved 225 respondents using structured interviews based on established scales [16].

Authors in [17] examined the impact of stay-at-home orders on mental health, physical activity, screen time, and alcohol consumption among social work students. The results showed a significant increase in anxiety, depression, and screen time during the pandemic, highlighting the need for increased student support and implications for social work education [17]. In [18], the impact of the COVID-19 pandemic on university students in Malaysia was examined, focusing on

anxiety levels and influencing factors. A cross-sectional survey with 983 respondents employed Zung's self-rating anxiety questionnaire, revealing varying anxiety levels. Financial constraints, remote learning, and uncertainty about academics and careers were major stressors. In Northern New Jersey, a survey of 162 undergraduate students identified factors associated with mental health burdens during the pandemic using multivariable regression analysis [19]. Additionally, a study on 233 healthcare-related course students assessed lifestyle changes, mental health, and educational impacts [20].

Authors in [21] utilized EEG for dementia diagnosis and ML algorithms for predicting incidents in healthcare settings. An Artificial Neural Network model achieved 94.4% accuracy in classifying Lewy body dementia, while incident prediction models trained on 476 event reports exhibited a high accuracy of 93% on cross-validation [22]. Many researchers engaged ML and DL to accelerate COVID-19 detection, prevention, and treatment [23]. Authors in [24] aimed to detect dementia in MCI patients using EEG abnormalities. In [25], a DL model was employed to assess the mental state of social media users based on their posts. It accurately identified posts related to depression, anxiety, bipolar disorder, and BPD with high accuracy rates adopting Convolutional Neural Networks (CNNs). A system that detects depression through voice analysis using ensemble averaging of 50 1d-CNNs, fine-tuning model parameters through experimentation, and employing random sampling to enhance the training dataset was proposed in [26].

The main contributions of the current research article are:

- ML models were deployed to classify college students in the age group of 16 to 25 as stressed or not stressed.
- Accuracy and derived conclusions were compared across the ML models put into service.

Figure 1 displays the followed methodology.

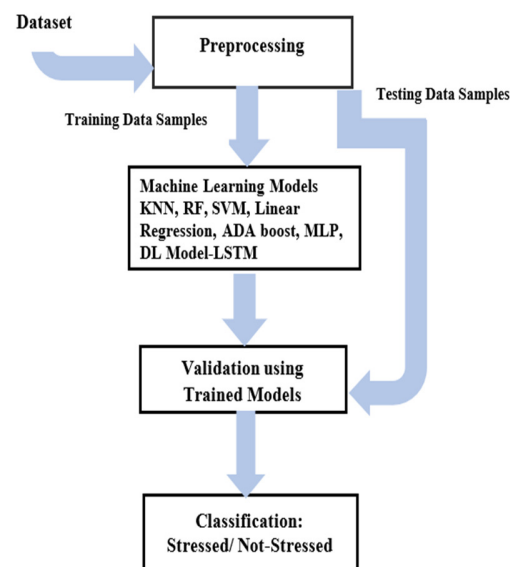


Fig. 1. The methodology used.

II. DATASET CREATION

We utilized our own real-world dataset, generated from answered questionnaires from student participants. The dataset consists of behavioral data with 10 attributes. The dataset was composed of the following fields:

- A1: Do you immediately respond when someone is calling your name?
- A2: Do you have difficulty in maintaining eye contact?
- A3: Do you respond positively when something is required from you?
- A4: Do you have common interests with parents/staff?
- A5: Have you pretended to study, work, or read from the laptop or mobile phone?
- A6: Do you show any interest in your peers company?
- A7: Do you comfort your peers when in difficulty?
- A8: Do you recall initial and earlier sessions?
- A9: Do you use simple gestures?
- A10: Do you stare blankly at nothing?

The total number of answered questionnaires was 1054, out of which 326 were considered as stressed and 728 as not stressed, based on their answers.

III. THE CONSIDERED MODELS

This article proposes models for predicting and detecting the mental health state using ML and DL models, as well as solutions to challenges encountered in previous attempts. Six different ML models and a DL model were considered and compared. Conclusions were drawn based on the prediction outcomes. Support Vector Machine (SVM), Random Forest (RF), AdaBoost, Logistic Regression (LR), k-Nearest Neighbor (KNN) algorithm, and Multi-Layer Perceptron (MLP) are the ML models used whereas LSTM is the DL model employed. The proposed ML and DL models, which aim to classify whether a student is stressed or not are described below.

A. Machine Learning Models

1) Support Vector Machine

SVM is an example of a supervised ML technology that can be used to solve problems of classification and regression. It is typically employed in classification projects.

2) Random Forest Algorithm

A group learning algorithm, RF utilizes a random sampling technique. RF can be deployed for classification, regression, and other purposes.

3) AdaBoost Algorithm

Adaptive Boosting (AdaBoost) is an approach to ensemble modeling that utilizes many relatively weak classifiers to create a single, robust one. A model is built using weak series models. AdaBoost was the first algorithm created to improve binary classification accuracy.

4) k-Nearest Neighbor Classifier

The KNN technique is a straightforward method for storing all known instances and categorizing new cases in accordance with a similarity metric (e.g. distance functions). Since the early 1970s, KNN's non-parametric technique has been implemented for statistical estimation and pattern recognition. KNN is a simple ML technique that may be used for both classification and regression.

5) Logistic Regression

LR, commonly known as the logit model or logit regression, is a statistical model employed to make educated guesses about the likelihood of an event occurring in the presence of known background information. LR relies on binary data, which can either be "true" (1) or "false" (0).

6) Multi-Layer Perceptron

An MLP is a feed-forward ANN that generates a set of outputs from a set of inputs. An MLP is characterized by several layers of input nodes connected as a directed graph between the input and the output layers.

B. The Long Shot Term Memory Deep Learning Model

The purpose of LSTM is to model and forecast data sequences. It is appropriate for tasks where the sequence of the data is important since it can capture temporal dependencies in the input data. Long-term information retention is possible in the memory cells found in LSTM networks. These memory cells consist of many gates that control the information flow in addition to a cell's state. LSTM is characterized by input, output, and forget gates. The forget gate establishes which data from the preceding time step ought to be ignored or forgotten. It creates a forget vector by multiplying the current input by the prior concealed state element by element. Which new data from the current time step should be added to the cell state is decided by the input gate. It creates an input vector by using the current input and the prior concealed state as inputs. The output gate is responsible for selecting which data from the cell state should be sent to the following concealed state. It creates an output vector, or the LSTM cell's output, by using the current input and the prior hidden state as inputs. The following equations fulfil the specific function assigned to each gate [27].

$$F_i = \sigma(W_f \cdot h_{i-1} + W_f \cdot x_t + b_f) \quad (1)$$

$$I_i = \sigma(W_i \cdot h_{t-1} + W_i \cdot x_t + b_i) \quad (2)$$

$$O_i = \tanh(W_o \cdot h_{i-1} + W_o \cdot x_t + b_o) \quad (3)$$

$$C_i = f_i * c_{t-1} + I_i \cdot \tanh(W_c \cdot [h_{i-1}, x_i] + b_c) \quad (4)$$

$$H_t = O_i * \tanh(c_t) \quad (5)$$

where F_i is the forget gate at timestep i , I_i is the input gate at timestep i , O_i is the output gate at timestep i , c_{t-1} is the state of the previous cell, h_{t-1} is the state of the previous hidden cell, C_i is the state of the current cell, H_t is the state of the current hidden cell, σ is the sigmoid function, \tanh is the hyperbolic tangent function, b_f, b_i, b_o are the biases at the forget, input, and output gates, and W_f, W_i, W_o the respective weights. The hidden layers of the proposed model had 100, 50, and 30 nodes.

C. Design of the Proposed Models

In the above-proposed model, the dataset has been pre-processed to remove redundant data and was split into training and testing sets with split ratios of 80:20 and 70:30. The training dataset has been fed into the ML models (KNN, RF, SVM, LR, AdaBoost, and MLP) and into the DL LSTM model followed by the testing that has been carried out using trained models with testing samples. The confusion matrices for the ML classifiers for the 70:30 split ratio are depicted in Figures 2-7.

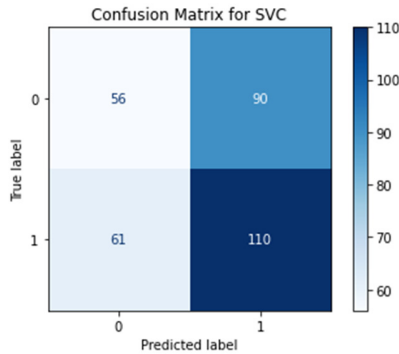


Fig. 2. Confusion matrix of the SVC model.

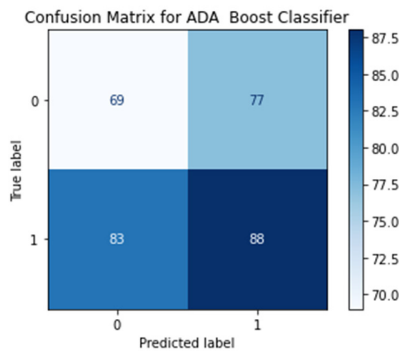


Fig. 3. Confusion matrix of the AdaBoost model.

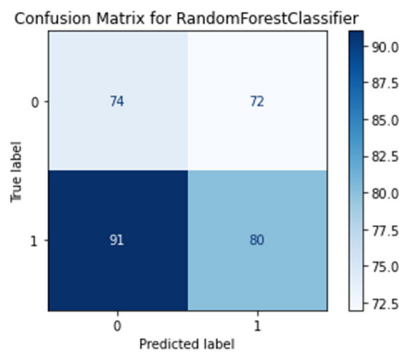


Fig. 4. Confusion matrix of the RF model.

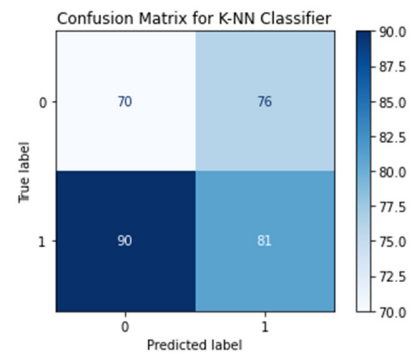


Fig. 5. Confusion matrix of the KNN model.



Fig. 6. Confusion matrix of the LR model.

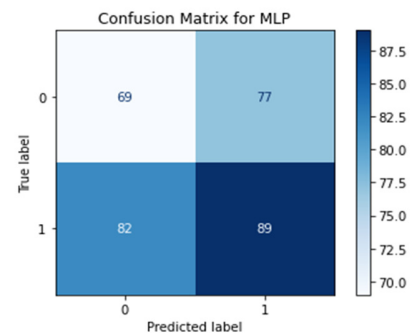


Fig. 7. Confusion matrix of the MLP model.

The training data set was used to train the seven models. The LSTM model had 100, 50, and 30 nodes in each hidden layer. The trained models were tested on the testing dataset. The pre-trained ML and DL model outputs are portrayed in Tables I and II. If the testing accuracy score is lower than the training accuracy score, the following calculations are performed:

Combined _Result =

$$\sum (testing_accuracy_score_i \times prediction_of_algo_i) \quad (6)$$

It is observed in Tables I and II that the best accuracy (100%) was obtained for LSTM. The DL LSTM model (with 500 epochs) outperformed the ML models in both considered split ratios. The ROC curves for the proposed models are in Figures 8-14. The accuracy of the proposed models is compared in Figure 15.

TABLE I. PERFORMANCE COMPARISON OF THE CONSIDERED MODELS WITH 80% TRAINING-20% TESTING DATASET SPLIT RATIO

Model Name	Precision	Recall	F1-score	Training accuracy	Testing accuracy
SVM/SVC	0.55	0.57	0.56	0.54	0.48
ADA Boost	0.42	0.48	0.45	0.69	0.49
RF	0.54	0.43	0.48	0.52	0.46
KNN	0.55	0.47	0.50	0.70	0.47
LR	0.55	0.57	0.56	0.55	0.49
MLP	0.58	0.52	0.55	0.93	0.52
DL model-LSTM	1.0	1.0	1.0	1.0	1.0

TABLE II. PERFORMANCE COMPARISON OF THE CONSIDERED MODELS WITH 70% TRAINING-30% TESTING DATASET SPLIT RATIO

Model Name	Precision	Recall	F1-score	Training accuracy	Testing accuracy
SVM/SVC	0.55	0.64	0.59	0.53	0.52
ADA Boost	0.45	0.47	0.46	0.71	0.50
RF	0.52	0.46	0.49	0.90	0.48
KNN	0.51	0.47	0.49	0.72	0.48
LR	0.55	0.67	0.61	0.54	0.43
MLP	0.56	0.55	0.56	0.98	0.52
DL model-LSTM	1.0	1.0	1.0	1.0	1.0

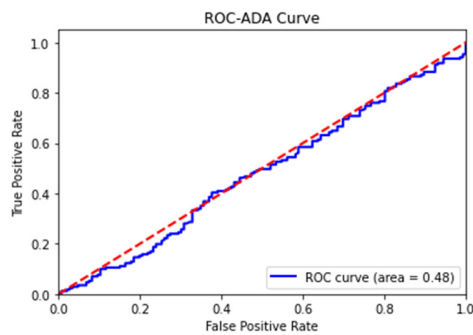


Fig. 8. ROC curve of the AdaBoost classifier.

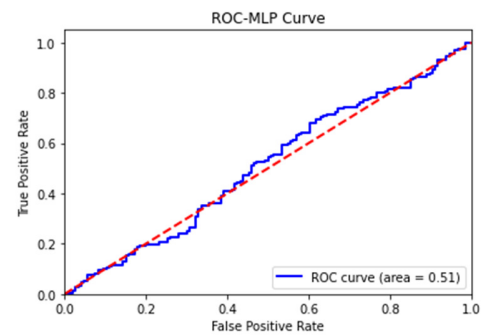


Fig. 11. ROC curve of the MLP.

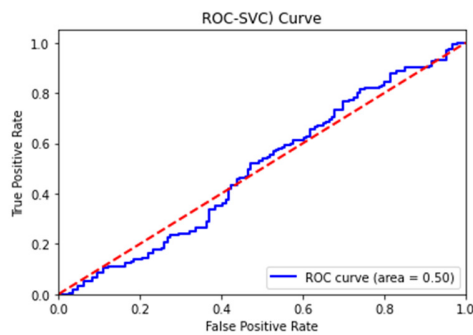


Fig. 9. ROC curve of the SCV classifier.

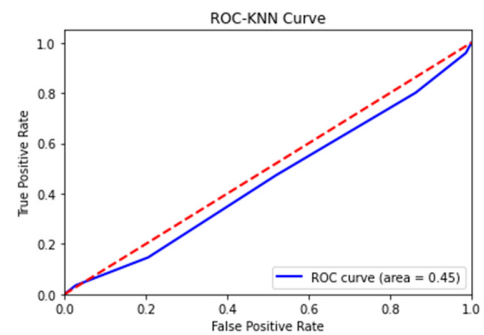


Fig. 12. ROC curve of the KNN classifier.

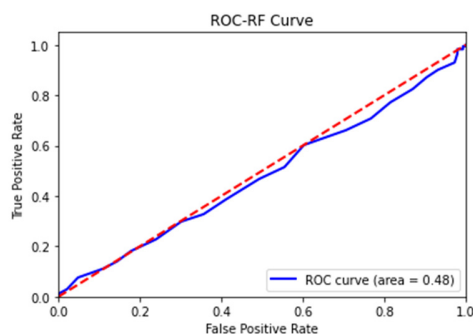


Fig. 10. ROC curve of the RF.

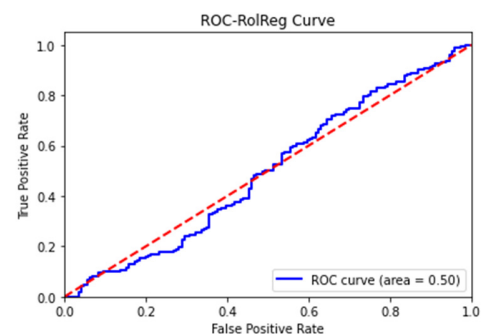


Fig. 13. ROC curve of the LR.

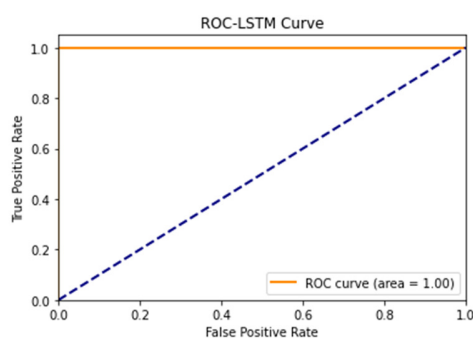


Fig. 14. ROC curve of the LSTM model.

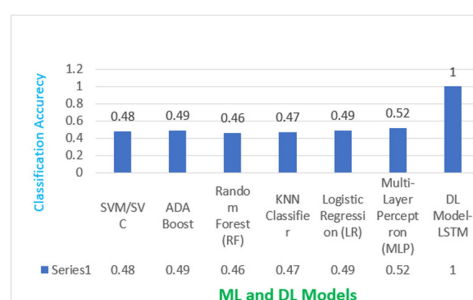


Fig. 15. Accuracy comparison.

IV. CONCLUSION

Considering the COVID-19 pandemic, this study suggests applying realistic Deep Learning (DL) and Machine Learning (ML) models with the purpose of identifying mental stress in students. Not only do these models aim to identify indicators of mental stress, but also to avert negative consequences. The specific models have the potential to save lives by greatly influencing the efficacy of therapies and treatments and quickly identifying severe cases of mental stress. The main goal was to create an efficient DL Long Short-Term Memory (LSTM) model. To accomplish this, a sophisticated architecture with hidden layers made up of 100, 50, and 30 nodes was used. This model was proven to perform better than other established ML models, such as Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbor (KNN), Logistic Regression (LR), AdaBoost, and Multi-Layer Perceptron (MLP), through extensive testing and evaluation. The study indicates that these models have a lot of potential in helping young people who are experiencing mental stress. It is anticipated that the results of the current study and the debate that follows will make a substantial contribution to the field and offer insightful information for further research. With the suggested hidden layer architecture, the deep learning LSTM model notably achieves a remarkable accuracy rate of 100%, demonstrating its potential to completely transform mental health interventions and diagnostics.

REFERENCES

- [1] K. Moghe, D. Kotecha, and M. Patil, "COVID-19 and Mental Health: A Study of its Impact on Students in Maharashtra, India." medRxiv, Feb. 05, 2021, <https://doi.org/10.1101/2020.08.05.20160499>.
- [2] P. Bhakat and K. Das, "Status of mental health among college and university students during first and second wave of COVID-19 outbreak in India: A cross-sectional study," *Journal of Affective Disorders*

Reports, vol. 12, Apr. 2023, Art. no. 100494, <https://doi.org/10.1016/j.jadr.2023.100494>.

- [3] S. Sengupta, S. Mugde, and G. Sharma, "An Exploration of Impact of COVID 19 on mental health -Analysis of tweets using Natural Language Processing techniques." medRxiv, Aug. 04, 2020, <https://doi.org/10.1101/2020.07.30.20165571>.
- [4] S. Deb, S. Kar, S. Deb, S. Biswas, A. A. Dar, and T. Mukherjee, "A Cross-Sectional Study on Mental Health of School Students during the COVID-19 Pandemic in India," *Data*, vol. 7, no. 7, Jul. 2022, Art. no. 99, <https://doi.org/10.3390/data7070099>.
- [5] R. Sharma, S. D. Pagadala, P. Bharti, S. Chellappan, T. Schmidt, and R. Goyal, "Assessing COVID-19 Impacts on College Students via Automated Processing of Free-form Text." arXiv, Dec. 16, 2020, <https://doi.org/10.48550/arXiv.2012.09369>.
- [6] H. Yasmin, S. Khalil, and R. Mazhar, "Covid 19: Stress Management among Students and its Impact on Their Effective Learning," *International Technology and Education Journal*, vol. 4, no. 2, pp. 65–74, 2020.
- [7] S. Ray, "Mental and Psychosocial Health: A Post-COVID Concern in India," *Neurology India*, vol. 70, no. 5, pp. 2116–2120, 2022, <https://doi.org/10.4103/0028-3886.359196>.
- [8] S. Grover *et al.*, "Psychological impact of COVID-19 lockdown: An online survey from India," *Indian Journal of Psychiatry*, vol. 62, no. 4, pp. 354–362, 2020, https://doi.org/10.4103/psychiatry.IndianJPsychiatry_427_20.
- [9] S. Ray, V. Goswami, and C. M. Kumar, "Stress-The hidden pandemic for school children and adolescents in India during COVID-19 era," *Current Psychology (New Brunswick, N.J.)*, pp. 1–10, Feb. 2022, <https://doi.org/10.1007/s12144-022-02827-3>.
- [10] Y.-J. Zhao *et al.*, "Post COVID-19 mental health symptoms and quality of life among COVID-19 frontline clinicians: a comparative study using propensity score matching approach," *Translational Psychiatry*, vol. 12, no. 1, pp. 1–7, Sep. 2022, <https://doi.org/10.1038/s41398-022-02089-4>.
- [11] T. Zhang, A. M. Schoene, S. Ji, and S. Ananiadou, "Natural language processing applied to mental illness detection: a narrative review," *npj Digital Medicine*, vol. 5, no. 1, pp. 1–13, Apr. 2022, <https://doi.org/10.1038/s41746-022-00589-7>.
- [12] A. P. Chaudhary, N. S. Sonar, J. Tr, M. Banerjee, and S. Yadav, "Impact of the COVID-19 Pandemic on the Mental Health of College Students in India: Cross-sectional Web-Based Study," *JMIRx med*, vol. 2, no. 3, 2021, Art. no. e28158, <https://doi.org/10.2196/28158>.
- [13] K. Chaturvedi, D. K. Vishwakarma, and N. Singh, "COVID-19 and its impact on education, social life and mental health of students: A survey," *Children and Youth Services Review*, vol. 121, Feb. 2021, Art. no. 105866, <https://doi.org/10.1016/j.childyouth.2020.105866>.
- [14] A. Mahapatra and P. Sharma, "Education in times of COVID-19 pandemic: Academic stress and its psychosocial impact on children and adolescents in India," *International Journal of Social Psychiatry*, vol. 67, no. 4, pp. 397–399, Jun. 2021, <https://doi.org/10.1177/0020764020961801>.
- [15] S. Tang, M. Xiang, T. Cheung, and Y.-T. Xiang, "Mental health and its correlates among children and adolescents during COVID-19 school closure: The importance of parent-child discussion," *Journal of Affective Disorders*, vol. 279, pp. 353–360, Jan. 2021, <https://doi.org/10.1016/j.jad.2020.10.016>.
- [16] S. Patra, B. K. Patro, and S. P. Acharya, "COVID-19 lockdown and school closure: Boon or bane for child mental health, results of a telephonic parent survey," *Asian Journal of Psychiatry*, vol. 54, Dec. 2020, Art. no. 102395, <https://doi.org/10.1016/j.ajp.2020.102395>.
- [17] S. A. Lawrence, J. Garcia, C. Stewart, and C. Rodriguez, "The mental and behavioral health impact of COVID-19 stay at home orders on social work students," *Social Work Education*, vol. 41, no. 4, pp. 707–721, May 2022, <https://doi.org/10.1080/02615479.2021.1883582>.
- [18] S. Sundarasan *et al.*, "Psychological Impact of COVID-19 and Lockdown among University Students in Malaysia: Implications and Policy Recommendations," *International Journal of Environmental Research and Public Health*, vol. 17, no. 17, Sep. 2020, Art. no. 6206, <https://doi.org/10.3390/ijerph17176206>.

- [19] A. Kecojevic, C. H. Basch, M. Sullivan, and N. K. Davi, "The impact of the COVID-19 epidemic on mental health of undergraduate students in New Jersey, cross-sectional study," *PLoS ONE*, vol. 15, no. 9, Sep. 2020, Art. no. e0239696, <https://doi.org/10.1371/journal.pone.0239696>.
- [20] N. Gadi, S. Saleh, J.-A. Johnson, and A. Trinidad, "The impact of the COVID-19 pandemic on the lifestyle and behaviours, mental health and education of students studying healthcare-related courses at a British university," *BMC Medical Education*, vol. 22, no. 1, Feb. 2022, Art. no. 115, <https://doi.org/10.1186/s12909-022-03179-z>.
- [21] G. Anuradha and D. N. Jamal, "Classification of Dementia in EEG with a Two-Layered Feed Forward Artificial Neural Network," *Engineering, Technology & Applied Science Research*, vol. 11, no. 3, pp. 7135–7139, Jun. 2021, <https://doi.org/10.48084/etasr.4112>.
- [22] K. Koklonis, M. Sarafidis, M. Vastardi, and D. Koutsouris, "Utilization of Machine Learning in Supporting Occupational Safety and Health Decisions in Hospital Workplace," *Engineering, Technology & Applied Science Research*, vol. 11, no. 3, pp. 7262–7272, Jun. 2021, <https://doi.org/10.48084/etasr.4205>.
- [23] S. A. A. Biabani and N. A. Tayyib, "A Review on the Use of Machine Learning Against the Covid-19 Pandemic," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 8039–8044, Feb. 2022, <https://doi.org/10.48084/etasr.4628>.
- [24] G. Anuradha, N. Jamal, and S. Rafiammal, "Detection of dementia in EEG signal using dominant frequency analysis," in *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, Chennai, India, Sep. 2017, pp. 710–714, <https://doi.org/10.1109/ICPCSI.2017.8391806>.
- [25] J. Kim, J. Lee, E. Park, and J. Han, "A deep learning model for detecting mental illness from user content on social media," *Scientific Reports*, vol. 10, no. 1, Jul. 2020, Art. no. 11846, <https://doi.org/10.1038/s41598-020-68764-y>.
- [26] A. Vázquez-Romero and A. Gallardo-Antolín, "Automatic Detection of Depression in Speech Using Ensemble Convolutional Neural Networks," *Entropy*, vol. 22, no. 6, Jun. 2020, Art. no. 688, <https://doi.org/10.3390/e22060688>.
- [27] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5929–5955, Sep. 2020, <https://doi.org/10.1007/s10462-020-09838-1>.

A Cybersecurity Awareness Model for the Protection of Saudi Students from Social Media Attacks

Gaseb Alotibi

Department of Computer and Information Technology, University of Tabuk, Saudi Arabia
galotaibi@ut.edu.sa (corresponding author)

Received: 22 February 2024 | Revised: 9 March 2024 | Accepted: 12 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7123>

ABSTRACT

Social engineering addresses a broad category of techniques aiming to persuade someone to reveal data or perform actions for criminal purposes, such as disclosing personal information about a particular target. Cybersecurity awareness is required to raise people's understanding of how these social engineering techniques are being used and so their capacity to exploit them. To accomplish this objective, primary focus is given to educating and training individuals on how to recognize such incidents and respond to them effectively. To protect people against social engineering threats, various cybersecurity models and approaches have been proposed. There are, however, a few differences between these models, since they are developed for specific purposes. Thus, the main objective of this study is to develop a cybersecurity awareness model specifically designed for Saudi students to protect them from social engineering attacks. The design science methodology was utilized in this study. The proposed model consists of four main stages: education and training, developing policies and guidelines, improving Saudi schools' security, as well as monitoring and evaluation. The model introduced can ensure the safety and privacy of students, teachers, and staff across different social platforms.

Keywords-cybersecurity; cybersecurity awareness; social media engineering; Saudi Arabia; design science

I. INTRODUCTION

Cybersecurity awareness refers in the enhancement of people's knowledge and comprehension of potential social engineering threats and risks [1]. As a form of attack in cybersecurity, social engineering involves exploiting human vulnerabilities by using influence, persuasion, deception, manipulation, and induction to breach cyberspace security. Cyberspace elements, such as infrastructure, data, resources, users, and operations must be secure in terms of confidentiality, integrity, availability, controllability, and auditability. The purpose of social engineering is to exploit human vulnerability to breach cyberspace security through social interactions [2, 3]. In the realm of cybersecurity, security awareness plays a crucial role in safeguarding individuals and organizations against potential security threats and risks. It empowers individuals to make informed decisions and take proactive measures to protect their digital assets. This paper discusses the concept of security awareness, its significance in the field of cybersecurity, and its impact on individuals and companies. Many organizations are automating their processes to offer their customers cheaper, faster, and easier ways to access their services due to the extensive developments in communication technology. During the last few years, the use of technology has increased exponentially, especially in the field of communications. Globally, over 7 billion people subscribe to mobile phones and more than 2 billion people use the Internet

[4]. Approximately 294 billion emails and 5 billion mobile messages are exchanged every day [4, 5]. Technology deployment in developing countries, as well as its continued spread in developed countries, is expected to increase these numbers.

Globally, organizations are increasingly adopting automation as a means of delivering more efficient services. As technology has become a part of everyday life, many of their daily tasks are accomplished through mobile devices. These tasks include shopping, banking, and entertainment. All these services are managed by using communication technologies for effective information exchange. While these technologies are becoming increasingly employed, crimes connected with them also increase. Cybercrime is the process in which computer networks are utilized for the purpose of committing illegal acts [6]. Because most businesses and organizations rely more and more on communication technologies, it is of utmost importance that the information being exchanged is secure and unauthorized use is prevented.

Saudi Arabia has experienced a recent significant increase in the implementation of communication technologies, the Internet, and mobile technologies. This country is one of the fastest-developing countries in the Middle East. More than 18 million Internet users exist in the country, which means that approximately 66% of the population has access to the Internet, which equals to more than 18 million people. Among all,

Facebook and X (Twitter), are the most deployed social media platforms [7]. Many people use the Internet to buy products online. The total amount of e-commerce sales in the country is about \$520 million, with 39% of Internet users purchasing products online [8]. In the Kingdom of Saudi Arabia, there has been a boom in the utilization of smartphones while the penetration of the Internet is relatively new. Due to the lack of understanding and information about security measures that can be taken, it can be assumed that there is a lack of knowledge regarding cybersecurity. Literature mostly consists of studies conducted to assess cybersecurity awareness in developed Western countries. The Kingdom of Saudi Arabia, however, has a vastly different culture, social attitude, language, government regulations, and understanding of security importance.

Therefore, in this study, the realization of the students in Saudi schools when it comes to security awareness is discussed along with the security measures these students take to increase their security. Then, a comprehensive cybersecurity awareness model is developed for Saudi schools to effectively address security concerns related to the use of social media, adopting the design science approach. The model consists of four main stages: education and training of Saudi students, developing policies and guidelines, improving Saudi school security, and monitoring and evaluation. By addressing these areas, the model aims to create a safe and secure learning environment for students in Saudi schools.

The main contribution of the development of a cybersecurity awareness model for Saudi students is the establishment of a tailored approach which addresses specific cybersecurity challenges faced by Saudi students on a daily basis. With the assistance of this model, students can navigate the digital world with confidence and security as they enhance their digital literacy, promote responsible behavior, and contribute to the government's efforts to ensure cybersecurity in a national level. The model was created implementing the design science approach. It is a systematic and rigorous approach that could be adopted when developing models and solutions based on the principles of modeling and analysis [9]. The model can evolve and respond to threats and changes in the educational landscape, making it both robust and adaptable.

A. Related Work

This section provides a review of the related work focusing on cybersecurity awareness in Saudi Arabia. Due to the rise of cyberattacks in the developed world, monitoring cybercrimes has attracted a great deal of scientific attention. Several recent approaches of assessing cybercrime knowledge either globally or focused on the Saudi Arabian cybercrime scene are discussed below.

Authors in [10] examined the Saudi national level of cybersecurity awareness through a quantitative, online study employing 629 participants (70% males and 30% females). It was found that although the participants had adequate IT knowledge, they did not have a high level of awareness regarding cybercrime, cybersecurity practices, and their role in ensuring the integrity of information on the Internet and the safety of computers and data. Authors in [11] determined

whether 116 employees working for governmental and private organizations in Riyadh were aware of phishing emails and how they should handle them. It was reported that they were mostly Saudi people who had not been employed in any companies working in the IT field. The survey indicated that the following data types were gathered: employee demographics, administrative specifics, employee understanding of the idea of email phishing, and employee awareness of the company's anti-phishing plan. The authors suggested that anti-phishing training programs should be put in place to raise awareness of the risks associated with phishing, given the low level of knowledge about this practice. Providing sufficient training to employees of an organization regarding email phishing is very important, as this is the most convenient approach to initiate these types of assaults.

Several studies have been conducted on the current risk of cybercrime in specific regions and the current level of awareness about these risks. In [12], 132 undergraduate students from Saudi Arabia's Alnamas district were surveyed regarding their knowledge of cybersecurity. According to the results, 69.6% of cybercrimes were committed via social media, with 57% of these crimes being sexual in nature, whereas 15% of the respondents had experienced cybercrime, and 80.7% wanted to receive training to improve their understanding regarding this matter. A survey of 633 Saudi public members was conducted in [13] in order to determine the level of information security knowledge. As part of the survey, respondents were asked about security awareness, password use, updates, data backups, and password management (changes and sharing). A total of 2325 people participated in the online survey administered in [14] that measured the level of information security awareness. The authors found that 35% of the participants were aware of general information security, 37% were aware of password security, 38% were aware of wireless network security, 40% were aware of social networking security, and 44% were aware of cloud storage security.

Measurements have been conducted on several aspects of awareness. Using three scenarios in terms of phishing email attacks, the authors in [15] attempted to determine the level of the awareness regarding phishing emails. The study employed embedded links, attachments, and social engineering to solicit sensitive information. Evaluations were made according to failure percentage, distribution by email types, and distribution by classes. The study found a failure rate of 80% for the embedded links and 40% for the attachments and requests of sensitive information. The authors reported that 38% of the embedded links, 50% of the attachments, and 46% of sensitive information failed to be delivered within the specified time frame for distribution by email. Finally, their study investigated several undergraduate students, from freshmen to seniors, from different levels of education. The authors reported that as many as 10% of all students had faced phishing attacks whereas as many as 70% of them students were victims.

By creating an online questionnaire and disseminating it among 161 individuals, authors in [16] explored whether conceptual or procedural knowledge had a positive effect on computer awareness. Positive results were found when both

conceptual and procedural knowledge were applied to prevent further phishing attacks. Using the Technology Acceptance Model (TAM) as a basis for Open-Source Cloud Computing (OSCC), authors in [17] studied a model based on TAM in the Iraqi environment in order to determine possible improvements of organization awareness. A total of 385 participants were included in the study, and 500 questions were asked during a period of 5 months during which the survey was conducted. According to the results, OSCC adoption relies upon mediation between perception and intention, between perception and goals, and between attitude and goals. Authors in [18], using a quantitative questionnaire divided into structured and unstructured questions, investigated the level of Information Security Awareness (ISA) in Greek Information Technology students. It was a two-month-long study with 87 participants. According to the results, a good level of awareness was associated with a good level of behavioral patterns, and the relationship between the level of awareness and behavioral patterns was studied. To examine whether cloud computing will be accepted by Malaysian university students, authors in [19] selected 45 students from Malaysian universities and scored them on at least 150 questions regarding readiness, perception, knowledge, and security awareness. The purposes of the study were to measure the level of security awareness and to obtain general knowledge about cloud computing. The findings confirmed the lack of awareness concerning the latter.

To investigate the cognitive use of cloud computing in the educational environment, authors in [20] used the Theory of Planned Behavior (TPB) in conjunction with the Theory of Knowledge Creation (TKC). A total of 240 people participated in that study, with a response rate of 91.95%. The study focused on attitudes and perceptions of cloud computing, privacy and security perceptions, behavioral control understanding, and final goals associated with the employment of cloud computing. The researchers deployed one-sample statistics, one-sample Kolmogorov-Smirnov tests [21] and linear interpolation to assess the experiment outcomes. In [22], the authors looked into various ways to spread information security knowledge within a business, such as interactive films, internal training sessions, screen savers, email, and social media, with the goal of enhancing end-user behavior and awareness in the context of phishing attempts.

Various studies have been reported that investigate and recognize cybercrime [22, 31-40], data cracks [23, 41-50], and other digital risks [24, 25, 41-58].

II. METHODOLOGY

A comprehensive cybersecurity awareness model for Saudi schools is developed in this study using the design science approach. Generally, design science is known as a systematic and rigorous approach. It aids in creating models and solutions that incorporate modeling and analysis [9, 60, 61]. It is a powerful research methodology that combines theory and practice to create and evaluate innovative artifacts or systems [62]. By associating problem definition, design, evaluation, and implementation, researchers can address real-world challenges and contribute to technological advancements. Therefore, the method adopted in this paper comprised the following five steps, as shown in Figure 1.

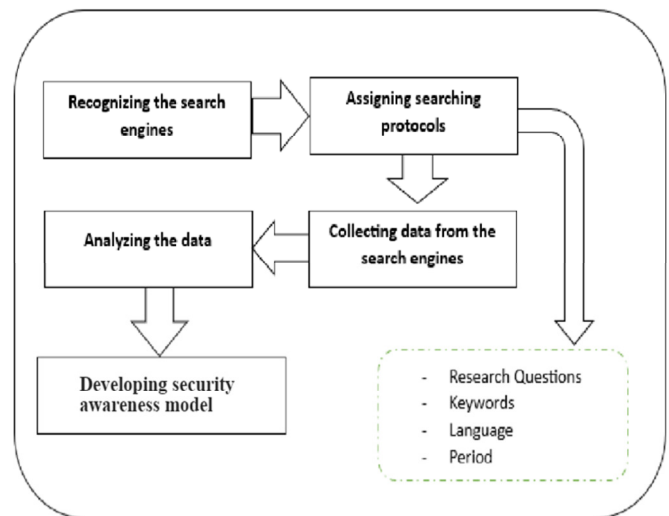


Fig. 1. Development process

1) Recognizing Search Engines

In this step, popular search engines, i.e., IEEE Xplore, Web of Science, Scopus, Springer, and Google Scholar were identified. Each engine offers unique features and sources that scholars can use to locate relevant academic literature and scholarly articles.

2) Assigning Search Protocols

Search protocols to identify articles relevant to our study were defined. The protocols that must be adhered to for the survey to be successful include questions, keywords, the language of the survey, and time constraints. To collect valuable information about topics related to cybersecurity, security awareness, and Saudi schools, the protocols presented below were followed. First, the following research questions were developed to ensure that the search will be focused and targeted:

- How can cybersecurity awareness be effectively promoted in Saudi schools?
- What are the benefits of developing a cybersecurity model for protecting Saudi school students against social engineering on social media?

3) Data Collection from the Search Engines

This step involved gathering data from the search engines based on the search protocols defined above. The data collected from the search engines are displayed in Table I. A total of 9 articles were extracted from Scopus, 8 from IEEE Xplore, 12 from Web of Science, 2,163 from Springer Link, and 1,200 from Google Scholar. Table I and Figure 2 exhibit the results of the search.

4) Analyzing Data

In this stage, an analysis was conducted on the 3392 papers that were retrieved from the search engines. This study was performed to make sure the data utilized for the research were correct and relevant, and to exclude any irrelevant sources. Books, articles, book chapters, and reports that were judged as

non-relevant to the research, were not considered. This rigorous selection procedure was carried out with the intention of excluding sources that would have biased findings and failed to advance the goals of the study. In this way, only relevant and reliable sources would be implemented. Using only the most relevant data enhances the credibility and validity of the research results. Consequently, 50 articles relating to Saudi Arabia's cybersecurity awareness are the final considered data set of this study.

TABLE I. PAPERS COLLECTED

Search engines	Keywords	Time	Totals
Scopus	"Cybersecurity"; "Security awareness"; "Saudi Arabia"	2010-2024	200
IEEE Explorer			250
Web of Science			199
Springer Link			2163
Google Scholar			1200

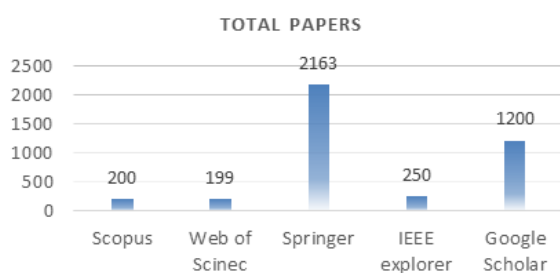


Fig. 2. Data collected from the search engines.

5) Developing a Security Awareness Model

This step involves developing a cybersecurity awareness model in relation to the use of social engineering on social media as a means of promoting cybersecurity knowledge in Saudi schools. After analyzing the 50 considered studies, 4 stages were identified to constitute the developed model (Figure 3):

a) Education and Training of Saudi Students

The purpose of this stage is to provide students, teachers, and other school staff with mandatory information and services to recognize potential risks associated with the usage of social media engineering. Three main tasks should be accomplished during this stage:

- **Adjusting Privacy Settings:** User privacy settings play a key role in safeguarding personal information. For students and teachers in Saudi Arabia to be protected against the misuses of social media platforms and online applications, privacy settings on these platforms and applications should be adjusted. Their understanding of what information is public and what information can be shared selectively with the public is critical. Furthermore, they should be informed about the consequences of sharing personal information, such as their telephone numbers and addresses, with third parties.
- **Safe Behavior:** Saudi children should be taught how to carefully surf on the Internet. Many precautions may be necessary to protect one's privacy, but among the most important measures and tips that people should take notice

are: not to share private information with anyone outside their network, to stay away from suspicious connections/links, and to be careful while linking with unfamiliar people on social media. Furthermore, individuals should make sure that their passwords are strong, and they should be fully aware of any phishing attempts.

- **Threat Identification and Response:** Students and teachers in Saudi Arabia should be educated about scams and how to avoid them. It is important to teach them the risks associated with disclosing private information to unfamiliar people, as well as how to identify unnecessary texts that demand personal or private details. In addition, helpful information should also include ways to recognize suspect behavior and potential risks on websites and to distinguish the real from the fake ones.

b) Developing Policies and Guidelines

During this stage, there will be policies and guidelines that can assist possible threat evaluation.

- **Establishing Acceptable Use Policies:** Students and teachers at Saudi schools are being taught, through this course, how to use social networking tools for their schoolwork. To ensure acceptable usage, clear guidelines must be established. These policies should cover a wide range of topics, including:
 - **Purpose:** By highlighting the learning value of the social networking tools, the policymakers should distinctly state the intention of those tools.
 - **Content:** It is recommended that the use of social media tools as well as the content that is generated through them be outlined in guidelines that define what types of content are acceptable and appropriate for utilization.
 - **Usage Limitations:** Appropriate guidelines and policies must be identified and established to identify the constraints that can be imposed on social networking tools to make them effective and efficient.
 - **Safety Measures:** Several factors contribute to the preservation of private information online, including the recording of any suspicious activities, highlighting the value of online protection, and ensuring that personal details are protected.
 - **Intellectual Property:** As guidelines and policies are created, they should emphasize the importance for customers of social media tools to see the value in protecting their intellectual rights.
 - **Consequences:** It is crucial for students to be aware that non-compliance can lead to disciplinary actions, loss of privileges, and expulsion.
- **Outlining Consequences for Misuse:** Creating applicable management procedures for social engineering tools is significant, as is identifying the abuses that will be imposed on those who harm them maliciously. Such consequences to further encourage liable tool use and deter improper action.

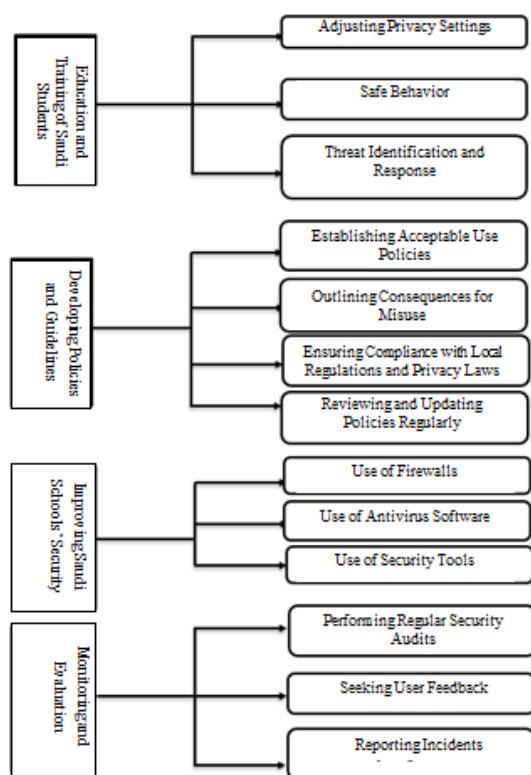


Fig. 3. Schematic of the developed cybersecurity awareness model.

- **Ensuring Compliance with Local Regulations and Privacy Laws:** Ensuring compliance with local rules and privacy laws is crucial when formulating guidelines and policies pertaining to the usage of social engineering media in Saudi schools. Schools are forced to follow the laws that regulate the former by creating school regulations. There are four tasks in this step:

- **Compliance Training:** This teacher and staff training program includes a thorough education on social engineering tools. It helps individuals comprehend pertinent privacy regulations and guidelines, as well as how to apply them.
- **Data Protection Measures:** Encrypting all data communicated when using social media is one of the many strong security methods that may be adopted to protect sensitive information, such as student grades, personal details, etc.
- **Regular Audits:** By routinely examining and evaluating security and privacy policies, we can make sure that they are in accordance with local privacy and security requirements. Schools should carry out these audits to make sure that all local laws, rules, policies, and guidelines are followed.
- **Data Breach Response Plan:** When schools expose sensitive data, unauthorized parties may gain access to sensitive information. Consequently, schools need to prepare a plan of action if there is a breach concerning the personal data of their students.

- **Reviewing and Updating Policies Regularly:** To ensure the effectiveness of the policies and guidelines, it is important to regularly review and update them. This includes considering any new security concerns or emerging threats that may require adjustments to the guidelines. Constant review and updating ensures that the social media use policies, remain relevant, and address the changing needs of Saudi schools.

c) Improving Saudi Schools' Security

Expanding security methods, precisely targeting social engineering channels in Saudi schools, is the third stage of upgrading school security in the country. To accomplish this, it is essential to implement technical tools and security measures to reassure that potential threats can be controlled. This stage consists of three security measures.

- **Use of Firewalls:** Any network, including those utilized in schools, needs a firewall to be secure. Schools can prevent potential problems by configuring firewall rules and protocols to monitor and block all kinds of unauthorized network traffic. Students can prevent unauthorized access to their computers and malicious activities, such as hacking attempts, if they take such measures.
- **Use of Antivirus Software:** Anti-malware software can be employed as a complementary feature to the antivirus software to protect computers from malicious software (malware). It is recommended that school computer systems be frequently updated with antivirus software so that potential threats cannot access them through social engineering and that potential threats are obstructed from emerging via those means. Additionally, antivirus software delivers real-time protection against known and future threats that are likely to appear soon.
- **Use of Security Tools:** School administrators should be aware that, beyond firewalls and antivirus software, some other security tools, such as intrusion detection systems and intrusion prevention systems are available to make the school campuses safer. It is possible to detect and mitigate potential security breaches by implementing these tools, which will improve their overall security.

d) Monitoring and Evaluation

Ensuring the successful implementation of security measures and periodically evaluating the efficacy of the cybersecurity awareness model are the main objectives of this phase. There are three primary duties involved:

- **Performing Regular Security Audits:** Conducting routine security audits is essential to monitoring and assessment. A thorough examination of several factors, including social engineering technologies, is necessary to carry out security audits of schools in an efficient manner.
- **Seeking User Feedback:** If educators actively seek input from students and staff, they may be able to provide them with insightful comments about how their security awareness model is being implemented.

- **Reporting Incidents:** Schools should create strong incident reporting mechanisms as soon as they become aware of social media engineering and put them timely into place.

III. FINDINGS AND DISCUSSION

This part addresses the study's findings, the efficacy of the created cybersecurity awareness model, and the resolution of the research issues. The preceding segment elucidated the principal phases of the cybersecurity awareness framework designed to safeguard Saudi schoolchildren against social engineering via social media. The model gives students the tools they need to navigate the digital world safely and responsibly by teaching them about privacy settings, cybersecurity best practices, risk awareness, and responsible digital citizenship. This program helps to create a more secure and safe online learning environment for Saudi Arabian students. Based on the study's findings, the established research questions are addressed below.

A. How can Sybersecurity Awareness be Effectively Promoted in Saudi Schools?

Security awareness is essential to ensuring the wellbeing and safety of Saudi school employees, teachers, and students. Here are some strategies Saudi schools can deploy to successfully raise students' awareness of security issues:

- **Comprehensive Security Policies:** To encourage security awareness and foster a culture of safety in schools, complete security policies must be developed, put into place, and strictly enforced. These policies should involve a wide range of security topics, including physical security measures, cyber security, emergency planning, and access control.
- **Security Training Programs:** Teachers and staff members will be better equipped to recognize possible threats and learn how to respond to them in a timely, effective, and professional manner if regular security training programs are implemented in schools and universities. These seminars ought to include several subjects, such as recognizing suspicious activities, reporting events, and adhering to emergency procedures.
- **Awareness Campaigns:** Launching awareness programs that can attract a lot of interest and attention is one way to promote security awareness in Saudi schools. There are numerous campaign formats available to emphasize the value of security precautions and inform pupils of possible hazards. Posters, pamphlets, and digital media content are examples of these campaigns' collateral.
- **Engaging External Experts:** Some students can interact with external experts by attending seminars or presentations, in addition to receiving insightful advice and useful knowledge from them. Examples of these experts include cybersecurity specialists and law enforcement officials. With these experts' contributions, the learning process will be improved and made more pertinent as they can offer examples and best practices that come from actual circumstances.

- **Parent Involvement:** To promote the efforts being made in this area, educators can encourage parents of students enrolled in Saudi schools to take part in security awareness seminars. Schools can utilize a variety of strategies, such as hosting parent workshops or providing them with instructional materials, to make sure that parents are aware of the security measures in place and can reinforce them at home as well.
- **Technology Integration:** Technology can significantly enhance school security awareness if it is used on a regular basis as part of a regular school security program. To protect an organization from security breaches, one should install surveillance cameras, implement access control systems, and employ cybersecurity software.

B. Advantages of the developed Cybersecurity Model

The advantages of the developed cybersecurity model for Saudi school students are:

- Improved online safety
- Awareness raising
- Enhanced digital literacy
- Developed critical thinking skills
- Established ethical online behaviors
- Preventing measures against cybercrime

A comparative analysis between the developed model and the existing models can be seen in Table II.

TABLE II. COMPARATIVE ANALYSIS BETWEEN THE DEVELOPED AND EXISTING MODELS

Existing models	Proposed Model			
	Educating and training of Saudi students	Developing policies and guidelines	Improving Saudi schools' security	Monitoring and evaluating
[10]	x	√	√	x
[11]	x	√	√	x
[12]	x	√	√	√
[13]	x	√	√	√
[14]	x	x	x	x
[15]	x	x	x	√
[16]	x	√	x	x
[17]	x	√	x	√
[18]	x	x	x	x
[19]	x	x	x	x
[20]	x	√	x	√
[21]	x	x	x	x
[22]	x	x	x	√

The developed cybersecurity awareness model is complete and covers a wider range of awareness security phases than the existing model. For example, authors in [10, 11] covered developing policies and guidelines, and the issue of improving Saudi schools' security, whereas authors in [12, 13] covered the following three phases: developing policies and guidelines, improving Saudi schools' security, and the monitoring and evaluation. The education and training of Saudi students were not covered by the existing models.

IV. CONCLUSION

A security awareness program enhances understanding and knowledge of potential security threats and risks, as well as the ability to mitigate them. One of its primary objectives is to educate and train individuals so that they can recognize and respond effectively to various cybersecurity threats that may occur, including phishing attacks, malware infections, and social engineering attempts that aim to manipulate users. Different approaches and models have been proposed to address the issue of security awareness in different fields. The purpose of this study was to develop a comprehensive cybersecurity awareness model for Saudi school students so that they can effectively address security concerns surrounding the use of social media. However, several security challenges need to be addressed because of the rapid growth of social media engineering. There is a need to guarantee the safety of students and educators. This study applied the design science approach to create a comprehensive cybersecurity awareness model. The developed model incorporated four stages: education and training of Saudi students, developing policies and guidelines, improving Saudi schools' security, and monitoring and evaluation. Several tasks and activities associated with each stage were defined. The developed model for Saudi school students will reassure the safety and privacy of students, teachers, and staff as well as it will promote responsible and secure online behavior. As far as is known, this one is the first study on the topic directly focused on Saudi Arabia. The implementation of the developed cybersecurity awareness model in Saudi schools is the aim of future work in order to evaluate its effectiveness.

REFERENCES

- [1] A. Parsaei, "Awareness and Social Engineering-Based Cyberattacks," *International Journal of Reliability, Risk and Safety: Theory and Application*, vol. 7, no. 1, pp. 31–36, Feb. 2024, <https://doi.org/10.22034/IJRRS.2024.7.1.4>.
- [2] Z. Wang, H. Zhu, P. Liu, and L. Sun, "Social engineering in cybersecurity: a domain ontology and knowledge graph application examples," *Cybersecurity*, vol. 4, no. 1, Aug. 2021, Art. no. 31, <https://doi.org/10.1186/s42400-021-00094-6>.
- [3] A. Alshammari, "A Novel Security Framework to Mitigate and Avoid Unexpected Security Threats in Saudi Arabia," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11445–11450, Aug. 2023, <https://doi.org/10.48084/etasr.6091>.
- [4] N. Sandjojo, M. Zuhriyanto, and I. W. W. Pradnyana, "The Effects of Fear of Cybercrime and Information Systems Security Policy on National Vigilance," in *International Conference on Informatics, Multimedia, Cyber and Information System*, Jakarta, Indonesia, Nov. 2020, pp. 195–200, <https://doi.org/10.1109/ICIMCIS51567.2020.9354283>.
- [5] A. Cetrulo, A. Sbardella, and M. E. Virgillito, "Vanishing social classes? Facts and figures of the Italian labour market," *Journal of Evolutionary Economics*, vol. 33, no. 1, pp. 97–148, Jan. 2023, <https://doi.org/10.1007/s00191-022-00793-4>.
- [6] J. Liu, Y. Xiao, S. Li, W. Liang, and C. L. P. Chen, "Cyber Security and Privacy Issues in Smart Grids," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 4, pp. 981–997, 2012, <https://doi.org/10.1109/SURV.2011.122111.00145>.
- [7] M. Bardus, A. Keriabian, M. Elbejjani, and S. Al-Hajj, "Assessing eHealth literacy among internet users in Lebanon: A cross-sectional study," *Digital Health*, vol. 8, Jan. 2022, Art. no. 20552076221119336, <https://doi.org/10.1177/20552076221119336>.
- [8] A. S. Alqahtani, "Factors Influencing the Adoption of E-commerce in Saudi Arabia: Study of Online Shopping," Ph.D. dissertation, Flinders University, Adelaide, South Australia, 2016.
- [9] A. Al-Dhaqm, S. Razak, and S. H. Othman, "Model Derivation System to Manage Database Forensic Investigation Domain Knowledge," in *IEEE Conference on Application, Information and Network Security*, Langkawi, Malaysia, Nov. 2018, pp. 75–80, <https://doi.org/10.1109/AINS.2018.8631468>.
- [10] F. Alotaibi, S. Furnell, I. Stengel, and M. Papadaki, "A survey of cybersecurity awareness in Saudi Arabia," in *11th International Conference for Internet Technology and Secured Transactions*, Barcelona, Spain, Dec. 2016, pp. 154–158, <https://doi.org/10.1109/ICITST.2016.7856687>.
- [11] N. Innab, H. Al-Rashoud, R. Al-Mahawes, and W. Al-Shehri, "Evaluation of the Effective Anti-Phishing Awareness and Training in Governmental and Private Organizations in Riyadh," in *21st Saudi Computer Society National Computer Conference*, Riyadh, Saudi Arabia, Apr. 2018, pp. 1–5, <https://doi.org/10.1109/NCG.2018.8593144>.
- [12] E. I. M. Zayid and N. A. A. Farah, "A study on cybercrime awareness test in Saudi Arabia - Alnamas region," in *2nd International Conference on Anti-Cyber Crimes*, Abha, Saudi Arabia, Mar. 2017, pp. 199–202, <https://doi.org/10.1109/Anti-Cybercrime.2017.7905290>.
- [13] A. Alarifi, H. Tootell, and P. Hyland, "A study of information security awareness and practices in Saudi Arabia," in *International Conference on Communications and Information Technology*, Hammamet, Tunisia, Jun. 2012, pp. 6–12, <https://doi.org/10.1109/ICCITTechnol.2012.6285845>.
- [14] A. Alzahrani and K. Alomar, "Information Security Issues and Threats in Saudi Arabia: A Research Survey," *International Journal of Computer Science Issues*, vol. 13, no. 6, pp. 129–135, Nov. 2016, <https://doi.org/10.20943/01201606.129135>.
- [15] R. C. Dodge, C. Carver, and A. J. Ferguson, "Phishing for user security awareness," *Computers & Security*, vol. 26, no. 1, pp. 73–80, Feb. 2007, <https://doi.org/10.1016/j.cose.2006.10.009>.
- [16] N. A. G. Arachchilage and S. Love, "Security awareness of computer users: A phishing threat avoidance perspective," *Computers in Human Behavior*, vol. 38, pp. 304–312, Sep. 2014, <https://doi.org/10.1016/j.chb.2014.05.046>.
- [17] H. A. Albaroodi, M. Abomaali, and S. Manickam, "Iraqi's Organizations Awareness to Prompt Open Source Cloud Computing (OSCC) in Their Service: A Study," in *International Conference on Advances in Cyber Security*, Penang, Malaysia, Dec. 2020, pp. 305–319, https://doi.org/10.1007/978-981-15-2693-0_22.
- [18] A. P. Filippidis, C. S. Hilas, G. Filippidis, and A. Politis, "Information security awareness of greek higher education students — Preliminary findings," in *7th International Conference on Modern Circuits and Systems Technologies*, Thessaloniki, Greece, Dec. 2018, pp. 1–4, <https://doi.org/10.1109/MOCASST.2018.8376578>.
- [19] S. S. Md Kassim, M. Salleh, and A. Zainal, "Cloud Computing: A General User's Perception and Security Awareness in Malaysian Polytechnic," in *Pattern Analysis, Intelligent Security and the Internet of Things*, A. Abraham, A. K. Mada, and Y.-H. Choo, Eds. New York, NY, USA: Springer, 2015, pp. 131–140.
- [20] Z. Asadi, M. Abdekhoda, and H. Nadrian, "Cloud computing services adoption among higher education faculties: development of a standardized questionnaire," *Education and Information Technologies*, vol. 25, no. 1, pp. 175–191, Jan. 2020, <https://doi.org/10.1007/s10639-019-09932-0>.
- [21] F. J. Massey Jr., "The Kolmogorov-Smirnov Test for Goodness of Fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, Mar. 1951, <https://doi.org/10.1080/01621459.1951.10500769>.
- [22] J. Abawajy, "User preference of cyber security awareness delivery methods," *Behaviour & Information Technology*, vol. 33, no. 3, pp. 237–248, Mar. 2014, <https://doi.org/10.1080/0144929X.2012.708787>.
- [23] A. M. R. Al-Dhaqm, S. H. Othman, S. Abd Razak, and A. Ngadi, "Towards adapting metamodeling technique for database forensics investigation domain," in *International Symposium on Biometrics and Security Technologies*, Kuala Lumpur, Malaysia, Aug. 2014, pp. 322–327, <https://doi.org/10.1109/ISBAST.2014.7013142>.

- [24] A. Al-Dhaqm, S. Razak, R. A. Ikuesan, V. R. Kebande, and S. Hajar Othman, "Face Validation of Database Forensic Investigation Metamodel," *Infrastructures*, vol. 6, no. 2, Feb. 2021, Art. no. 13, <https://doi.org/10.3390/infrastructures6020013>.
- [25] S. Abd Razak, N. H. Mohd Nazari, and A. Al-Dhaqm, "Data Anonymization Using Pseudonym System to Preserve Data Privacy," *IEEE Access*, vol. 8, pp. 43256–43264, 2020, <https://doi.org/10.1109/ACCESS.2020.2977117>.
- [26] A. Aldhaqm, S. A. Razak, S. H. Othman, A. Ali, and A. Ngadi, "Conceptual Investigation Process Model for Managing Database Forensic Investigation Knowledge," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 12, no. 4, pp. 386–394, Feb. 2016, <https://doi.org/10.19026/rjaset.12.2377>.
- [27] M. Ngadi, R. Al-Dhaqm, and A. Mohammed, "Detection and prevention of malicious activities on RDBMS relational database management systems," *International Journal of Scientific & Engineering Research*, vol. 3, no. 9, pp. 1–10, Oct. 2012.
- [28] A. Ali, S. A. Razak, S. H. Othman, and A. Mohammed, "Extraction of Common Concepts for the Mobile Forensics Domain," in *International Conference of Reliable Information and Communication Technology*, Johor Bahru, Malaysia, Apr. 2017, pp. 141–154, https://doi.org/10.1007/978-3-319-59427-9_16.
- [29] A. Ali, S. Razak, S. Othman, and M. Arafat, "Towards Adapting Metamodeling approach for the Mobile Forensics Investigation Domain," in *International Conference on Innovation in Science and Technology*, Kuala Lumpur, Malaysia, Apr. 2015, pp. 364–367.
- [30] M. A. Saleh, S. Hajar Othman, A. Al-Dhaqm, and M. A. Al-Khasawneh, "Common Investigation Process Model for Internet of Things Forensics," in *2nd International Conference on Smart Computing and Electronic Enterprise*, Cameron Highlands, Malaysia, Jun. 2021, pp. 84–89, <https://doi.org/10.1109/ICSCEE50312.2021.9498045>.
- [31] B. Zawali, R. A. Ikuesan, V. R. Kebande, S. Furnell, and A. A-Dhaqm, "Realising a Push Button Modality for Video-Based Forensics," *Infrastructures*, vol. 6, no. 4, Apr. 2021, Art. no. 54, <https://doi.org/10.3390/infrastructures6040054>.
- [32] A. Al-Dhaqm *et al.*, "Digital Forensics Subdomains: The State of the Art and Future Directions," *IEEE Access*, vol. 9, pp. 152476–152502, 2021, <https://doi.org/10.1109/ACCESS.2021.3124262>.
- [33] A. Aldhaqm, S. A. Razak, and S. H. Othman, "Common Investigation Process Model for Database Forensic Investigation Discipline," in *International Conference on Innovation in Science and Technology*, Kuala Lumpur, Malaysia, Apr. 2015, pp. 297–300.
- [34] F. M. Alotaibi, A. Al-Dhaqm, and Y. D. Al-Otaibi, "A Novel Forensic Readiness Framework Applicable to the Drone Forensics Field," *Computational Intelligence and Neuroscience*, vol. 2022, 2022, Art. no. 8002963, <https://doi.org/10.1155/2022/8002963>.
- [35] F. M. Ghabban, I. M. Alfadli, O. Ameerbakhsh, A. N. AbuAli, A. Al-Dhaqm, and M. A. Al-Khasawneh, "Comparative Analysis of Network Forensic Tools and Network Forensics Processes," in *2nd International Conference on Smart Computing and Electronic Enterprise*, Cameron Highlands, Malaysia, Jun. 2021, pp. 78–83, <https://doi.org/10.1109/ICSCEE50312.2021.9498226>.
- [36] O. Ameerbakhsh, F. M. Ghabban, I. M. Alfadli, A. N. AbuAli, A. Al-Dhaqm, and M. A. Al-Khasawneh, "Digital Forensics Domain and Metamodeling Development Approaches," in *2nd International Conference on Smart Computing and Electronic Enterprise*, Cameron Highlands, Malaysia, Jun. 2021, pp. 67–71, <https://doi.org/10.1109/ICSCEE50312.2021.9497935>.
- [37] A. A. Alhussan, A. Al-Dhaqm, W. M. S. Yafooz, A.-H. M. Emara, S. Bin Abd Razak, and D. S. Khafaga, "A Unified Forensic Model Applicable to the Database Forensics Field," *Electronics*, vol. 11, no. 9, Jan. 2022, Art. no. 1347, <https://doi.org/10.3390/electronics11091347>.
- [38] F. M. Alotaibi, A. Al-Dhaqm, Y. D. Al-Otaibi, and A. A. Alsewari, "A Comprehensive Collection and Analysis Model for the Drone Forensics Field," *Sensors*, vol. 22, no. 17, Jan. 2022, Art. no. 6486, <https://doi.org/10.3390/s22176486>.
- [39] W. M. S. Yafooz, A. Al-Dhaqm, and A. Alsaedi, "Detecting Kids Cyberbullying Using Transfer Learning Approach: Transformer Fine-Tuning Models," in *Kids Cybersecurity Using Computational Intelligence Techniques*, W. M. S. Yafooz, H. Al-Aqrabi, A. Al-Dhaqm, and A. Emara, Eds. New York, NY, USA: Springer, 2023, pp. 255–267.
- [40] A. A. Alhussan, A. Al-Dhaqm, W. M. S. Yafooz, S. B. A. Razak, A.-H. M. Emara, and D. S. Khafaga, "Towards Development of a High Abstract Model for Drone Forensic Domain," *Electronics*, vol. 11, no. 8, Jan. 2022, Art. no. 1168, <https://doi.org/10.3390/electronics11081168>.
- [41] I. M. Alfadli, F. M. Ghabban, O. Ameerbakhsh, A. N. AbuAli, A. Al-Dhaqm, and M. A. Al-Khasawneh, "CIPM: Common Identification Process Model for Database Forensics Field," in *2nd International Conference on Smart Computing and Electronic Enterprise*, Cameron Highlands, Malaysia, Jun. 2021, pp. 72–77, <https://doi.org/10.1109/ICSCEE50312.2021.9498014>.
- [42] A. Al-Dhaqm, S. H. Othman, W. M. S. Yafooz, and A. Ali, "Review of Information Security Management Frameworks," in *Kids Cybersecurity Using Computational Intelligence Techniques*, W. M. S. Yafooz, H. Al-Aqrabi, A. Al-Dhaqm, and A. Emara, Eds. New York, NY, USA: Springer, 2023, pp. 69–80.
- [43] M. Salem, S. H. Othman, A. Al-Dhaqm, and A. Ali, "Development of Metamodel for Information Security Risk Management," in *Kids Cybersecurity Using Computational Intelligence Techniques*, W. M. S. Yafooz, H. Al-Aqrabi, A. Al-Dhaqm, and A. Emara, Eds. New York, NY, USA: Springer, 2023, pp. 243–253.
- [44] A. Al-Dhaqm, W. M. S. Yafooz, S. H. Othman, and A. Ali, "Database Forensics Field and Children Crimes," in *Kids Cybersecurity Using Computational Intelligence Techniques*, W. M. S. Yafooz, H. Al-Aqrabi, A. Al-Dhaqm, and A. Emara, Eds. New York, NY, USA: Springer, 2023, pp. 81–92.
- [45] M. Saleh *et al.*, "A Metamodeling Approach for IoT Forensic Investigation," *Electronics*, vol. 12, no. 3, Jan. 2023, Art. no. 524, <https://doi.org/10.3390/electronics12030524>.
- [46] A. Ali, S. A. Razak, S. H. Othman, R. R. Marie, A. Al-Dhaqm, and M. Nasser, "Validating Mobile Forensic Metamodel Using Tracing Method," in *Advances on Intelligent Informatics and Computing*, F. Saeed, F. Mohammed, and F. Ghaleb, Eds. New York, NY, USA: Springer, 2021, pp. 473–482.
- [47] D. S. A. Baras, S. H. Othman, A. Al-Dhaqm, and R. Z. R. M. Radzi, "Information Security Management Metamodel (ISMM) Validation and Verification through Frequency-based Selection Technique," in *International Conference on Data Science and Its Applications*, Bandung, Indonesia, Oct. 2021, pp. 292–297, <https://doi.org/10.1109/ICoDSA53588.2021.9617527>.
- [48] A. M. R. Al-Dhaqm, "Simplified Database Forensic Investigation Using Metamodeling Approach," Ph.D. dissertation, University of Technology Malaysia, Johor, Malaysia, 2019.
- [49] V. R. Kebande and I. Ray, "A Generic Digital Forensic Investigation Framework for Internet of Things (IoT)," in *4th International Conference on Future Internet of Things and Cloud*, Vienna, Austria, Aug. 2016, pp. 356–362, <https://doi.org/10.1109/FiCloud.2016.57>.
- [50] V. Kebande and H. S. Venter, "Requirements for Achieving Digital Forensic Readiness in the Cloud Environment using an NMB Solution," in *11th International Conference on Cyber Warfare and Security*, Boston, MA, USA, Mar. 2016, pp. 1–9.
- [51] V. R. Kebande and H. S. Venter, "A comparative analysis of digital forensic readiness models using CFRaaS as a baseline," *WIREs Forensic Science*, vol. 1, no. 6, 2019, Art. no. e1350, <https://doi.org/10.1002/wfs2.1350>.
- [52] A. Al-Dhaqm, S. A. Razak, R. A. Ikuesan, V. R. Kebande, and K. Siddique, "A Review of Mobile Forensic Investigation Process Models," *IEEE Access*, vol. 8, pp. 173359–173375, 2020, <https://doi.org/10.1109/ACCESS.2020.3014615>.
- [53] A. Al-Dhaqm *et al.*, "Categorization and Organization of Database Forensic Investigation Processes," *IEEE Access*, vol. 8, pp. 112846–112858, 2020, <https://doi.org/10.1109/ACCESS.2020.3000747>.
- [54] A. Al-Dhaqm, S. A. Razak, K. Siddique, R. A. Ikuesan, and V. R. Kebande, "Towards the Development of an Integrated Incident Response Model for Database Forensic Investigation Field," *IEEE Access*, vol. 8, pp. 145018–145032, 2020, <https://doi.org/10.1109/ACCESS.2020.3008696>.

- [55] V. R. Kebande, R. A. Ikuesan, N. M. Karie, S. Alawadi, K.-K. R. Choo, and A. Al-Dhaqm, "Quantifying the need for supervised machine learning in conducting live forensic analysis of emergent configurations (ECO) in IoT environments," *Forensic Science International: Reports*, vol. 2, Dec. 2020, Art. no. 100122, <https://doi.org/10.1016/j.fsir.2020.100122>.
- [56] V. R. Kebande, R. A. Ikuesan, and N. M. Karie, "Review of Blockchain Forensics Challenges," in *Blockchain Security in Cloud Computing*, K. M. Baalamurugan, S. R. Kumar, A. Kumar, V. Kumar, and S. Padmanaban, Eds. New York, NY, USA: Springer, 2022, pp. 33–50.
- [57] V. R. Kebande and K.-K. R. Choo, "Finite state machine for cloud forensic readiness as a service (CFRaaS) events," *Security and Privacy*, vol. 5, no. 1, 2022, Art. no. e182, <https://doi.org/10.1002/spy2.182>.
- [58] S. Makura, H. S. Venter, V. R. Kebande, N. M. Karie, R. A. Ikuesan, and S. Alawadi, "Digital forensic readiness in operational cloud leveraging ISO/IEC 27043 guidelines on security monitoring," *Security and Privacy*, vol. 4, no. 3, 2021, Art. no. e149, <https://doi.org/10.1002/spy2.149>.
- [59] V. R. Kebande, N. M. Karie, R. A. Ikuesan, and H. S. Venter, "Ontology-driven perspective of CFRaaS," *WIREs Forensic Science*, vol. 2, no. 5, 2020, Art. no. e1372, <https://doi.org/10.1002/wfs2.1372>.
- [60] F. Alotaibi, A. Al-Dhaqm, and Y. D. Al-Otaibi, "A Conceptual Digital Forensic Investigation Model Applicable to the Drone Forensics Field," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11608–11615, Oct. 2023, <https://doi.org/10.48084/etasr.6195>.
- [61] A. S. Alraddadi, "A Survey and a Credit Card Fraud Detection and Prevention Model using the Decision Tree Algorithm," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11505–11510, Aug. 2023, <https://doi.org/10.48084/etasr.6128>.
- [62] A. Al-Dhaqm, W. M. S. Yafooz, S. H. Othman, and A. Ali, "Database Forensics Field and Children Crimes," in *Kids Cybersecurity Using Computational Intelligence Techniques*, W. M. S. Yafooz, H. Al-Aqrabi, A. Al-Dhaqm, and A. Emara, Eds. New York, NY, USA: Springer, 2023, pp. 81–92.

Assessing Real-Time Health Impacts of outdoor Air Pollution through IoT Integration

Pradeep Mullangi

Department of ECE, Shri Vishnu Engineering College for Women, Bhimavaram, India
pradeepm999@gmail.com

K. M. V. Madan Kumar

Department of CSE, Vignan Institute of Technology and Science, Hyderabad, Telangana State, India
madankukunuri@gmail.com

Gera Vijaya Nirmala

Department of ECE, CVR College of Engineering, Hyderabad, Telangana State, India
vijaya.4b8@gmail.com

Ramesh Chandra Aditya Komperla

Geico, USA
akomperla@gmail.com

Nagalinagam Rajeswaran

Electrical and Electronics Engineering, Malla Reddy College of Engineering, Secunderabad, India
rajeswarann@gmail.com (corresponding author)

Amar Y. Jaffar

Computer and Network Engineering Department, College of Computing, Umm Al-Qura University, Makkah, Saudi Arabia
ayjaafar@uqu.edu.sa

Abdullah Alwabli

Department of Electrical Engineering, College of Engineering and Computing in Alqunfudah, Umm Al-Qura University, Mecca 21955, Saudi Arabia
aswabli@uqu.edu.sa

Saeed Faisal Malky

Department of Electrical and Electronics Engineering, University of Jeddah, Jeddah, Saudi Arabia
sfmalki@uj.edu.sa

Received: 30 January 2024 | Revised: 19 February 2024 and 5 March 2024 | Accepted: 13 March 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.6981>

ABSTRACT

Air pollution constitutes a significant global challenge in both public health and the environment, particularly for countries undergoing industrialization and transitioning from low- to middle-income economies. This study aims to investigate the feasibility and effectiveness of a real-time air quality prediction system based on data collected from Internet of Things (IoT) sensors to help people and public institutions track and manage atmospheric pollution. The primary objective of this study was to investigate whether an IoT-based approach can provide accurate and continuous real-time air quality forecasting. The standard dataset provided by the Indian government was analyzed using regression, traditional Long-Short-Term Memory (LSTM), and bidirectional LSTM (BLSTM) models to evaluate their performance on

multivariate air quality features. The results show that the proposed BLSTM model outperformed the other models in minimizing RMSE errors and avoiding overfitting.

Keywords-*Internet of Things; air pollution; LSTM; health controllers*

I. INTRODUCTION

As the world faces numerous challenges in the field of public health, the widespread issue of atmospheric pollution remains a central concern due to its pervasive impact and cross-border implications [1-2]. Atmospheric pollution, characterized by the presence of harmful substances, such as particulate matter, nitrogen dioxide, and sulfur dioxide, originates from various sources, including industrial emissions, vehicular exhaust, and biomass burning. These pollutants are linked to a wide range of health problems, from respiratory to cardiovascular diseases and cancers, contributing to an estimated 11.65% of global mortality rates. Despite ongoing efforts to mitigate these effects, traditional monitoring and control mechanisms often lack real-time responsiveness and localized accuracy. In recent decades, the world has experienced significant economic advances and urban growth. However, these developments have also been accompanied by serious environmental challenges, such as air pollution. This escalating environmental crisis has adversely affected public health and exacerbated the severity of climate change [3]. Furthermore, the increasing number of vehicles, industrial operations, and other sources of combustion engines has intensified the environmental health crisis. Government bodies play a key role in mitigating these problems, as they strive to combat air pollution by implementing national and global policies and regulations that promote sustainable growth [4]. However, effective management of these environmental issues requires a unified approach from all participants. This includes government entities, organizations, industrial sectors, corporations, local communities, and individuals. The investigation of air quality has become a focal point in recent years [5]. The World Health Organization (WHO) has identified air pollution as a major factor in global disease proliferation, responsible for more than 7 million premature deaths each year [6-7]. These data underscore the immediate need to reduce exposure to air pollution and protect public health, and, as a consequence, air quality monitoring devices have gained significant attention, as they can measure temperature, humidity, and air pollution levels.

In an era marked by rapid industrialization and urbanization, the rising levels of atmospheric pollution represent an urgent call to action [8]. This study underscores the gravity of the situation, emphasizing the need for comprehensive strategies to mitigate air pollution and its adverse health impacts. This involves stringent regulations on industrial and vehicular emissions, promoting renewable energy sources, and raising public awareness of the health risks associated with poor air quality [9]. The rapid escalation in air pollution, driven by intense industrialization and urban expansion, urgently calls for immediate countermeasures [10]. This study illuminates the critical nature of this issue and advocates for the creation of comprehensive strategies to reduce air pollution and its harmful health consequences. These strategies involve strict control over industrial and vehicular

emissions, endorsement of renewable energy sources, and raising public awareness regarding the health risks associated with poor air quality. Indoor air pollution alone accounts for roughly 4.1% of global mortality rates. In 2019, the death toll from indoor air pollution exceeded 600,000. A viable solution to this problem lies in the realm of Internet of Things (IoT) monitoring systems [11]. Equipped with cutting-edge sensor technology, these systems can persistently track and collect data on a variety of indoor or outdoor air quality indicators, such as levels of pollutants like CO₂, CO, particulate matter, and volatile organic compounds. Once these data are collected, they can be transmitted to a cloud-based platform for processing and interpretation. When applying machine learning and predictive analysis algorithms to these data, it becomes feasible to generate instant information on air quality and project future patterns [12].

In recent years, the advent of IoT technologies combined with advanced machine learning models has opened new avenues to address air quality concerns. In [13], sensor data were used to perform time-series prediction on air quality parameters and pollutants by applying linear regression. In [14], a systematic review of deep learning models was conducted for time-series air quality forecasting. In [15], supervised models were used to predict the concentration of air pollutants in multiple locations of a city by using spatial-temporal relationships. In [16], air pollution in India, which is exacerbated by rapid urbanization and transportation development, was investigated. This study noted a discrepancy between current air quality monitoring systems and actual pollutant exposure at ground level, where humans directly inhale vehicle emissions. This study presented a real-time monitoring system using sensors to detect key pollutants from vehicular emissions. Leveraging a deep learning-based LSTM algorithm, the system forecasts pollutant levels, assisting decision-making to improve air quality and allowing citizens to accordingly plan their activities. This study also compared predicted ground-level pollutants with ambient air quality levels, highlighting the system's practical value. In [17], a comparative study of various statistical and deep learning methods was conducted to forecast pollution trends of PM_{2.5} and PM₁₀. The results showed that auto-regressive and Seasonal Auto-Regressive Integrated Moving Average (SARIMA) outperformed deep-learning methods on a limited dataset in Kolkata, India. In [18], a predictive model for pollutant emissions was presented for an airport, based on the number of takeoff and landing cycles. In [19], various shallow, deep, and hybrid learning models were reviewed to determine their advantages and limitations.

Recent studies have explored various IoT-based models for air quality monitoring, employing technologies ranging from simple sensor networks to complex machine learning algorithms. Although these approaches offer significant information, they often lack real-time prediction capabilities or do not adequately account for localized pollution sources. This

research gap underscores the need for a more dynamic and adaptable solution that is capable of providing accurate, real-time air quality predictions. The current study seeks to bridge this gap by introducing a novel IoT framework enhanced with bidirectional LSTM models to offer a promising avenue for comprehensive and timely air quality monitoring. The main objective of this study is to offer a scalable solution that can adapt to varying environmental conditions, thus enhancing public health responses by integrating environmental monitoring with public health initiatives.

II. MATERIALS AND METHODS

A. Datasets

This study used the dataset collected and distributed by the Central Pollution Control Board (CPCB) in conjunction with the Ministry of Environment, Forests, and Climate Change [20]. The specific dataset was created utilizing IoT and cloud technologies. The CPCB maintains an extensive database of pollution levels and closely collaborates with state pollution control boards and other government bodies to enact and monitor environmental legislation and regulations and increase public awareness of environmental issues. To collect real-time air quality data, the CPCB has deployed a network of IoT-enabled sensors at diverse locations. The data collected are stored in a cloud-based time-series database, facilitating access to air quality information across different regions. The publicly accessible real-time Air Quality Index (AQI) data can be used to issue warnings and provide air quality evaluations for various areas. Moreover, these data can be exploited for research purposes and time series analysis. The dataset encapsulates 42,000 entries and 16 variables, which include meteorological attributes, such as Benzene, Toluene, Xylene, PM2.5, PM10, NO₂, NO, NO_x, NH₃, CO, SO₂, O₃, AQI_Bucket, and AQI, along with City and Date. The data, collected on an hourly basis from various stations in different states, span five years, from January 1, 2015, to July 31, 2020. The real-time IoT-enabled dataset, stored in a cloud-based database, offers air quality information for diverse locations and is publicly accessible for research and time series analysis. This study utilized a computing environment powered by an AMD Ryzen 5 4500U@2.38GHz processor with 16GB of RAM. All experiments were run using Jupyter Notebook operating on Windows 11 Pro 64-bit. Python 3.6.5 was employed along with multiple open-source libraries, such as Pandas and NumPy. The setup also incorporated libraries, like Matplotlib, Statsmodel, and Sklearn, to optimize the experimentation process.

B. Methodology

Predicting air quality has become crucial in providing early warnings and managing urban air pollution. The objective is to forecast fluctuations in the PM2.5 air pollution index at specific monitoring points over a given timeframe. The observation period is set to one hour, a standard determined by terrestrial air quality monitoring stations. Figure 2 displays a representative example of air pollution data, such as PM2.5 levels, between 2015 and 2020. The task of predicting PM2.5 concentrations can be defined as follows: Given a certain time T , the objective is to predict the PM2.5 concentration values

$P_{i,T+1}$ at time $T+1$ or $P_{i,T+n}$ at time $T+n$, taking into account the historical air-quality time series data $AQD = \{AQD_{i,t} | i \in 0, t = 1, 2, 3 \dots T\}$ from the past [21]. AQD stands for historical air quality-related data, O represents the total observation points, and AQD encompasses not only PM2.5, but also other air quality-related time series data, such as pressure, temperature, and wind speed.

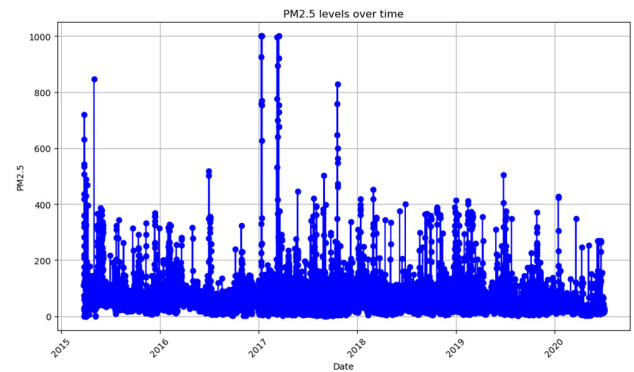


Fig. 1. Air pollution data between 2015 and 2020.

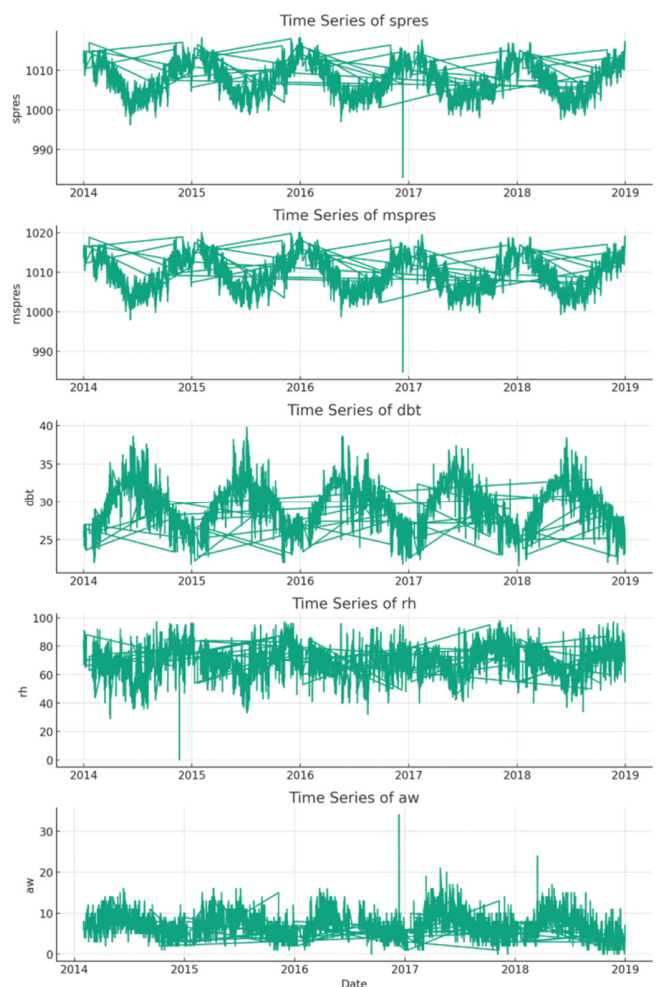


Fig. 2. Average weather pressure, humidity, temperature, and windfall between 2014 to 2019.

As noticed in Figure 2, air quality data typically contain real-valued PM2.5 pollutants, while some datasets may also include CO₂ and PM10. Alongside pollutant data, meteorological observation data play a significant role in determining air quality. For example, high wind speed tends to decrease PM2.5 concentration, elevated humidity often exacerbates air pollution, and high atmospheric pressure generally leads to better air quality. These characteristics are vital for predicting air quality. The crux of air quality prediction lies in how to process and capture the spatial-temporal features of the aforementioned air quality data items. Looking at the PM2.5 data for a month's observation data points, e.g. from 01/01/2010 to 01/31/2010, it is evident that there is contextual information among the observation points in the PM2.5 and wind speed time series. The historical state exerts some influence on the evolution of future trends. In other words, adjacent data points and periodic intervals of air quality time series data typically exhibit a strong correlation with each other.

C. Overview of the Deep Air Quality Forecasting Framework

The proposed deep-learning architecture is a modified Bidirectional LSTM (BLSTM) model, designed to capture the spatial-temporal dependencies of air quality-related time series data. Given the correlations among local trend characteristics and the long-term dependencies of multivariate air quality time series data, particularly PM2.5, the time series data are interrelated with other air quality data, and these factors are intrinsically interconnected. Figure 3 offers a visual representation of this deep air quality forecasting framework. This study focuses on the BLSTM for Air Quality Forecasting (BL-AQF) model to process multiple one-dimensional time series data and effectively grasp the spatial-temporal characteristics of various air quality indicators.

The primary phase of this study encompassed sequential training models to extract local trend attributes and potential spatial associations from multistation air quality readings. The model does not analyze the features of each time series in isolation. Instead, it simultaneously processes all the time series data collected from each monitoring point spread across various stations. Subsequently, the extracted features, comprising the local trend features of each station data and the potential spatial correlation features of multistation data, from numerous one-dimensional CNNs are concatenated and fed into a specific BLSTM. This BLSTM learns spatial-temporal dependency features from both past and future contexts, concurrently using time series in both forward and backward directions [22].

The computational components of a typical BLSTM are:

$$i_t = \sigma(U^{(i)}x_t + W^{(i)}h_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(U^{(f)}x_t + W^{(f)}h_{t-1} + b_f) \quad (2)$$

$$o_t = \sigma(U^{(o)}x_t + W^{(o)}h_{t-1} + b_o) \quad (3)$$

$$\tilde{s}_t = \tanh(U^{(c)}x_t + W^{(c)}h_{t-1} + b_c) \quad (4)$$

$$s_t = f_t \cdot s_{t-1} + i_t \cdot \tilde{s}_t \quad (5)$$

$$h_t = o_t \cdot \tanh(s_t) \quad (6)$$

In the equations, the input gate determines the influx of new information into the memory cell. The symbol f_t stands for the forget gate, dictating the volume of data to be removed. Meanwhile, o_t serves as the output gate, deciding the measure of data to be relayed either to the subsequent step or directly to the output. \tilde{s}_t functions as a neuron equipped with a self-repeating cell akin to RNNs. The memory cell within the LSTM block denoted s_t , is an aggregation of two components. The initial segment is derived from the prior internal memory state s_{t-1} and the forget gate f_t . The subsequent part is determined by the element-wise product of the self-repeating state \tilde{s}_t and the input gate [23]. A limitation of conventional LSTMs is their ability to harness only the antecedent context of sequential data [24]. In contrast, BLSTM can interpret time series information in a dual-directional manner deploying two distinct hidden layers. Subsequently, the data from both directions are merged and advanced to the output stage. Put differently, BLSTM cyclically assesses time series data in two directions: the forward layer goes from $t = 1$ to T , while the backward layer operates in reverse form, from $t = T$ to 1.

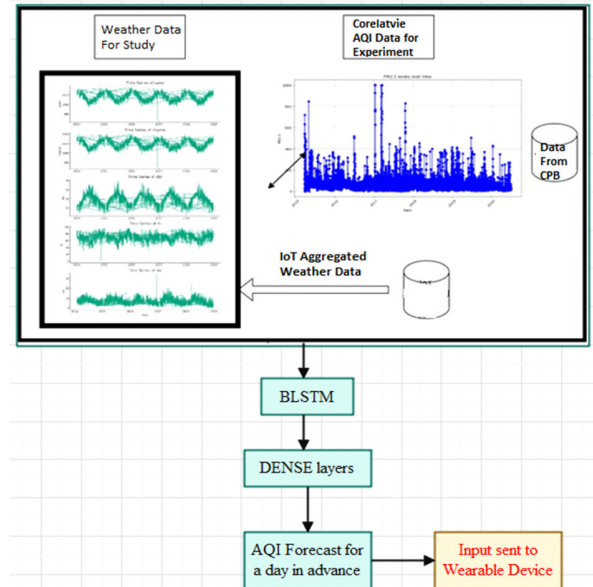


Fig. 3. Architecture of the AQI monitoring system.

III. RESULTS AND DISCUSSION

A. Understanding the Influence of Traditional Machine Learning in Air Quality Forecasts

Linear regression, a foundational statistical method, plays a pivotal role in air quality projections by drawing a clear correlation between factors, such as temperature variations, wind dynamics, and moisture levels with air quality index and offering an intuitive way to predict future atmospheric conditions. Its straightforward nature and easy-to-follow logic have cemented its status among environmental analysts. Beyond the linear approach, the Decision Tree (DT) offers a structured method to dissect data based on specific criteria. For example, it can evaluate air quality differences that stem from fluctuating temperatures. This approach excels in highlighting

intricate connections and dependencies among various environmental elements. Not only does the tree-like representation bring clarity, but also enhances the comprehensibility of the model logic. Building upon the foundation of DT, Random Forest (RF) introduces an ensemble technique that amalgamates insights from numerous trees. This collective approach mitigates the typical pitfalls of individual trees, such as oversensitivity to data discrepancies. When predicting air quality, it ensures a balanced and comprehensive understanding, harnessing the insights of multiple evaluations for a more refined prediction.

The Mean Squared Error (MSE) of the linear regression was 4.963×10^{-34} while R^2 was 1.0. The MSE for linear regression, being extraordinarily close to zero, and the R^2 value indicate almost perfect predictions that match the actual data. Although this might seem impressive, such perfect scores can be indicative of overfitting, where the model might be too closely tailored to the training data and may not perform as well on new unseen data. For DT, MSE was 2.914×10^{-7} and R^2 was 0.9983. DT had a slightly higher MSE compared to linear regression. However, its MSE is still very low, indicating good prediction accuracy. The R^2 value is slightly less than 1, indicating that the model explains approximately 99.83% of the variance in the dependent variable. This is an excellent score, but the model might be complex and could risk overfitting, given the nature of DT. For RF, MSE was 2.05986×10^{-7} and R^2 was 0.99988. RF, being an ensemble method, combines multiple DTs to produce predictions. Its MSE is lower than that of the DT, exhibiting higher accuracy. The R^2 value is very close to 1, suggesting that the model explains approximately 99.99% of the variance in the dependent variable. This model seems to provide a balanced trade-off between complexity and accuracy.

However deep-learning algorithms are better for AQI prediction due to:

- Complex relationships: Air quality is influenced by a variety of factors, including pollutants, weather patterns, industrial activities, and more. Deep learning can capture intricate, nonlinear relationships among these variables more effectively than shallow algorithms
- Feature extraction: Deep learning models, especially neural networks, can automatically extract and learn important features from raw data, removing the need for manual

feature engineering, which can be required for shallow models.

- Handling large datasets: Air quality datasets can be vast, containing data from multiple sensors over long periods. Deep learning models excel when trained on large datasets, allowing them to generalize better.
- Temporal dependencies: RNNs and LSTM networks, subsets of deep learning, are well suited for time-series data, such as air quality indices, as they can remember past information to influence future predictions.
- Regularization techniques: Deep learning models come with a variety of regularization methods, like dropout, which can prevent overfitting, especially when dealing with complex datasets.

B. Results for LSTM-based AQI Prediction

Figure 4 illustrates that RMSE in both the training and validation sets decreases as the number of epochs increases for LSTM. This suggests that the model is learning and refining its predictions with more training. Over the epochs, the RMSE values seem to be stabilizing, indicating that the model is reaching a point of convergence. The gap between the training and validation RMSE appears minimal, suggesting that the model generalizes well and does not overfit the training data. Concerning MSE, both the training and validation MSE values seem to be decreasing as the number of epochs increases. This trend reaffirms that the model is improving its predictions with continued training. Similarly to the RMSE plot, the MSE values appear to stabilize as the epochs progress, suggesting model convergence. Since RMSE is the square root of MSE, the trends observed here are consistent with those seen in the RMSE plot. Considering Mean Absolute Error (MAE), both the training and validation values decrease as the epochs increase, further supporting the assertion that the model is making progress and refining its predictions. As with the previous two metrics, the MAE values seem to level off as the epochs progress, indicating that the model might be reaching an optimal state. The small gap between the training and validation curves suggests good model generalization. All three metrics (RMSE, MSE, MAE) demonstrate a decreasing trend over the epochs, which is a clear sign that the LSTM model is learning and enhancing its predictive capabilities with each training epoch.

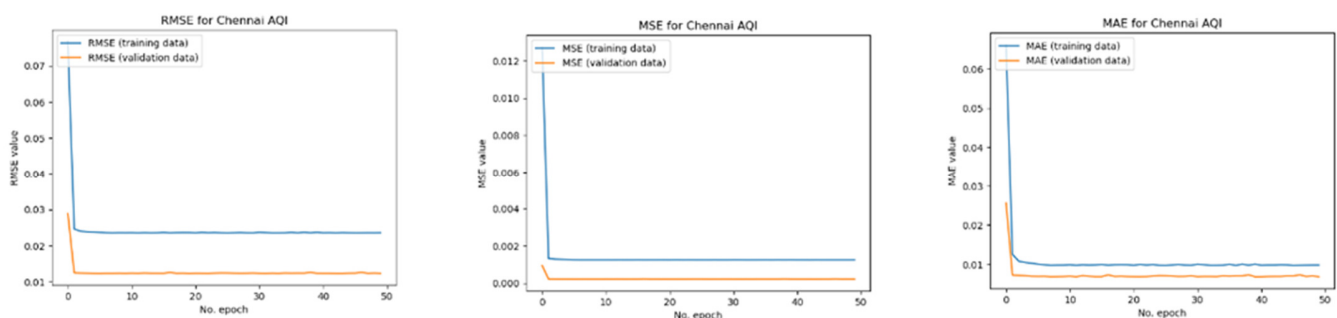


Fig. 4. RMSE, MSE, and MAE results for LSTM-based evaluation.

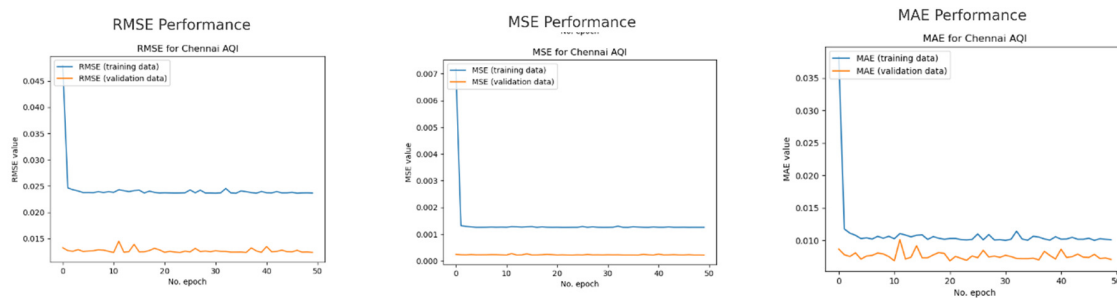


Fig. 5. RMSE, MSE, and MAE of the BLSTM model.

C. Results for BLSTM AQI Prediction

Figure 5 provides a visual representation of the model's performance across epochs in terms of RMSE for both training and validation. Both the training and validation RMSE values rapidly decline during the initial epochs, indicating that the model is learning and improving its prediction accuracy. After the rapid decrease, the RMSE values for both training and validation start to stabilize. This suggests that the model is converging and that further training might not yield significant improvements in prediction accuracy. There is a noticeable gap between the training and validation RMSE values, but it does not appear to significantly widen as training progresses. This is a positive sign, as it suggests that the model is not overfitting the training data. The model shows good generalization, since the validation RMSE is relatively close to the training RMSE, and there is no significant divergence between them over the epochs. Stabilization of the RMSE values indicates that the model may have reached its optimal performance for the given architecture and data. In conclusion, the BLSTM model appears to perform well for multistep AQI forecasting based on the provided RMSE values. Further domain-specific insights and comparisons with actual AQI values would provide a more comprehensive evaluation.

Figure 6 compares the performance of both the BLSTM and the regular LSTM models across epochs. In Epoch 1, BLSTM starts with a training RMSE of 0.1749 and a validation RMSE of 0.0857, while LSTM starts with a training RMSE of 0.2446 and a validation RMSE of 0.1496. BLSTM starts with a better initial performance compared to LSTM. Both models show a rapid decrease in RMSE during the initial epochs, indicating learning and improvements in prediction accuracy. Throughout the epochs, the BLSTM model consistently achieves a lower RMSE than the regular LSTM for both training and validation, showcasing its superior performance. After the initial rapid decrease, the RMSE values for both models begin to stabilize. This indicates that the models reach a point of convergence, where further training might not significantly improve RMSE. BLSTM seems to stabilize at a lower RMSE value compared to LSTM, further emphasizing its efficiency for this task. The gap between the training and validation RMSE for both models is relatively small, suggesting good generalization capabilities on unseen data. However, the BLSTM's validation RMSE remains consistently lower, indicating potentially better generalization for this specific task. BLSTM consistently outperforms regular LSTM in terms of RMSE across all epochs for both training and

validation. This can be attributed to its ability to leverage both past and future context in sequence data, making it particularly effective for time series predictions, such as AQI forecasting. Both models demonstrate a converging trend as training progresses, with BLSTM demonstrating a more favorable trajectory. Both models exhibit good generalization, but the BLSTM shows a slight edge in this dataset, given its consistently lower validation RMSE. Overall, for air quality prediction, BLSTM appears to be a more suitable choice than regular LSTM, on the basis of the RMSE values provided.

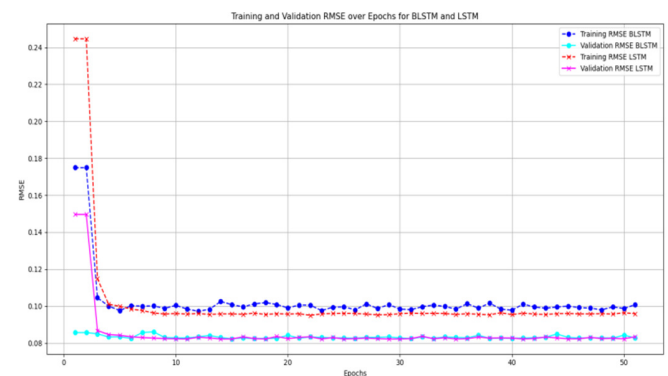


Fig. 6. Comparative performance of LSTM and BLSTM for AQI prediction.

Figure 7 reveals that the actual and predicted values appear to be in close alignment for most of the data points. This suggests that the BLSTM model has captured the underlying patterns and trends in the data quite well. There are certain points where the predicted values deviate from the actual values. These deviations, however, appear to be minimal and not too frequent, indicating that the model performance is generally consistent. The model seems to capture both the peaks and troughs of the actual data, suggesting that it has learned the seasonality or cyclic behavior, if any, present in the dataset. In many sections of the chart, the actual and predicted curves overlap, which is a strong indicator of the model's accuracy in those intervals. BLSTM appears to perform quite effectively for the AQI prediction task, as the close alignment between the actual and predicted values signifies a high degree of accuracy. The ability of the model to capture major trends, peaks, and troughs in the data emphasizes its robustness.

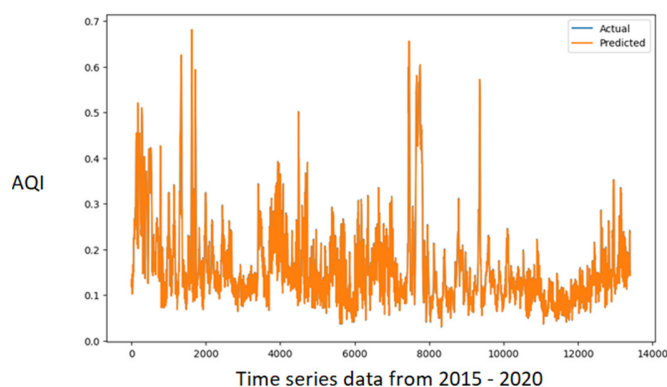


Fig. 7. Actual vs predicted time series data results by BLSTM.

IV. CONCLUSIONS

This study introduces a groundbreaking IoT-based framework that uses BLSTM for precise prediction of air quality. The findings demonstrate a significant improvement in prediction accuracy over traditional methods, avoiding overfitting and underscoring the potential of integrating advanced machine learning techniques with IoT technologies for air pollution monitoring. The application of this study extends beyond academic interest, offering tangible benefits for urban planning, public health strategies, and environmental protection initiatives. By providing accurate real-time air quality data, the proposed BLSTM model supports informed decision-making, allowing timely interventions that can mitigate the adverse health impacts of air pollution and improve the quality of life in urban environments. Looking ahead, the integration of the proposed IoT framework with smart city infrastructure represents an exciting avenue for research and development. Such advances could lead to the creation of highly adaptive and responsive environmental monitoring systems, capable of predicting and managing air quality issues more effectively. Furthermore, exploring the model's performance across a broader range of environmental conditions will be critical to ensure its applicability and reliability on a global scale. The limitations of this study are particularly based on the need for extensive data collection and validation in diverse environmental settings. Future research should aim to address these challenges, seeking to refine and expand the model's capabilities to ensure its effectiveness and scalability. In conclusion, this research contributes a novel approach to the challenge of air quality prediction, highlighting the synergy between IoT technologies and machine learning. Continuous advances in this field grow the potential for impactful real-world applications, offering a beacon of hope for addressing the global issue of air pollution.

REFERENCES

- [1] N. Harkat, A. Rahmane, and I. Bendjemila, "The Impact of Industrial Air Pollution on the Urban Environment of Setif: Modeling and Mapping of Total Suspended Particles," *Engineering, Technology & Applied Science Research*, vol. 12, no. 6, pp. 9431–9439, Dec. 2022, <https://doi.org/10.48084/etasr.5256>.
- [2] H. Ritchie and M. Roser, "Indoor Air Pollution," *Our World in Data*, Mar. 2024, [Online]. Available: <https://ourworldindata.org/indoor-air-pollution>.
- [3] E. Fazakas, I. A. Neamtiu, and E. S. Gurzau, "Health effects of air pollutant mixtures (volatile organic compounds, particulate matter, sulfur and nitrogen oxides) – a review of the literature," *Reviews on Environmental Health*, Mar. 2023, <https://doi.org/10.1515/reveh-2022-0252>.
- [4] C. Li *et al.*, "The role of mental health professionals in the climate crisis: an urgent call to action," *International Review of Psychiatry*, vol. 34, no. 5, pp. 563–570, Jul. 2022, <https://doi.org/10.1080/09540261.2022.2097005>.
- [5] H. Wu, Y. Li, Y. Hao, S. Ren, and P. Zhang, "Environmental decentralization, local government competition, and regional green development: Evidence from China," *Science of The Total Environment*, vol. 708, Mar. 2020, Art. no. 135085, <https://doi.org/10.1016/j.scitotenv.2019.135085>.
- [6] F. Ekici, G. Orhan, Ö. Gümüş, and A. B. Bahce, "A policy on the externality problem and solution suggestions in air transportation: The environment and sustainability," *Energy*, vol. 258, Nov. 2022, Art. no. 124827, <https://doi.org/10.1016/j.energy.2022.124827>.
- [7] M. O. Raimi, Z. Adio, O. O. Emmanuel, Timothy Kayode Samson, B. S. Ajayi, and T. J. Ogunleye, "Impact of Sawmill Industry on Ambient Air Quality: A Case Study of Ilorin Metropolis, Kwara State, Nigeria," *Energy and Earth Science*, vol. 3, no. 1, 2020, <https://doi.org/10.2139/ssrn.3586971>.
- [8] D. L. Mendoza *et al.*, "Air Quality and Behavioral Impacts of Anti-Idling Campaigns in School Drop-Off Zones," *Atmosphere*, vol. 13, no. 5, May 2022, Art. no. 706, <https://doi.org/10.3390/atmos13050706>.
- [9] X. Zhang *et al.*, "Linking urbanization and air quality together: A review and a perspective on the future sustainable urban development," *Journal of Cleaner Production*, vol. 346, Apr. 2022, Art. no. 130988, <https://doi.org/10.1016/j.jclepro.2022.130988>.
- [10] P. Kokate, S. Sadistap, and A. Middey, "Atmospheric CO2 Level Measurement and Discomfort Index Calculation with the use of Low-Cost Drones," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11728–11734, Oct. 2023, <https://doi.org/10.48084/etasr.6230>.
- [11] G. Marques and R. Pitarma, "mHealth: Indoor Environmental Quality Measuring System for Enhanced Health and Well-Being Based on Internet of Things," *Journal of Sensor and Actuator Networks*, vol. 8, no. 3, Sep. 2019, Art. no. 43, <https://doi.org/10.3390/jsan8030043>.
- [12] Q. Guo *et al.*, "Applications of artificial intelligence in the field of air pollution: A bibliometric analysis," *Frontiers in Public Health*, vol. 10, Sep. 2022, <https://doi.org/10.3389/fpubh.2022.933665>.
- [13] R. Kumar, P. Kumar, and Y. Kumar, "Time Series Data Prediction using IoT and Machine Learning Technique," *Procedia Computer Science*, vol. 167, pp. 373–381, Jan. 2020, <https://doi.org/10.1016/j.procs.2020.03.240>.
- [14] N. Zaini, L. W. Ean, A. N. Ahmed, and M. A. Malek, "A systematic literature review of deep learning neural network for time series air quality forecasting," *Environmental Science and Pollution Research*, vol. 29, no. 4, pp. 4958–4990, Jan. 2022, <https://doi.org/10.1007/s11356-021-17442-1>.
- [15] A. Gilik, A. S. Ogrenici, and A. Ozmen, "Air quality prediction using CNN+LSTM-based hybrid deep learning architecture," *Environmental Science and Pollution Research*, vol. 29, no. 8, pp. 11920–11938, Feb. 2022, <https://doi.org/10.1007/s11356-021-16227-w>.
- [16] J. Srishtishree, S. Mohana Kumar, and C. Shetty, "Air Quality Monitoring with IoT and Prediction Model using Data Analytics," in *Innovations in Computer Science and Engineering: Proceedings of 7th ICICSE*, H. S. Saini, R. Sayal, R. Buyya, and G. Aliseri, Eds. Singapore: Springer, 2020, pp. 535–544.
- [17] P. Nath, P. Saha, A. I. Middya, and S. Roy, "Long-term time-series pollution forecast using statistical and deep learning methods," *Neural Computing and Applications*, vol. 33, no. 19, pp. 12551–12570, Oct. 2021, <https://doi.org/10.1007/s00521-021-05901-2>.
- [18] D. A. Pamplona and C. J. P. Alves, "Civil Aircraft Emissions Study and Pollutant Forecasting at a Brazilian Airport," *Engineering, Technology & Applied Science Research*, vol. 10, no. 1, pp. 5217–5220, Feb. 2020, <https://doi.org/10.48084/etasr.3227>.

-
- [19] H. Liu, G. Yan, Z. Duan, and C. Chen, "Intelligent modeling strategies for forecasting air quality time series: A review," *Applied Soft Computing*, vol. 102, Apr. 2021, Art. no. 106957, <https://doi.org/10.1016/j.asoc.2020.106957>.
- [20] M. Mondal, "Air Quality Index." [Online]. Available: <https://kaggle.com/code/mahadevmondal/air-quality-index>.
- [21] Q. B. Jamali *et al.*, "Analysis of CO₂, CO, NO, NO₂, and PM Particulates of a Diesel Engine Exhaust," *Engineering, Technology & Applied Science Research*, vol. 9, no. 6, pp. 4912–4916, Dec. 2019, <https://doi.org/10.48084/etasr.3093>.
- [22] Z. Zhang and M. Tao, "Deep Learning for Wireless Coded Caching With Unknown and Time-Variant Content Popularity," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 1152–1163, Oct. 2021, <https://doi.org/10.1109/TWC.2020.3030973>.
- [23] M. S. Balamurugan and R. Manojkumar, "Study of short term rain forecasting using machine learning based approach," *Wireless Networks*, vol. 27, no. 8, pp. 5429–5434, Nov. 2021, <https://doi.org/10.1007/s11276-019-02168-3>.
- [24] C. Zhang *et al.*, "Satellite UV-Vis spectroscopy: implications for air quality trends and their driving forces in China during 2005–2017," *Light: Science & Applications*, vol. 8, no. 1, Nov. 2019, Art. no. 100, <https://doi.org/10.1038/s41377-019-0210-6>.