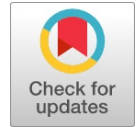


# Breast Cancer Prediction Based on Feature Extraction using Hybrid Methodologies

G. Rajasekaran, C. Sunitha Ram



**Abstract-** The breast cancer prediction is essential for effective treatment and management of the disease. Using data mining techniques to develop predictive models can assist in identifying patients at high risk of developing breast cancer, allowing for early detection and treatment. Early detection has been shown to improve patient outcomes and survival rates. The proposed system for breast cancer prediction involves two main techniques: Linear Discriminant Analysis (LDA) based feature extraction and hyperparameter tuned LSTM-XGBoost based hybrid modelling. The LDA is used to extract the features from the input data that can be trained using a hybrid model such as LSTM and XGBoost. The hyperparameters of both models are optimized using cross-validation techniques to achieve high accuracy in breast cancer prediction. Overall, this proposed system has achieved an accuracy and efficiency of breast cancer prediction than existing

**Keywords:** Breast Cancer prediction, Feature extraction, Linear Discriminant Analysis, Long Short-Term Memory, XGBoost, classification, performance accuracy

## I. INTRODUCTION

Breast cancer is originated in the breast tissue that generally in the lobules and ducts of the breast. It is the most common cancer among women worldwide and the second most common cancer overall. Breast cancer can also occur in men, although it is rare. Breast cancer can spread to other parts of the body, including the lymph nodes, bones, liver, and lungs. The spread of cancer from one part of the body to another is called metastasis. The likelihood of metastasis depends on the cancer's size and location, cancer's stage, and the individual's health. The spreading rate of breast cancer varied based on breast cancer stages at diagnosis [1]. In general, breast cancer that is detected early and treated promptly has a lower rate of spreading than cancer that is diagnosed at a later stage. The spread of breast cancer can be slowed or even stopped with appropriate treatment, which may include chemotherapy, surgery, targeted therapy or radiation therapy.

Earlier breast cancer prediction using data mining techniques can play a crucial role in improving patient outcomes [2]. Early detection can lead to timely treatment, which is essential for successful management of the disease. In addition to improving patient outcomes, early prediction can also reduce healthcare costs and enhance the life's quality for patients. Data mining techniques can help identify patterns and trends in large amounts of data that can be predicted the likelihood of breast cancer development. By leveraging these techniques, healthcare professionals can take a proactive approach to patient care and provide timely interventions that can lead to better patient outcomes. Overall, the importance of early prediction of breast cancer using data mining cannot be overstated and is a critical component of breast cancer management [3].

Developing a machine learning algorithm for predicting breast cancer using data mining is a valuable application in healthcare [4][5]. The first step in creating such an algorithm is to collect a relevant dataset containing information about patients with and without breast cancer. Popular datasets for this task include the Breast Cancer Wisconsin (BCW) and Wisconsin Diagnostic Breast Cancer (WDBC) dataset [6]. Once the dataset is obtained, data pre-processing methods are implemented to clean and prepare the data for analysis. Next, feature selection methods can be used to measure the prediction of breast cancer. Various machine learning algorithms can then be applied, such as logistic regression, decision trees and support vector machines (SVM), to develop a predictive model. The performance metrics like precision, accuracy, F1-score and recall. Finally, the system can be deployed and tested on new data to assess its generalization ability. The main purpose of this method is to propose a hybrid model that utilizes meta heuristic optimization method to enhance an accuracy of breast cancer prediction.

The paper is organized as follows: Chapter II discusses a literature review of breast cancer prediction techniques. Chapter III provides background information on the Pelican Optimization Algorithm. Chapter IV describes the proposed hybrid model for breast cancer prediction. Chapter V presents the results; Chapter VI presents conclusions of the study.

## II. RELATED WORK

In order to address the issue of overfitting and enhances a prediction's accuracy, the author E. O. Jessica et al [7] utilized LDA to extract relevant features. By reducing the dataset features, the model was able to generate new, more useful features while eliminating the original ones. This approach was effective in improving the accuracy of the predictions.

Manuscript received on 23 April 2023 | Revised Manuscript received on 06 May 2023 | Manuscript Accepted on 15 May 2023 | Manuscript published on 30 May 2023.

\*Correspondence Author(s)

**G. Rajasekaran\***, Research Scholar, Department of Computer Science and Engineering, SCSVMV University, Kancheepuram (Tamil Nadu), India. E-mail: [gtrajasekaran@gmail.com](mailto:gtrajasekaran@gmail.com), ORCID ID: <https://orcid.org/0009-0008-8042-9267>

**Dr. C. Sunitha Ram**, Assistant Professor, Department of Computer Science and Engineering, SCSVMV University, Kancheepuram (Tamil Nadu), India. ORCID ID: <https://orcid.org/0000-0002-0618-8660>

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

# Breast Cancer Prediction Based on Feature Extraction using Hybrid Methodologies

In [8], R. R. Rath et al developed a hierarchical multi-model approach using a hierarchical clustering technique to assess the similarity between decision trees. The approach involved selecting sample trees from distinct groups to construct a hierarchical randomization model with less overlap and higher accuracy. The study applied this approach to the WDBC database and utilized samples from the UCI repository. The study by S. Pravesjit et al. [9] introduces a hybrid algorithm that combines Particle Swarm Optimization (PSO) and Rao algorithms for classifying breast cancer data. The Rao algorithm is utilized to update the velocity, and the most efficient particle classification values are employed to determine the new particle position.

S. Kayikci et al. [10] have developed a multimodal deep learning (DL) model that facilitates decision from multiple sources. The study employs three different classifiers and achieves high accuracy rates of 82% for decision trees, 88% for SVM and 90% for random forests. P. Liu et al. [11] have introduced a gradient boosting algorithm, named EXSA, which is optimized for survival analysis with ties to predict the progression of breast cancer. The method employs the XGBoost method and the Cox proportional hazards model in survival analysis.

In [12], R. Lupat et al. introduce a DL algorithm named as Moanna, that is implemented a multi-omics data for a breast cancer prediction. It is based on semi-supervised Autoencoder linked to a multi-task learning to combine a copy number, gene expression and somatic mutation data. Nathiya S et al. [13] presented a SVM-ANN optimized method for classification tasks. The results show that the SVM-ANN optimized algorithm performed well with an accuracy rate of 97 percent, demonstrating its effectiveness for classification tasks.

C.-H. Yang et al [14] developed a FuzzyDeepCoxPH, that combines a fuzzy logic with DL algorithms and conventional model. The system measured a high-risk of mutation variants and candidate genes with cancer mortality. By integrating the strengths of multiple approaches, FuzzyDeepCoxPH can effectively analyze and interpret complex genomic data to identify potential risk factors and cancer mortality prediction accuracy.

N N. Arya et al. [15] developed a DL based predictive models for a breast cancer prediction in a stacked ensemble network. This method leverages available multi-modal datasets to enhance accuracy. One of the key advantages of the approach is the two-stage model architecture. In the first stage, a convolutional neural network (CNN) for feature extraction, and in the second stage, stack-based ensemble model is performed for a classification. This architecture allows the model to effectively capture complex relationships between features and enhance a breast cancer prediction accuracy.

A.-M. Yang et al. [16] propose a novel twin SVM algorithm, which utilizes a dependent nearest neighbor (DNN) weighting approach to improve the accuracy of the epsilon-TSVM model. The proposed approach includes a local information mining function that helps to identify the optimal dependency region domain, and the solution of the improved model is derived. The authors also employ a cuckoo algorithm to determine the optimal parameters of DNN, further improving the accuracy of the model.

D. Sun et al. [17] propose a novel Multimodal Deep Neural Network (MDNNMD) that integrates multi-dimensional data for the prognosis prediction of breast cancer. The proposed method's architecture is designed to leverage the strengths of DL algorithms and the fusion of multiple data sources. The authors utilize a multi-modal approach to combine data from different sources, including clinical, imaging, and genomic data. R. Mendonca-Neto et al [18] propose an outlier-based gene selection (OGS) method for selecting relevant genes that can efficiently and effectively classify breast cancer subtypes. The OGS method applies outlier detection techniques to identify genes that have a significant impact on classification accuracy.

M. Poryahya et al. [19] implemented a multiomics clustering based on network analysis and the Wasserstein distance from optimal mass transport. This distance metric is chosen due to its desirable geometric properties that are well-suited for machine learning and clustering applications. A novel breast cancer risk assessment model called BCRAM has been proposed by A. Li et al. [20] This model relies on epidemiological factors, making it more suitable for populous countries such as China compared to gene-based models. BCRAM presented the similarities between epidemiological factors to identify individuals with a huge risk of developing breast cancer, who share similarities with breast cancer patients.

## III. BACKGROUND

### Dataset Description

The UCI repository is a well-known dataset used for diagnosing breast cancer. It classifies data into normal and cancer-affected categories, based on whether the tumor is benign or malignant. In this work, the WDBC was utilized, consisting of 569 patterns from 32 patients. [Table -1](#) displays several attributes provided for all patterns, with attributes 4-13 being divided into mean, standard error, and worst result categories, each containing 30 attributes.

### Linear Discriminant Analysis

LDA is a useful method for feature extraction in breast cancer data analysis. The objective is to minimize an input data dimensionality while preserving class separability, which can help to improve the accuracy of breast cancer classification.

In the data set, LDA works by identifying the most relevant features that can distinguish between benign and malignant stages. The first step is to pre-process the data and select the most relevant features using methods such as feature selection or dimensionality reduction.

**Table -1 : Dataset Attributes**

S. No.	ATTRIBUTES
1	ID number
2	Benign result
3	Malignant result
4	Radius
5	Perimeter
6	Texture

7	Smoothness
8	Area
9	Compactness
10	Concave points
11	Concavity
12	Facial dimension
13	Symmetry

After selecting the features, LDA is applied to the dataset to find a selected features linearly that maximizes the variation among the two classes of cancers. This process helps to identify the most discriminative features that can distinguish between normal and abnormal stages. The resulting projection represents the new set of features that are more informative and discriminative for breast cancer classification. This new set of features can then be used as input to a hybrid model to predict the likelihood of a cancer being benign or malignant.

*Pelican Optimization Algorithm*

Solving optimization problems is a fundamental challenge across various scientific disciplines, and optimization is an important aspect of this task. A novel stochastic optimization model named as Pelican Optimization Algorithm (POA), has been developed to address this challenge. The POA is inspired by the pelicans of hunting behavior in nature, and its design is based on the simulation of this behavior. In this algorithm, search agents behave like pelicans in search of food sources. The POA has a mathematical model that can be used to solve optimization problems.

Pelicans are skilled hunters due to their intelligent hunting behavior and strategy. The optimization process was conducted in two phases, which involved winging on the water surface and moving towards prey. To start the optimization process, population are initialized randomly within an upper and lower bound using Equation (1). This ensures that the algorithm explores the entire search space and avoids getting stuck in a local minimum.

$$p_{i,j} = LB_j + rand(UB_j - LB_j), i=1,2...N \quad (1)$$

Where,  $p_{i,j}$  is the candidate solution and N is total population

*Moving towards Prey (Exploration Phase)*

Exploration phase involves identifying the prey's location and moving towards it. By modelling this strategy, the POA gains an availability to scan the search space and explore different areas. The prey's location is randomly generated in the search space in POA. These concepts, along with the pelican's hunting strategy, are expressed mathematically in Equation (2).

$$p_{i,j} = \begin{cases} p_{i,j} + rand(x_j - Ip_{i,j}), F_p < F_i \\ p_{i,j} + rand(p_j - p_j), else \end{cases} \quad (2)$$

Where,  $p_{i,j}$  is the new position of pelican. i and rand are random numbers varies from 1 to 2.  $F_p$  is the objective function.

POA utilizes effective updating to ensure that a pelican's new position is updated only if it improves the objective function's value. This technique prevents the algorithm from

moving towards non-optimal areas. Equation (3) is used to model this process.

$$P_i = \begin{cases} P_i^{new1}, F_p < F_i \\ P_i, else \end{cases} \quad (3)$$

Where,  $P_i^{new1}$  is the new position of pelican

*Winging on the Water Surface (Exploitation Phase)*

Equation (4) mathematically simulates the hunting behavior of pelicans in the second phase, where it spread their wings on the water surface to move the fish upwards and collect it in their throat pouch. It enables pelicans to hunt many fishes in the attacked area. Therefore, the proposed POA can meet to better points in the hunting area.

$$P_i^{new2} = p_{i,j} + R \left(1 - \frac{t}{T}\right) \cdot (2 \cdot rand - 1) p_{i,j} \quad (4)$$

Where,  $P_i^{new2}$  is the new position of pelican. Equation (5) models the use of effective updating in the second phase of POA, where new pelican positions are accepted or rejected based on their ability to improve the objective function's value. This technique ensures that the algorithm does not move towards non-optimal areas.

$$P_i = \begin{cases} P_i^{new2}, F_p < F_i \\ P_i, else \end{cases} \quad (5)$$

Once all population is initialised according to the first and second phases, the algorithm evaluates the new population status and objective function values to determine the best candidate solution. This solution is then updated to reflect the new information obtained during the optimization process.

**IV. PROPOSED OPTIMIZED HYBRID MODEL**

The proposed system for breast cancer prediction is based on two powerful learning techniques: LDA based feature extraction and hyperparameter tuned LSTM-XGBoost based hybrid modelling. The LDA model can extract the most relevant features from an input data, and the hybrid LSTM-XGBoost model to provide a final prediction based on the extracted features as shown in Figure.

By optimizing the hyperparameters of both models, the proposed system achieves high accuracy in breast cancer prediction. This system has the potential to enhance an accuracy and efficiency of breast cancer prediction .



## Breast Cancer Prediction Based on Feature Extraction using Hybrid Methodologies

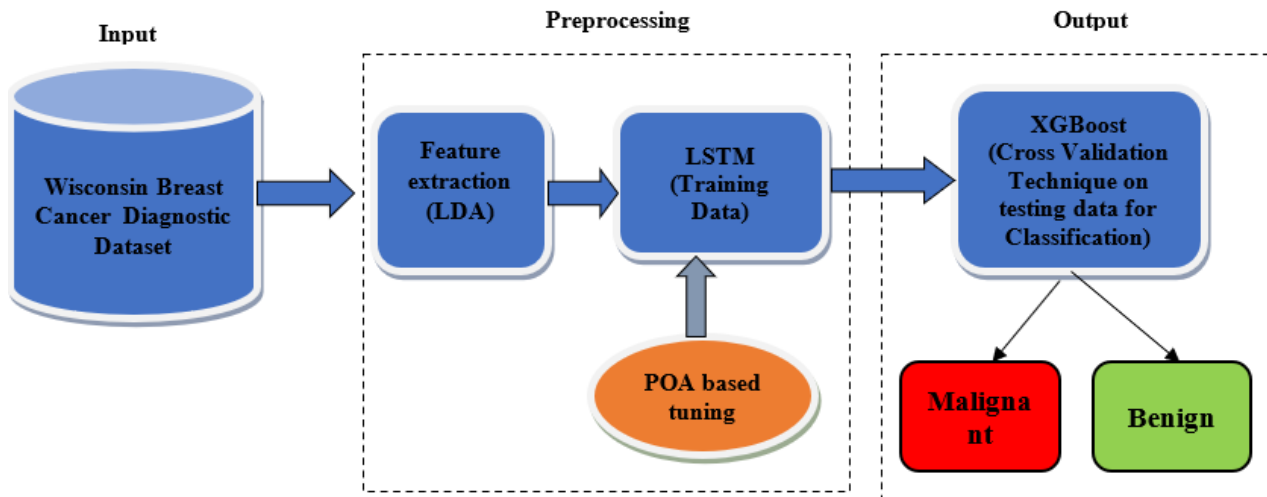


Fig. 1 . Architecture of Breast Cancer Prediction using Hybrid LSTM – XGBoost Model

The steps involved in proposed prediction model explained below:

**Data Pre-processing:** Pre-process the input data and select the most relevant features using LDA.

**Split the data:** Split the data into testing and training sets.

**LSTM Modelling:** Train an LSTM model on the training data, with hyperparameters such as the number of layers, number of neurons, and activation functions tuned using cross-validation techniques.

**XGBoost Modeling:** Train an XGBoost model on the output of the LSTM model from step 3, with hyperparameters namely maximum depth, learning rate and number of trees tuned using cross-validation techniques.

**Evaluation:** the proposed model performance Evaluation based on the metrics such as accuracy, precision, recall, and F1 score.

**Hyperparameter Tuning:** Further optimize the hyperparameters of the hybrid model using POA

### 4.1 Architecture of hybrid LSTM-XGBoost model

This technique combines the strengths of two different machine learning models: the LSTM neural network and the XGBoost algorithm. The proposed model includes following layers for prediction.

**Input Layer:** The input layer receives the sequential data, such as time series data, and passes it to the LSTM layer.

**LSTM Layer:** The LSTM layer is a type of Recurrent Neural Network (RNN) that can process sequential data and learn the temporal dependencies between data points. The LSTM layer consists of several memory cells, each with an output gate, input gate and a forget gate. The input gate controlled the data flow into the cell, the output gate controlled the data flow out of the cell, and the forget gate-controlled data retention in the cell. The LSTM layer produces a set of features that capture the temporal patterns in the input data.

**Dropout Layer:** The dropout layer is used to prevent overfitting. It randomly drops out some of the connections among the neurons in the LSTM layer during training, which

helps to reduce the dependence on specific features and improve generalization.

**Time Distributed Dense Layer:** The time distributed dense layer applied a dense layer to every time step of the LSTM output. This layer helps to transform the LSTM output into a format that can be fed into the XGBoost algorithm.

**XGBoost Layer:** The XGBoost layer is an implementation of the XGBoost algorithm, which is a gradient boosting machine learning technique that can handle both regression and classification problems. The XGBoost layer provides an output of the time distributed dense layer as input and performs the final prediction task.

### 4.2 Hyperparameter optimization

Hyperparameter optimization is an important step in building a hybrid LSTM-XGBoost model. It involves tuning the various hyperparameters of the model to find the optimal configuration that maximizes its performance. The hyperparameters of the models are

#### LSTM hyperparameters:

**Number of LSTM units:** This parameter controls the capacity of the LSTM layer and can range from 10 to 500 or more.

#### Dropout rate:

This parameter determines the percentage of connections to be randomly dropped out during training and can range from 0 to 1.

#### Learning rate:

This parameter controls the step size of the weight updates during training and can range from 0.0001 to 0.1 or more.

#### Number of LSTM layers:

This parameter determines the number of LSTM layers in the model and can range from 1 to 5 or more.

#### XGBoost hyperparameters:

**Maximum Tree depth:** This parameter controls the depth of the decision tree and can range from 3 to 10 or more.

#### Number of estimators:

This parameter controls the number of trees in the XGBoost ensemble and can range from 10 to 500 or more.

Table – II : Overall Metrics results

Techniques	Precision (%)	Recall (%)	Accuracy (%)	F1 score (%)	RMSE	MAE
Optimized hybrid LSTM-XGBoost model (OHLSTM-XM)	97.70	98.91	97.07	97.74	0.989	0.976
LSTM-XGBoost model without optimization (LSTM-XMWO)	95.12	96.0	95.13	95.28	1.967	1.98
CNN-LSTM	95.98	95.51	94.98	94.43	2.83	2.56
LSTM	93.16	94.0	93.41	94.12	2.987	3.512
RF-SVM	91.70	93.35	91.13	91.00	3.458	3.589

**Learning rate:**

This parameter controls the step size of the weight updates during training and can range from 0.0001 to 0.1 or more.

**Subsample ratio:** This parameter determines the percentage of samples to be randomly selected for each tree and can range from 0.5 to 1.

involves incrementing the number of iterations, updating the population, and returning to updating state.

An optimal value of these hyperparameters may depend on the dataset. By tuning the hyperparameters with POA, the hybrid model can attain better model performance and improve the accuracy of predictions.

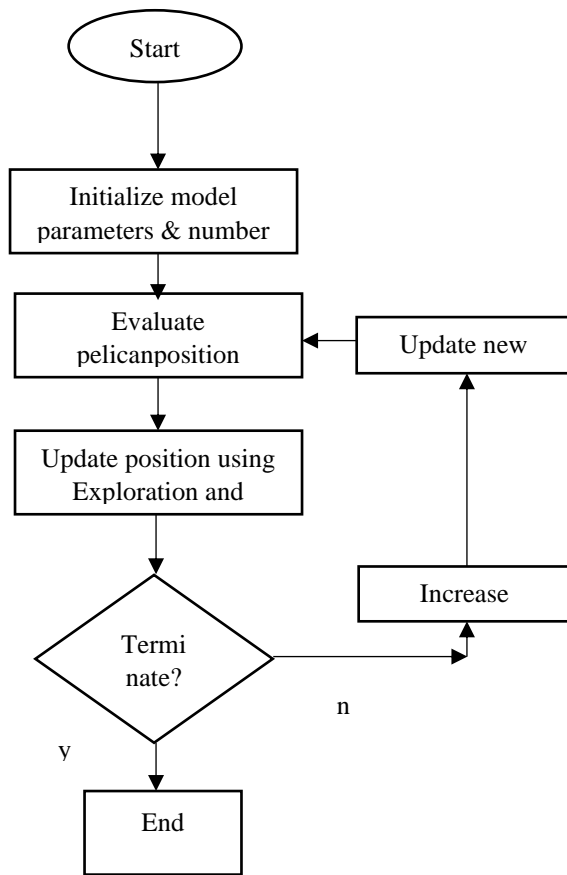


Fig. 2. Hyperparameter Tuning

The first step in this process is to initialize the model parameters. Next, the parameters for the population-based optimization algorithm (POA) are initialized, including the number of Pelican. Then, random positions and random numbers of optimization are set. Using the initial population, a LSTM-XGBoost model is created. The error rate is then calculated to evaluate the fitness function. The best solution with the least fitness in the population is

obtained. The algorithm continues to iterate through the process until the stopping criteria are satisfied, at which point it returns the best hyperparameter values. Each iteration

**V. RESULTS & DISCUSSION**

The performance result of proposed method using the optimized hybrid LSTM-XGBoost model are discussed in this section. The method was executed and evaluated for feature extraction, fusion, selection, and classification using various hybridization techniques. The WDBC dataset with 32 attributes was used for both training and testing, with 65% of the data used for training and 35% for testing. Classification performance was evaluated using precision, accuracy, F1 score, recall, root mean squared error (RMSE), and mean absolute error (MAE) as metrics. To demonstrate its effectiveness, the performance of the proposed method was compared with that of conventional methods, such as RF-SVM, LSTM, CNN-LSTM, and LSTM- XGBoost without POA.

The performance metrics are evaluated using the following equations.

$$\text{Precision} = \frac{T^+}{T^+ + F^+} \tag{6}$$

$$\text{Recall} = \frac{T^+}{T^+ + F^-} \tag{7}$$

$$\text{Accuracy} = \frac{T^+ + T^-}{T^+ + F^+ + T^- + F^-} \tag{8}$$

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Recall} + \text{precision}} \tag{9}$$

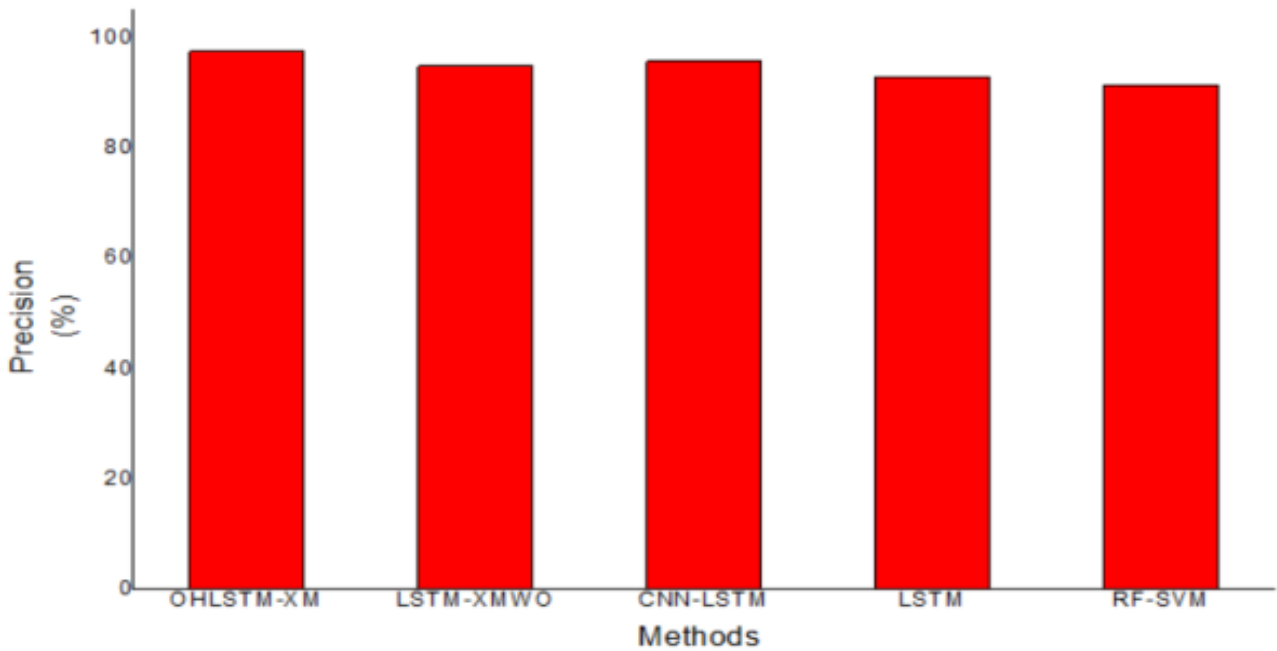
$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{a=1}^n (\text{predicted cases} - \text{overall cases})^2} \tag{10}$$

$$\text{MAE} = \sqrt{\frac{1}{n} \sum_{a=1}^n |\text{predicted cases} - \text{overall cases}|} \tag{11}$$

Where, T<sup>+</sup> represents the accurately positive, T<sup>-</sup> denotes the accurately negative, F<sup>+</sup> indicates the inaccurate positive result, F<sup>-</sup> denotes the inaccurate negative result, n indicates overall patients.

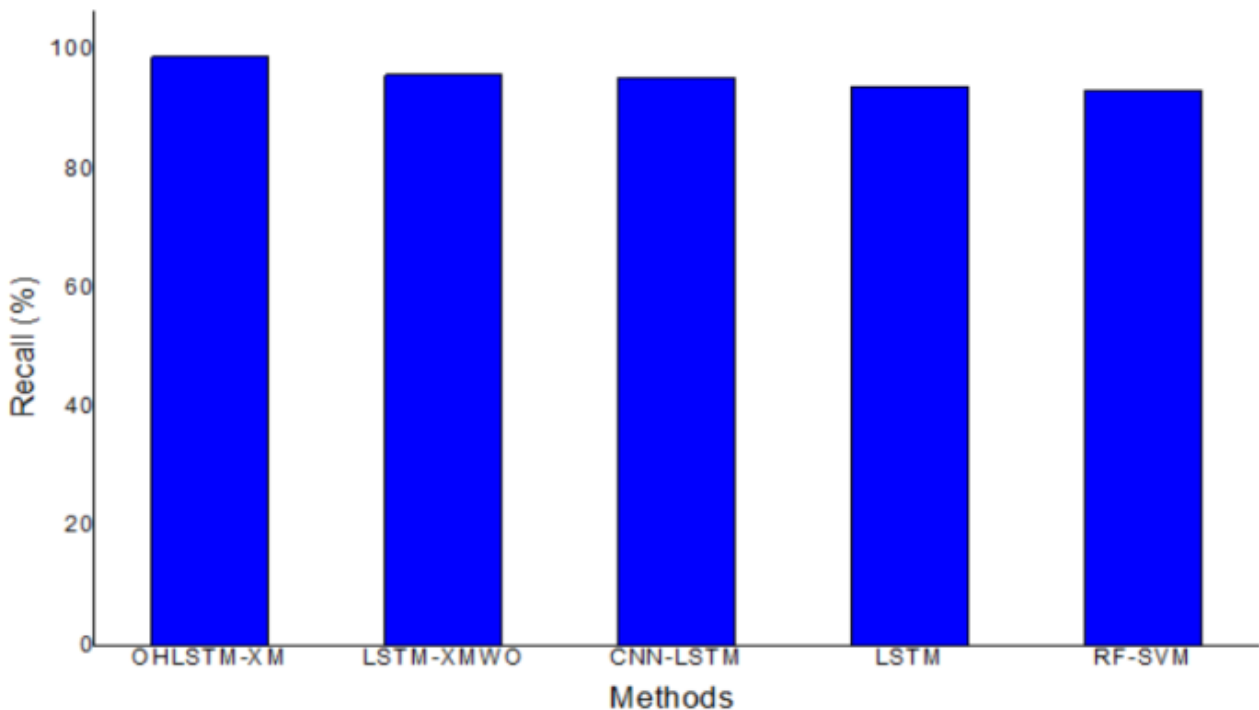
Table 2 displays the classification outcomes for both the proposed hybrid method and the conventional methods. The precision, recall, accuracy, and F1 score of the proposed hybrid method are 97.70%, 98.91%, 97.07%, and 97.74%, respectively. These results demonstrate that the proposed model outperforms the conventional methods in terms of classification effectiveness.

## Breast Cancer Prediction Based on Feature Extraction using Hybrid Methodologies



**Fig. 3. Precision analysis**

The precision results of the proposed hybrid method and the conventional methods, OHLSTM-XM, LSTM-XMWO, CNN-LSTM, LSTM, and RF-SVM, are depicted in Figure 3. The precision scores attained by these techniques are 97.70%, 95.12%, 95.98%, 93.16%, and 91.70%, respectively. The proposed hybrid method outperformed better than the conventional methods in terms of precision.



**Fig. 4. Recall analysis**

The recall results of hybrid method and the conventional methods, OHLSTM-XM, LSTM-XMWO, CNN-LSTM, LSTM, and RF-SVM, are presented in Figure 4. The recall scores attained by these techniques are 98.91%, 96%, 95.51%, 94%, and 93.55%, respectively. The results showed that the hybrid method performed well than the conventional methods in terms of recall.

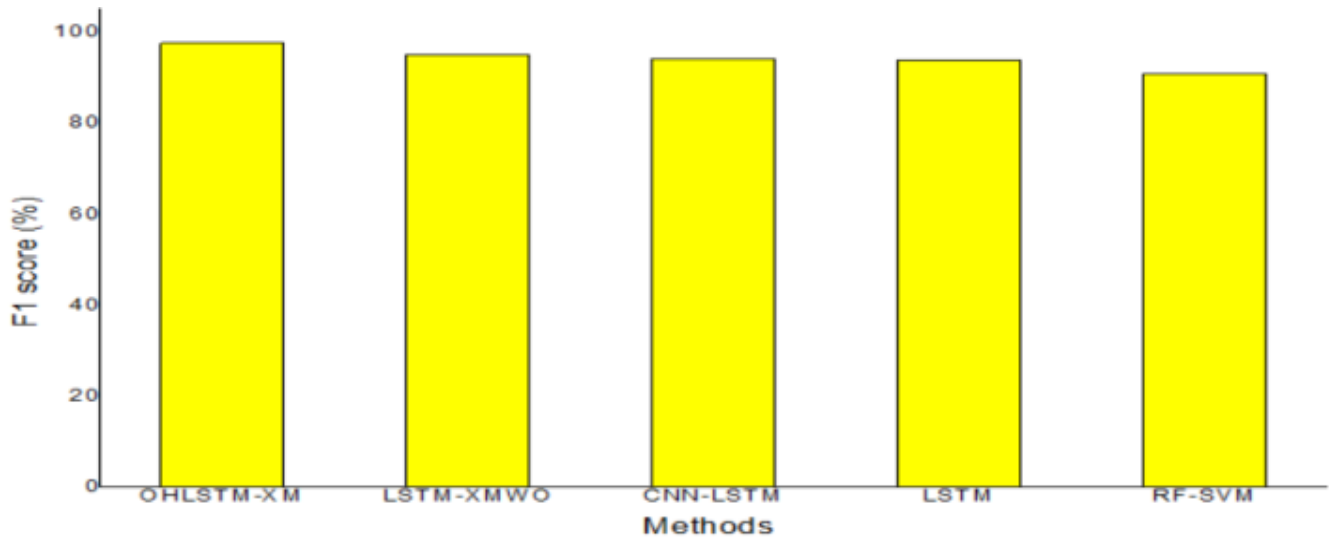


Fig. 5. Accuracy Analysis

Figure 5 illustrates the accuracy results of the proposed hybrid method and the conventional methods, OHLSTM-XM, LSTM-XMWO, CNN-LSTM, LSTM, and RF-SVM. The accuracy scores attained by these techniques are 97.07%, 95.13%, 94.98%, 93.41%, and 91.13%, respectively. Therefore, proposed method attained maximum accuracy than existing.

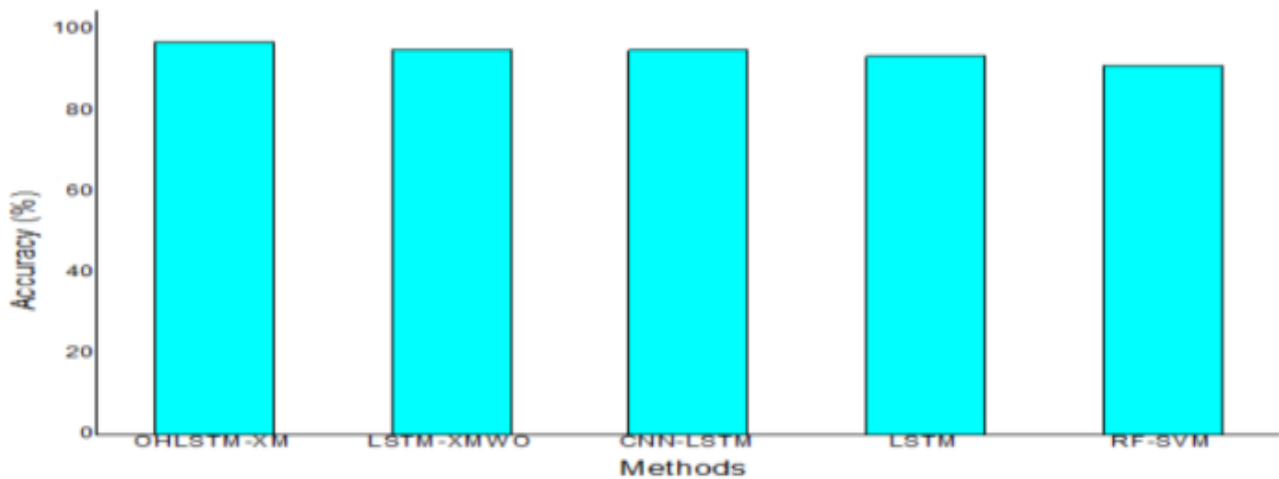


Fig. 6. F1-Score Rate Analysis

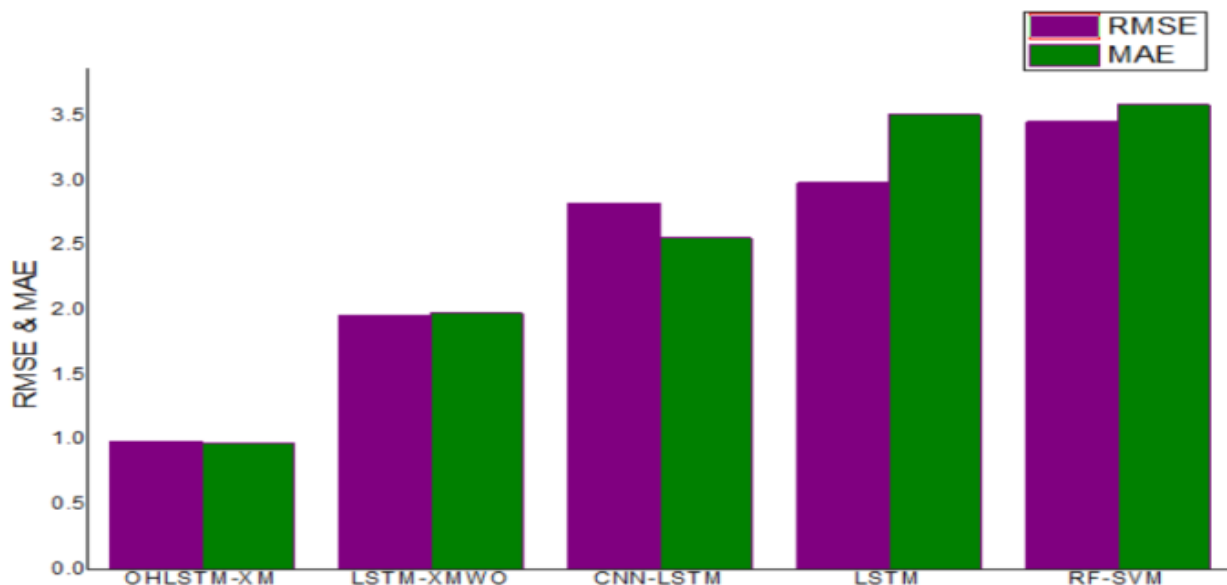


Fig. 7. Error Rate Analysis

# Breast Cancer Prediction Based on Feature Extraction using Hybrid Methodologies

According to Figure 6, the F1-score resulted a hybrid method and the conventional methods, OHLSTM-XM, LSTM-XMWO, CNN-LSTM, LSTM, and RF-SVM, are presented. The F1-score values achieved by these techniques are 97.77%, 95.28%, 94.43%, 93.41%, and 91.13%, respectively. These results suggest that the proposed method outperformed an existing methods in terms of F1-score.

Figure 7 displays the RMSE and MAE validation results for both the proposed hybrid method and the conventional methods. The results demonstrate that the proposed hybrid method has achieved lower RMSE and MAE errors compared to the conventional methods, indicating its superior performance.

Finally, the results showed that the proposed hybrid model achieved a high accuracy of 97.07 %, outperforming other traditional and deep learning models. The LDA based feature extraction technique helped to identify the most relevant features from the input data, which led to better model performance. The hyperparameter tuning of both LSTM and XGBoost further improved the performance of the hybrid model. These results suggest that the proposed system has the potential to be a valuable tool in breast cancer diagnosis and treatment.

## VI. CONCLUSION

In recent years, machine learning techniques have been widely used for breast cancer prediction, and in this work, discussed two such techniques: LDA based feature extraction and hyperparameter tuned LSTM-XGBoost based hybrid model. In the context of breast cancer prediction, LDA can be used to extract the most relevant features from the input data, which can then be used to train a classifier for breast cancer prediction. The hyperparameter tuned LSTM-XGBoost based hybrid model used for accurate prediction. By optimizing the hyperparameters of both models using cross-validation techniques, the hybrid model can achieve high accuracy in breast cancer prediction. Overall, both LDA based feature extraction and hyperparameter tuned LSTM-XGBoost based hybrid modeling are promising techniques for prediction in breast cancer.

## DECLARATION

Funding/ Grants/ Financial Support	No, I did not receive.
Conflicts of Interest/ Competing Interests	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval and consent to participate with evidence.
Availability of Data and Material/ Data Access Statement	Not relevant.
Authors Contributions	All authors have equal participation in this article.

## REFERENCES

- Jemal A, Murray T, Ward E, Samuels A, Tiwari RC, Ghafoor A, Feuer EJ, Thun MJ. Cancer statistics, 2005. CA: a cancer journal for clinicians. 2005 Jan 1;55(1):10-30. [\[CrossRef\]](#)
- M. L. Santilli et al., "Self-Supervised Learning For Detection Of Breast Cancer In Surgical Margins With Limited Data," 2021 IEEE

- 18th International Symposium on Biomedical Imaging (ISBI), Nice, France, 2021, pp. 980-984, doi: 10.1109/ISBI48211.2021.9433829. [\[CrossRef\]](#)
3. Fu, P. Liu, J. Lin, L. Deng, K. Hu and H. Zheng, "Predicting Invasive Disease-Free Survival for Early Stage Breast Cancer Patients Using Follow-Up Clinical Data," in IEEE Transactions on Biomedical Engineering, vol. 66, no. 7, pp. 2053-2064, July 2019, doi: 10.1109/TBME.2018.2882867. [\[CrossRef\]](#)
4. H. Kutrani and S. Eltalhi, "Decision Tree Algorithms for Predictive Modeling in Breast Cancer Treatment," 2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA), Sabratha, Libya, 2022, pp. 223-227, doi: 10.1109/MI-STA54861.2022.9837762. [\[CrossRef\]](#)
5. Yifan, L. Jialin and F. Boxi, "Forecast Model of Breast Cancer Diagnosis Based on RF-AdaBoost," 2021 International Conference on Communications, Information System and Computer Engineering (CISCE), Beijing, China, 2021, pp. 716-719, doi: 10.1109/CISCE52179.2021.9445847. [\[CrossRef\]](#)
6. M. Sabha and B. Tugrul, "Breast Cancer Prediction Using Different Classification Algorithms with Various Feature Selection Strategies," 2021 5th International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia, 2021, pp. 18-23, doi: 10.1109/ICICoS53627.2021.9651867. [\[CrossRef\]](#)
7. O. Jessica, M. Hamada, S. I. Yusuf and M. Hassan, "The Role of Linear Discriminant Analysis for Accurate Prediction of Breast Cancer," 2021 IEEE 14th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoc), Singapore, Singapore, 2021, pp. 340-344, doi: 10.1109/MCSoc51149.2021.00057. [\[CrossRef\]](#)
8. R. R. Rath, S. K. Swain, K. Pooranapriya, J. K. M. Deivakani and V. Avasthi, "Breast Cancer Prediction using Deep Network Model with Multi-Modal Data Fusion," 2022 International Conference on Edge Computing and Applications (ICECAA), Tamilnadu, India, 2022, pp. 1192-1197, doi: 10.1109/ICECAA55415.2022.9936361. [\[CrossRef\]](#)
9. S. Pravesjit, P. Longpradit, K. Kantawong, R. Pengchata and N. Oul, "A Hybrid PSO with Rao Algorithm for Classification of Wisconsin Breast Cancer Dataset," 2021 2nd International Conference on Big Data Analytics and Practices (IBDAP), Bangkok, Thailand, 2021, pp. 68-71, doi: 10.1109/IBDAP52511.2021.9552152. [\[CrossRef\]](#)
10. S. Kayikci and T. Khoshgoftaar, "A Stack Based Multimodal Machine Learning Model for Breast Cancer Diagnosis," 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Ankara, Turkey, 2022, pp. 1-5, doi: 10.1109/HORA55278.2022.9800004. [\[CrossRef\]](#)
11. P. Liu, B. Fu, S. X. Yang, L. Deng, X. Zhong and H. Zheng, "Optimizing Survival Analysis of XGBoost for Ties to Predict Disease Progression of Breast Cancer," in IEEE Transactions on Biomedical Engineering, vol. 68, no. 1, pp. 148-160, Jan. 2021, doi: 10.1109/TBME.2020.2993278. [\[CrossRef\]](#)
12. R. Lupat, R. Perera, S. Loi and J. Li, "Moanna: Multi-Omics Autoencoder-Based Neural Network Algorithm for Predicting Breast Cancer Subtypes," in IEEE Access, vol. 11, pp. 10912-10924, 2023, doi: 10.1109/ACCESS.2023.3240515. [\[CrossRef\]](#)
13. [Nathiya S. S](#) and S. G., "SVM-ANN Optimized Algorithm for the Classification of Breast Cancer Data as Benign and Malignant," 2022 Smart Technologies, Communication and Robotics (STCR), Sathyamangalam, India, 2022, pp. 1-7, doi: 10.1109/STCR55312.2022.10009301. [\[CrossRef\]](#)
14. C. -H. Yang, S. -H. Moi, M. -F. Hou, L. -Y. Chuang and Y. -D. Lin, "Applications of Deep Learning and Fuzzy Systems to Detect Cancer Mortality in Next-Generation Genomic Data," in IEEE Transactions on Fuzzy Systems, vol. 29, no. 12, pp. 3833-3844, Dec. 2021, doi: 10.1109/TFUZZ.2020.3028909. [\[CrossRef\]](#)
15. N. Arya and S. Saha, "Multi-Modal Classification for Human Breast Cancer Prognosis Prediction: Proposal of Deep-Learning Based Stacked Ensemble Model," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 19, no. 2, pp. 1032-1041, 1 March-April 2022, doi: 10.1109/TCBB.2020.3018467. [\[CrossRef\]](#)





16. M. Yang, Y. Han, C. -S. Liu, J. -H. Wu and D. -B. Hua, "D-TSVR Recurrence Prediction Driven by Medical Big Data in Cancer," in IEEE Transactions on Industrial Informatics, vol. 17, no. 5, pp. 3508-3517, May 2021, doi: 10.1109/TII.2020.3011675. [[CrossRef](#)]
17. Sun, M. Wang and A. Li, "A Multimodal Deep Neural Network for Human Breast Cancer Prognosis Prediction by Integrating Multi-Dimensional Data," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 16, no. 3, pp. 841-850, 1 May-June 2019, doi: 10.1109/TCBB.2018.2806438. [[CrossRef](#)]
18. R. Mendonca-Neto, Z. Li, D. Fenyö, C. T. Silva, F. G. Nakamura and E. F. Nakamura, "A Gene Selection Method Based on Outliers for Breast Cancer Subtype Classification," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 19, no. 5, pp. 2547-2559, 1 Sept.-Oct. 2022, doi: 10.1109/TCBB.2021.3132339. [[CrossRef](#)]
19. M. Pouryahya et al., "wCluster: A Novel Integrative Network-Based Clustering of Multiomics for Subtype Analysis of Cancer Data," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 19, no. 3, pp. 1472-1483, 1 May-June 2022, doi: 10.1109/TCBB.2020.3039511. [[CrossRef](#)]
20. Li et al., "BCRAM: A Social-Network-Inspired Breast Cancer Risk Assessment Model," in IEEE Transactions on Industrial Informatics, vol. 15, no. 1, pp. 366-376, Jan. 2019, doi: 10.1109/TII.2018.2825345. [[CrossRef](#)]

### AUTHORS PROFILE



**G. Rajasekaran** is a Research Scholar in the Department of Computer Science and Engineering (CSE) in SCSVMV University, Enathur, Kancheepuram, Tamil Nadu. He was did Master of Engineering in CSE in SKP Engineering College, Tiruvannamalai, Affiliated to Anna University, Chennai during 2009-2011. He was did Bachelor of Engineering in CSE, Vellore Institute of Technology, Vellore, Affiliated to University of Madras during 2000-2003.



**Dr. C. Sunitha Ram** Working as Assistant Professor in the Department of Computer Science and Engineering (CSE) in SCSVMV University, Enathur, Kancheepuram, Tamil Nadu. Her educational qualification is B.E (CSE), M.E (CSE), Ph.D., Her area of interest is Pattern Recognition in Compiler Design and Computational Intelligence. Her published paper in various international journals are IJER, IJAER, IJRASE and IJARCET.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)/ journal and/or the editor(s). The Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP) and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.