INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
ISSN 2320-7345

PREDECTIVE ANALYTICS FRAMEWORK FOR INTELLIGENT DECISION SUPPORT SYSTEM TO DIABETES PREDICTION USING ACTIVE LEARNING

[1,2]Asaad Elsaadany Bedeir, and [3]Mohamed M. El Hadi

[1]Ph.D. Candidate in Sadat Academy for Management Sciences (SAMS), Department of Computer Sciences and Information Systems Sadat Academy for Management Sciences, Cairo, Egypt

[2]Department of Computer and Information Systems, Delta Higher Institute for Computers, 3568, Egypt

[3]Professor of Computer and Information Systems Department, Sadat Academy for Management Sciences (SAMS), Cairo, Egypt

Email: [1]asaad@dhic.edu.eg

Email: [3]mohamed.m.elhadi@gmail.com

Abstract: — currently, medical diagnoses of diseases based on Artificial Intelligence (AI) and Machine Learning (ML) algorithms are more interesting and required for rapid and precise detection. Furthermore, the Intelligent Decision Support System (IDSS) of healthcare systems using big data analytics is more helpful and effective. The utilization of Active Learning (AL) approaches is more crucial to enhance the diagnosis process for diabetes by integrating the experience and feedback of human–experts with lightly available labeled data. In this paper, we presented multi-labeled active learning algorithms with both classification and selection strategies to detect diabetes disease. Moreover, the proposed framework utilized four active learning algorithms called AUDI, RANDOM, MMC, and Adaptive. These AL algorithms are promising for multi-label selection of diabetes because of their ability to reduce the cost of querying the labeled selected data based on AL strategies. The evaluation results of the proposed AL framework prove the ability to classify diabetes with and without optimization of hyper-parameter values based on a grid search algorithm. The result indicated that for the optimal label ranking model, the selection approach is used over others due to accuracy in generalization of the learning model beyond the current data. In terms of the Recall without grid search optimization parameters, however, the selection technique was highlighted.

Keywords—***IDSS; Big Data Analytics; Active learning; Health Care; Diabetes prediction.***

I. INTRODUCTION

Healthcare systems are widely increased to solve and assist the specialist in the diagnosis, detection, and classification the diseases. As well as the utilization of Fog Computing (FC) with the Internet of Things (IoT) is promising to handle the variability and multi-labeled data[1]. Due to their success in a variety of industries, active learning procedures are growingly becoming innovative and interesting tools for interpreting healthcare data[2].

Unlabeled data is more popular than labeled data in the real world, according to studies. Because label gaining is often costly due to the sharing of human professionals, training a perfect prediction model with a major number of labeled examples is critical. To avoid this problem, AL chooses just the most useful examples for session tasks. One of the most prevalent approaches for generating information from poorly labeled data is active learning [3]–[5]. Its goal is to reduce the amount of effort spent annotating data by finding the most relevant instances. There are many unlabeled data, but labeling it is too expensive. When creating an active learning system, deciding on suitable criteria for selecting which cases are important querying is critical[6].

Accurately forecasting the development of diabetes disease is one of the most important and difficult challenges in modern medicine. In many developed countries, diabetes disease is the major cause of death. Especially pregnant females are more vulnerable to diabetes[7]–[9]. Currently, during the COVID-19 pandemic the study of machine learning techniques and their effect on the health of the patients based on regression and classification approaches are more interesting[10]–[14].

There are different methodologies and algorithms to predict different diseases called the predictive learning algorithms such as Linear Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), and Deep Neural Networks (DNN). The main purpose of these algorithms is to classify, predict, and analysis of the data. The utilization of ML-based analyses is to discover the relationships and correlation of the data for necessary training and testing. Therefore, when the information is abundant, and the labels are complex with maximum elapsed time, it is more appreciated to use AL algorithms for this purpose.

The proposed model in this research attempts to overcome the difficulty of memorizing learning models. If the model does not over fit the data instances, active learning is an excellent strategy. However, only train on samples that have a major impact on its performance. As a result, the objective was to create a model that generalizes the current facts, rather than memorization. Learning a model that performs well on training data but fails on new data is frequent in many circumstances, especially in high-dimensional environments. When the model undergoes "over fitting" or "under fitting" of the training data, the traditional ML approaches will not be able to classify and diagnose the diseases correctly. (i.e., it has simply memorized/unmemorized the data). As a result, five multi-label active learning selection procedures are used. MMC, Random, Adaptive, and AUDI are the five approaches. In each technique, the grid-search with the label ranking classifier is used as predictive modeling. There have been several distinct circumstances in which the system can be turned off. The AL technique is usually done numerous times as a rule of thumb (number of iterations). A test set and an evaluation metric are used to assess the performance of the base classifier. The complete model was tested on a dataset. Hence, the major highlighted contribution of this work is investigated as follows in the following points.

1. Employing AL for diagnosing diabetes also improves generalization over memorizing of the applied model;
2. Statistical analysis of diabetes data to facilitate the visualization process.
3. A comparison study of the four AL selection strategic methodologies via traditional ML algorithms was conducted.
4. The utilization of the grid search method to optimize the learned model's hyper parameters.

The remaining the paper structure is presented as follows; section II illustrates the related work of the current study; the proposed methodologies are investigated in section III; the experimental results and discussion is validated in section IV; finally, the conclusion and future directions are presented in section V.

II. RELATED WORK

Active learning is the more generalized methodology that can determine precisely the features effect compared with other learning algorithms. In the traditional Machine Learning (ML) approaches such as Support Vector Machine (SVM), Neural Networks (NNs), and Back Propagation (BP), Long Short Term Memory (LSTM) are memorization algorithms. The ability to transfer the memorization output to generalization output base AL techniques is more appreciated for diagnosing diabetes diseases. Therefore, in this paper, we highlighted and tracked the recent efforts for classifying diabetes based on the dataset.

Ordas et al. [7] presented diabetes detection algorithm-based DL and data augmentation using Vibrational Auto Encoder (VAE) to classify the diabetes dataset and they achieved an accuracy reached to 92.31% compared with the traditional Convolutional Neural Networks (CNN). A Fine Tuning Fuzzy KNN (TFKNN) classifier presented by Salem et al. [15] is introduced to diagnose diabetes for uncertainty membership function. They obtained 90.63%, 85.00%, 93.18%, and 94.13%, in terms of accuracy, specificity, precision, and the Area Under the Curve (AUC), respectively. Furthermore, they measured the time complexity of the K-Nearest Neighbor (KNN), and TFKNN) with a promising accuracy and minimum time complexity.

Moreover, a comparative review based on data mining algorithms is presented by Khan et al [16] topredict and detect diabetes. They recommended that for precise detection of diabetes, a hybrid technique is essentially required in the preprocessing stage to facilitate the diagnosis process. Furthermore, they recommended that dimensionality reduction of the applied features be required to decrease the elapsed time of the classification process. Elsadek et al [17] present a data mining technique to early detect diabetes disease using Random Forest (RF) and Multi-Layer Perceptron (MLP). Their dataset contains 16 attributes and 250 instances and the average accuracy achieved was 97.09% using both RF and MLP.

Chowdary and Kumar [9] present an effective method to detect diabetes based on Convolutional Long Short Term Memory (CLSTM) applied to the dataset and the results obtained are 78.60%, 78.00%, 77.20%, 82.00%, 72.00%, 93.70%, and 95.60% using Naïve Bayes (NB), SVM, MLP, K-means, LSTM, and CLSTM, respectively. Besides that SVM, Logistic Regression (LR), and Random Forest (RF) presented by Sangien et al[8] are utilized to detect 764 samples with 534 training and the remaining 230 for testing. Their algorithm used binary classification to classify dataset into diabetes and non-diabetes. They achieved 80.00%, 79.00%, and 78.00% accuracy for SVM, LR, and RF, respectively.

III. PROPOSED METHODOLOGY

A. *Fog Computing*

Generally, Fog Computing (FC) is a computing distributed paradigm that operates at the intermediary node between the client layer and the cloud layer in which low-power embedded processors are used[18]. The FC system operates at an intermediate layer between the user system and the cloud system. More precisely, in the proposed work the FC includes data segregation, categorization, prediction, and assessment. A general health service provided by the FC can be illustrated in Figure 1[19].

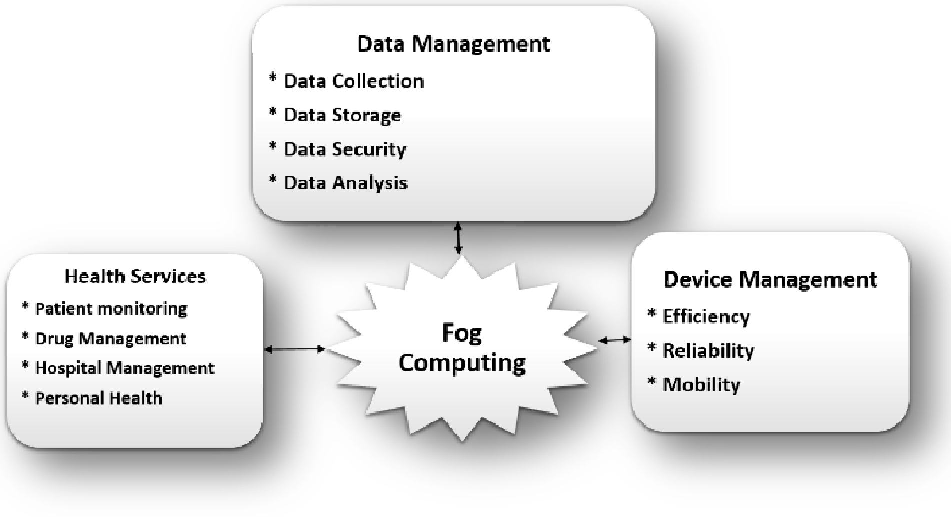

FIGURE 1. THE FOG COMPUTING LAYER FOR HEALTH SERVICES

The general structure of the proposed framework shown in Figure 2investigated that there are different Fog Computing layers has the following characteristics:

- There are three Fog Computing layers (FC), which operate as an intermediate layer between both users with data sources and the cloud computing layer.
- FC is a decentralized computing infrastructure that contains data, storage, computing, and applications
- FC enables short-term analysis, while cloud computing performs the long-term analysis.
- FC collects the data generated from different devices and sensors.

B. *Predictive Analytics Using Active Learning Algorithm*

Predictive analytics is the process of modeling historical electronic health records for predicting future events. Predictive modeling involves multiple steps. First, define the prediction target. Second, defining all the potentially relevant patients for this study, and third selecting which features are relevant for predicting the

target. Fourth, computing the model that maps the input features of the patient to the output target, and fifth we evaluate the model.
The proposed framework initially dealt with the combined dataset to produce the aggregated database. Then, the data are labeled and stored in the database that receives the newel updated labeled datasets.

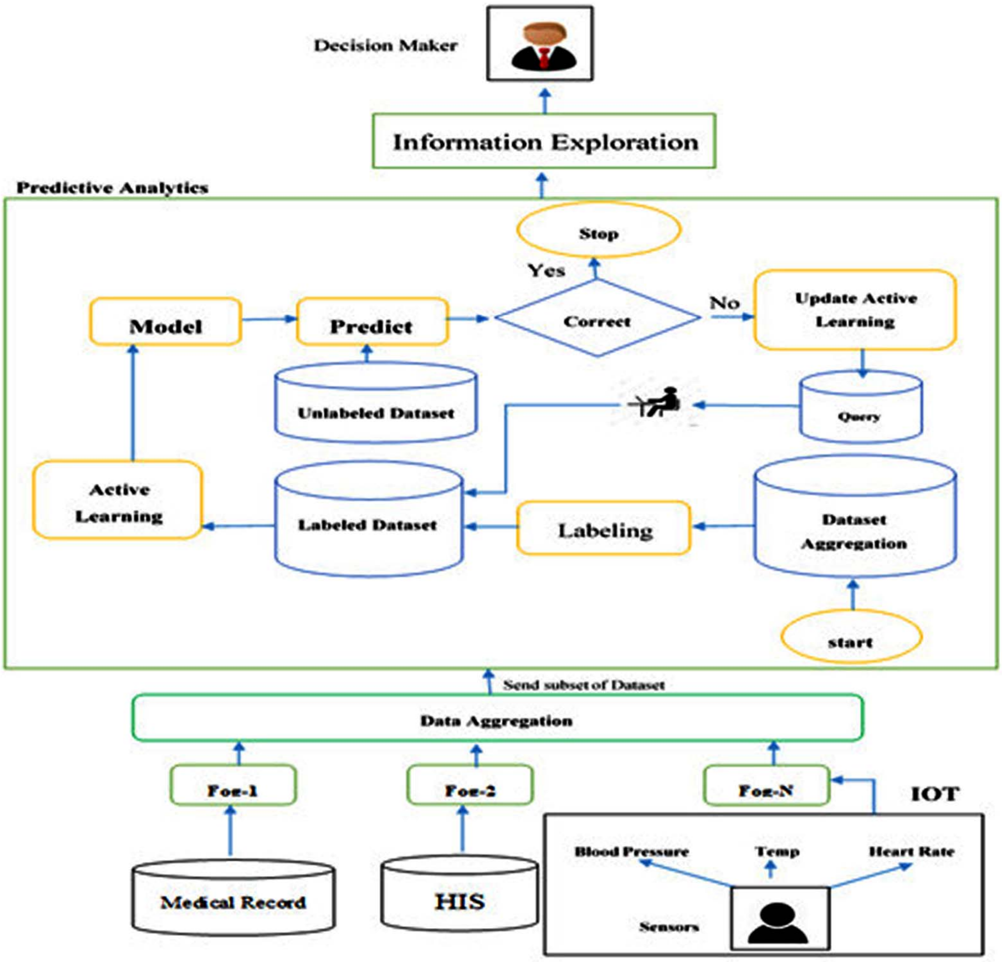

FIGURE 2. THE GENERAL STRUCTURE OF THE PROPOSED FRAMEWORK

This process is automatically repeated until the predetermined endpoints are realized. The model predicts the features of diabetes using an unlabeled dataset that determines the matched class labels and then updates the active learning process to the query, which produces the labeled database.
The construction of the acceptable criteria for determining the most valuable instances for the query is one of the most important components of active learning algorithms.
In this case, when two query selection criteria used by active learning are applied, then the active learning will reduce the uncertainty of unlabeled data. Hence, the newel-updated instances are more informative to ensure the ability of the active learning model to minimize the uncertainty of the model using a statistical approach, which accurately represents the unlabeled data of the input patterns.

C. Active Learning Strategies

There are several active learning strategies in use, including active learning with querying features,active learning with instance selection, active learning for multi-label data, and active learning at several costs. According to the diabetes disease dataset, we applied active learning algorithms for multi-label data. Active learning for multi-label data was applied in this research. In the next paragraphs, the four selection methods used (Adaptive, Random, MMC and AUDI) are defined. A comparison of the implemented algorithms. MMC algorithm, which chooses an instance to execute a loss reduction and confidence maximization query on all

labels. The random algorithm, Selects the instances or instance–label pairings at random. The adaptive algorithm, in which the maximum margin uncertainty and label cardinality inconsistency are employed to query all labels. AUDI algorithm, based on the degree of uncertainty and diversity, chooses an instance–label pair.
The strategy of the proposed active learning algorithm includes the following steps:

- Selection of instances from unlabeled data.
- Categorization of the selected instances.
- Choosing the appended instances to produce labeled data.
- Training the labeled dataset.
- Testing the labeled dataset.
- Evaluating the performance of the estimated classifier
- Repeating until ending condition.

IV. DISCUSSION AND RESULTS

A. Dataset Description

Pima Indians Diabetes Dataset(PIDD) is one of the most well-known and widely available datasets, containing a wealth of empirical data. It has been widely used in the latest state comparative research. All pregnant female patients with diabetic indications who are vulnerable to diabetes, regardless of demographic characteristics such as social position or previous surgery, were included in the study of patients with diabetes from the Dataset.
The collected dataset is publicly available at(https://www.kaggle.com/uciml/pima-indians-diabetes-database). The Dataset has 768 female diabetic patients with eight attributes, dividing it into two groups: 268 patients with diabetes and 500 normal patients without diabetes. Table I, summarize the description of the dataset structure containing the features with numeric and categorical type with a brief description. Moreover, Table II, details a statistical analysis of the applied features of the dataset including the count, mean, standard deviation (std), minimum (min), and Maximum (max) values.

B. Dataset Analysis and Visualization

To clarify the correlation between the applied features and their distribution, FIGURE3illustrates the scheming of all the features in the dataset. This Figure shows that a normal distribution is investigated in Glucose, Blood Pressure, and BMI. While other features are skewed positively and negatively.
This is may help to discover more interesting knowledge from the applied features. For the determination of outliers, a boxplot can be utilized to verify the assumption of the applied experiments as shown in Figure 4. However, in many ML models, feature standardization is unlikely to guarantee significant improvement. The outcomes of outlier rejection and imputation of missing values are combined in the correlation's confusion matrix. Figure 5shows the confusion matrix heat maps, which illustrate the correlation of features based on the applied dataset before and after processing. Visual representations of correlation matrices that show the connection between numerous variables are known as these. The correlation coefficient might be any value between -1 and 1. A correlation is a statistical word describing a relationship between two variables that is linear. It can also be referred to as a two-variable correlation measure. The purpose of this scenario is to establish a link between several variables and then arrange the data. The information was stored using a matrix data structure. Figure 5describes this on a feature-by-feature basis.

TABLE I. DESCRIPTION OF DATASET STRUCTURE

Feature	Type	Description
Pregnancies	Numeric	Number of times pregnant
Glucose	Numeric	Plasma glucose concentration
Blood Pressure	Numeric	Diastolic blood pressure
Skin Thickness	Numeric	Triceps skin fold thickness
Insulin	Numeric	serum insulin
BMI	Numeric	Body mass index
Diabetes Pedigree Function	Numeric	Diabetes pedigree function
Age	Categorical	Age

Vol.10 Issue 6, Pg.: 1-13
June 2022

TABLE II. A STATISTICAL ANALYSIS OF THE DATASET.

Features	Count	Mean	Std	Min	Max
Pregnancies	768	3.845	3.37	0	17
Glucose	768	120.895	31.973	0	199
Blood Pressure	768	69.105	19.356	0	122
Skin Thickness	768	20.536	15.952	0	99
Insulin	768	79.8	115.24	0	846
BMI	768	31.993	7.88416	0	67.1
Diabetes Pedigree Function	768	0.472	0.331329	0.078	2.42
Age	768	33.241	11.76	21	81

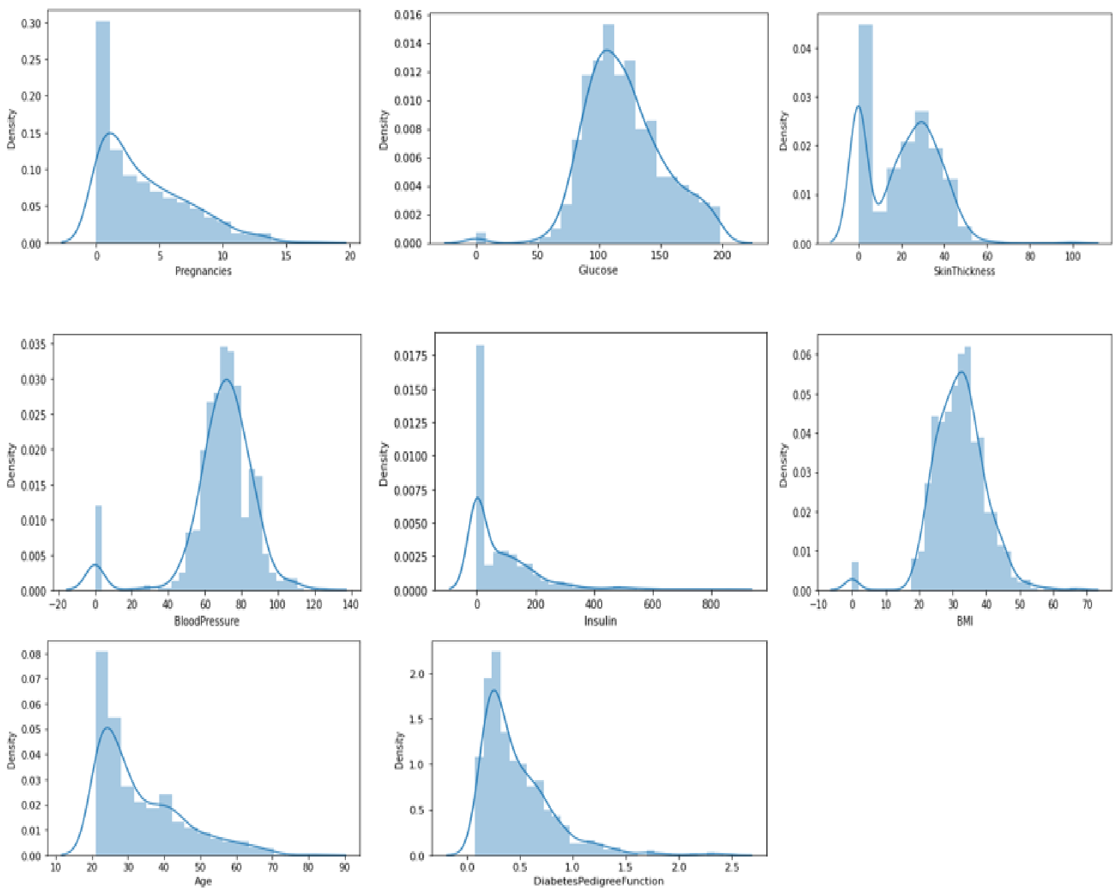

FIGURE3. THE STATISTICAL ANALYSIS DESCRIPTION OF THE DATASET

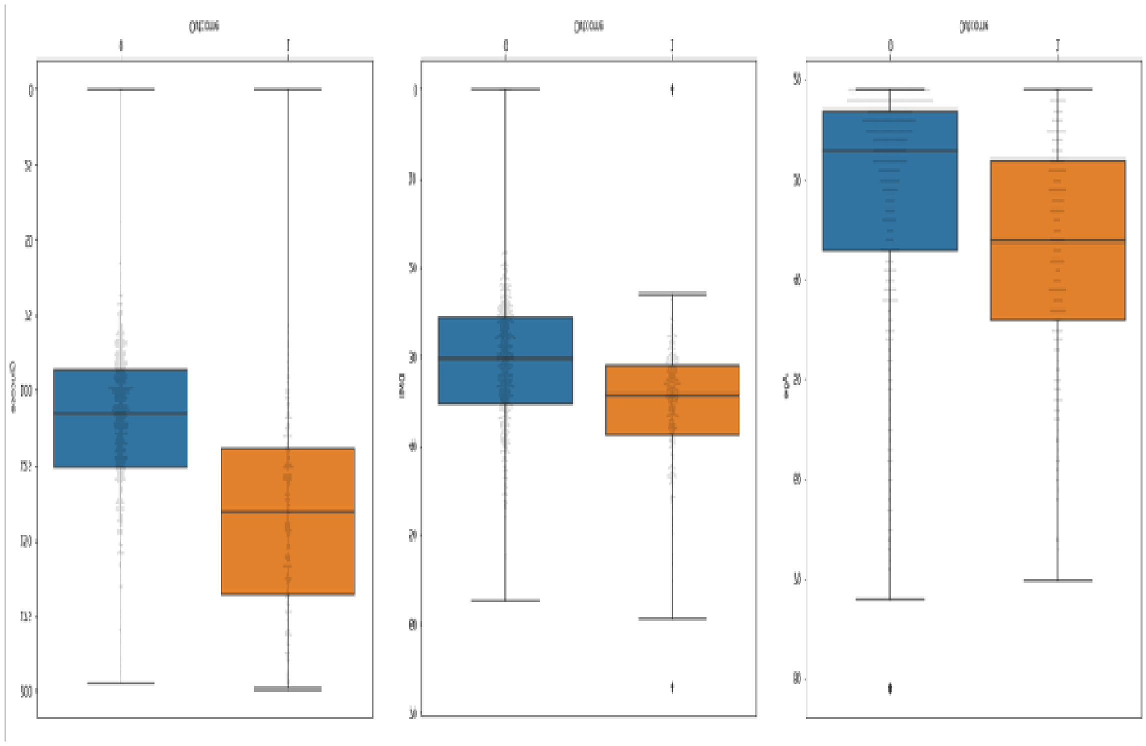

FIGURE 4. BOXPLOT SHOWING FEATURES (GLUCOSE, BMI, AGE) OF THERESULTING OUTLIERS.

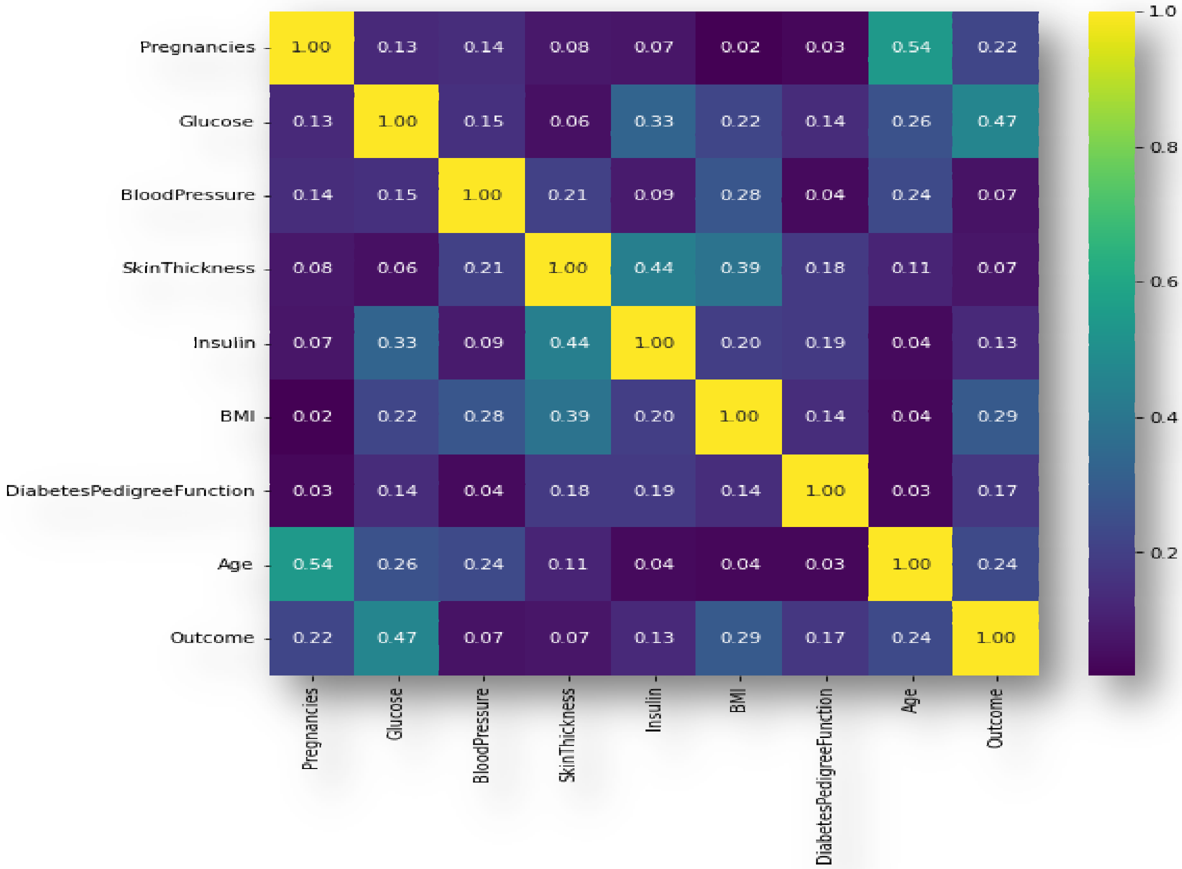

FIGURE 5. THE CONFUSION MATRIX OF DATASET FEATURES HEATMAP

C. *Results*

While the dataset is a hot medical topic, the experimental results were constructed to explore the accuracy, F-score and recall of the proposed framework to strengthen the recognition rate. The experiments consist of hyper-parameter optimization approaches based on the grid search algorithm. Now the memorization should be adapted to achieve the similarity of the new trained samples. Therefore, the generalization process can take the contest and perform more efficiently. Table III shows the average value of the four different selection AUDI, RANDOM, MMC, and Adaptive methods. We compare the achieved results using two scenarios: with and without the hyper-parameter grid search optimization process. Hence there is a slight improvement in the accuracy after hyper-parameter optimization using a grid search optimizer. Given that the number of quires and cost is 120. The results indicated that before grid search optimization the accuracies of AUDI, RANDOM, MMC, and Adaptive are 57.52%, 56.50%, 52.50%, and 60.00%, respectively. While the results achieved in terms of accuracy of AUDI, RANDOM, MMC, and Adaptive are 64.53%, 57.00%, 61.05%, and 61.22%, respectively. The results investigated in Figure 6 shows the accuracy of the proposed active learning methods before grid search optimization. While Figure 7 shows the achieved results of the proposed active learning using a grid search optimization algorithm. The results indicated that there is an enhancement of the accuracy after applying grid search optimization.

TABLE III. THE ACCURACY OF THE ACTIVE AUDI, RANDOM, MMC, AND ADAPTIVE LEARNING BEFORE AND AFTER GRID SEARCH OPTIMIZATION.

Before Grid Search Optimization		After Grid Search Optimization	
Method	Accuracy	Method	Accuracy
AUDI	57.52%	AUDI	**64.53%**
RANDOM	56.50%	RANDOM	57.00%
MMC	52.50%	MMC	61.05%
Adaptive	60.00%	Adaptive	61.22%

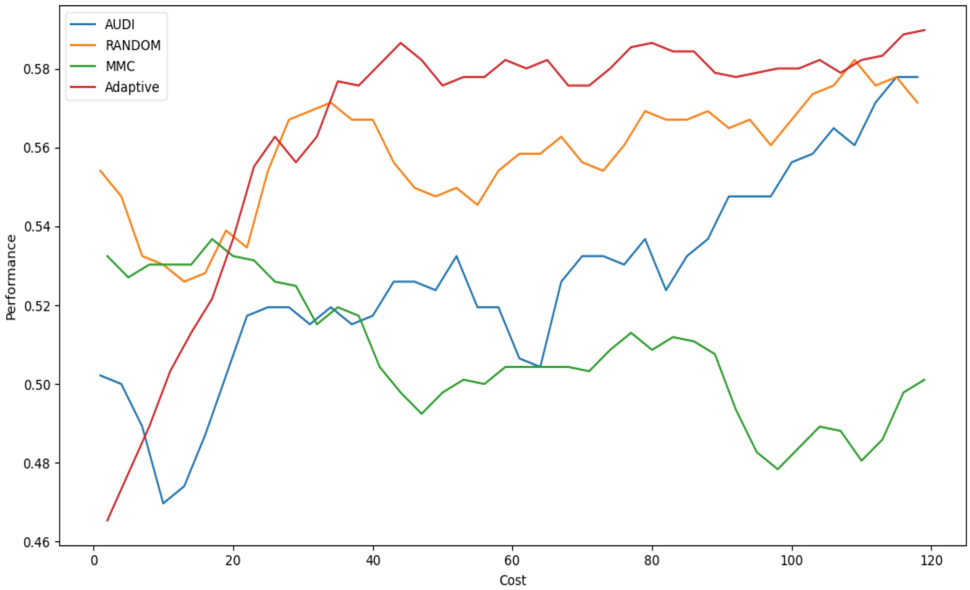

FIGURE 6. THE ACCURACY OF THE PROPOSED AUDI, RANDOM, MMC, AND ADAPTIVE FOR THE APPLIED FEATURES FOR THE COST = 20 AND WITHOUT GRID SEARCH OPTIMIZATION

Vol.10 Issue 6, Pg.: 1-13
June 2022

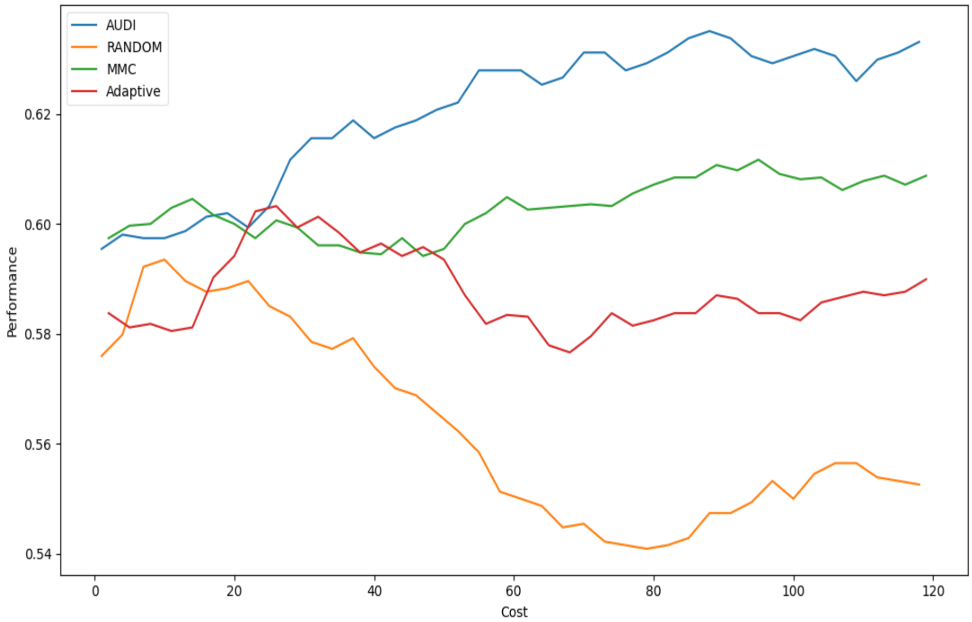

FIGURE 7. THE ACCURACY OF THE PROPOSED AUDI, RANDOM, MMC, AND ADAPTIVE FOR THE APPLIED FEATURES FOR THE COST = 20 IN THE PRESENCE OF GRID SEARCH OPTIMIZATION.

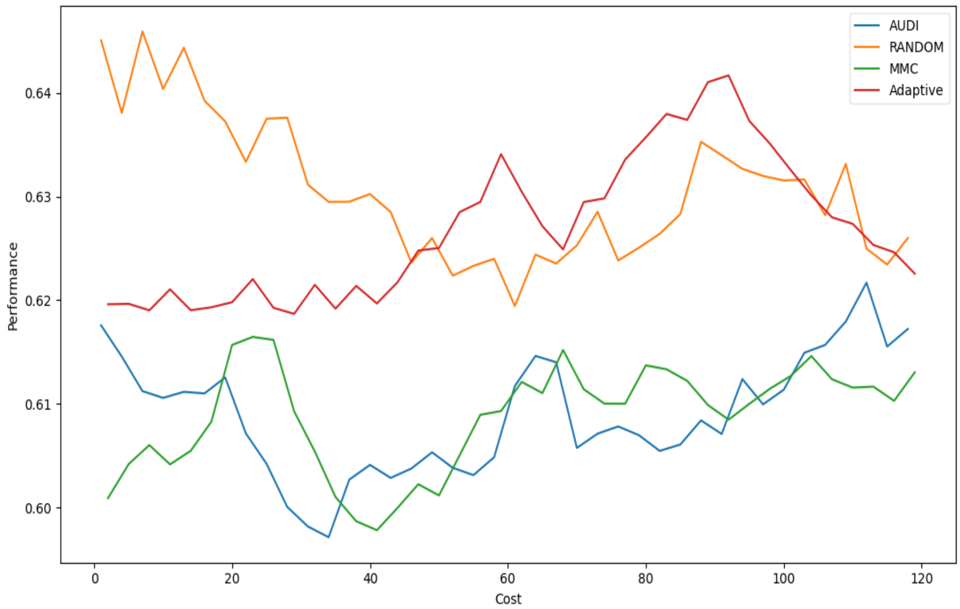

FIGURE 8. THE F1-SCORE OF THE PROPOSED AUDI, RANDOM, MMC, AND ADAPTIVE FOR THE APPLIED FEATURES FOR THE COST = 20 WITHOUT GRID SEARCH OPTIMIZATION.

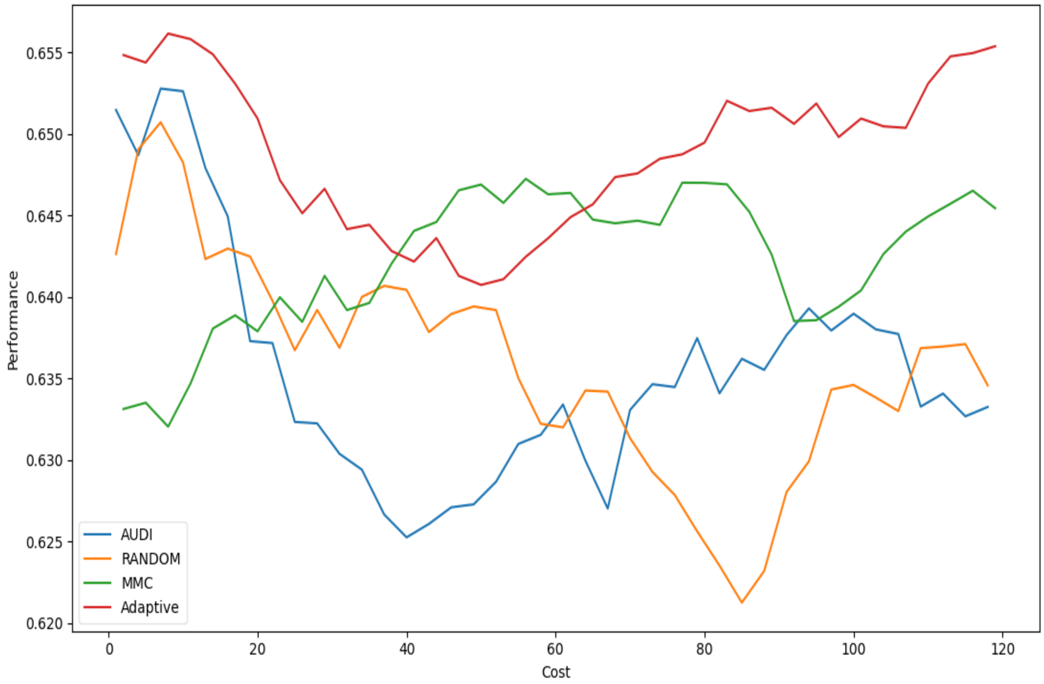

FIGURE 9. THE F1-SCORE OF THE PROPOSED AUDI, RANDOM, MMC, AND ADAPTIVE FOR THE APPLIED FEATURES FOR THE COST = 20 IN THE PRESENCE OF GRID SEARCH OPTIMIZATION.

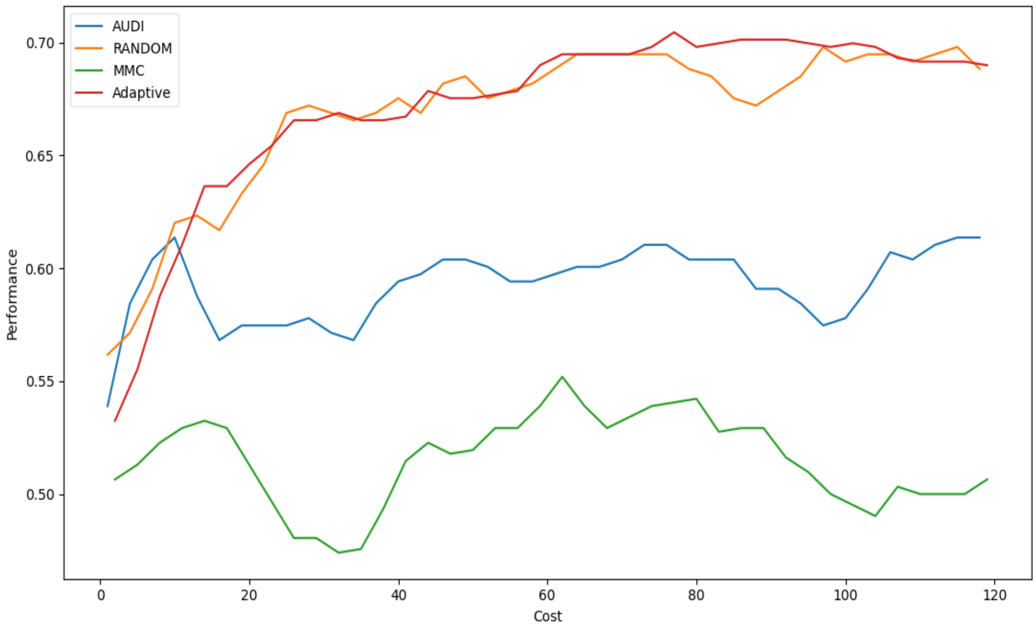

FIGURE 10. THE RECALL OF THE PROPOSED AUDI, RANDOM, MMC, AND ADAPTIVE FOR THE APPLIED FEATURES FOR THE COST = 20 WITHOUT GRID SEARCH OPTIMIZATION.

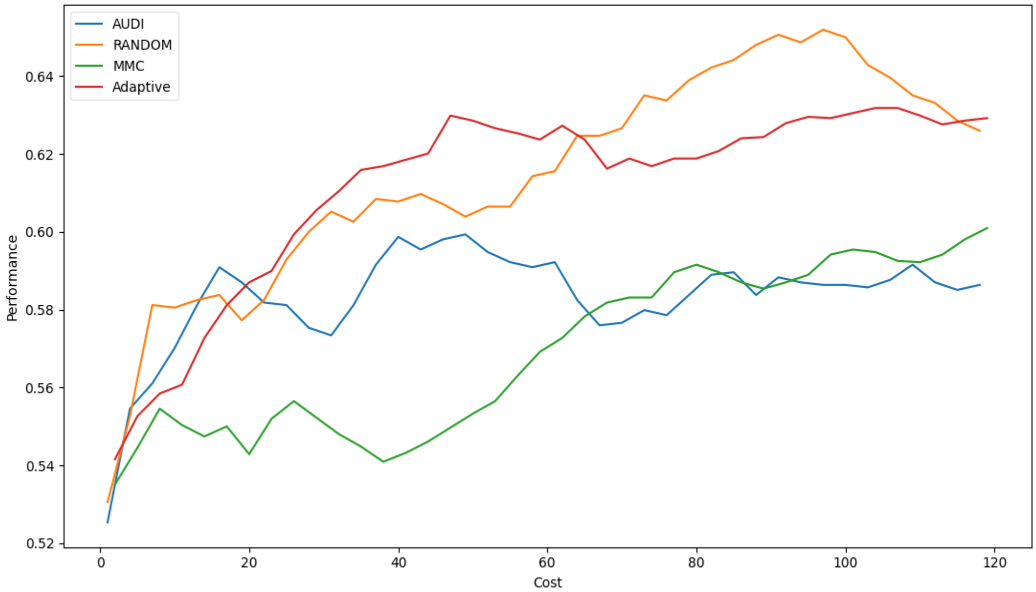

FIGURE 11. THE RECALL OF THE PROPOSED AUDI, RANDOM, MMC, AND ADAPTIVE FOR THE APPLIED FEATURES FOR THE COST = 20 IN THE PRESENCE OF GRID SEARCH OPTIMIZATION.

To ensure the reliability of the proposed active learning based on AUDI, RANDOM, MMC, and Adaptive, we further determine the F1-score and Recall using the two scenarios without and with grid search optimization as shown in Figure 8, Figure 9, Figure 10, and Figure 11, respectively. The summery of the obtained results based on F1-score and recall with and without grid search optimization Table IV.

TABLE IV. THE F1-SCORE, AND RECALL OF THE ACTIVE AUDI, RANDOM, MMC, AND ADAPTIVE LEARNING BEFORE AND AFTER GRID SEARCH OPTIMIZATION.

Before Grid Search Optimization			After Grid Search Optimization		
Method	F1-Score	**Recall**	Method	F1-Score	Recall
AUDI	61.57%	62.57%	AUDI	63.50%	59.05%
RANDOM	62.59%	**70.12%**	RANDOM	63.87%	63.71%
MMC	61.75%	53.17%	MMC	64.50%	61.50%
Adaptive	63.47%	70.00%	Adaptive	65.75%	64.05%

As shown in Table IV using the F1-score measure is considered one of the most essential evaluation measures in machine learning. It elegantly summarizes a model's prediction performance by integrating two previously competing criteria. F1-score after applying grid search optimization, there is an improvement in the performance based on AUDI, RANDOM, MMC, and the Adaptive active learning model. While the recall achieved lower results after grid search optimization. This is because recall is a metric that indicates how well the proposed model can detect relevant data. It's also known as True Positive Rate or Sensitivity. Moreover, the number of true positives divided by the total number of elements that genuinely belong to the positive class is produced lower results using grid search optimization. Overall, after grid search optimization, using active learning accuracy and F1-score are favorites compared with recall. The results reveal that for the optimal label ranking model, the choice approach is used over others due to its accuracy in generalizing the learning model beyond the present data. In the case of the F-score utilizing optimum parameters, however, the selecting method was underlined.

The effect tracking of the grid search technique's outputs was included in the overall design tests. The hyper parameters of the various active learning selection approaches were tuned using grid search. The accuracy or F1-score was used to represent the optimization's fitness function. In addition, the recall was measured to determine the best active learning selection strategy. In terms of accuracy optimization, The Adaptive approach used the fewest number of queries to compute the exact cost of MMC, and it was more significant than the others by one

Vol.10 Issue 6, Pg.: 1-13
June 2022

step, with 51.43 percent accuracy. Figure shows the comparative study of the proposed active learning model based on AUDI, RANDOM, MMC, and Adaptive without applying a grid search optimizer. While in Figure 13, a comparative study of using AUDI, RANDOM, MMC, and Adaptive in the presence of grid search optimization.

V. CONCLUSION AND FUTURE WORK

Diabetes disease is a current challenge for specialists in medical and health informatics. datasets are most commonly used to evaluate and track diabetes in female patients, especially in pregnancy. In this paper, we utilized an active learning model based on four classifications ML AUDI, RANDOM, MMC, and Adaptive methods. Through the proposed active learning, we are interested in using the generalization process instead of the memorization model. Here, in classical classification algorithms, more cost will be produced without the generalization of applied features. Therefore, this work presents four multi label active learning selection strategies, AUDI, RANDOM, MMC, and Adaptive to request the most significant iterative data to minimize the labeling cost. As well, with the utilization of a grid search optimizer, an enhancement of the accuracy and the F-score is achieved compared with the recall results. Corresponding to the results obtained, the proposed learning model generalizes further than the applied data sample. The results indicated that before grid search optimization RANDOM active learning achieves56.50%, 62.59%, and 70.12% of accuracy, F-score, and recall, respectively. While Adaptive active learning achieved F1-score 65.75% after grid search optimization. The achieved results are not promising but the generalization is also important for the specialist to tackle the cost problem and the elapsed time. In the future, a predictive model with fine-tuning the hyper parameter values using meta heuristic algorithms are essentially required to enhance the classification results of the active learning model.

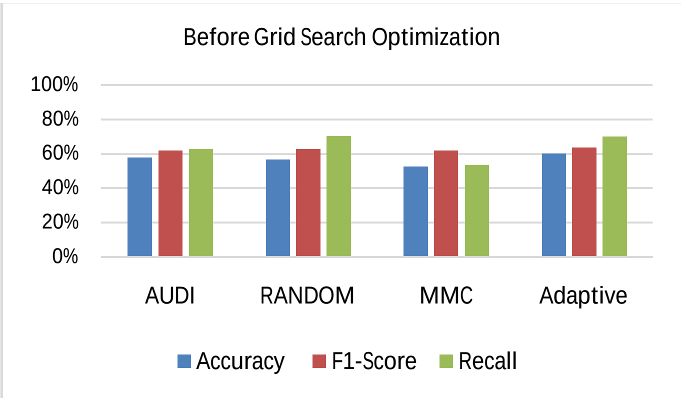

FIGURE 12. THE COMPARISON BETWEEN THE PROPOSED ACTIVE LEARNING BEFORE GRID SEARCH OPTIMIZATION.

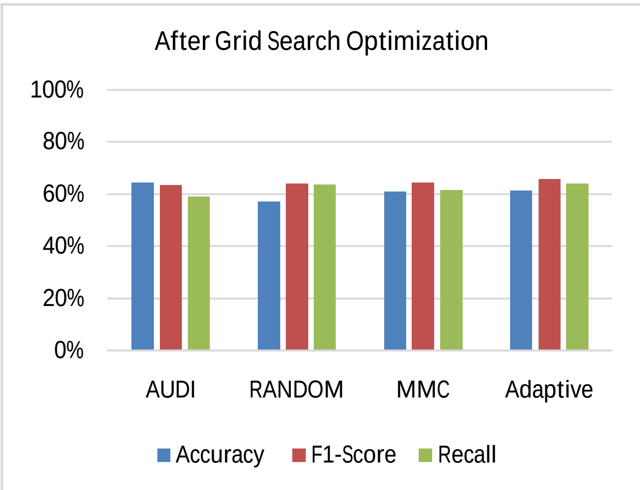

FIGURE 13. THE COMPARATIVE BETWEEN THE PROPOSED ACTIVE LEARNING AFTER GRID SEARCH OPTIMIZATION.

REFERENCES

1) [M. N. Birje and S. S. Hanji, "Internet of things based distributed healthcare systems: a review," *J. Data Inf. Manag.*, vol. 2, no. 3, pp. 149–165, 2020.
2) MG Abdelhaleem Hussein, M Ali, MY Shams, T Medhat, "HEALTHCARE SYSTEM USING A MEDICAL SERVICE ROBOT BASED ON INTERNET OF THINGS." ARPN Journal of Engineering and Applied Sciences, 2021
3) I. M. El-Hasnony, O. M. Elzeki, A. Alshehri, and H. Salem, "Multi-label active learning-based machine learning model for heart disease prediction," *Sensors*, vol. 22, no. 3, p. 1184, 2022.
4) R. A. Hunter, R. R. Pompano, and M. F. Tuchler, "Alternative Assessment of Active Learning," in *Active Learning in the Analytical Chemistry Curriculum*, ACS Publications, 2022, pp. 269–295.
5) A. M. Wiles and S. L. Simmons, "Establishment of an engaged and active learning community in the biology classroom and lab with Discord," *J. Microbiol. Biol. Educ.*, pp. e00334-21, 2022.
6) J. W. Cho, D.-J. Kim, Y. Jung, and I. S. Kweon, "Mcdal: Maximum classifier discrepancy for active learning," *IEEE Trans. Neural Netw. Learn. Syst.*, 2022.
7) M. T. García-Ordás, C. Benavides, J. A. Benítez-Andrades, H. Alaiz-Moretón, and I. García-Rodríguez, "Diabetes detection using deep learning techniques with oversampling and feature augmentation," *Comput. Methods Programs Biomed.*, vol. 202, p. 105968, 2021.
8) T. Sangien, T. Bhat, and M. S. Khan, "Diabetes Disease Prediction Using Classification Algorithms," in *Internet of Things and Its Applications*, Springer, 2022, pp. 185–197.
9) P. B. K. Chowdary and D. R. U. Kumar, "An Effective Approach for Detecting Diabetes using Deep Learning Techniques based on Convolutional LSTM Networks," *IJACSA Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 4, 2021.
10) M. Y. Shams, O. M. Elzeki, M. Abd Elfattah, L. M. Abouelmagd, A. Darwish, and A. E. Hassanien, "Impact of COVID-19 Pandemic on Diet Prediction and Patient Health Based on Support Vector Machine," in *Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2021*, 2021, pp. 64–76.
11) M. Y. Shams, O. M. Elzeki, L. M. Abouelmagd, A. E. Hassanien, M. Abd Elfattah, and H. Salem, "HANA: a healthy artificial nutrition analysis model during COVID-19 pandemic," *Comput. Biol. Med.*, vol. 135, p. 104606, 2021.
12) O. M. Elzeki, M. Shams, S. Sarhan, M. Abd Elfattah, and A. E. Hassanien, "COVID-19: a new deep learning computer-aided model for classification," *PeerJ Comput. Sci.*, vol. 7, p. e358, 2021.
13) O. M. Elzeki, M. Abd Elfattah, H. Salem, A. E. Hassanien, and M. Shams, "A novel perceptual two layer image fusion using deep learning for imbalanced COVID-19 dataset," *PeerJ Comput. Sci.*, vol. 7, 2021.
14) M. E. ElAraby, O. M. Elzeki, M. Y. Shams, A. Mahmoud, and H. Salem, "A novel Gray-Scale spatial exploitation learning Net for COVID-19 by crawling Internet resources," *Biomed. Signal Process. Control*, vol. 73, p. 103441, 2022.
15) H. Salem, M. Y. Shams, O. M. Elzeki, M. Abd Elfattah, J. F Al-Amri, and S. Elnazer, "Fine-Tuning Fuzzy KNN Classifier Based on Uncertainty Membership for the Medical Diagnosis of Diabetes," *Appl. Sci.*, vol. 12, no. 3, p. 950, 2022.
16) F. A. Khan, K. Zeb, M. Alrakhami, A. Derhab, and S. A. C. Bukhari, "Detection and prediction of diabetes using data mining: a comprehensive review," *IEEE Access*, 2021.
17) S. N. Elsadek, L. S. Alshehri, R. A. Alqhatani, Z. A. Algarni, L. O. Elbadry, and E. A. Alyahyan, "Early Prediction of Diabetes Disease Based on Data Mining Techniques," in *International Conference on Computational Intelligence in Data Science*, 2021, pp. 40–51.
18) R. Priyadarshini, R. K. Barik, and H. Dubey, "Deepfog: fog computing-based deep neural architecture for prediction of stress types, diabetes and hypertension attacks," *Computation*, vol. 6, no. 4, p. 62, 2018.
19) D. K. K. Reddy, H. S. Behera, J. Nayak, A. R. Routray, P. S. Kumar, and U. Ghosh, "A Fog-Based Intelligent Secured IoMT Framework for Early Diabetes Prediction," in *Intelligent Internet of Things for Healthcare and Industry*, Springer, 2022, pp. 199–218.