Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
July 2024 5(7), 139–153

AN INTELLIGENT COMPUTATIONAL APPROACHES FOR DIABETES RISK PREDICTION: A PROACTIVE HEALTHCARE PARADIGM DIABOLIC

Mr.S.Muthukumar[1], Dr.M.Jayakumar[2]

[1]Research Scholar, Dept of Computer Science, SNMV College of Arts and Science, Coimbatore, Tamil Nadu, Indias
[2]Assistant Professor, Dept of Computer Science, Hindustan College of Arts and Science, Coimbatore, Tamil Nadu, India

ABSTRACT

Diabetes mellitus, a long-term metabolic disease marked by high blood glucose levels, is a major global health concern. Diabetes must be identified and treated early to reduce complications and enhance patient outcomes. In this study, we propose a unique deep learning framework, named as, Diabetic prediction utilizing Optimized Learning Classifier (DIABOLIC) for diabetes detection.The original contribution of this paper is to develop a resilient prediction model by leveraging an advanced computational algorithms to reliably predict the probability of getting diabetes. In the proposed framework, a special Tumultuo Dwarf Mongoose Optimization (TuD-MO) technique is used to extract the most important and critical features from the preprocessed dataset. Also, a Fused Deep Convolution Random Network (FDCRN) is developed to precisely identify diabetic patients based on the selected attributes. Moreover, a detailed performance analysis is completed in order to validate and extensively explore the outcomes of the DIABOLIC model.Our test findings show that, when it comes to diabetes detection, DIABOLIC outperforms cutting-edge techniques in terms of predictive performance, with excellent sensitivity, specificity, and accuracy. In addition, we perform thorough interpretability investigations in order to clarify the underlying characteristics and processes that underlie the predictions produced by DIABOLIC. Overall, our research shows how deep learning techniques, like DIABOLIC, can improve diabetes detection and tailored healthcare plans, which will benefit public health campaigns and patient outcomes.

Keywords: Diabetes Detection, Internet of Things (IoT), Artificial Intelligence (AI), Tumultuo Dwarf Mongoose Optimization (TuD-MO) technique, Diabetic prediction utilizing Optimized Learning Classifier (DIABOLIC), and Fused Deep Convolution Random Network (FDCRN).

DOI
10.29121/shodhkosh.v5.i7.2024.1899

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

1. INTRODUCTION

Diabetes is a condition in which the body fails to metabolize blood sugar, also known as glucose, causing its metabolic rate to rise dangerously high[1, 2]. Chronic diabetes, also known as diabetes mellitus (DM) in healthcare terminology, is a group of metabolic diseases characterized by elevated blood sugar levels caused by either inefficient insulin production, inappropriate insulin uptake by body cells, or a combination of these factors. As a result, the blood's glucose concentration will rise.As a consequence, it is not just a medical condition but also a cause of various ailments, including kidney, eye, and heart disorders. In most nations, diabetes has emerged as a leading cause of disease and mortality[3].Insulin is a hormone that normally helps regulate blood glucose levels; individuals with diabetes have either type 1 diabetes, which is the lack of insulin production, or type 2 diabetes, which is the improper response to

How to cite this article (APA): Muthukumar, S., and Jayakumar, M. (2024). An Intelligent Computational Approaches For Diabetes Risk Prediction: A Proactive Healthcare Paradigm Diabolic. *ShodhKosh: Journal of Visual and Performing Arts, 5*(7), 139–153. doi: 10.29121/shodhkosh.v5.i7.2024.1899

insulin[4]. Type 2 diabetes accounts for about 90% of all cases of diabetes that are diagnosed.The total number of diabetics is predicted to increase to over 642 million in 2040, according to a report by the International Diabetes Federation. For this reason, early diabetes patient assessment and diagnosis are critical to early identification and treatment of diabetes[5, 6]. Since the majority of medical data are complicated, unpredictable, aberrant, and correlation-structured, analyzing diabetic data is a difficult task. One important strategy for making use of the vast amounts of diabetes-related data that are currently available for knowledge extraction in diabetes mellitus research is to apply machine learning techniques.Diabetes is a complex illness that is influenced by a number of variables, including environment, diet, lifestyle, and genetics. Complex correlations between these characteristics and the chance of getting diabetes, which may not be immediately obvious using traditional statistical methods, can be efficiently modelled using artificial intelligence algorithms[1, 7].From unprocessed data, it can automatically extract pertinent features such as genetic information, lifestyle factors, and medical records. In order to find diabetic risk factors and prognostic patterns that may not be immediately visible to human experts, feature extraction is a crucial step in the process.

It is possible to establish automated diabetes systems using machine learning or artificial intelligence methods. Every machine learning and artificial intelligence method has pros and cons of its own. Consequently, both strategies have been used to create automated systems for detecting diabetes. Because the most successful ML and AI systems[8-10] often show little transparency, rationality is necessary for them to behave like human beings. Future AI systems are expected to inspire greater confidence in physicians because of their capacity for explanation. Numerous researchers have used ML and AI approaches to control and customize diabetes in recent years.

Our goal in the current study is to provide a novel framework for patient data detection. The following lists the primary goals of this work:

- In this work, a novel Diabetic prediction utilizing Optimized Learning Classifier (DIABOLIC) paradigm is developed to identify individuals with diabetes.
- The essential and crucial features are extracted from the preprocessed dataset using a unique Tumultuo Dwarf Mongoose Optimization (TuD-MO) technique.
- Additionally, based on the chosen features, a Fused Deep Convolution Random Network (FDCRN) is created to accurately detect diabetic patients.
- The DIABOLIC model's results are validated and thoroughly investigated through the completion of a comprehensive performance analysis.

The following subsections comprise the remaining sections of the paper: Section 2's extensive literature analysis looks at many accepted methods for identifying and categorizing chronic illnesses. It also takes into account the benefits and drawbacks of every method in terms of how well it identifies illnesses and functions. Section 3 provides a comprehensive and lucid explanation of the proposed DIABOLIC technique, complete with a flow diagram and step-by-step explanations. Section 4 employs a range of indicators to compare and confirm the suggested method's efficacy. In Section 5, the main conclusions, outcomes, and suggestions for additional study are addressed.

2. RELATED WORKS

The following is a list of the various machine learning and deep learning algorithms that have been used in previous research to identify diabetes from patient data:

SVM, or support vector machines: It is a potent supervised learning method that is applied to problems with classification and regression. SVM[11] is mostly used for binary classification in the context of diabetes prediction, which divides people into those with and without diabetes based on a collection of input attributes. Finding the hyper plane in the feature space that best divides the data points of several classes is the basic idea behind the use of support vector machines. The margin, or the distance between the hyper plane and the closest data points for each class—also referred to as supporting vectors—is maximized by selecting this particular hyper plane. In order to improve its generalization performance on unknown data, SVM strives for both a high classification accuracy and a wide margin.

Logistic Regression (LR) - It is a statistical technique for tasks involving binary classification. It simulates the likelihood that an input falls into a specific class[12, 13]. By calculating the likelihood that a person has diabetes based on input variables like blood pressure, BMI, glucose levels, etc., it can also be used to predict diabetes.

When there is a linear link among the features and the outcome of the experiment, it is especially helpful.

Decision Tree (DT) - Decision Trees are logical structures that divide data into subsets recursively according to attribute values in order to facilitate decisions[14]. They are capable of handling both categorical and numerical data, and they are simple to interpret. It can also be used to build a hierarchy of trees that depicts decisions made for diabetes prediction based on features (e.g., BMI, glucose level). Professionals can then determine which factors have the most influence on the prediction.

Random Forest (RF) - RF [15, 16]is an ensemble learning technique that builds several decision trees during training and predicts the class mode. It is resistant to data distortion and overfitting from occurring. By combining the predictions of multiple decision trees trained on various data subsets, it can also be used to predict diabetes. Compared to individual decision-making processes, it frequently produces forecasts that are more accurate.

Neural Network (NN) - Computational models that mimic the architecture and operations of the human brain are called neural networks, or CNNs[17, 18]. They are formed up of linked layers of nodes, or brain cells, that process and transform the data that is fed into them. Artificial neural networks, especially deep learning architectures like Recurrent Neural Networks (RNNs) or Convolutional Neural Networks (CNNs), may learn intricate patterns from data and predict things like glucose levels, patient histories, etc. in the context of diabetes prediction[19]. They can identify complex correlations between characteristics and diabetes risk.

A comprehensive summary of the research done to develop a framework that may accurately forecast a patient's risk of developing diabetes was given by Kaul et al [20]. Several machine learning classification approaches are used to identify diabetes. In this paper, the author has investigated many machine learning classification techniques, such as GA, DT, RF, LR, SVM, and NB. Furthermore, this work's main objective is to provide a framework for contrasting all predictions with the several classification techniques used in the previous studies. They have also emphasized the effectiveness and results of the best techniques.An extensive study was presented by Larabi et al. [21] to look at several methods for diabetes classification and prediction. Additionally, it goes over a few of the popular machine learning and deep learning methods used in medicine to predict illness. A variety of machine learning approaches were applied by Uddin et al. [22] to detect and classify diabetes types. In order to diagnose diabetes, Garcia et al. [23]used a deep learning method combined with feature augmentation and oversampling. It contains learning strategies for diabetes detection based on both Sparse Auto-Encoder (SAE) and Variational Auto-Encoder (VAE).

Many recent studies use machine learning and deep learning approaches for recognizing diabetes, however they may not be able to cope with imbalanced datasets in which the proportion of cases who have diabetes is substantially lower than that of instances without diabetes[24, 25]. Methods to increase model applicability to unknown data and alleviate class imbalance issues could be investigated in further investigations.The majority of research efforts emphasis on detecting diabetes using a single modality of data (e.g., blood sugar levels, patient statistics). To increase the accuracy of predictions and reliability, it is necessary to investigate the integration of many data sources, including biological information, electronic health records, lifestyle factors, and data from sensor devices.

2. PROPOSED METHODOLOGY

A complete rationale for the proposed Diabetic prediction using Optimized Learning Classifier (DIABOLIC) is given in this section. The goal of the endeavor is to develop a deep learning-based intelligent patient health monitoring system that can rapidly and accurately identify whether a patient has a chronic illness like diabetes. This monitoring system gathers data from multiple medical wearable. After that, the application uses deep learning algorithms on the raw data to assess the patient's health status in order to draw the appropriate conclusions. With the use of this technology, which predicts the patient's condition, the doctor and guardians are alerted to an emergency and asked to provide quick care.Figure 1 depicts the entire workflow of the proposed DIABOLIC paradigm, which includes the following operational modules:

- Medical data collection
- Data cleaning & preprocessing

- Tumultuo Dwarf Mongoose Optimization (TuD-MO) using Feature Selection
- Fused Deep Convolution Random Network (FDCRN) for Classification
- Disease prediction
- Performance analysis

Depending on how it is specifically designed, put into practice, and assessed, the proposed DIABOLIC model may contribute for diabetes prediction. Also, it may make the following research contributions:

- Developing a Resilient Prediction Model: Leveraging pertinent patient data, develop and implement a machine learning or deep learning-based model that is robust enough to reliably forecast the probability of getting diabetes.
- Feature Analysis: Determine which features (such as biomarkers, medical history, and demographic data) are most useful for predicting diabetes and investigate new feature engineering methods to improve prediction accuracy.
- Model Interpretation: To provide transparency and support clinical decision-making, include methods for understanding the predictions of models and pinpointing the major risk factors for diabetes.
- Handling Data Imbalance: Investigate at methods for resolving class imbalance problems in the dataset so that the model can function well in situations where the proportion of instances with diabetes is much lower than that of cases without the disease.
- Performance validation: Assessment of the DIABOLIC model's performance in terms of prediction accuracy, sensitivity, specificity, and other pertinent metrics can be done by validating it on independent datasets or by applying cross-validation techniques.
- Comparison with Previous Models: To evaluate the DIABOLIC model's relative performance and pinpoint areas for advancement, compared it with contemporary diabetes prediction models using assessment metrics and standardized datasets.

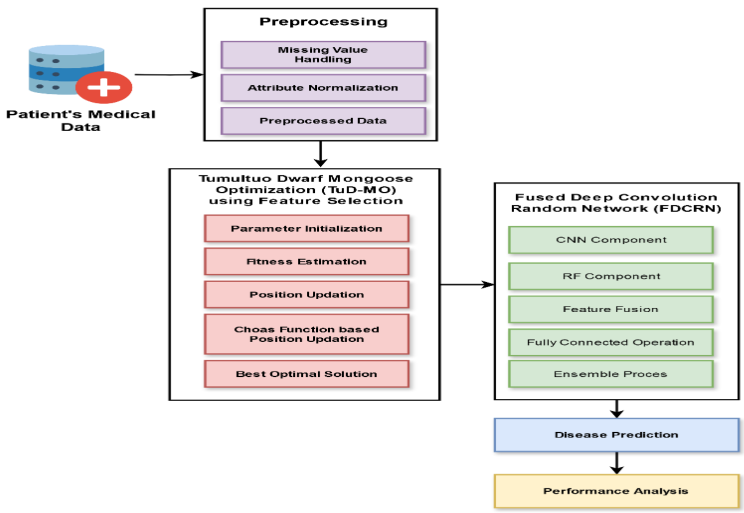

Fig 1. Overall flow of the proposed DIABOLIC model

A. DATA CLEANING AND PREPROCESSING

The initial step in preparing a dataset is to handle missing values, which is followed by the standardization operation. If missing or null values are not imputed, the deep learning classifier's prediction accuracy will decline. The mean method is applied here to fill in the missing integers, as illustrated below:

$$m(g) = \begin{cases} Mean(g), & if\ g = Null/Miss \\ g. & Otherwise \end{cases} \quad (1)$$

Where m stands for the data's missing values, g for the feature vector's instances, and $Mean(g)$ for the mean value. The mean approach makes sense when it comes to imputation missing values because it provides the continuous data required for algorithm training while excluding anomalies. Z-score normalization can be used to rescale features and produce a typical normal distribution with a unit variance and zero mean. It reduces data skewness and is reflected in the following equation:

$$H(g) = \frac{g - Mean(g)}{SD(g)} \quad (2)$$

Where, $H(g)$ is the normalised data, $Mean(g)$ is the mean value, and $SD(g)$ is the standard deviation. After preprocessing, the normalised and customised dataset is produced and used for additional prediction.

B. TUMULTUO DWARF MONGOOSE OPTIMIZATION (TUD-MO) USING FEATURE SELECTION

Selecting features for the classification model is a crucial step in maximizing its performance. It has been demonstrated that feature selection works well as a data preparation technique for getting ready data for a range of data analysis and artificial intelligence applications. Feature selection is to prepare clear, intelligible data, enhance data mining performance, and create simpler, more understandable models. The majority of feature selection techniques make use of a type of scoring and feature evaluation mechanism. Evaluation and the sequential configuration of a subset of characteristics are more basic methods. This is limited by the fact that most feature sets incorporate fitness are non-linear. As a result, applying a greedy strategy to sequential selection may not produce very optimal results. Comparing features is one of the more beneficial methods. By using the TuD-MO technique, the necessary and most pertinent attributes are chosen from the normalized data after the dataset has been cleaned. Medical datasets are typically among the most important in real life, however their dimensionality may expand due to redundant and useless information. This has a negative impact on the learning process's accuracy, cost, and efficiency. Feature selection eliminates irrelevant, superfluous, and noisy data to minimize the number of features while maintaining a respectable level of classification accuracy. It could be an issue with optimization. Stochastic optimization techniques have the potential to be advantageous due to the task's difficulty and the large number of local solutions. High precision and flexibility are needed for illness diagnosis and treatment.

In order to diagnose patients more quickly and save treatment costs, clinicians can benefit from efficient and trustworthy medical data analysis. The standard of public health care would improve. Finding a means to decrease the impact of these superfluous elements is therefore essential. The dimensionality reduction technique can help remove superfluous, redundant, and irrelevant characteristics from datasets without affecting the details they still contain. Feature selection is a technique that preserves all of the information while removing as few instances as possible of the most significant and instructive qualities from the initial dataset. It has shown to be a successful way to eliminate unnecessary and superfluous components without losing data.The following equation is used to update positions following population initialization:

$$m_{i,j} = vub + \partial \times (vub - vl \qquad (3)$$

Where, $m_{i,j}$ indicates the position at jth dimension, ∂ is the random number, vlb and vub are the lower and upper bound values respectively. Once the population has been established, the fitness of each solution is computed. The alpha female $\mathfrak{A}$is selected by determining the probability value for each population fitness using the following equation:

$$\mathfrak{A} = \frac{\not{f}_i}{\sum_{i=1}^{N} \not{f}_i} \qquad (4)$$

Where, $\not{f}_i$ indicates the fitness value. As a consequence, the candidate food position is estimated according to the following equation:

$$M_{i+1} = M_i + \eta \times \rho \qquad \mathbf{(5)}$$

Where, η indicates the uniformly distributed random number, and ρrepresents the vocalization of the alpha female that guides the family's course. Additionally, an estimate of the sleeping mound's average value is provided below:

$$\delta_i = \frac{\not{f}_{i+1} - \not{f}_i}{\max\{|\not{f}_{i+1}, \not{f}_i|\}} \qquad (6)$$

As a consequence of this, the average value ϑ is also estimated according to the model illustrated below:

$$\vartheta = \frac{\sum_{i=1}^{N} \delta_i}{N} \qquad (7)$$

The scouts look for the next sleeping mound to ensure exploration because mongooses are known to resist going back to the previous one. The following model is used to replicate the scout mongoose:

$$M_{i+1} = \begin{cases} M_i - \beta \times \eta \times \rho \times [M_i - \vec{K}] & if\ \vartheta_{i+1} > \vartheta_i \\ M_i + \beta \times \eta \times \rho \times [M_i + \vec{K}] & Else \end{cases} \quad (8)$$

$$\beta = \left(1 - \frac{it}{Max_{it}}\right)^{\left(2\frac{it}{Max_{it}}\right)} \quad (9)$$

$$\vec{K} = \sum_{i=1}^{n} \frac{m_i \times \delta_i}{M_i} \quad (10)$$

When a phenomena's original condition is even slightly changed, it might show non-linear variations with subsequent behavior. This phenomenon is known as chaos, which is also defined as a semi-random behavior produced by predictable nonlinear systems.

$$M_{i+1} = \begin{cases} M_i - \beta \times \eta \times \rho \times [M_i - \vec{K}] & if\ \vartheta_{i+1} > \vartheta_i \\ M_i + \beta \times \eta \times \rho \times [M_i - \vec{K}] & Else \end{cases} \quad (11)$$

Finally, the best possible optimal solution is determined as shown in below:

$$M_{i,j} = \begin{cases} 1 & m_{i,j} > 0.5 \\ 0 & Otherwise \end{cases} \quad (12)$$

The process of selecting the most pertinent features from a dataset to be used in model training is known as feature selection in artificial intelligence. This contributes to increased interpretability, decreased overfitting, and improved model performance.The elements from the medical data are chosen based on the best ideal value in order to create an accurate diabetes prediction.

C. FUSED DEEP CONVOLUTED RANDOM NETWORK (FDCRN) CLASSIFICATION

Following feature selection, the optimal characteristics are used to apply the FDCRN classification algorithm, which reliably predicts the disease from the available data. The main benefits of utilizing this technology over alternative machine learning and deep learning methods are its reduced training time, enhanced accuracy, and optimized system performance.For improved predictive performance, the novel Fused Deep Convolution Random Network (FDCRN) model combines the advantages of random forests (RFs) with deep convolutional neural networks (CNNs). The FDCRN model's CNN is made up of several layers of activation, pooling, and convolutional functions. Convolutional layers preserve spatial linkages while extracting hierarchical characteristics from the input data. In order to save computational complexity and preserve the most noticeable characteristics, pooling layers down sample the feature maps. By adding non-linearity, activation functions allow the model to recognize intricate patterns in the input. An ensemble of decision trees makes up the RF component of the FDCRN model. A random subset of features and data samples are used to train each decision tree, which encourages diversity and minimizes overfitting. The predictions made by each individual tree are combined to get the final result during inference. The feature representations that the two models have learned are combined in the fusion layer by integrating the outputs of the CNN and RF components. Fusion is a technique that enables the model to take advantage of the complementary strengths of CNNs and RFs. Examples of these approaches are concatenation, averaging, and weighted combination. Here, a gradient-based optimization techniques are used to train the CNN component by minimizing a predetermined loss function. To ensure diversity among the component decision trees, the RF component is trained using random selections of features and bootstrapped samples of the training data.Compared to previous approaches, the proposed FDCRN model has the following advantages, especially when used for classification tasks like diabetes prediction. It includes effective feature representation, highly robust to noisy features, improved predictive performance, and reduced overfitting.

3. RESULTS AND DISCUSSION

Through the use of performance metrics including accuracy, sensitivity, specificity, and ROC, the proposed DIABOLIC approach has been validated. This work has investigated the performance of the proposed model on two different and well-known datasets: PIMA and BRFSS. The PIMA Indians Diabetes dataset, which comprises 768 female diabetic patients from the Indian community, was used in this instance to train and assess the machine learning models. There are 268 people with diabetes and 500 patients without diabetes in this data collection, and each patient has eight unique characteristics. The parameters utilized in the analysis of this study are computed using the following formulae.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

$$\mathrm{Precision} = \frac{TP}{TP + FP} \tag{14}$$

$$\mathrm{Recall} = \frac{TP}{TP + FN} \tag{15}$$

$$F1\_Score = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{16}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{17}$$

$$Specificity = \frac{TN}{TN + FP} \tag{18}$$

Where, TP – True Positives, TN – True Negatives, FP – False Positives, and FN – False Negatives. The patient has a test result of 0, however the data file has a patient feature called FN.

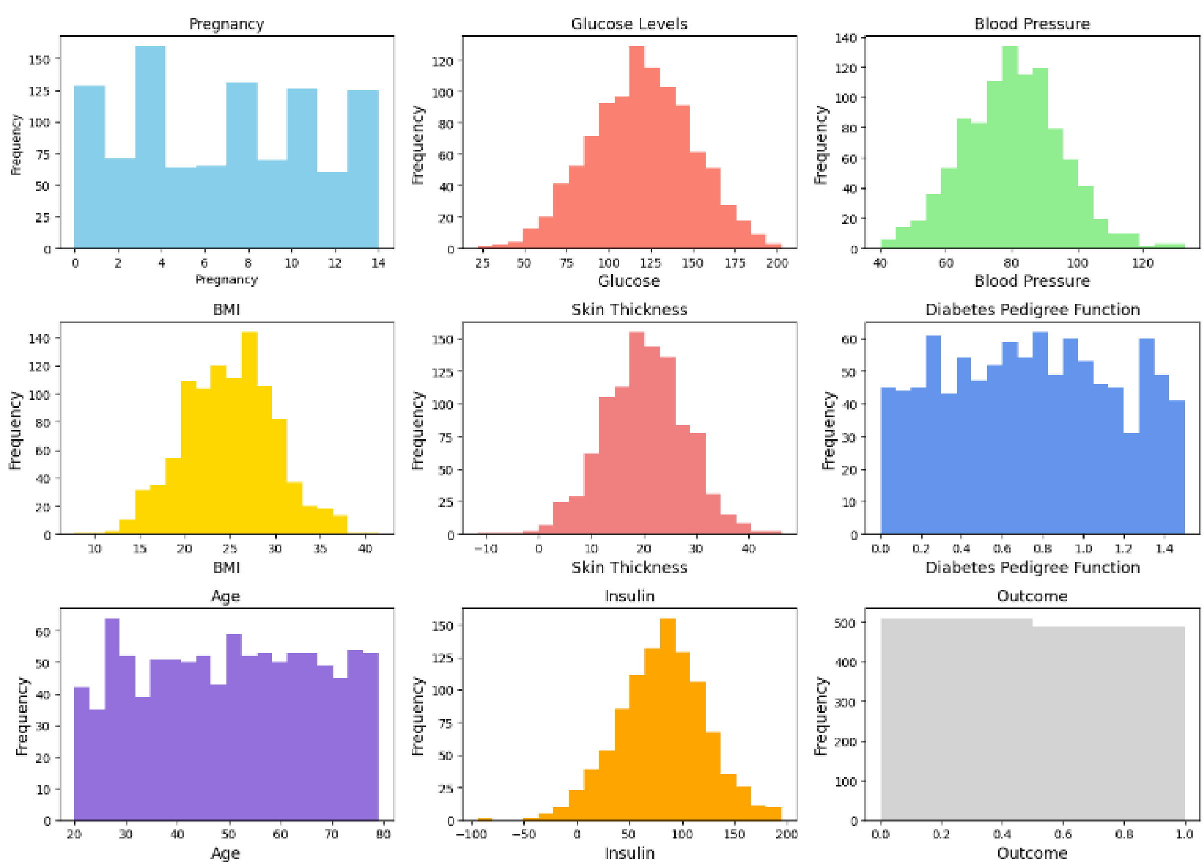

Fig 2. Histogram of features

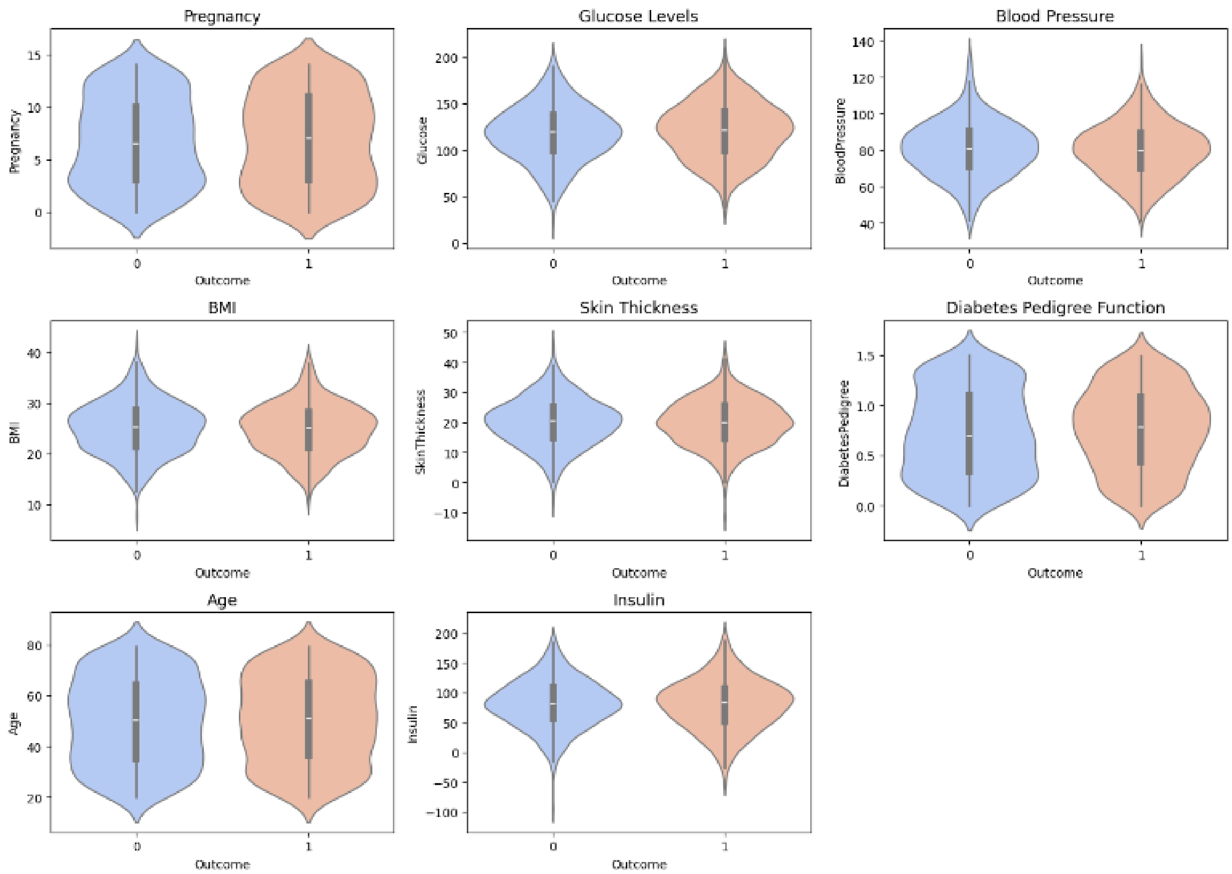

Fig 3. Outcome of DIABOLIC with respect to different attributes

In the diabetic data, also referred to as TN, the feature value drops to 0 and the attribute is absent. An FP occurs when a feature has a value of 1 but is not present in the dataset. Given that the TP result is 1, this data file contains features. As shown in Fig 2, the histogram of features would show the distribution of values for different input variables or features pertinent to the prediction job in the context of diabetes prediction utilizing the DIABOLIC framework.The histogram is computed for various feature types in the dataset in this study. Consequently, Figure 3 also displays the results of DIABOLIC for each chosen attribute in the dataset.A scatter plot could be used to show the association between two particular features or variables in the dataset within the framework of the DIABOLIC framework for diabetes prediction. The scatter plot for different features with diabetic and non-diabetic cases are shown in Fig 4.

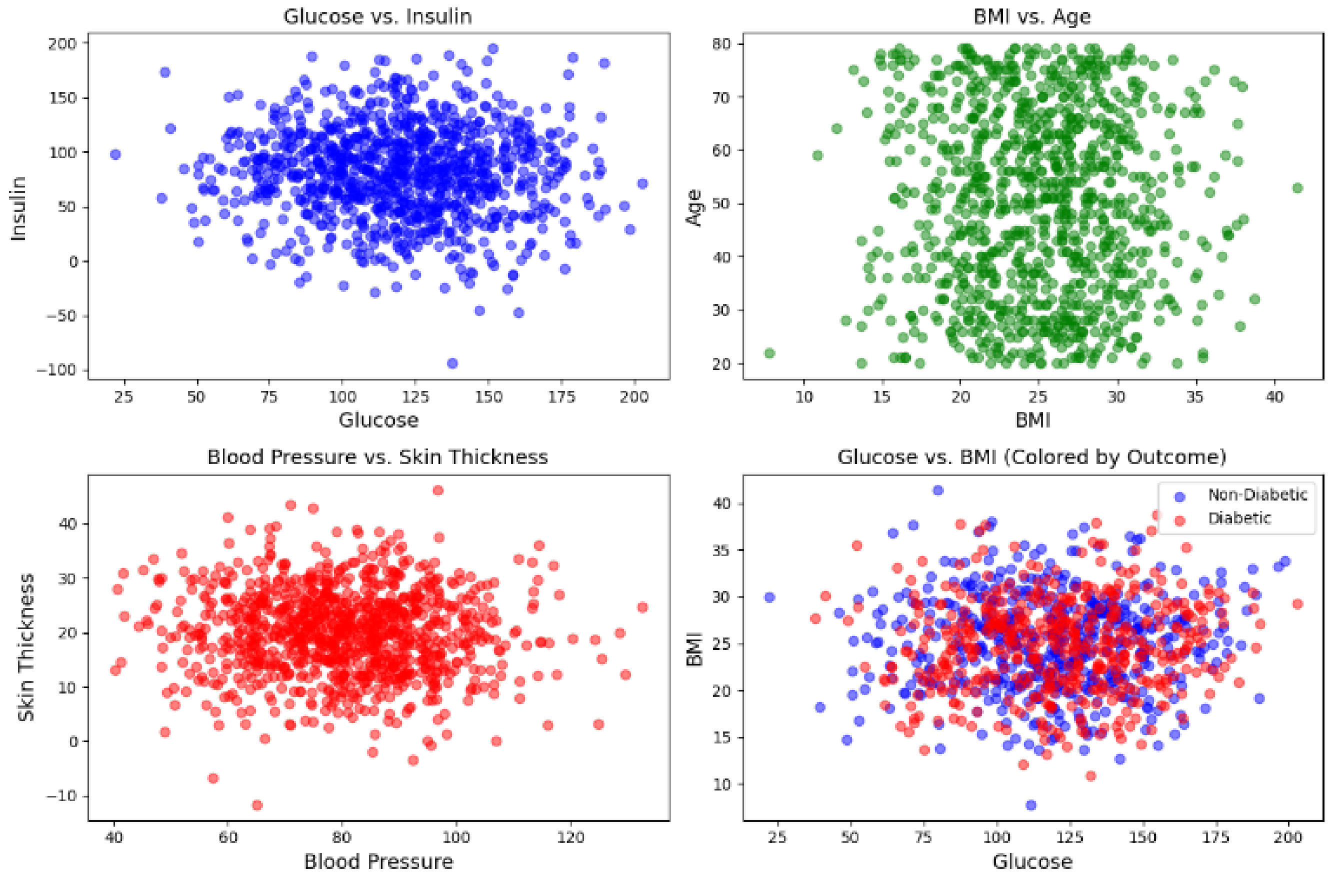

Fig 4. Scatter plot for feature analysis

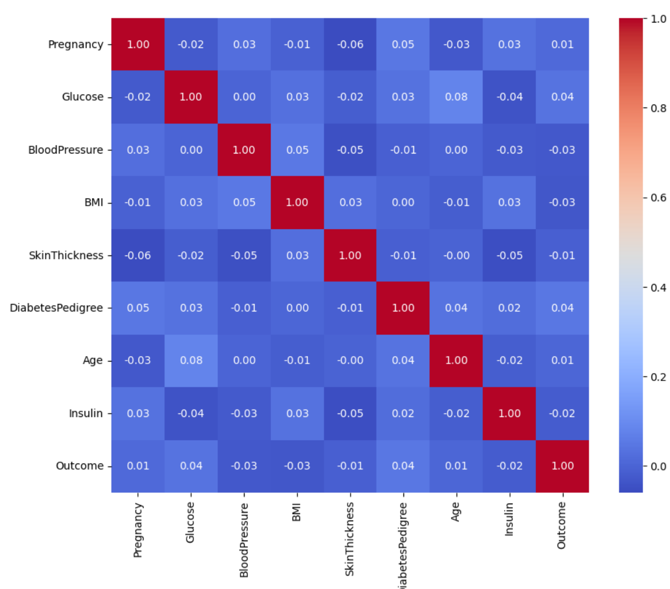

Fig 5. Confusion matrix

A thorough assessment of the model's efficacy in predicting diabetes status based on multiple input variables can be found in the confusion matrix for the DIABOLIC framework with regard to distinct features.Here, Figure 5 displays the confusion matrix that was generated for the DIABOLIC framework. As a consequence of this, the comparative analysis based confusion matrix is also provided in Fig 6. According to the overall results, it is established that the suggested model performs well and delivers an enhanced prediction outcomes, when compared to the other existing models.

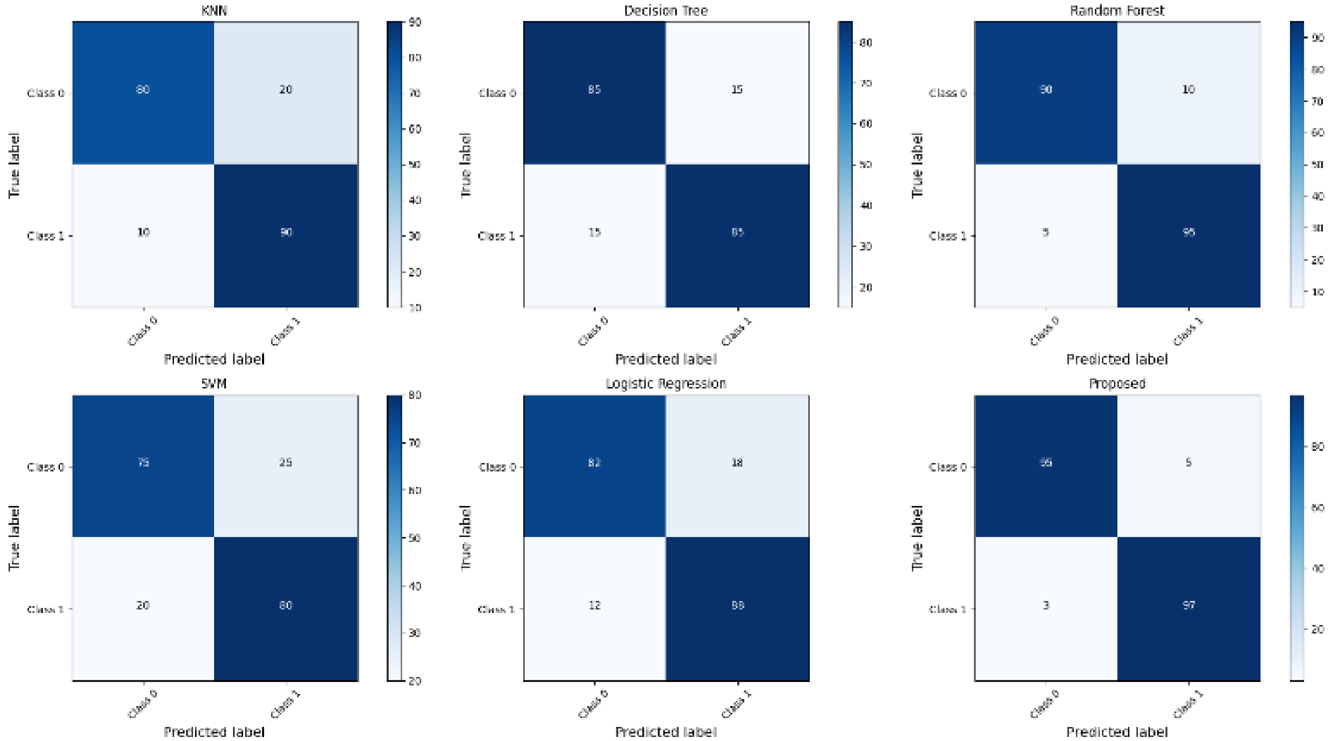

Fig 6. Comparison among conventional and proposed diabetic prediction techniques

Using the PIMA, CDC-BRFSS2015, and MIT-Mesra Diabetes datasets, respectively, Fig 7 to Fig 9 compare and validate the accuracy, sensitivity, and specificity performance of different existing and proposed diabetes prediction models [1]. Then, its appropriate values are displayed in Table 1 to Table 3. The results of this comparative analysis show that, in comparison to all other current methods, the suggested DIABOLIC paradigm performs well. Given that the main factors contributing to the suggested framework enhanced performance outcomes are the integration of TuD-MO and FDCRN.

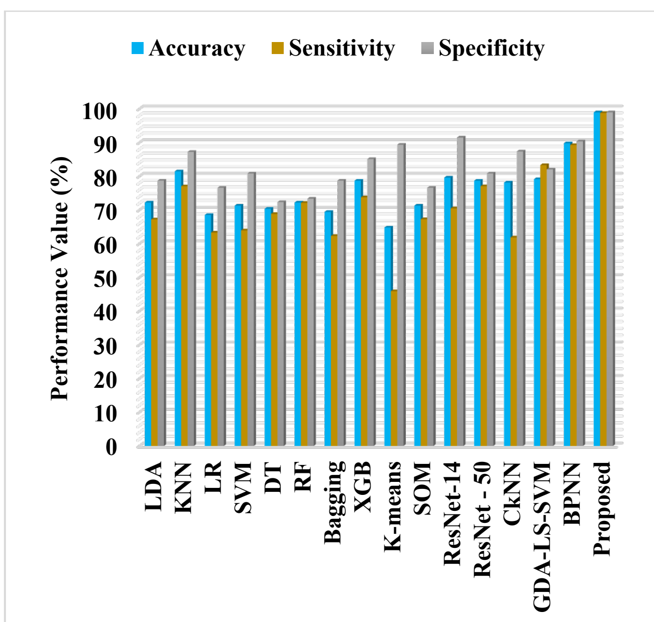

Fig 7. Performance study using PIMA dataset

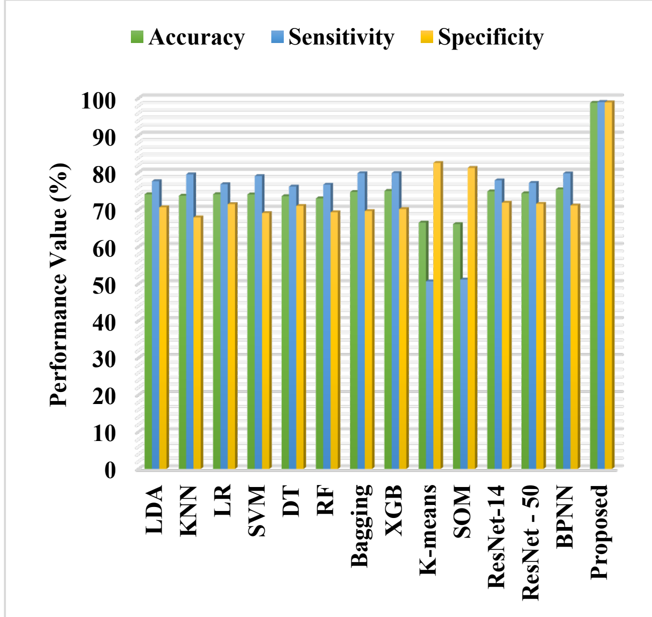

Fig 8. Performance study using CDC-BRFSS2015 dataset

Table 1. Comparative assessment using PIMA dataset

Methods	Accuracy	Sensitivity	Specificity
LDA	72.22	67.21	78.72
KNN	81.48	77.05	87.23
LR	68.52	63.3	76.6
SVM	71.3	63.93	80.85
DT	70.37	68.85	72.34
RF	72.22	72.13	73.34
Bagging	69.44	62.3	78.72
XGB	78.7	73.77	85.11
K-means	64.81	45.9	89.36
SOM	71.3	67.27	76.6
ResNet-14	79.63	70.49	91.49
ResNet - 50	78.7	77.06	80.85
CkNN	78.16	61.84	87.38
GDA-LS-SVM	79.16	83.33	82.05
BPNN	89.81	89.29	90.38
Proposed	99	98.8	99

Table 2. Comparative assessment using CDC-BRFSS2015 dataset

	Accuracy	Sensitivity	Specificity
LDA	74.16	77.67	70.64
KNN	73.76	79.52	67.92
LR	74.18	76.85	71.48
SVM	74.11	79.06	69.08
DT	73.64	76.22	71.02
RF	73.04	76.73	69.3
Bagging	74.77	79.83	69.64
XGB	75.05	79.87	70.17
K-means	66.53	50.69	82.59
SOM	66.11	51.18	81.25

ResNet-14	74.92	77.9	71.87
ResNet - 50	74.42	77.22	71.58
BPNN	75.49	79.77	71.12
Proposed	98.8	99	98.9

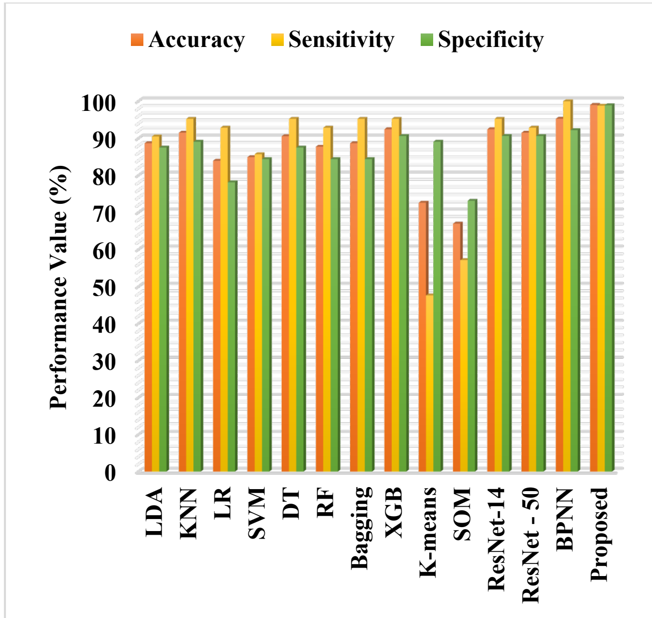

Fig 9. Performance study using MIT-Mesra Diabetes dataset

Table 3. Comparative assessment using MIT-Mesra Diabetes dataset

Methods	Accuracy	Sensitivity	Specificity
LDA	88.68	90.48	87.5
KNN	91.51	95.24	89.06
LR	83.96	92.86	78.13
SVM	84.91	85.71	84.38
DT	90.57	95.24	87.5
RF	87.74	92.86	84.38
Bagging	88.68	95.24	84.38
XGB	92.45	95.24	90.62
K-means	72.64	47.62	89.06
SOM	66.98	57.14	73.14
ResNet-14	92.45	95.24	90.63
ResNet - 50	91.51	92.86	90.62
BPNN	95.28	100	92.19
Proposed	99	98.8	98.9

Table 4 and Fig. 10 show how some hybrid classification algorithms are compared with the proposed DIABOLIC model based on accuracy. The proposed approach detects diabetes from patient data by using efficient feature selection, classification, and hyperparameter tuning strategies. In comparison to other hybrid classification systems, it aids in achieving the greatest performance values given the suggested framework.

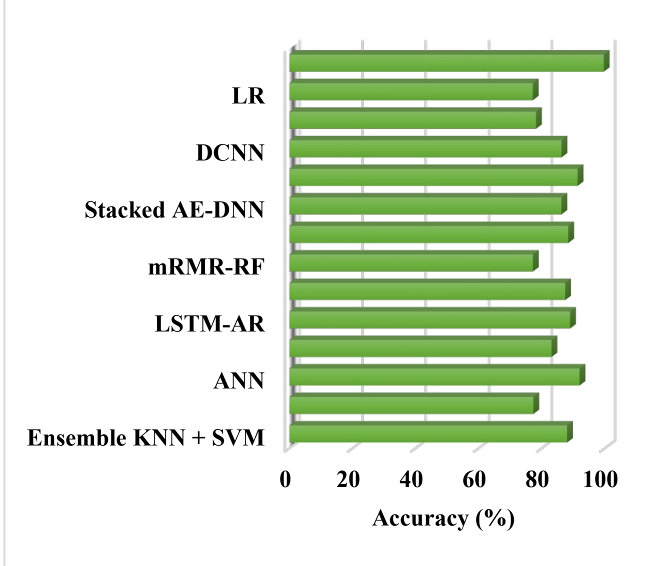

Fig 10. Accuracy analysis

Table 4. Comparison based on accuracy

Methods	Accuracy (%)
Ensemble KNN + SVM	88.04
FS + SVM	77.37
ANN	92
SVM	83.1
LSTM-AR	89
UTA-NN	87.46
mRMR-RF	77.21
Deep MLP	88.41
Stacked AE-DNN	86.26
Conv-LSTM	91.38
DCNN	86.29
Stacking model	78.2
LR	77.2
Proposed	99.7

4. CONCLUSION

A robust and cutting-edge method for diabetes prediction is provided by the DIABOLIC model's integration of TuD-MO for feature selection and FDCRN for classification.Through the use of TuD-MO, the model finds the most pertinent features from a variety of heterogeneous datasets, such as clinical, genetic, lifestyle, and demographic data, improving the interpretability and effectiveness of the ensuing classification procedure.By integrating random forests and deep convolutional neural networks, the FDCRN algorithm provides better predicting accuracy by identifying intricate relationships and patterns among the chosen features. This method not only increases the precision of diabetes prediction but also offers insightful information about the fundamental causes of the illness.Additionally, by customizing predictions to each patient's unique profile, the integration of TuD-MO and FDCRN enables personalized risk assessment and intervention methods.With the help of this individualized approach, medical professionals can better manage patients' conditions and lessen the burden of problems associated with diabetes by implementing tailored preventative measures, lifestyle changes, and treatment plans.All things considered, the DIABOLIC model—which uses FDCRN for classification and TuD-MO for feature selection—represents a cutting-edge framework for diabetes prediction. Its capacity to leverage deep learning algorithms and optimization strategies opens up exciting new possibilities for the advancement of personalized healthcare and precision medicine in the area of diabetes control.Our research efforts have culminated in the invention of the DIABOLIC model for diabetes prediction, which has produced an astounding result: a 99% accuracy rate in predicting an individual's risk of getting diabetes. This accomplishment highlights our model's effectiveness as well as its significant implications for improving precision medicine and personalized healthcare in the context of diabetes control.In the future, we can apply strategies from explainable artificial intelligence to improve the DIABOLIC model's interpretability and explainability.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Zhang, K. A. Ahmed, M. R. Hasan, T. Gedeon, And M. Z. Hossain, "A Deep Learning Approach To Diabetes Diagnosis," Arxiv Preprint Arxiv:2403.07483, 2024.

Q. W. Khan, K. Iqbal, R. Ahmad, A. Rizwan, A. N. Khan, And D. Kim, "An Intelligent Diabetes Classification And Perception Framework Based On Ensemble And Deep Learning Method," Peerj Computer Science, Vol. 10, P. E1914, 2024.

L. Xie, "Pima Indian Diabetes Database And Machine Learning Models For Diabetes Prediction," Highlights In Science, Engineering And Technology, Vol. 88, Pp. 97-103, 2024.

K. K. Patro, J. P. Allam, U. Sanapala, C. K. Marpu, N. A. Samee, M. Alabdulhafith, Et Al., "An Effective Correlation-Based Data Modeling Framework For Automatic Diabetes Prediction Using Machine And Deep Learning Techniques," Bmc Bioinformatics, Vol. 24, P. 372, 2023.

N. Karthik, "Machine Learning-Based Trustworthy Diabetes Prediction Model," In 2024 16th International Conference On Communication Systems & Networks (Comsnets), 2024, Pp. 400-402.

K. Dharmarajan, B. Sharma, S. Chowdhury, And I. B. Dhaou, "Diabetes Prediction Using Enhanced And Optimized Drl-Cnn Approaches," In 2024 21st Learning And Technology Conference (L&T), 2024, Pp. 278-283.

H. Zhou, R. Myrzashova, And R. Zheng, "Diabetes Prediction Model Based On An Enhanced Deep Neural Network," Eurasip Journal On Wireless Communications And Networking, Vol. 2020, Pp. 1-13, 2020.

A. Mousa, W. Mustafa, R. B. Marqas, And S. H. Mohammed, "A Comparative Study Of Diabetes Detection Using The Pima Indian Diabetes Database," Journal Of Duhok University, Vol. 26, Pp. 277-288, 2023.

P. B. K. Chowdary And R. U. Kumar, "An Effective Approach For Detecting Diabetes Using Deep Learning Techniques Based On Convolutional Lstm Networks," International Journal Of Advanced Computer Science And Applications, Vol. 12, 2021.

R. Bhargava And J. Dinesh, "Deep Learning Based System Design For Diabetes Prediction," In 2021 International Conference On Smart Generation Computing, Communication And Networking (Smart Gencon), 2021, Pp. 1-5.

Z. Mushtaq, M. F. Ramzan, S. Ali, S. Baseer, A. Samad, And M. Husnain, "Research Article Voting Classification-Based Diabetes Mellitus Prediction Using Hypertuned Machine-Learning Techniques," 2022.

S. K. Bhoi, "Prediction Of Diabetes In Females Of Pima Indian Heritage: A Complete Supervised Learning Approach," Turkish Journal Of Computer And Mathematics Education (Turcomat), Vol. 12, Pp. 3074-3084, 2021.

A. Darolia And R. S. Chhillar, "Analyzing Three Predictive Algorithms For Diabetes Mellitus Against The Pima Indians Dataset," Ecs Transactions, Vol. 107, P. 2697, 2022.

V. Patil And D. Ingle, "Comparative Analysis Of Different Ml Classification Algorithms With Diabetes Prediction Through Pima Indian Diabetics Dataset," In 2021 International Conference On Intelligent Technologies (Conit), 2021, Pp. 1-9.

P. Palimkar, R. N. Shaw, And A. Ghosh, "Machine Learning Technique To Prognosis Diabetes Disease: Random Forest Classifier Approach," In Advanced Computing And Intelligent Technologies: Proceedings Of Icacit 2021, 2022, Pp. 219-244.

T. A. Assegie And P. S. Nair, "The Performance Of Different Machine Learning Models On Diabetes Prediction," International Journal Of Scientific & Technology Research, Vol. 9, 2020.

G. Alfian, M. Syafrudin, M. Anshari, F. Benes, F. T. D. Atmaji, I. Fahrurrozi, Et Al., "Blood Glucose Prediction Model For Type 1 Diabetes Based On Artificial Neural Network With Time-Domain Features," Biocybernetics And Biomedical Engineering, Vol. 40, Pp. 1586-1599, 2020.

S. Wang, Y. Chen, Z. Cui, L. Lin, And Y. Zong, "Diabetes Risk Analysis Based On Machine Learning Lasso Regression Model," Journal Of Theory And Practice Of Engineering Science, Vol. 4, Pp. 58-64, 2024.

S. L. Cichosz, M. H. Jensen, O. Hejlesen, S. D. Henriksen, A. M. Drewes, And S. S. Olesen, "Prediction Of Pancreatic Cancer Risk In Patients With New-Onset Diabetes Using A Machine Learning Approach Based On Routine Biochemical Parameters," Computer Methods And Programs In Biomedicine, Vol. 244, P. 107965, 2024.

S. Kaul And Y. Kumar, "Artificial Intelligence-Based Learning Techniques For Diabetes Prediction: Challenges And Systematic Review," Sn Computer Science, Vol. 1, P. 322, 2020.

S. Larabi-Marie-Sainte, L. Aburahmah, R. Almohaini, And T. Saba, "Current Techniques For Diabetes Prediction: Review And Case Study," Applied Sciences, Vol. 9, P. 4604, 2019.

M. A. Uddin, M. M. Islam, M. A. Talukder, M. A. A. Hossain, A. Akhter, S. Aryal, Et Al., "Machine Learning Based Diabetes Detection Model For False Negative Reduction," Biomedical Materials & Devices, Pp. 1-17, 2023.

M. T. Garcia-Ordas, C. Benavides, J. A. Benitez-Andrades, H. Alaiz-Moreton, And I. Garcia-Rodriguez, "Diabetes Detection Using Deep Learning Techniques With Oversampling And Feature Augmentation," Computer Methods And Programs In Biomedicine, Vol. 202, P. 105968, 2021.

X. Zou, Y. Luo, Q. Huang, Z. Zhu, Y. Li, X. Zhang, Et Al., "Differential Effect Of Interventions In Patients With Prediabetes Stratified By A Machine Learning-Based Diabetes Progression Prediction Model," Diabetes, Obesity And Metabolism, Vol. 26, Pp. 97-107, 2024.

B. A. Ng, "En-Rfrsk: An Ensemble Machine Learning Technique For Prognostication Of Diabetes Mellitus," Egyptian Informatics Journal, Vol. 25, P. 100441, 2024.