scientific reports

Check for updates

OPEN

Efficient diagnosis of diabetes mellitus using an improved ensemble method

Blessing Oluwatobi Olorunfemi[1], Adewale Opeoluwa Ogunde[1], Ahmad Almogren[2], Abidemi Emmanuel Adeniyi[3], Sunday Adeola Ajagbe[4], Salil Bharany[5], Ayman Altameem[6], Ateeq Ur Rehman[7,8,9,10]✉, Asif Mehmood[11]✉ & Habib Hamam[12,13,14,15]

Diabetes is a growing health concern in developing countries, causing considerable mortality rates. While machine learning (ML) approaches have been widely used to improve early detection and treatment, several studies have shown low classification accuracies due to overfitting, underfitting, and data noise. This research employs parallel and sequential ensemble ML approaches paired with feature selection techniques to boost classification accuracy. The Pima India Diabetes Data from the UCI ML Repository served as the dataset. Data preprocessing included cleaning the dataset by replacing missing values with column means and selecting highly correlated features using forward and backward selection methods. The dataset was split into two parts: training (70%), and testing (30%). Python was used for classification in Jupyter Notebook, and there were two design phases. The first phase utilized J48, Classification and Regression Tree (CART), and Decision Stump (DS) to create a random forest model. The second phase employed the same algorithms alongside sequential ensemble methods—XG Boost, AdaBoostM1, and Gradient Boosting—using an average voting algorithm for binary classification. Evaluation revealed that XG Boost, AdaBoostM1, and Gradient Boosting achieved classification accuracies of 100%, with performance metrics including F1 score, MCC, Precision, Recall, AUC-ROC, and AUC-PR all equal to 1.00, indicating reliable predictions of diabetes presence. Researchers and practitioners can leverage the predictive model developed in this work to make quick predictions of diabetes mellitus, which could save many lives.

Keywords Classification, Diabetes Mellitus, Gradient boosting, Random forest, XG boost

Context and motivation

Diabetes Mellitus (DM) is a disease that prevents the body from receiving adequate energy from the foods consumed by humans. It is a chronic condition marked by abnormally high blood glucose levels[1]. This is caused by a lack of insulin synthesis or the body's inability to adequately utilize the insulin. When diabetes is detected early, it can be managed. It spreads rapidly, and thus, according to World Health Organization, DM will increase by the end of 2023 It was predicted that DM would affect around 425 million people between the ages of 20 to 79 years, and this figure was projected to increase to 629 million by 2045[2].

[1]Department of Computer Science, Faculty of Natural Sciences, Redeemer's University, Ede, Osun state, Nigeria. [2]Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11633, Saudi Arabia. [3]College of Computing and Communication Studies, Bowen University, Iwo, Nigeria. [4]Department of Computer Science, University of Zululand, Kwadlangezwa 3886, South Africa. [5]Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India. [6]Department of Natural and Engineering Sciences, College of Applied Studies and Community Services, King Saud University, Riyadh 11543, Saudi Arabia. [7]School of Computing, Gachon University, Seongnam-si 13120, Republic of Korea. [8]Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu, India. [9]Applied Science Research Center, Applied Science Private University, Amman, Jordan. [10]University Center for Research and Development, Chandigarh University, Mohali 140413, Punjab, India. [11]Department of Biomedical Engineering, College of IT Convergence, Gachon University, 1342 Seongnamdaero, Sujeong-gu, Seongnam-si 13120, Republic of Korea. [12]Faculty of Engineering, Université de Moncton, Moncton, NB E1A3E9, Canada. [13]School of Electrical Engineering, University of Johannesburg, Johannesburg 2006, South Africa. [14]International Institute of Technology and Management (IITG), Av. Grandes Ecoles, Libreville 1989, BP, Gabon. [15]Bridges for Academic Excellence-Spectrum, Tunis 1002, Centre-Ville, Tunisia. ✉email: ateqrehman@gmail.com; asif@gachon.ac.kr

There are three types of DM that exist. These include gestational diabetes, type 1 diabetes, and type 2 diabetes. Type 1 diabetes is also called insulin-dependent diabetes mellitus or juvenile-onset diabetes[3]. This is majorly found in young people below the age of 30 years. Here, the pancreas' beta cells that control insulin production are damaged due to the autoimmune system. The initial stage of Type 2 Diabetes, when the cells do not react to insulin, is insulin resistance. As the disease worsens, there might not be enough insulin available. Previously, this illness was referred to as "adult-onset diabetes" (NIDDM). The main causes are an excessive amount of body weight and a lack of activity. Pregnant women who acquire high blood glucose levels without a history of diabetes are said to have gestational diabetes. Although there is currently no cure for diabetes, it may be managed by maintaining healthy blood sugar levels by regular exercise, a well-balanced diet, and the use of the right medications. The following long-term effects of DM include: (1) Cataracts and retinopathy (gradual damage to the eye that might result in blindness). (2) Kidneys, including kidney failure and renal disease. (3) Neuropathy in the nerves (gradual damaging of nerves). (4) Ulcers, infections, gangrene, etc. on the feet. (5) The circulatory system. (6) Arterial hardening, cardiovascular disease, and stroke.

Artificial intelligence (AI), which includes machine learning (ML) and deep learning, has been widely used to predict[4], detect[5], and categorize[6] diseases, including diabetes[7,8]. Among these techniques, ensemble learning-based systems, a subset of ML, involve methods that generate multiple models, which are then combined effectively to enhance prediction accuracy and robustness. These systems leverage the strengths of individual models to minimize errors and improve overall performance, making them particularly effective in complex tasks such as disease diagnosis. Ensemble methods enhance the accuracy and strength of building a prediction model by combining a collection of base classifiers[9,10]. The two categories of ensemble methods are sequential ensemble techniques and parallel ensemble approaches. Base learners are produced sequentially via sequential ensemble techniques, such as Adaptive Boosting (AdaBoost). The ensemble method uses a variety of techniques, including Random Forest (RF), Bagging, AdaBoost, XG Boost, Light, and Stacking[11,12]. The experimental results, group learning (also known as "Ada boosting") performs better than solo learning. The continual generation of basis learners encourages dependence amongst base learners. The performance of the model is then enhanced by increasing the weights of previously misrepresented learners. Similar ensemble approaches build base learners using a framework akin to RF, for example. To promote basis learner independence, parallel techniques are employed to produce base learners in parallel. Feature selection is the procedure of reducing the amount of input variables while developing a predictive model[13]. Reducing the amount of input variables helps to lower modeling's computing cost and, in certain cases, enhance model performance. By merging parallel and sequential ensemble approaches with feature selection strategies to reduce the prediction issues, the goal of this research is to obtain a more accurate forecast.

According to[14], the number of patients with diabetes increased dramatically in the last ten years. The modern human lifestyle is suspected to be the primary cause of this rise. Three types of errors can occur in the current medical diagnosing procedure. The first is false-negative kind error- this is a situation where a test result indicates that a diagnosed patient does not have diabetes where he/she has it. The second one is the false-positive type. Here, the test result revealed that a patient without DM has the disease. The third form occurs when a system is unable to diagnose a particular case due to a lack of knowledge extraction from previous data. A patient's type may be projected to be unclassified. Traditional ML methods have failed to predict the presence of DM accurately from data but marginal accuracy. Hence, this research proposes to use ensemble methods to achieve greater prediction accuracy.

Scientific contribution

- Development of an Ensemble-Based Prediction Model: To increase the precision of diabetes mellitus prediction, the authors have created a novel ensemble-based model that makes use of both parallel and sequential techniques. This model greatly improves prediction performance by integrating several learning algorithms.
- Incorporation of Forward and Backward Feature Selection Techniques: The study makes use of forward and backward feature selection approaches, in particular. Through the identification and utilization of highly significant characteristics, these techniques helped the model become more predictive by reducing noise, underfitting, and overfitting.
- Comprehensive Validation with High Classification Accuracy: Using sequential ensemble approaches such as AdaBoostM1, XG Boost, and Gradient Boosting, the proposed model was verified on the Pima Indian Diabetes dataset, with classification accuracies of up to 100%. This shows how effectively the strategy works in clinical situations.
- Benchmarking Against Existing Methods: A thorough comparison between the suggested approaches and conventional machine learning algorithms is provided by the research. This benchmarking shows notable gains in precision, recall, and other performance parameters over current approaches.
- Using Ensemble Methods to Address Base-Level Errors: The work tackles common problems with single predictor models, such as residual errors and model independence, at the base level by combining parallel and sequential ensemble methods.

These contributions substantially advance the capabilities of predictive modelling for diabetes, delivering a comprehensive tool for early detection and control of the condition.

Structure of the article

The article is divided into five key sections to provide a thorough review of the research. First Section: Introduction sets the stage by presenting the context, motivation, and scientific contributions of the study. Second Section: Literature Review critically examines current research and methodologies, highlighting both

achievements and limitations in the field. Third section: Methodology details the ensemble methods and feature selection techniques employed, explaining how data was collected, processed, and analysed. Fourth section: Implementation and Results describes the practical application of the proposed methods and presents the results achieved, including performance metrics and model evaluations. Finally, fifth section: Conclusions summarizes the findings, examines the consequences of the research, and proposes areas for future work, relating back to the original research aims and the gaps noted in the literature.

Literature review

Current research

Recently, several studies have been undertaken to investigate the early detection of diabetes mellitus, and the algorithms utilized to do so are described in this section. Many biomedical problems, such as heart disease, breast cancer survival, diabetes, and so on, have been predicted and diagnosed using ML algorithms[15,16]. Recent work by[50] delved into the miRNA-mRNA-TF regulatory network in diabetic nephropathy (DN), identifying 163 differentially expressed miRNAs (DEMs) and 188 differentially expressed genes (DEGs). Their study highlighted 37 hub genes and four key transcription factors (EP300, SP100, NR6A1, JDP2), shedding light on DN pathogenesis and uncovering potential therapeutic targets. Additionally, the involvement of pathways such as MAPK, PI3K-Akt, and metabolic pathways in DN provided a deeper understanding of the molecular mechanisms driving diabetic complications. Expanding on therapeutic approaches for diabetes[51], investigated the effects of Liuwei Dihuang Decoction (LWDHT) on insulin resistance in Type 2 Diabetes Mellitus (T2DM) rats. Their findings revealed that LWDHT improved insulin sensitivity by modulating the PI3K/Akt signalling pathway and key liver proteins like IRS2 and Akt, highlighting its potential as a herbal intervention for managing T2DM. In glioma research[52] eveloped a 15-lncRNA signature associated with TGF-β signaling. This signature not only served as a predictor of poor prognosis but also correlated with changes in the immune microenvironment, with AGAP2-AS1 identified as a critical contributor to glioma progression. Notably, the signature demonstrated predictive capabilities for improved responses to immunotherapy, paving the way for personalized glioma treatment strategies. Similarly[53], explored quercetin's therapeutic potential in diabetic encephalopathy (DE), using Goto-Kakizak rats as their model. The study showed that quercetin mitigated cognitive impairments and suppressed ferroptosis in hippocampal neurons via the Nrf2/HO-1 signalling pathway, offering a promising direction for managing DE.

In a related context[54], emphasized the pivotal role of selenium and selenoproteins in maintaining oxidative balance as well as in glucose and lipid metabolism. Their review proposed selenium supplementation as a viable strategy for preventing and managing diabetes by targeting oxidative stress pathways, thereby contributing to a growing repertoire of metabolic interventions.

In the context of healthcare systems[55], investigated diabetes-related avoidable hospitalizations in Sichuan Province, China. Their findings revealed a significant association between inadequate primary healthcare resources and higher hospitalization rates, emphasizing the critical need to improve healthcare access, particularly in rural areas, to enhance diabetes management outcomes.

The application of machine learning in healthcare has been transformative, with techniques like AdaBoost paving the way for advancements in predictive diagnostics. Initially recognized as the first accurate boosting methodology for binary classification, AdaBoost was employed to predict breast cancer survival outcomes, assisting medical practitioners in decision-making processes[17]. Similarly, in diabetes, boosting algorithms have been applied for early diagnosis and predicting complications, showcasing their versatility across medical domains.

Among the most widely used classification algorithms in diabetes research are Support Vector Machines (SVM), Radial Basis Function, Multi-Layer Perceptron, and Multi-Level Counter Propagation Networks[18]. For instance, the ARTMAP-IC Structured Model was combined with a General Regression Neural Network (GRNN) to enhance accuracy on the PIMA Indian Diabetes dataset. Temurtas et al. advanced this by employing a multilayer neural network structure on the same dataset, achieving improved diagnostic precision[19].

Further advancements include leveraging the Levenberg-Marquardt (L.M.) technique and a probabilistic neural network structure to improve the accuracy of identifying heart disease in diabetic patients. Given that heart disease remains the leading cause of mortality among individuals with diabetes, a range of data mining approaches, including Nave Bayes and SVM algorithms, has been utilized to predict and manage this condition[20]. Interestingly, these methodologies were initially employed in breast cancer prediction[15], demonstrating their adaptability and potential across diverse medical challenges. A small number of sub-networks were originally identified as illness indicators before being utilized to categorize metastasis. Researchers used the decision tree C4.5 algorithm[21], bagging with the decision tree C4.5 technique, and bagging with the Nave Bayes algorithm[19], to predict cardiac illness. In ML and data mining, several metrics have been used. Precision, recall, F-measure, and ROC space are all critical performance indicators. These factors were used to calculate the optimum accuracy.

Diabetes and cancer are well-known chronic diseases that share a compound link, such as when the human body's glucose level rises to a diverging level. Diabetes is the result[16]. As a result, some recommender systems based on adaptive and personalized based insulin on Kalman filter theory have been presented to determine the classification accuracy in PIMA Indian dataset[22]. Multimedia retrieval systems have been developed and put into use in several applications. The discussion focuses on web-based front ends, specialized multimedia analytic techniques, meta-data annotation, and multimedia databases. Projects and systems that currently use MPEG-7 or offer additional retrieval functionality were given specific attention. Analysis of the number of semantics that may be utilized and defined was crucial in addition to MPEG-7 usage[23]. Other classification algorithms, on the other hand, have been developed, but their effectiveness has suffered because of an unsuitable dataset balance, since most classes have represented them. As a result, they may be classified using balanced classification algorithms like RHS-Boost[24]. This method is used to determine the best accuracy and prediction

for misbalanced datasets. Precision selection approaches have been combined with methods such as the Modified Spline Smooth Support Vector Machine (MS-SSVM) to improve accuracy. A 10-fold cross-validation approach, comprising accuracy, sensitivity, and specificity, has been proposed for this vector machine technology. ANFIS and Principal Component Analysis are two methods that can be used to detect diabetes individuals. The method of[25] can be combined with PCA. It was used to partition the dataset's eight characteristics into four groups in the first stage. The authors of[26] employed J48, KNN, filtered classifier, Artificial Neural Network (ANN), Nave Bayes (N.B), CV parameter selection, Xero, and basic cart to classify the diabetes dataset. This study did not employ cross-validation. With a score of 77.01%, N.B. has a good degree of accuracy.

Authors in[26] used a Bayesian classification technique that showed promise when compared to previous research, requires little training data, and has an accuracy rate of above 87%. Disease prediction does not appear to be particularly reliable due to a lack of data, i.e., a lack of samples. The number of characteristics must be lowered to improve categorization. KNN, Nave Bayes, Decision Tree, RF, SVM, and Logistic Regression were proposed by the authors of[27], which focused on recall, accuracy and produced a better result, however, the filtering criteria may be improved. Komi et al.[28] employed ELM, ANN LR, GMM, and SVM; their work used fewer sample data. The algorithm was compared to other technologies such as artificial neural networks. The method provided better accuracy than the other classifiers. Less amount of sample data was used in the study. Sai et al.[29] employed the weighted voting approach to combine the multiple ensemble and base classifiers into a new ensemble prediction model. The tenfold validation approach was used to assess the effectiveness of the system, and the results were derived using the percentage of occurrences that were properly and wrongly categorised. The basic classifier fusion model and the ensemble-based prediction model for diabetes were contrasted. The outcome also shown that a traditional ensemble prediction model can deliver superior predictive performance. By exploring how to extend the ensemble of the ensemble classifier technique and reduce the computing amount to produce minimum error values and greater accuracy, they increased the interpretability and enhanced the predictions.

Miled et al.[30] investigated 11,558 cardiac patient cases from 15 to 25 different institutes and controlled them. In ML, they employed the RF to predict dementia. Random forest classification accuracy in heart disease was calculated by the authors to be 77.43%.

Reza et al.[58] proposed a stacking ensemble method for the classification of diabetes patients using both the PIMA Indian Diabetes dataset and additional local healthcare data. Their approach incorporated both classical and deep neural network (NN) models to combine multiple classifiers' predictions, improving the overall accuracy and robustness of diabetes detection. The study found that stacking with three neural network architectures achieved the highest performance, with an accuracy of 95.50%, precision of 94%, recall of 97%, and an F1-score of 96% using 5-fold cross-validation (CV) on simulation data. The stacking model also outperformed the traditional ML algorithms used with the PIMA Indian Diabetes dataset, achieving better results with cross-validation and train-test splits. The ensemble approach demonstrates significant potential for improving early diabetes detection, benefiting both healthcare and machine learning applications.

A new ensemble feature engineering technique for machine learning-based diabetes detection was proposed by Rustam et al.[56]. A dependable feature extraction method was created by combining Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) models. The study overcome problems like small datasets and poor accuracy that are commonly present in current methodologies by integrating many datasets, increasing the dataset's size. The authors compared features from the CNN, LSTM, and Extra Tree Classifier models. With a Random Forest model and CNN-LSTM-based features, the ensemble approach yielded an exceptional accuracy score of 0.99 in the experimental data. Further testing using k-fold cross-validation confirmed that the proposed approach outperformed existing models, producing more reliable and accurate results for the identification of diabetes. With an emphasis on the early identification of Type 2 Diabetes Mellitus (T2DM) through machine learning, Faustin and Zou[57] suggested an improved version of the genetic algorithm (GA) to boost feature selection in a highly dimensional dataset. The problem of high processing costs and declining classification accuracy brought on by high-dimensional datasets is addressed in the work. In its crossover operator, the enhanced GA uses a two-step procedure: first, it defines a variable slice point, and then it uses feature frequency scores to determine whether to switch genes. With an accuracy of 97.5%, precision of 98%, recall of 97%, and F1-score of 97%, our method significantly improved the classification performance of the Pima Indian Diabetes dataset.

A useful framework for early diabetes detection via ensemble machine learning (ML) models was created by Saihood and Sonuç[59]. Data preparation, hyperparameter tweaking, and the use of ensemble techniques like bagging, boosting, and stacking are the three steps of the suggested approach. The Pima Indians Diabetes Database was used by the framework to conduct tests. The study showed that diabetes diagnosis accuracy was greatly increased by ensemble approaches, with the maximum accuracy of 97.50% being attained by the support vector machine (SVM) and stacked random forest (RF) models. Boosting and bagging, two other ensemble techniques, also demonstrated great accuracy rates. The study found that effective data preparation, calibrated models, and ensemble approaches are essential for improving diabetes diagnosis performance.

In order to diagnose diabetes, Daza et al.[60] suggested a stacking ensemble method that combines four models. The Diabetes Dataset, which has 768 patient records, was used in the study. Preprocessing techniques included cross-validation and data balancing using an oversampling technique. Four combined models based on stacking were proposed after seven separate algorithms were tried. Sensitivity, F1-Score, and Precision were all around 91.5% for the top-performing model, stacking 1 A (Logistic Regression with oversampling), which also had an accuracy of 91.5%. Additionally, the model obtained a good 97% ROC curve score. The study found that the stacking ensemble strategy outperformed individual models in diabetes diagnosis, leading to a considerable improvement.

Yadav and Pal[31] used ML to examine 1,025 heart disease occurrences and 14 factors. They examined three sequential ensemble approaches—XG Boost, AdaBoostM1, and Gradient Boosting—against Reduced Error Pruning, J48, and Decision Stump. A good matrix for heart disease features was provided by Pearson Correlation in the total study for ML prediction. Chi-Square, a feature selection technique, generated pertinent scores with related properties. Both feature selection techniques aided in the prediction of ML algorithms. This study's research was separated into two sections: the first concentrated on Decision Stump, Reduced Error Pruning, and J48 and generated an RF ensemble approach; the second focused on J48, Reduced Error Pruning, and Decision Stump and produced an RF ensemble method. To address Decision Stumps Reduced Error Pruning and J48, part two of the experiment used sequential ensemble approaches in the following order: AdaBoost M1, XG Boost, and Gradient Boosting.

Further reading on other types of diabetes milletus prediction can be found in the published articles by Zhou et al.[42] and Liang et al.[43].

Research gap

As discussed in Sect. 1.1 and 2.1, there have been notable advances in machine learning for disease prediction; nonetheless, there are still some gaps that this study aims to fill. First of all, problems like noisy data, overfitting, and underfitting cause many of the predictive models in use today to suffer from significant variability in diagnostic accuracy. These difficulties reduce the models' dependability in clinical settings when great precision is essential. Second, although many different algorithms have been investigated for the prediction of diabetes, there has been little integration of sequential and parallel ensemble approaches. By combining predictions from several algorithms, these methods could improve the resilience of the model. Thirdly, the literature currently in publication has not made full use of feature selection's ability to increase prediction model accuracy. Most studies do not fully integrate forward and backward feature selection techniques, which can significantly optimize the input feature set and improve model performance. Lastly, although many studies have employed ensemble methods, there is a need for more comprehensive research that not only compares these methods against traditional algorithms but also evaluates their synergy when combined. This research gap justifies the need for the current study's contributions, which focuses on using ensemble approaches and advanced feature selection strategies to produce a more accurate and reliable predictive model for diabetes mellitus.

Table 1 summarizes the state-of-the-art literature highlighting their strengths and limitations.

Methodology

This study proposes utilizing an ensemble model for diabetes attribute correlation strength that includes forward and backward features selection-based technique. The study adopted parallel and sequential ensemble approaches. The J48 method, Classification and Regression Tree (CART), and Decision Stump (DS) were used to produce a RF in the first experiment. The J48 algorithm, CART, and DS was used in the second phase of the experiment, along with three successive ensemble methods: XG Boost, AdaBoostM1, and Gradient Boosting.

The ultimate forecast was measured using average voting algorithms. The voting ensemble aggregates the decisions of various base classifiers for a specific classification or estimation with more options when it comes to integrating strategies to get the best categorization accuracy[32].

The algorithms used in this study are outlined below.

Algorithms used

1. 1. Random Forest (RF): An ensemble learning technique that generates the class mode for classification tasks by building several decision trees during training. It effectively handles a large number of features and prevents overfitting.
2. 2. J48 Decision Tree: A decision tree is generated from a dataset using this implementation of the C4.5 algorithm. It was selected because to its interpretability and effectiveness in managing both continuous and categorical data.
3. 3. Classification and Regression Tree (CART): A non-parametric decision tree approach that may be applied to both classification and regression tasks. It is straightforward and effective, with a clear mechanism for feature selection through its structure.
4. 4. Decision Stump (DS): A one-level decision tree that serves as a weak classifier. Its simplicity enables it to capture fundamental patterns and contribute to the ensemble's overall performance.
5. 5. XGBoost: A performance and speed-enhancing gradient boosting algorithm. It enhances model generalization and performs exceptionally well on binary classification tasks.
6. 6. AdaBoostM1: This adaptive boosting approach creates a strong classifier by combining many weak classifiers. By concentrating on cases that are incorrectly classified, it helps weak learners perform better.
7. 7. Gradient Boosting: This method builds models one after the other, fixing the mistakes of the earlier models. It successfully simulates intricate feature correlations and interactions.

Data collection and description

The Pima Indian dataset, which contains 768 instances and eight attributes, was used for this study's tests and analysis. It could be obtained at https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database. At the Gila River Indian Community Bank in Southern Arizona, Pima Indians have lived there since 1965. Lengthy research on diabetes and its repercussions involved this tribe. The population has the biggest diabetes dataset and the greatest prevalence of infectious diabetes in the world (50% of those under the age of 35)[33].

The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) is a government-funded research agency that focuses on kidney disease and diabetes[34]. As seen in Table 2, there are 768 instances in the

Name and author	Dataset	Methodology	Results and accuracy	Limitations	Future scope
Yadav and Pal[31]	UCI Repository	J48, Decision Stump, REP, RF, Gradient Boosting, AdaBoost M1, XGBoost	RF (Parallel) = 100%, XGBoost (Sequential) = 98.05%	Limited to ensemble methods	Explore hybrid models combining sequential and parallel approaches
Kumari, Kumar, and Mittal[39]	PIMA diabetes dataset (UCI)	RF, Logistic Regression, Naive Bayes	79.08% (PIMA), 97.02% (breast cancer)	Focused on soft voting ensembles	Extend to other medical datasets and more classifiers
Tewari and Dwivedi[32]	UCI dataset	JRIP, OneR, Decision Table, Boosting, Bagging	Bagging = 98%	Limited feature selection methods	Investigate more advanced feature selection techniques
Ghosh et al. 2021	PIMA Indians diabetes dataset	Gradient Boosting, SVM, AdaBoost, RF with and without MRMR feature selection	RF = 99.35% with MRMR	High complexity in MRMR feature selection process	Simplify feature selection and test on other datasets
Atif, Anwer, and Talib[44]	PIMA Indians dataset, Early-Stage Diabetes	Hard Voting Classifier (Logistic Regression, Decision Tree, SVM)	81.17% (PIMA), 94.23% (Early-Stage Diabetes)	Limited voting scheme to hard voting	Explore soft voting or weighted voting for improved results
Rashid, Yaseen, Saeed, and Alasaady [45]	PIMA Indians Diabetes Dataset (PIDD)	Decision Tree, Logistic Regression, KNN, RF, XGBoost	81% after standardization and imputation	Focused only on ensemble voting techniques	Test other ensemble strategies such as bagging or boosting
Zhou, Xin, and Li [46]	PIMA Indian diabetes dataset	Boruta feature selection, K-Means + + clustering, stacking ensemble learning	98%	High computational cost in clustering	Reduce computation and test scalability
Kawarkhe and Kaur[47]	PIMA Indians	CatBoost, LDA, LR, RF, GBC with preprocessing techniques	90.62%	Limited to specific preprocessing techniques	Broaden the preprocessing techniques and methods
Reza, Amin, Yasmin, Kulsum, and Ruhi[48]	PIMA Indian diabetes dataset, local healthcare	Stacking ensemble with classical and deep neural networks	77.10% (PIMA), 95.50% (simulation)	Limited to stacking approaches	Explore other ensemble methods or hybrid approaches
Thongkam et al.[17]	Breast Cancer Dataset	AdaBoost	Improved prediction and diagnosis	Initially used only for breast cancer	Extend to other medical conditions
Velu and Kashwan[18]	Various Datasets	SVM, Radial Basis Function, Multi-Layer Perceptron, and Multi-Level Counter Propagation Network.	High accuracy in various applications	Complexity in model selection	Test different combinations and optimizations
Temurtas et al.[19]	PIMA-diabetes illness dataset	Multilayer Neural Network	Improved accuracy	Focused on PIMA dataset	Apply to other chronic disease datasets
Ayo at al.[20]	Heart Disease Dataset	Levenberg–Marquardt approach, Probabilistic Neural Network, Naive Bayes, SVM	High accuracy in diagnosing cardiac disease	Limited to cardiac disease prediction	Broaden to include other comorbidities
Farvaresh and Sepehri[21]	Various medical datasets	Decision Tree C4.5, Bagging with C4.5, and Naive Bayes.	Improved prediction of cardiac illness	Initial focus on cardiac illness	Expand to other diseases and datasets
Kalman Filter Theory[22]	PIMA Indian dataset	Adaptive and personalized insulin recommendation	Enhanced classification accuracy	Focused on insulin recommendation systems	Broaden to other therapeutic recommendations
Ajagbe et al.[23]	Various applications	Multimedia analytic techniques, meta-data annotation, MPEG-7	Improved semantic analysis	Limited to MPEG-7	Explore alternative multimedia retrieval frameworks
Gong and Kim,[24]	Misbalanced Datasets	RHS-Boost for balanced classification	High accuracy and prediction	Designed for misbalanced datasets	Apply to other datasets and test alternative balancing methods
Purnami et al.[25]	Diabetes detection	ANFIS and PCA	Enhanced detection	Initial partitioning approach	Broaden to include other feature extraction methods
Rani and Jyothi[26]	Diabetes dataset	Bayesian Classification, J48, KNN, Filtered Classifier, ANN, Naive Bayes	77.01% accuracy	Lack of cross-validation	Implement cross-validation and expand dataset usage
Zheng et al.[27]	Various datasets	KNN, Naive Bayes, Decision Tree, RF, SVM, Logistic Regression	Improved recall and accuracy	Filtering criteria could be improved	Enhance feature selection and parameter tuning
Komi et al.[28]	Sample datasets	ELM, ANN, LR, GMM, SVM	Better accuracy with fewer samples	Less amount of sample data	Increase sample data and test on more complex datasets
Sai et al.[29]	Diabetes dataset	Weighted voting approach for ensemble prediction models	Enhanced predictive performance	Focused on ensemble prediction model	Explore ensemble expansion and optimization
Rustam et al.[56]	Multiple datasets	Ensemble of CNN and LSTM for feature extraction, Random Forest model for prediction	Accuracy score of 0.99 using CNN-LSTM features with Random Forest	Limited by dataset size, generalizability issues in existing approaches	Explore other ensemble models, improve dataset diversity, real-world applicability
Faustin and Zou et al.[57]	Pima Indian Diabetes Dataset	Genetic Algorithm (GA) enhanced with a two-step crossover operator for feature selection.	Accuracy: 97.5%, Precision: 98%, Recall: 97%, F1-score: 97%	Premature convergence due to insufficient population diversity in GA.	Apply the improved GA to other datasets, refine crossover technique.
Reza et al.[58]	PIMA Indian Diabetes dataset, Local healthcare data	Stacking ensemble method combining classical and deep neural network models for diabetes classification.	Stacking ensemble with NN architectures: Accuracy 95.50%, Precision 94%, Recall 97%, F1-score 96%	Limited to dataset used in the study; may need more diverse data for generalization.	Explore further with other datasets, and apply to real-time healthcare systems.
Saihood and Sonuç[59]	Pima Indians Diabetes Database	Ensemble machine learning models: Bagging, boosting, and stacking with hyperparameter tuning and data preprocessing.	Stacking (RF & SVM): 97.50% accuracy, Bagging (RF): 97.20%, Boosting (XGB): 97.10%	Framework limited to Pima Indians dataset; real-world data may vary.	Extend framework to include more diverse datasets and explore real-time applications in healthcare.
Daza et al.[60]	Diabetes Dataset (768 patient records)	Stacking ensemble approach using 7 base algorithms; oversampling to balance the dataset and cross-validation for model training.	Best accuracy: 91.5%, Sensitivity: 91.6%, F1-Score: 91.49%, Precision: 91.5%, ROC Curve: 97%.	Performance depends on the dataset and oversampling method.	Improve the model's generalizability by testing on other datasets and enhancing model

Table 1. Summary of the state-of-the-art literature.

Attribute	Description	Type
Pregnant	Pregnancy frequency	Real
Plasma	Plasma glucose levels during an oral glucose tolerance test	Real
Diastolic	Blood pressure diastolic (mm/Hg)	Real
Triceps	Triceps skinfold measurement (mm)	Real
Serum Insulin	Serum insulin after two hours (µU/ml)	Real
BMI	Body Mass Index, or BMI (kg/m2)	Real
Pedigree Fun	Pedigree function for diabetes	Real
Age(years)	Age of patient	Real
Diabetes	Status (0-Healthy, 1-Diabetes)	Discrete

Table 2. Pima Indian diabetic patient data.

Pima Indian Diabetes Dataset, each with eight input attributes and one output attribute. All input characteristics with 768 female samples only offer numeric data values, and the first character shows the number of pregnant patients. The body's glucose levels are the second feature. The third feature is the diastolic blood pressure measurement in millimetres of mercury (mmHg). Skin thickness in millimetres is the fourth attribute. The 5th, 6th, and 7th attributes define the overall quantity of insulin produced. Equation 1 defines the body mass index (BMI)

$$BMI = \frac{Patient' sweight\ in\ Kg}{Patients\ height\ in\ meter} \tag{1}$$

of infected patients and the structure of the diabetic family, respectively the patients' present age is the final characteristic stated. The dataset was analysed using the suggested classification algorithms, parallel and sequential ensemble methods, and feature selection strategies to develop the diabetes prediction model.

System architecture and flowchart

At this point, all of the experimental combinations were created using ML methods. The data description part explains several diabetes-related problems and their associated attributes, as well as visualizes their distribution using Python classification tools. For diabetic attribute correlation strength, forward and backward feature selection-based algorithms were applied. This study proposes utilizing an ensemble model for diabetes attribute correlation strength that includes forward and backward features selection-based technique. The study adopted parallel and sequential ensemble approaches. The J48 method, CART, and Decision Stump (DS) were used to produce a RF in the first experiment. The J48 algorithm, CART, and DS was used in the second phase of the experiment, along with three successive ensemble methods: XG Boost, AdaBoostM1, and Gradient Boosting. Average voting algorithms was used to measure the final Prediction. To reach the highest level of classification accuracy, it offered flexibility in combination tactics[32]. Figure 1 illustrates this explanation of the system design.

The flowchart for the prediction system is depicted in Fig. 2 and explained as follows. The dataset was inputted, pre-processed, the selection of high features was made using backward and forward features selection technique. If diabetes was found in the processed data, then the system will signify the type of diabetes, and if the processed data is normal, the system execution stops.

Techniques for selecting features

In ML, feature selection is used to identify the ideal set of characteristics for building accurate models of the phenomena under study.

Reasons for using feature selection

To simplify models and make them easier for academics and practitioners to understand, feature selection is used. By limiting overfitting, it helps shorten training periods, minimize the curse of dimensionality, and enhance generalization (formally, lowering variance).

Wrapper methods

Wrapper techniques need a method for searching the space for all viable feature subsets and evaluating each one's quality by training and testing a classifier on it. We use a ML technique to discover features and attempt to tailor it to a dataset. All possible feature combinations are evaluated to the evaluation criterion using a greedy search technique. In many cases, wrapper methods are more accurate predictors than filter methods.

Forward feature selection

An iterative method for choosing the variable that performs the best relative to the objective is called "forward feature selection." The next step is to select a second variable that, when paired with the first, yields the greatest results. The predefined condition is not satisfied until this procedure is used.

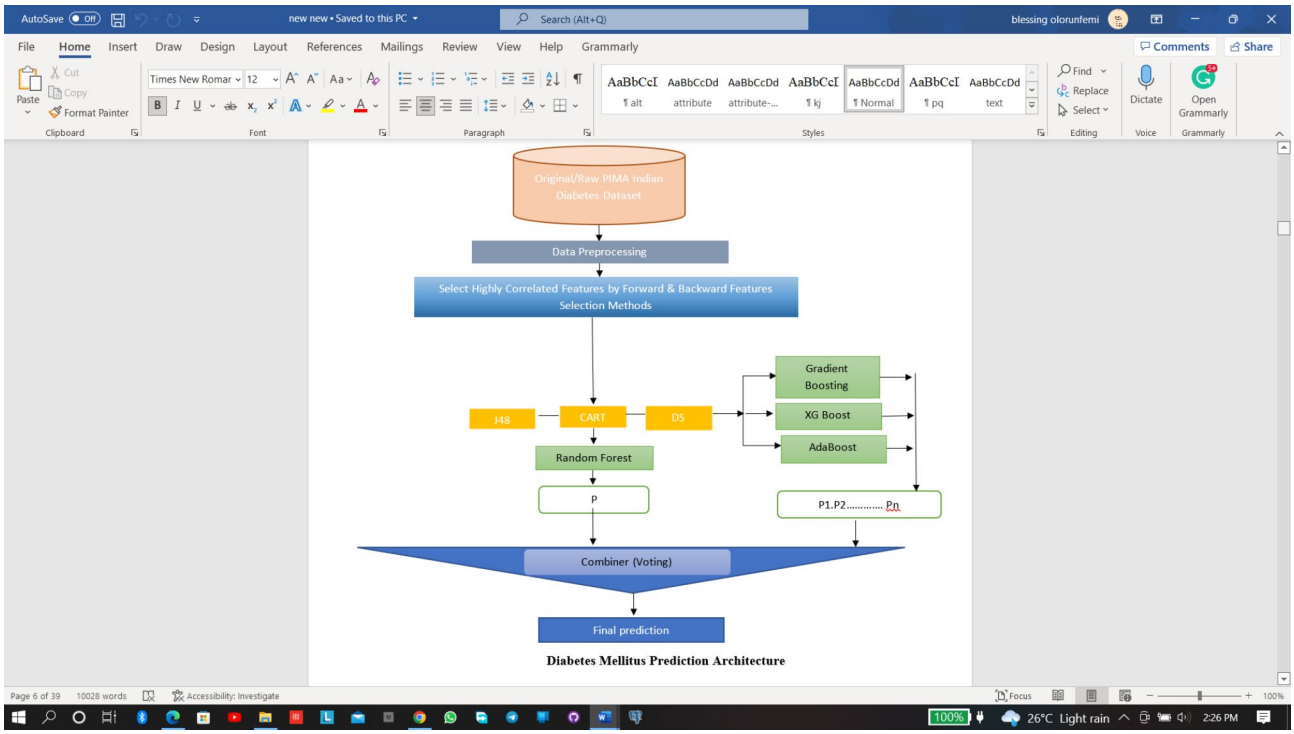

Fig. 1. Diabetes mellitus prediction architecture.

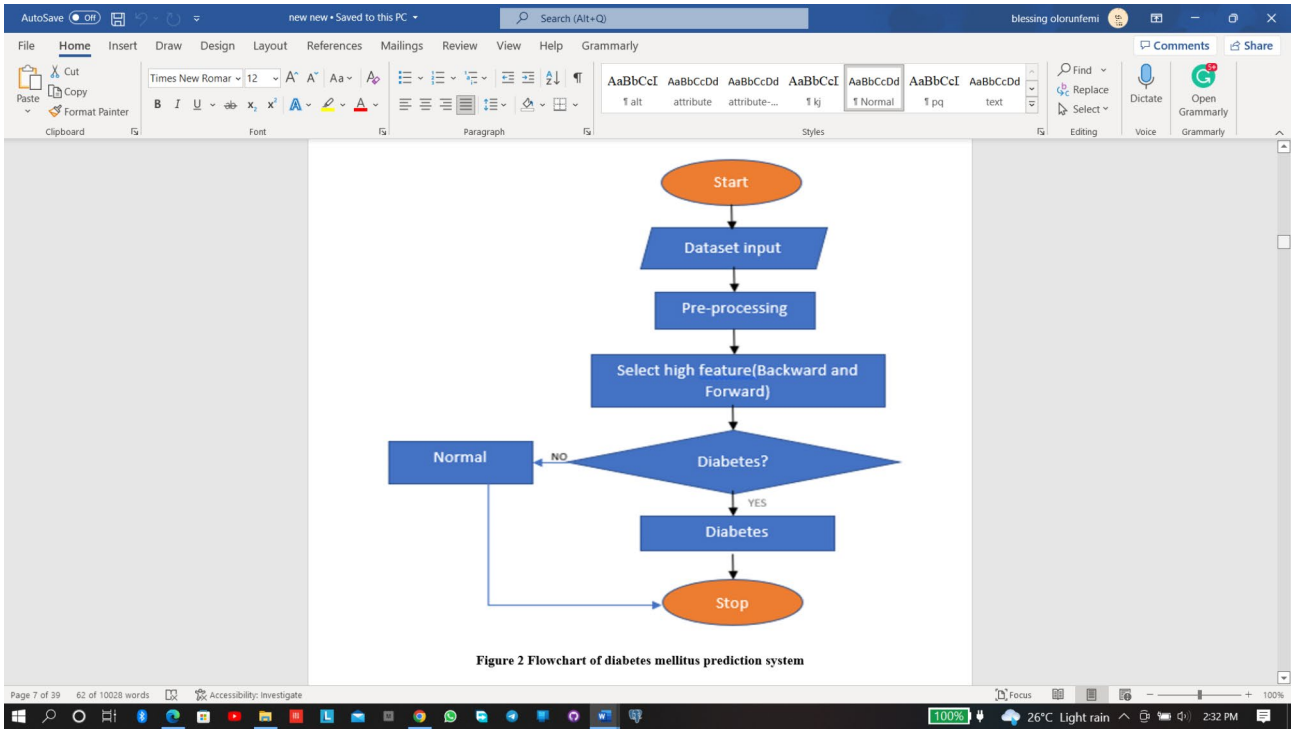

Fig. 2. Flowchart of diabetes mellitus prediction system.

Backward feature selections
The Forward Feature Selection technique is diametrically opposed to this notion. The process begins with all of the available features and work its way up. The model variable with the highest evaluation measure value is then selected. This technique is carried out until the predefined condition is met. This approach is also known as Sequential Feature Selection, in addition to the one discussed above.

Parallel ensemble method
Similar ensemble techniques, like as a RF, produce base learners in the same manner. Employing parallel generation of foundation learners is one technique to strengthen the independence of base learners.

Sequential ensemble method
Base learners are developed sequentially using adaptive boosting and other sequential ensemble approaches (AdaBoost). Basic learners are prompted to rely on one another through their gradual progress. The performance of the Model is then improved by giving previously underrepresented learners more weight. J48, Decision Stump, CART, AdaBoostM1 ensemble method, Gradient boosting ensemble method and XG Boost Ensemble Method were used in the work[35–38].

Evaluation metrics

The Confusion Matrix, which lists the right and wrong predictions, was the first performance indicator used to assess the model. Each of the two output classes is represented by the numbers 0 and 1, where 0 denotes no diabetes and 1 denotes diabetes. The confusion matrix's diagonal values display accurate predictions, whereas the off-diagonal values reflect incorrect classifications. The terms used in the matrix are:

i. True Positives (TP): are cases in which the model correctly predicted the positive class (diabetes).
ii. True Negatives (TN): Situations in which the model correctly predicted the negative class (no diabetes).
iii. False Positives (FP): Cases in which the model predicted diabetes when none existed.
iv. False Negatives (FN): When diabetes was present, the model mispredicted that it wasn't.

These terms allow to derive several important metrics for evaluating the model's performance:

1. Accuracy, which is the percentage of correctly identified occurrences (both positive and negative) out of all instances, gauges the model's overall efficacy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

2. Precision assesses how well the model detects positive diabetes patients while reducing false positives:

$$\text{Precision}\ \frac{TP}{TP + FP} \tag{3}$$

3. To reduce false negatives, recall (also known as sensitivity or true positive rate) assesses the model's ability to recognize all true positive cases (e.g. diabetes):

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

4. The F1 score is a balanced assessment of the model's performance, calculated as the harmonic mean of precision and recall. It's extremely useful for unbalanced datasets, like diabetes diagnosis:

$$\text{F1 Score} = 2^{*}\ \frac{\boldsymbol{Precision} \times \boldsymbol{Recall}}{\boldsymbol{Precision} + \boldsymbol{Recall}} \tag{5}$$

5. Matthews Correlation Coefficient (MCC) provides a comprehensive performance metric that evaluates the quality of binary classifications by considering all four terms from the confusion matrix (TP, TN, FP, FN). MCC is particularly valuable for datasets with significant class imbalance, offering a balanced assessment:

$$\text{MCC} = \frac{TP \times FP - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TP + FN)}} \tag{6}$$

MCC values range from − 1 (completely wrong classification) to + 1 (perfect classification), with 0 indicating no better than random prediction.

6. AUC-ROC measures the model's ability to differentiate between positive (diabetes) and negative (no diabetes) classes at different thresholds. A perfect classifier has an AUC of 1.00, but a random classifier scores 0.5.
7. The Area Under the Precision-Recall Curve (AUC-PR) provides a more detailed knowledge of the precision-recall trade-off, which is especially valuable when working with imbalanced datasets. A high AUC-PR indicates that the model maintains good precision and recall despite the class imbalance.

Implementation and results

As previously stated, the dataset comprises 9 characteristics and 768 instances that were utilized in the Model's development. The data was stored in the diabetes.csv CSV file. The plot in Fig. 3 shows that the data contains more cases without diabetes (0) than those with diabetes (1). The number of non-diabetics is 268 while the number of diabetic patients is 500.

Data pre-processing

This section describes the methods and findings gained during data preprocessing to achieve the best outcomes. This procedure is executed until the predefined condition is met.

Mean imputation

The technique of approximating missing values in a dataset using valid (non-missing) values is known as imputation. This method makes use of known dataset values to create correlations that can help identify missing values. The average of the remaining cases was used to fill in the gaps. The assumption of imputation is that the sample is statistically homogeneous.

Generalized imputation

The missing value is replaced by the average of all non-missing values for that variable. First, zeros were used instead of NaN since counting them was easier, and the zeros required had the right values. To fill these NaN values, the association with characteristics was recognized. The heatmap in Fig. 4 indicates greater associations

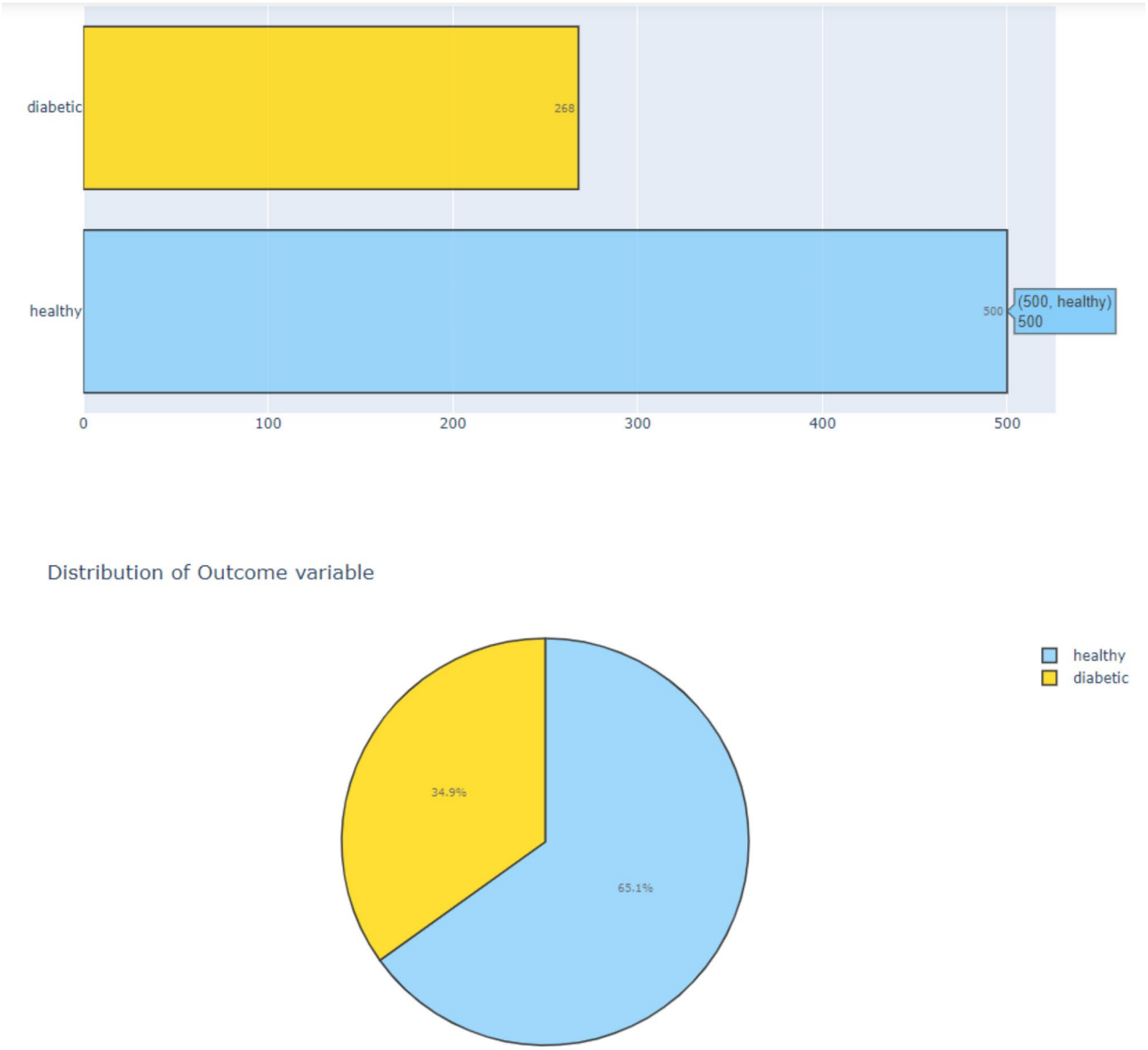

Fig. 3. Data distribution.

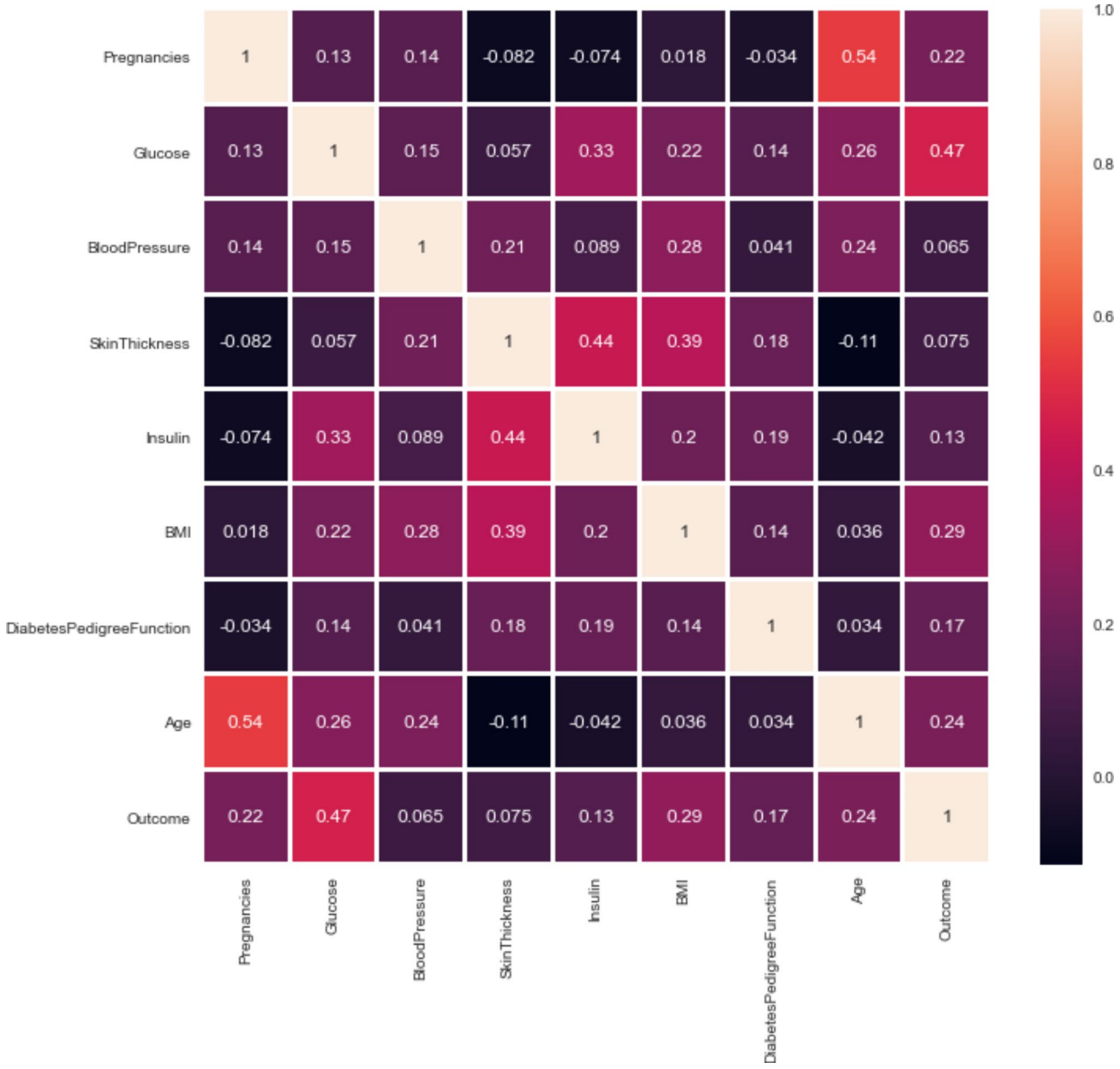

Fig. 4. Heatmap of feature (and outcome) correlations.

with brighter colours. The heatmap shows significant connections between the outcome variable and glucose levels, age, BMI, and number of pregnancies. There is a high association between features such as age and pregnancy, insulin and skin thickness.

Additionally, features were scaled to unit variance and the mean was removed using Standard Scaler to normalize them. The standard score for Sample X is computed as follows:

$$z = \frac{x - U}{S} \tag{7}$$

where u represents the mean of the training samples, or zero if mean = False, and std represents the standard deviation of the training samples, or one if std = False.

By calculating the relevant statistics on the samples from the training set, each feature is individually centred and scaled. The mean and standard deviation are then recorded and utilized to transform following data.

Implementation of forward and backward feature selection

Forward and backward strategies were used in this study to pick the relevant features for diabetes attribute correlation strength in the dataset, allowing the ML algorithm to be trained faster. By choosing the proper number of features, the model's complexity was minimized, which made it easier to read, as well as reduced over-fitting. Two trained subsets were created from the dataset. Create a model using a dataset (from the recently installed import Sequential Feature Selector and the sklearn library). Import LinearReggression because predicting the course of diabetes requires that the target variable and the independent variables be defined explicitly when working on regression issues. After eliminating the "Pregnancies," "Outcome," variable, X was employed as the independent variables in this case. Additionally, "OUTCOME," the goal variable, was chosen as Y.

The python function lreg, which stands for a linear regression model, is the first parameter in the forward feature selection technique used. The attribute *features* in the code specified how many features should be chosen. In this case, *features* was set to 4. Up until four features were chosen, four models were trained. When using the forward feature selection approach as opposed to the backward elimination method, the parameter *forward* was set to True, which entails training the forward feature selection model. When using the backward feature removal strategy, this option was set to False. When verbose = 2, the model summary was displayed after each iteration. Starting with the whole Model (containing all of the independent variables), backward elimination was performed by deleting the unimportant features with the highest p-value (higher significance level). Up to the last set of significant traits, this procedure was repeated. Figure 5 showed that the same set of features were re-selected in a backward elimination procedure, which included age, BMI, diabetes pedigree function and glucose.

Experimentations and results evaluation

Output of the model

The output of the Model, as well as the metrics used to evaluate it, were provided in this section. The parameters utilized to fine-tune the Model and the re-evaluation procedure in order to reach the best possible outcome were explained. Each category's efficacy was assessed using sensitivity (S.N.), specificity (S.P.), accuracy, and Matthew's correlation coefficient (MCC).

Predictive modelling

The model's performance was evaluated using training and testing accuracy scores, classification reports, and a confusion matrix. The confusion matrix was utilized to compute the actual positive rate (recall), accuracy, false-positive rate, AUC_ROC, and AUC_PR false-negative rate (miss rate). Figures 6, 7, 8 and 9; Table 2 show the classification report of the trained Model for the parallel method. When given the training set, the model obtained an accuracy of 98% using the parallel technique.

The confusion matrix in Fig. 6 gives us additional information about the Model's performance. It was found out that the Model attained a true positive of 48%; also, it was able to achieve a true negative for 131% of its predictions. However, 36% of its predictions wrongly predicted a sample to be diabetes when it was not. It was also observed that it has a false negative of 41, which means 41% of its predictions for a decision slump.

Figure 7 depicts the confusion matrix, which offers information about the model's performance. When J48 was implemented, the Model was found to have 90 true positives and 307 true negatives of its predictions. However, 26 of its predictions incorrectly indicated diabetes in a sample. It was also discovered that it has a false negative rate of 89, implying that 89 of its predictions were incorrectly identified as negative while they were diabetic.

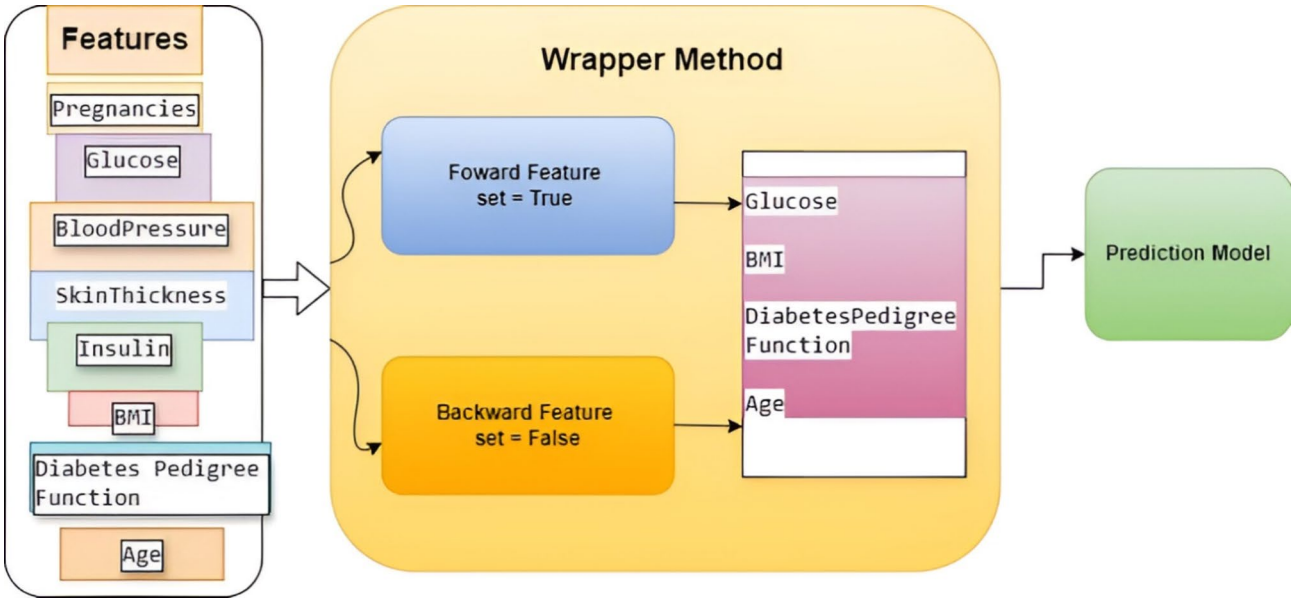

Fig. 5. Forward and backward feature selection prediction model.

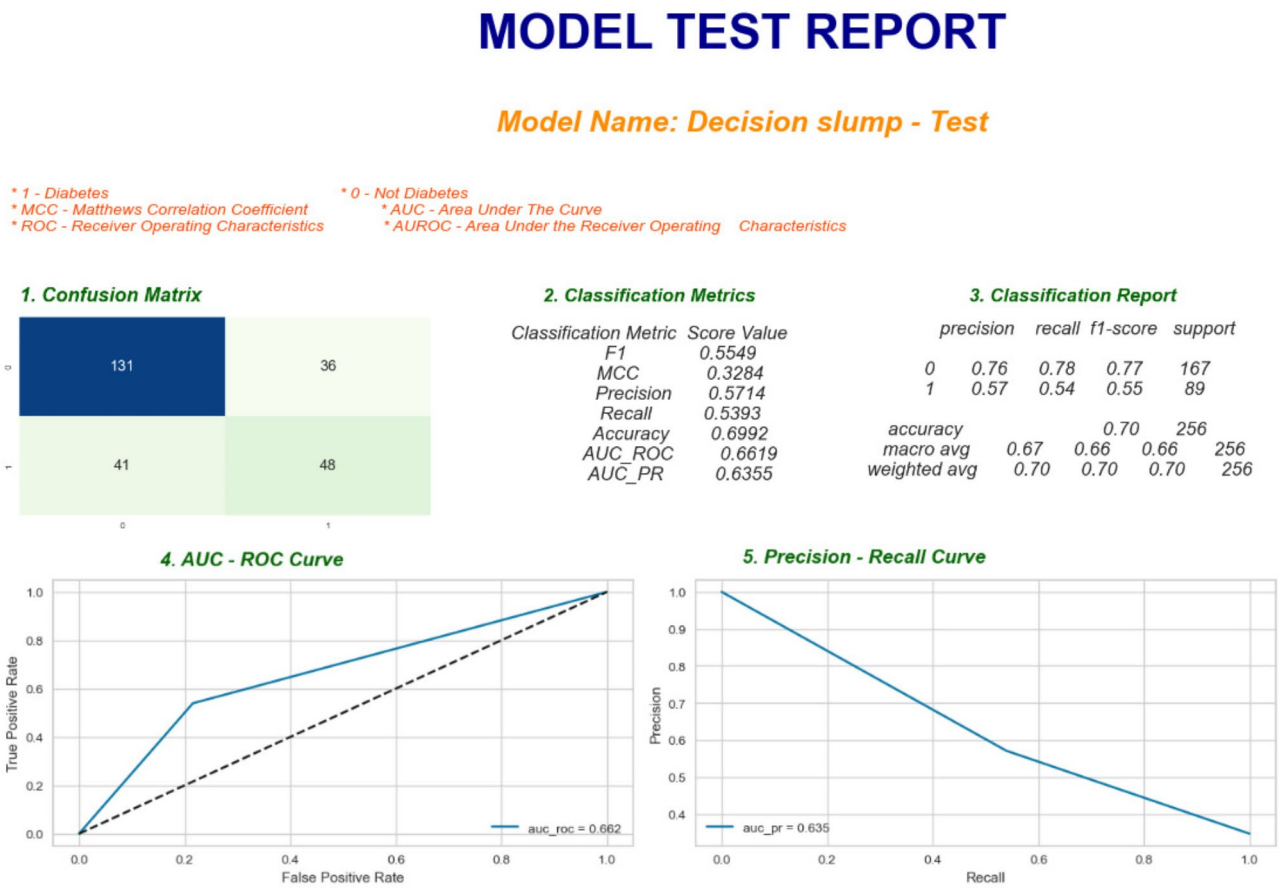

Fig. 6. Output of decision stump.

The confusion matrix in Fig. 8 depicts the Model's performance for the CART. The Model had a true positive of 97 and a true negative for 305 of its forecasts. However, 28 of its predictions wrongly predicted a sample to be diabetes. It was also observed that it has a false negative of 82, which means 82 of its predictions were predicted to be negative while in the actual sense they are positive (diabetic).

Figure 9 confusion matrix illustrates how well the model performed when Random Forest (RF) was applied. The Model was found to have 171 accurate positive predictions and 330 true negative predictions. 3 of its projections, however, were incorrect. It was also noted that it had a false negative of 8, meaning that it incorrectly predicted that 8 people would develop diabetes. This study concentrated on three algorithms: J48, Decision Stump, and CART to obtain an RF ensemble method. In comparison to the other three separate approaches, the RF ensemble methodology had a classification accuracy of 98%. Table 3 demonstrates that RF outperformed J48, CART, and Decision Stump algorithms in terms of f1-score, MCC, Precision, Recall, AUC ROC, and AUC PR.

In the second comparison, three algorithms (J48, CART, and Decision Stump) and three sequential ensemble techniques (AdaBoost, XG Boost, and Gradient Boosting) were investigated.

Hyperparameter tuning

Figures 10, 11, and 12 show that XGBoost, Adaboost, and Gradient Boosting all achieved 100% classification accuracies in the experiments conducted. This was made possible by the hyperparameter tuning presented here in this section. To improve the performance of machine learning models, hyperparameter tuning was conducted using RandomizedSearchCV. This technique enables systematic exploration of different hyperparameter combinations to identify the optimal settings for each model. Max_depth, learning rate, n_estimators, min_child_weight, subsample, and colsample_bytree were among the several hyperparameters investigated for the XGBoost classifier. The hyperparameter grid was selected with care to encompass a wide range of values, and the performance of each combination was assessed using 5-fold cross-validation. The best parameters were identified as max_depth = 26, learning rate = 0.3, n_estimators = 1000, subsample = 0.4. Similarly, for the AdaBoost and Gradient Boosting classifiers, hyperparameter tuning was performed. While the tuning of Gradient Boosting concentrated on max_depth, learning rate, n_estimators, and subsample, AdaBoost optimized learning_rate and n_estimators. This approach ensured that the models were configured with the most effective hyperparameters, contributing to improved classification accuracy and model performance.

The performance of the Model was provided by the confusion matrix in Figs. 10, 11, and 12. A true positive of 179 was recorded by the model, and 333 of its predictions were found to be true negatives. Conversely, it

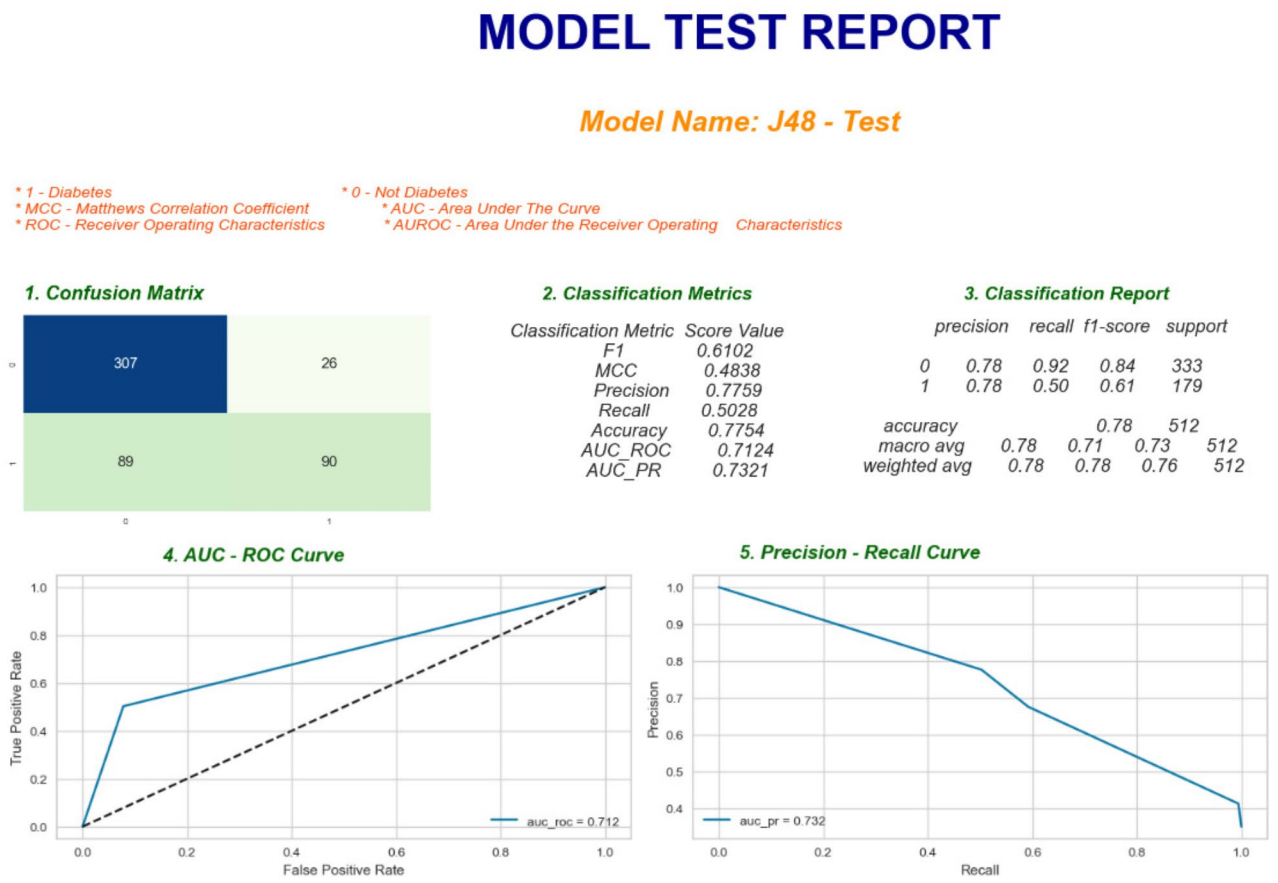

Fig. 7. Output of J48 prediction algorithm.

produced 0 false negatives and 0 false positives. To put it another way, its forecast accuracy is 100%. The f1-score, MCC, Precision, Recall, AUC_ROC, AND AUC_PR of XGboost[39–41]. Adaboost and Gradient boosting were higher when compared to the base leaners in Table 4.

Machine learning model experiments on eight (8) features

Eight features—pregnancies, blood pressure, glucose, skin thickness, insulin, BMI, diabetes pedigree function, and age—were used to compare the models' overall performance. Gradient Boosting was found to perform better than the other approaches in both testing and training. Notably, as forward and backward feature selection methods were not used in this analysis, they are not included in this comparison. Table 5 provides a summary of the findings.

The output of the final Prediction on voting ensemble method was computed to be 100% accurate, as shown in Fig. 13.

K-Fold cross-validation for model evaluation

During training and evaluation, K-Fold cross-validation was employed to guarantee the robustness and generalizability of the classification models. The method separates the dataset into K subsets, or folds, with the remaining K-1 folds acting as the training set and each fold as a validation set. To guarantee that every data point is present in both the training and validation sets, this procedure is carried out K times. A more reliable estimate of the model's performance is obtained by averaging the performance metrics across all folds.

In our work, 5-fold cross-validation was employed to assess different classifiers such as Random Forest, CART, J48, Decision Stump, AdaBoost, Gradient Boosting, and XGBoost. Several metrics were used to analyse each model's performance, including accuracy, precision, recall, F1-score, Matthew's correlation coefficient (MCC), AUC-ROC, and AUC-PR. Table 6 presents the performance metrics for each classifier after applying K-Fold cross-validation. The Random Forest classifier achieved an accuracy of 100%, demonstrating its exceptional ability to generalize to unseen data. Both AdaBoost and Gradient Boosting also recorded accuracy scores of 100%, indicating their effectiveness in classification tasks. However, the performance metrics varied among the different models, emphasizing the necessity of picking the proper model depending on the specific characteristics of the dataset, as shown in Table 6.

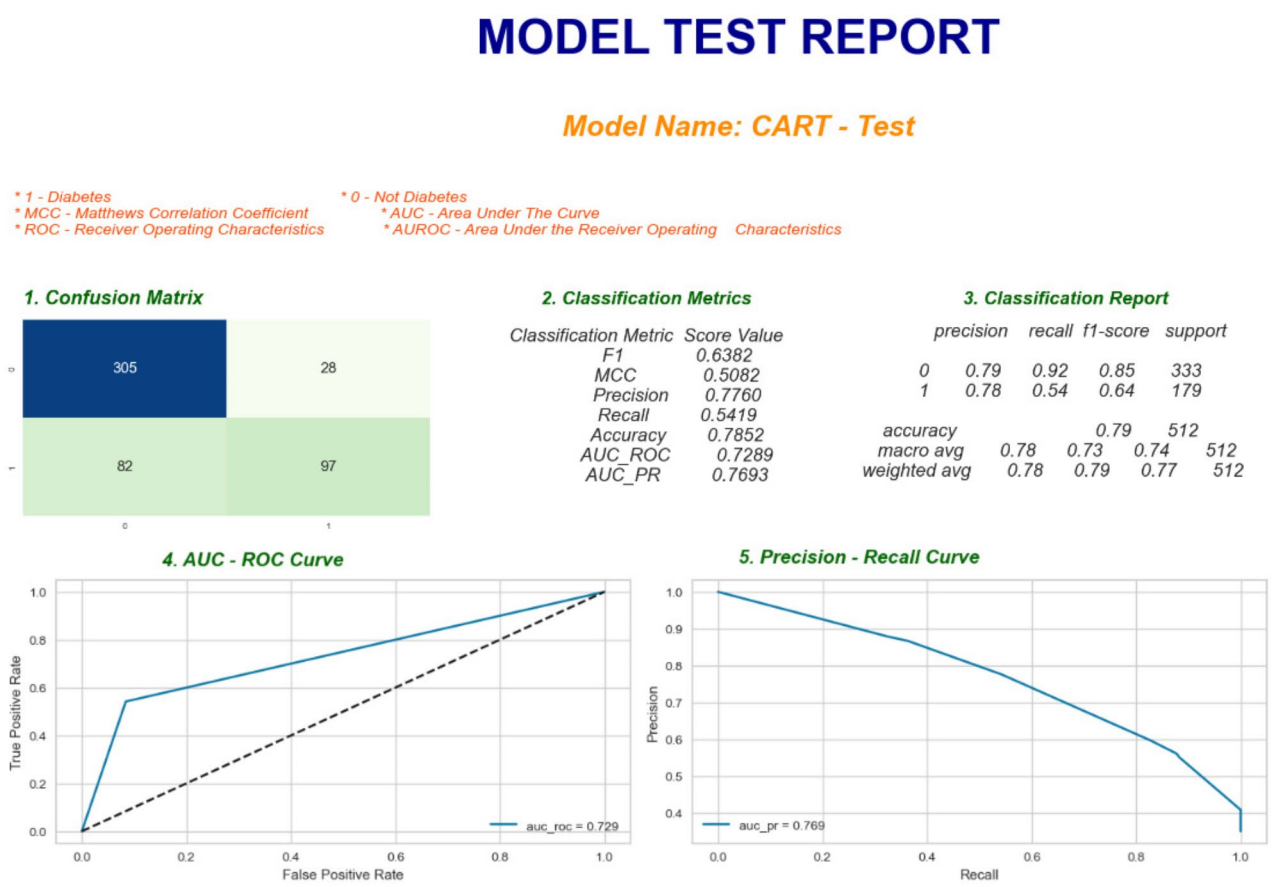

Fig. 8. Output of CART prediction algorithm.

Comparison of diabetes datasets and model performance

Two public diabetes datasets were evaluated: one from Rashid and Ahlam[49], titled "Diabetes Dataset" available on Mendeley Data, and the other from Kaggle called diabetes_prediction_dataset.csv. The proposed model in this work was tested against various machine learning models using both datasets after applying forward and backward feature selection. The results demonstrated consistent performance of the proposed model, indicating its robustness and ability to generalize well on new datasets as shown in Tables 7 and 8.

Evaluation of results

Based on the set of features used in building the model in this work, it is quite difficult to make direct comparisons with other studies. However, results from the most related studies were identified and used to substantiate the results in this work. The works used for the comparison were the ones where the authors used ML techniques in the discovery of DM on Pima Indians dataset. The summary of the result comparison is given in Table 9, which showed that the current study performed better than the existing related methods in terms of prediction accuracy.

Discussion of results

In this study, the high accuracy and performance metrics indicate that the sequential ensemble approaches considerably improved the model's ability to predict diabetes with high precision. The use of Forward and Backward feature selection methods ensured that the most relevant features were included in the model, contributing to its superior performance.

The results from our study demonstrate a significant advancement in diabetes diagnosis accuracy compared to previous research. For instance, the proposed study outperformed Yadav and Pal[31]. Their study achieved 100% accuracy with Random Forest (RF) in parallel ensemble methods and 98.05% with XGBoost in sequential ensembles. While these results are impressive, our sequential ensemble methods—XGBoost, AdaBoost, and Gradient Boosting—reached 100% accuracy, surpassing their performance. The perfect scores across all sequential models indicate a robust approach to diabetes diagnosis. Also in Kumari, Kumar, and Mittal[39], their ensemble soft voting classifier achieved 79.08% accuracy on the PIMA diabetes dataset, which is significantly lower than our results. Despite their use of ensemble methods combining RF, logistic regression, and Naive Bayes, the proposed study's sequential ensemble methods delivered a perfect classification accuracy of 100%, showcasing a substantial improvement.

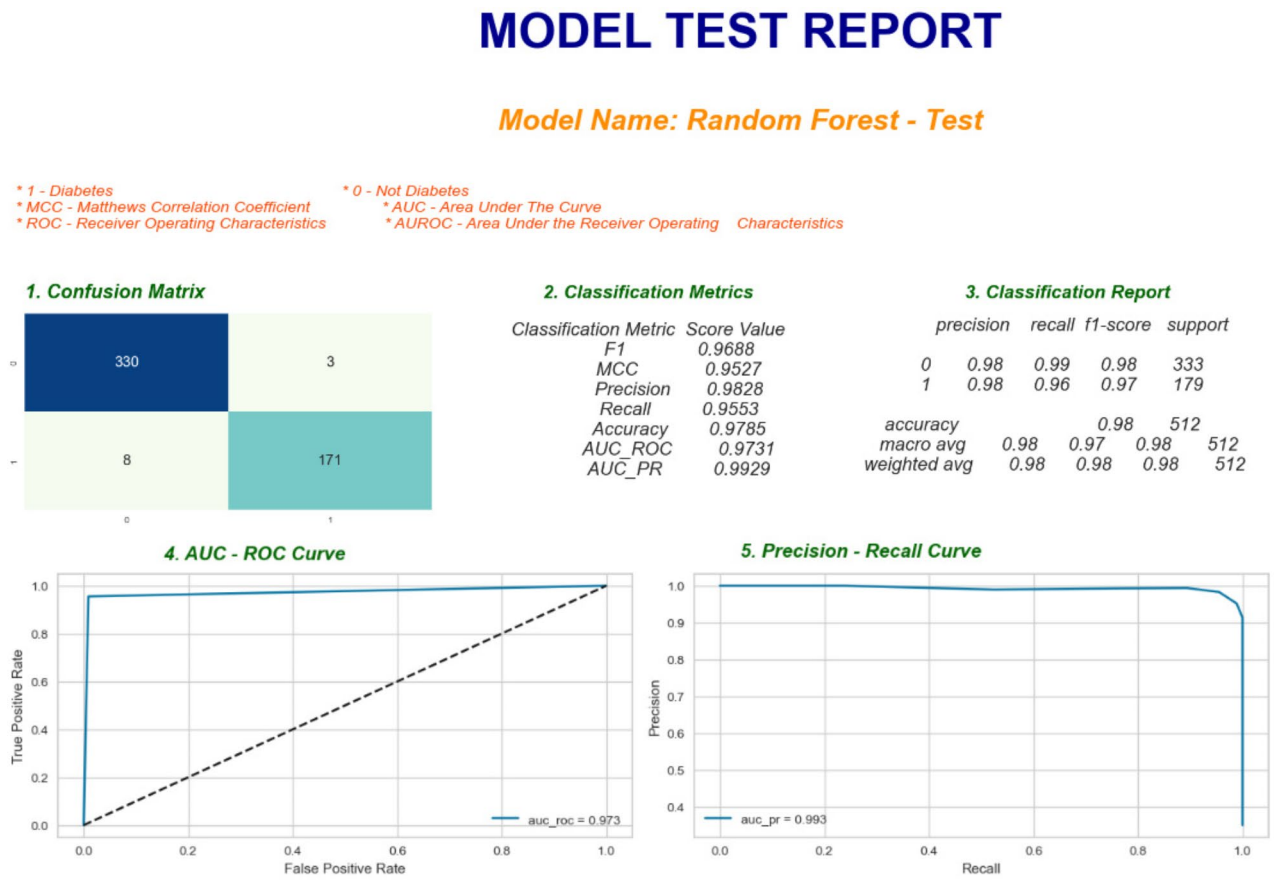

Fig. 9. Output of random forest (RF) prediction algorithm.

Statistical analysis	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MCC (%)	AUC_ROC (%)	AUC_PR (%)
Decision slump	69.92	57.14	53.93	55.49	32.84	66.19	66.55
J48	77.54	77.59	50.28	61.02	48.38	71.24	71.21
CART	78.52	77.60	54.19	63.82	50.82	72.89	76.93
RF	97.85	98.28	95.53	96.88	95.27	97.31	99.29

Table 3. Model classification report for parallel ensemble method.

Tewari and Dwivedi[32] reported 98% accuracy using a bagging ensemble method. Our results not only match but exceed their accuracy, with our sequential ensemble methods achieving perfect scores. This improvement highlights the effectiveness of our feature selection and ensemble approach in capturing and leveraging the data's underlying patterns. Moreover, Rashid et al.[45] achieved 81% accuracy with an ensemble voting classifier using a variety of models, including Decision Trees and XGBoost. This study's sequential ensemble methods significantly outperformed this with 100% accuracy. This difference underscores the potential of our approach to deliver more reliable predictions.

The stacking ensemble method with feature selection presented in Zhou, Xin, and Li[46] achieved 98% accuracy. Our sequential ensemble methods also reached 100% accuracy, demonstrating that our approach is not only competitive but superior in terms of classification performance. Kawarkhe and Kaur[47] achieved 90.62% accuracy with various ensemble classifiers. Our results of 100% accuracy with sequential ensemble techniques suggest a more effective model for diabetes diagnosis. Reza et al.[48] reported 77.10% accuracy on the PIMA dataset with a stacking ensemble method and 95.50% using deep neural networks. While their deep learning methods are impressive, but the sequential ensemble methods also achieved 100% accuracy, providing an alternative high-performance approach.

The proposed study demonstrates a clear advancement in diabetes diagnosis accuracy compared to the methods reported in previous research. The use of advanced sequential ensemble techniques, coupled with effective feature selection, has resulted in a model that achieves perfect classification accuracy. This indicates a significant improvement in diagnostic capabilities and sets a new benchmark for future research in this area.

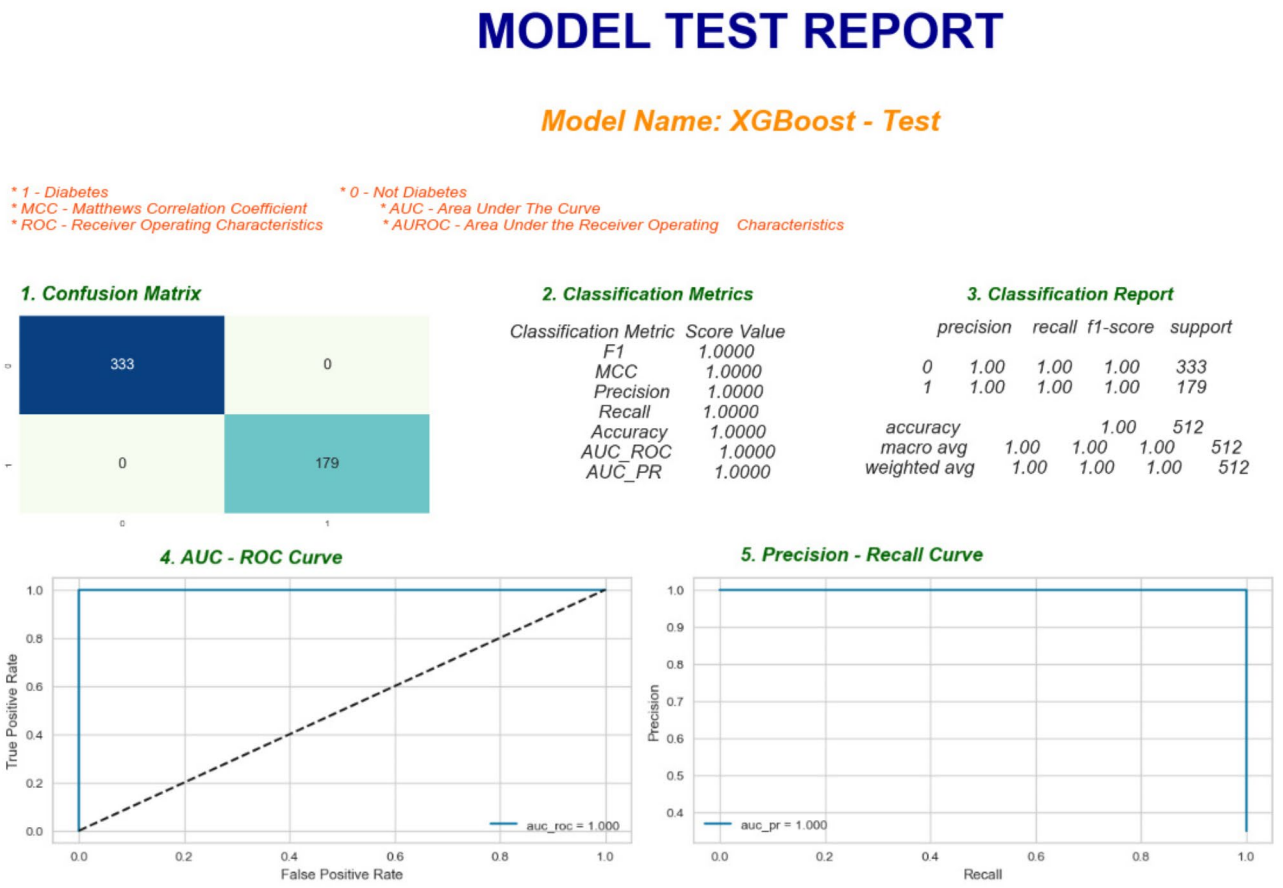

Fig. 10. Output of XG boost.

Conclusions

Recapitulation

Diabetes Mellitus is a serious health issue, and it affects people all over the globe. It is a diverse set of diseases characterized by a chronic increase in blood glucose levels. It rises because of the failure of the body to produce insulin that is enough to meet an individual's needs. Diabetes affects around 400 million people worldwide. Early detection and treatment of such diseases can save lives. This research work analysed diabetes mellitus dataset obtained from the Kaggle repository, which has nine attributes with 789 instances, to achieve the goal. The feature selection technique termed forward, and backward features was adopted in the study, and this generated valuable scores with comparable features in the whole analysis. The prediction was helped by both feature picking procedures. The trials in this study were divided into two groups: one that concentrated on J48, CART, and Decision Stump and produced a RF ensemble approach, and the other that concentrated on J48, CART, and Decision Stump. The classification accuracy for this parallel ensemble technique was 96%. The second portion of the experiment includes three sequential ensemble methods: AdaBoostM1, XG Boost, and Gradient Boosting together with J48, CART, and Decision Stump. The AdaBoostM1 X.G.Boost ensemble and Gradient Boosting ensemble techniques all produced superior results with 100% classification accuracy. In this work, the sequential ensemble models XG Boost, AdaBoostM1, and Gradient Boosting outperformed the fundamental learning techniques. The results from experiments conducted in this study showed that AdaBoostM1, XG Boost, and Gradient Boosting performed better than the base learning algorithms applied in this work. Researchers and developers can leverage on the predictive Model developed in this work to make quick predictions of diabetes mellitus, which could save many lives. This work introduced a novel combination of parallel and sequential ensemble methods with feature selection techniques for the prediction of DM, which played a vital role in resolving the problems of noisy data, over/underfitting, residual errors associated with base-level models. The research found that XGBoost, AdaBoostM1, and Gradient Boosting achieved 100% classification accuracy in diabetes detection. Feature selection improved model performance by focusing on the most relevant features. Sequential ensemble methods outperformed traditional algorithms, demonstrating superior performance metrics. The developed model offers a robust tool for accurate and efficient diabetes diagnosis, with potential practical applications in clinical settings. This study enhances diabetes detection by employing advanced ensemble methods and feature selection techniques with hyperparameter tuning to achieve high classification accuracy, offering a novel and more effective approach compared to existing methods, and providing valuable insights and benchmarks for future research and practical applications.

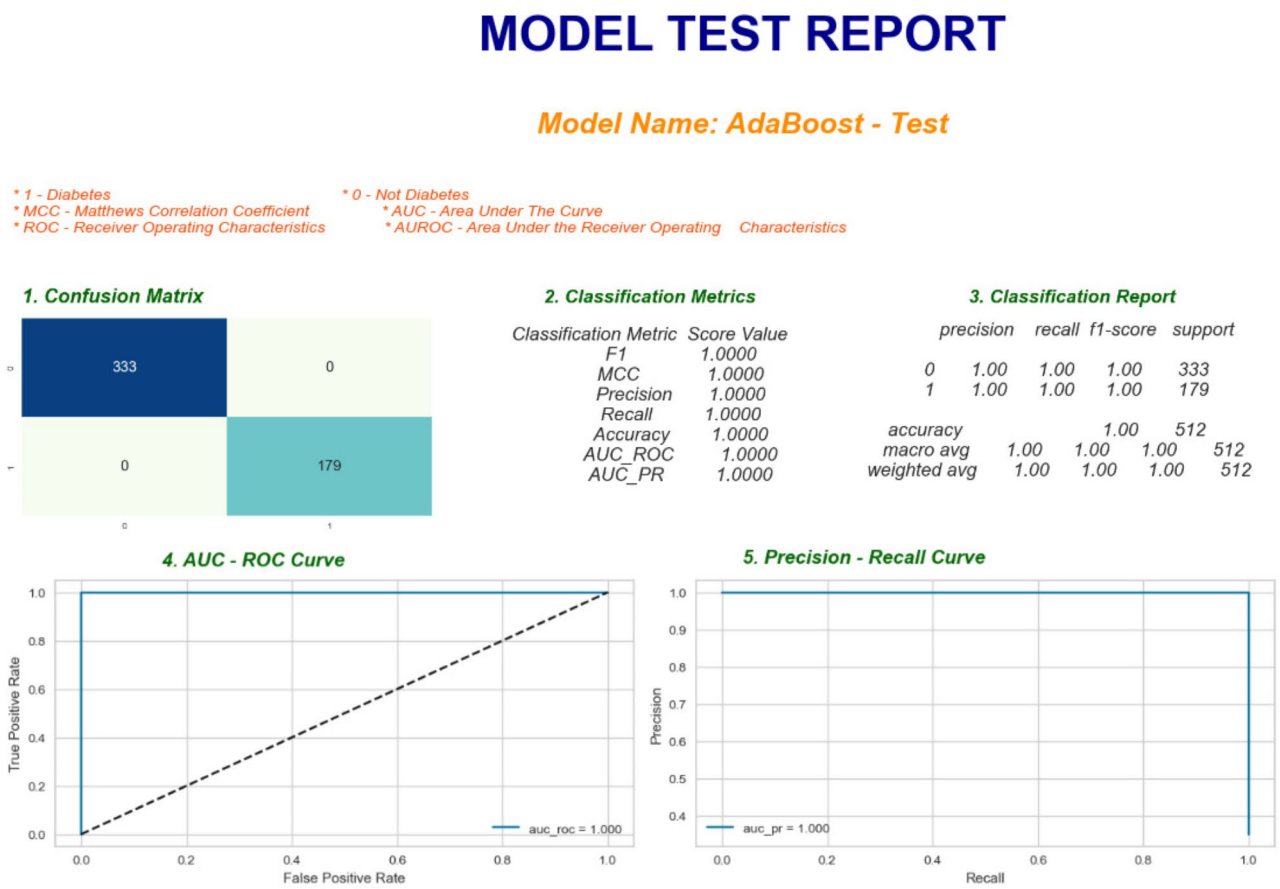

Fig. 11. Output of Ada BOOST.

Future work

The study, like others, has certain limitations that could guide future research efforts. Firstly, the Pima Indian Diabetes dataset may not fully capture the diversity of diabetic populations across different regions, which could limit the generalizability of the findings. Therefore, future studies should aim to evaluate the model using diverse, real-world datasets to assess its performance across various demographic groups and geographical regions. This would help improve the robustness of the model and its applicability to a broader population. Secondly, the study relied on a fixed set of features that may not comprehensively represent all relevant variables influencing diabetes risk. Future work could explore more advanced feature selection methods, such as Artificial Neural Network Hybrid Ensembles or Fuzzy-based models, to identify and incorporate additional factors that might improve prediction accuracy. Additionally, we suggest expanding the scope of feature engineering, possibly integrating time-series data or clinical histories where applicable. Another significant area for future research is the prediction of different types of diabetes. The current study does not distinguish between Type 1 and Type 2 diabetes, nor does it predict gestational diabetes. Exploring methods to classify different types of diabetes, potentially through the integration of more detailed datasets, could provide more specific and actionable insights for clinical practice. Furthermore, the deployment of the model into clinical decision support systems (CDSS) should be explored in future work. This would involve testing the model in real-time clinical settings and validating its effectiveness in assisting healthcare professionals with diagnostic and treatment decisions. Ensuring the model is interpretable and transparent for clinicians will also be critical. We recommend incorporating explainability techniques like SHAP or LIME, which could provide insights into how the model makes decisions, enhancing its trustworthiness. Lastly, ethical considerations, including the detection and reduction of biases, should be a priority in future work. Ensuring that the model provides fair and equitable predictions across different populations is essential to its adoption in healthcare setting.

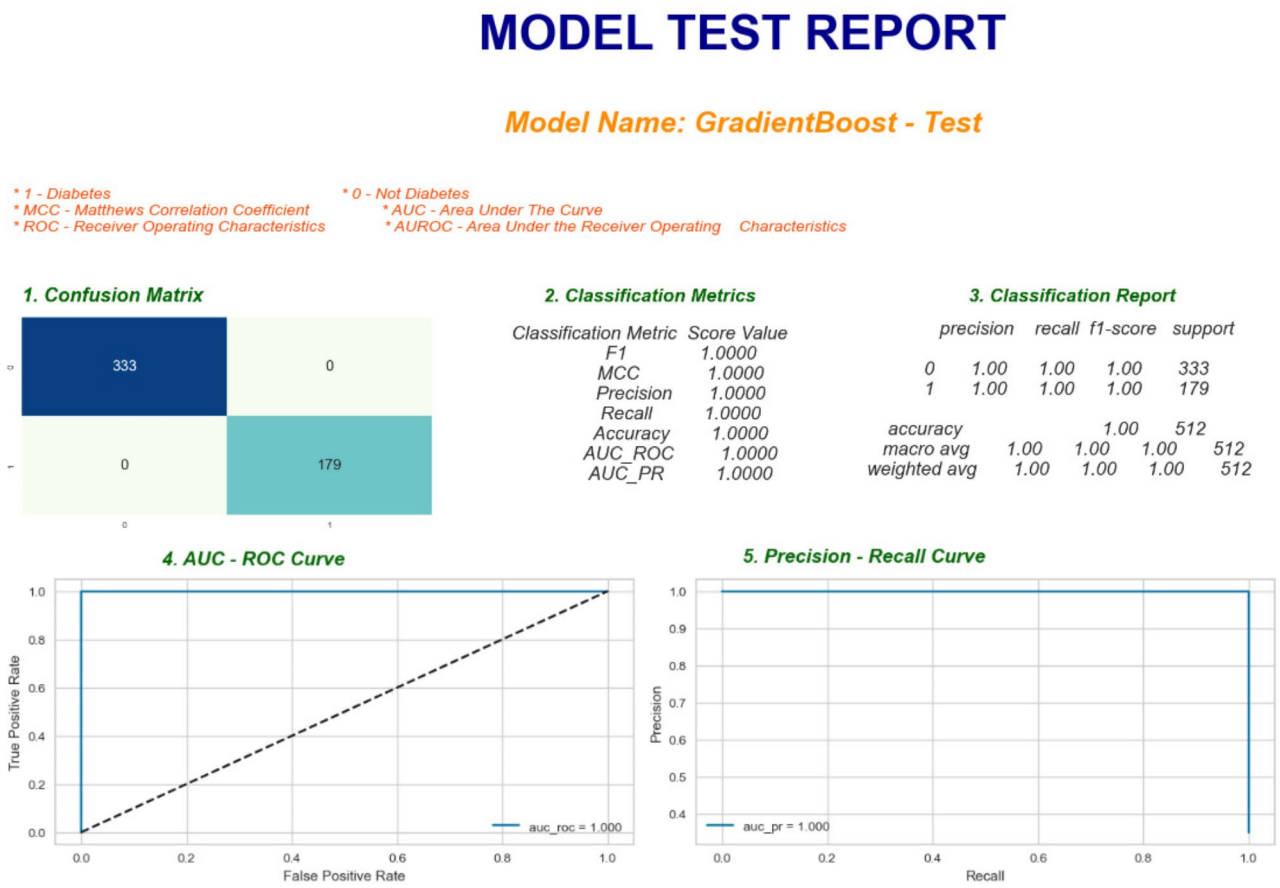

Fig. 12. Output of Gradient BOOST.

Statistical analysis	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MCC (%)	AUC_ROC (%)	AUC_PR (%)
XG boost	100	100	100	100	100	100	100
ADA boostingm1	100	100	100	100	100	100	100
Gradient boosting	100	100	100	100	100	100	100

Table 4. Model classification report for sequential.

Statistical analysis	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MCC (%)	AUC_ROC (%)	AUC_PR (%)
Decision slump	70.00	55.67	60.67	58.06	34.29	67.46	59.42
J48	73.83	62.50	61.80	62.15	42.15	71.02	64.47
CART	69.53	56.47	53.93	55.17	32.13	65.89	61.31
RF	75.79	64.21	68.54	66.30	47.49	74.09	71.38
XG boost	73.05	62.50	56.18	59.17	39.26	69.11	63.38
ADA Boostingm1	75.79	64.21	68.54	66.30	47.49	74.09	71.38
Gradient boosting	76.79	66.21	68.54	66.30	57.49	78.09	75.38

Table 5. (8) eight features classification model report without forward and backward feature techniques.

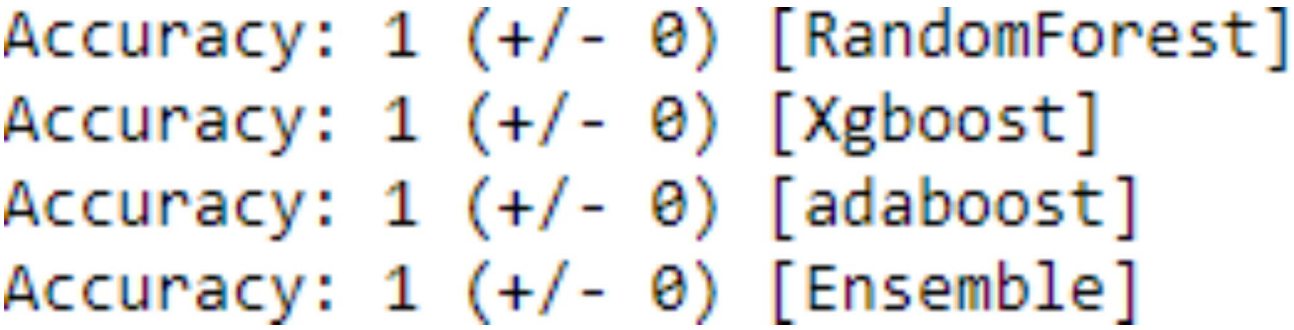

Fig. 13. Output of Ensemble method.

Statistical analysis	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MCC (%)	AUC_ROC (%)	AUC_PR (%)
Random forest	100	100	100	100	100	100	100
CART	78	0.79	0.90	0.84	50.46	84.42	70.80
J48	78	0.79	0.90	0.84	50.46	84.21	70.45
Decision stump	76	0.79	0.85	0.82	45.40	71.95	54.24
AdaBoost	81	0.81	0.92	0.86	56.51	89.24	80.66
Gradient boosting	100	100	100	100	100	100	100
XGBoost	100	100	100	100	100	100	100

Table 6. Performance metrics of classification models.

Model	Precision	Recall	F1-Score	Accuracy
Random forest	0.97	1.00	0.98	0.97
J48 (C4.5)	0.98	0.97	0.97	0.95
CART	0.98	0.97	0.97	0.95
Decision stump	0.95	1.00	0.98	0.95
XGBoost	0.97	1.00	0.98	0.97
AdaBoost	0.97	1.00	0.98	0.97
Gradient boosting	0.97	1.00	0.99	0.97
Proposed model	0.97	1.00	0.97	**0.98**

Table 7. Accuracy results for diabetes_prediction_dataset.Csv (Kaggle).

Model	Precision	Recall	F1-Score	Accuracy
Random forest	0.97	0.97	0.97	0.99
J48 (C4.5)	0.97	0.94	0.96	0.99
CART	0.95	0.97	0.96	0.98
Decision stump	0.50	1.00	0.67	0.88
XGBoost	0.97	0.97	0.97	0.99
AdaBoost	0.86	1.00	0.92	0.94
Gradient boosting	0.97	0.94	0.96	0.98
Proposed model	0.98	1.00	0.99	0.99

Table 8. Accuracy results for "diabetes dataset" (Mendeley data).

Name and author	Dataset	Method	Result
Yadav and Pal[31]	UCI Repository	J48, Decision Stump, REP, RF, Gradient Boosting, AdaBoost M1, XGBoost.	Parallel (RF = 100%) and Sequential Ensemble (XGBoost = 98.05%)
Kumari, Kumar and Mittal[39]	PIMA diabetes dataset (UCI)	The proposed ensemble soft voting classifier uses an ensemble of three machine learning methods to give binary classification, including RF, logistic regression, and Naive Bayes.	On the Pima Indians diabetes dataset, the ensemble soft voting classifier obtained 79.08% accurate results and 97.02% correct results on the breast cancer dataset.
Tewari and Dwivedi[32]	UCI dataset.	Methods for choosing features like the correlation matrix and Chi-square Rule-based classification algorithms: JRIP, OneR, and Decision Table Ensemble techniques: boosting and bagging	According to the result's summary. The Bagging Ensemble Method had the highest accuracy of 98%
Rashid, Yaseen, Saeed, and Alasaady[45]	PIMA Indians Diabetes Dataset (PIDD)	Ensemble voting classifier with Decision Tree, Logistic Regression, K-Nearest Neighbour, Random Forests, and XGBoost	Achieved 81% accuracy after applying standardization, imputation, and Local Outlier Factor (LOF)
Zhou, Xin, and Li[46]	PIMA Indian diabetes dataset	Boruta feature selection, K-Means + + clustering, stacking ensemble learning method	Achieved 98% accuracy
Kawarkhe and Kaur[47]	PIMA Indians	Ensemble Classifiers: CatBoost, LDA, LR, Random Forest, GBC with pre-processing techniques	Achieved 90.62% accuracy
Reza, Amin, Yasmin, Kulsum, and Ruhi[48]	PIMA Indian diabetes dataset and local healthcare data	Stacking ensemble with classical and deep neural network methods using multiple classification models	77.10% accuracy on PIMA dataset with CV protocol; 95.50% accuracy using deep NN stacking on simulation study
Proposed study	Kaggle (PIMA Indian)	J48, Decision Stump, CART, RF, AdaBoost, XGBoost, Gradient Boost (Forward and backward techniques)	**8 features (without Forward and backward features)** Decision slump 70.00 J48 73.83 CART 69.53 RF 75.79 XG BOOST73.05 ADA BOOSTINGM 75.79 GRADIENT BOOSTING 76.79 **with Forward and backward features** Parallel (RF = 98%) and Sequential Ensemble (XGBoost = 100%, AdaBoost = 100%, Gradient Boost = 100%)

Table 9. State-of-the-arts comparison of results with related studies.

Data availability

The dataset used in this study is publicly available on Kaggle at https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database.

Received: 4 June 2024; Accepted: 21 January 2025
Published online: 25 January 2025

References

1. Islam, M. T. et al. Diabetes mellitus prediction using different ensemble machine learning approaches. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1–7, IEEE. (2020).
2. World Health Organization 2020 Global report on diabetes, 2020. [Online]. Available: https://apps.who.int/iris/handle/10665/204871. [Accessed: May 2020].
3. Kumar, R. et al. Type1 & Type2. *World J. Pharm. Pharm. Sci.* **9**(10), 838–850 (2020).
4. Ouichka, O. et al. Deep Learning Models for Predicting Epileptic Seizures Using iEEG Signals. *Electronics* **11**(4). https://doi.org/10.3390/electronics11040605 (2022).
5. Bin Tufail, A. et al. Early-Stage Alzheimer's Disease Categorization Using PET Neuroimaging Modality and Convolutional Neural Networks in the 2D and 3D Domains. *Sensors* **22**, 12. https://doi.org/10.3390/s22124609 (2022).
6. Raza, A. et al. A Hybrid Deep Learning-Based Approach for Brain Tumor Classification. *Electronics* **11**(7). https://doi.org/10.3390/electronics11071146 (2022).
7. Kaushik, H. et al. Diabetic Retinopathy diagnosis from Fundus images using stacked generalization of deep models. *IEEE Access.* **9**, 108276–108292. https://doi.org/10.1109/ACCESS.2021.3101142 (2021).
8. Guefrachi, S. et al. Automated Diabetic Retinopathy Screening using Deep Learning. *Multimed. Tools Appl.* **81**. https://doi.org/10.1007/s11042-024-18149-4 (2024).
9. Behera, D. K., Dash, S., Behera, A. K. & Dash, C. S. K. Extreme gradient boosting and soft voting ensemble classifier for diabetes prediction. In *2021 19th OITS International Conference on Information Technology (OCIT)*, pp. 191–195, IEEE. (2021).
10. Wei, Y. et al. Exploring the causal relationships between type 2 diabetes and neurological disorders using a mendelian randomization strategy. *Medicine* **103**(46), e40412. https://doi.org/10.1097/MD.0000000000040412 (2024).
11. Li, Y. & Chen, W. A comparative performance assessment of ensemble learning for credit scoring. *Mathematics* **8**(10), 1756 (2020).
12. Zhao, X., Zhang, Y. & Yang, Y. & J. Pan. Diabetes-related avoidable hospitalisations and its relationship with primary healthcare resourcing in China: a cross-sectional study from Sichuan Province. *Health Soc. Care Commun.*, **30**(4), e1143–e1156. https://doi.org/10.1111/hsc.13522
13. Awotunde, J. B. et al. Explainable Machine Learning (XML) for Multimedia-Based Healthcare Systems: Opportunities, Challenges, Ethical and Future Prospects. In *Explainable Machine Learning for Multimedia Based Healthcare Applications*, pp. 21–46. (2023).
14. Mujumdar, A. & Vaidehi, V. Diabetes prediction using machine learning algorithms. *Procedia Comput. Sci.* **165**, 292–299 (2019).
15. Zhang, W., Zeng, F., Wu, X., Zhang, X. & Jiang, R. A comparative study of ensemble learning approaches in the classification of breast cancer metastasis. In *2009 International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing*, pp. 242–245, IEEE. (2009).
16. Kalaiselvi, C. & Nasira, G. M. A new approach for diagnosis of diabetes and prediction of cancer using ANFIS. In *2014 World Congress on Computing and Communication Technologies*, 188–190, IEEE. (2014).
17. Thongkam, J., Xu, G. & Zhang, Y. AdaBoost algorithm with random forests for predicting breast cancer survivability. In *2008 IEEE International Joint Conference on Neural Networks*, pp. 3062–3069, IEEE. (2008).

18. Velu, C. M. & Kashwan, K. R. Multi-Level counter propagation network for diabetes classification. In *2013 International Conference on Signal Processing, Image Processing & Pattern Recognition*, pp. 190–194, IEEE. (2013).
19. Temurtas, H., Yumusak, N. & Temurtas, F. A comparative study on diabetes disease diagnosis using neural networks. *Expert Syst. Appl.* **36** (4), 8610–8615 (2009).
20. Ayo, F. E., Ogundokun, R. O., Awotunde, J. B., Adebiyi, M. O. & Adeniyi, A. E. Severe acne skin disease: A fuzzy-based method for diagnosis. In *Computational Science and Its Applications–ICCSA 2020: 20th International Conference*, pp. 320–334 (Springer International Publishing, 2020).
21. Farvaresh, H. & Sepehri, M. M. A data mining framework for detecting subscription fraud in telecommunication. *Eng. Appl. Artif. Intell.* **24**(1), 182–194 (2011).
22. Torrent-Fontbona, F. Adaptive basal insulin recommender system based on Kalman filter for type 1 diabetes. *Expert Syst. Appl.* **101**, 1–7 (2018).
23. Ajagbe, S. A., Adegun, A. A., Mudali, P. & Adigun, M. O. Performance of machine learning models for pandemic detection using COVID-19 dataset. In *2023 IEEE AFRICON,* , pp. 1–6, Nairobi, Kenya, IEEE. https://doi.org/10.1109/AFRICON55910.2023.10293525 (2023).
24. Shen, Y. et al. SGLT2 inhibitor empagliflozin downregulates miRNA-34a-5p and targets GREM2 to inactivate hepatic stellate cells and ameliorate non-alcoholic fatty liver disease-associated fibrosis. *Metabolism* **146**, 155657. https://doi.org/10.1016/j.metabol.2023.155657 (2023).
25. Purnami, S. W., Zain, J. M. & Embong, A. A new expert system for diabetes disease diagnosis using modified spline smooth support vector machine. In *Computational Science and Its Applications–ICCSA* 2010, pp. 83–92 (Springer Berlin Heidelberg, 2010).
26. Rani, A. S. & Jyothi, S. Performance analysis of classification algorithms under different datasets. In *3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 1584–1589, IEEE. (2016).
27. Zheng, T. et al. A machine learning-based framework to identify type 2 diabetes through electronic health records. *Int. J. Med. Informatics*. **97**, 120–127 (2017).
28. Komi, M., Li, J., Zhai, Y. & Zhang, X. Application of data mining methods in diabetes prediction. In *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, pp. 1006–1010, IEEE. (2017).
29. Sai, M. J. et al. An Ensemble of Light Gradient Boosting Machine and adaptive boosting for prediction of Type-2 diabetes. *Int. J. Comput. Intell. Syst.* **16** (1), 14 (2023).
30. Miled, Z. B. et al. Predicting dementia with routine care EMR data. *Artif. Intell. Med.* **102**, 101771 (2020).
31. Yadav, D. C. & Pal, S. Analysis of heart disease using parallel and sequential ensemble methods with feature selection techniques: heart disease prediction. *Int. J. Big Data Analytics Healthc. (IJBDAH)*. **6** (1), 40–56 (2021).
32. Tewari, S. & Dwivedi, U. D. A comparative study of heterogeneous ensemble methods for the identification of geological lithofacies. *J. Petroleum Explor. Prod. Technol.* **10** (5), 1849–1868 (2020).
33. Dey, M. R., Khan, M. H. & Rahman, M. F. Diabetes mellitus prediction using hybrid ensemble models. In *2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, pp. 206–210, IEEE. (2019).
34. Kumar, P. & Saha, S. Diabetes prediction using ensemble classification algorithms. In *IEEE Calcutta Conference (CALCON)*, pp. 1–5, IEEE. (2020).
35. Islam, M. S. & Hossain, K. A. R. T. A hybrid approach for the prediction of heart disease using fuzzy logic and machine learning, In *2018 7th International Conference on System Engineering (ICSE)*, pp. 11–16 (IEEE, 2018).
36. Yadav, Y., Kumar, P. & Bhatt, S. Heart disease prediction using hybrid model of Random Forest and K-Nearest Neighbor. *Int. J. Eng. Res. Technol.* **9** (8), 751–756 (2020).
37. Rathi, S. K., Ghosh, S. & Kumar, S. Hybrid Ensemble Method for Effective Classification of Heart Disease. In *2022 8th International Conference on Advances in Computing, Communication and Control (ICAC3)*, pp. 1–6, IEEE. (2022).
38. Awotunde, J. B. et al. On evaluating the relationship between machine learning model accuracy and complexity in healthcare. *Health Inform. Sci. Syst.* **9** (1), 14 (2021).
39. Yao, J. Y., Jha, M. K. & Sun, J. J. An adaptive hybrid algorithm for diabetes prediction. *J. Biomed. Inform.* **126**, 103958 (2022).
40. Adetunji, A. & Adebayo, O. M. Diabetes prediction using machine learning: A review. In *2020 International Conference on Innovative Research and Development (ICIRD)*, pp. 1–6, IEEE. (2020).
41. Shaik, M. M. D. I. et al. An ensemble machine learning framework for prediction of diabetes. In *2022 International Conference on Emerging Trends in Engineering and Technology (ICETET)*, pp. 1–5, IEEE. (2022).
42. Zhou, Y., Chai, X., Yang, G., Sun, X. & Xing, Z. Changes in body mass index and waist circumference and heart failure in type 2 diabetes mellitus. *Front. Endocrinol.* **14**, 1305839. https://doi.org/10.3389/fendo.2023.1305839 (2023).
43. Liang, D. et al. Burden of type 1 and type 2 diabetes and high fasting plasma glucose in Europe, 1990–2019: a comprehensive analysis from the global burden of disease study 2019. *Front. Endocrinol.* **14**, 1307432. https://doi.org/10.3389/fendo.2023.1307432 (2023).
44. Atif, M., Anwer, F. & Talib, F. An ensemble learning approach for effective prediction of diabetes mellitus using hard voting classifier. *Indian J. Sci. Technol.* **15** (39), 1978–1986. https://doi.org/10.17485/IJST/v15i39.1520 (2022).
45. Rashid, M. M., Yaseen, O. M., Saeed, R. R. & Alasaady, M. T. An improved ensemble machine learning approach for diabetes diagnosis. *Pertanika J. Sci. Technol.* **32** (3), 1335–1350. https://doi.org/10.47836/pjst.32.3.19 (2024).
46. Zhou, H., Xin, Y. & Li, S. A diabetes prediction model based on Boruta feature selection and ensemble learning. *PubMed BMC Bioinf.* **24** (1), 224. https://doi.org/10.1186/s12859-023-05300-5 (Jun. 2023).
47. Kawarkhe, M. & Kaur, P. Prediction of diabetes using diverse ensemble learning classifiers. *Procedia Comput. Sci.* **235**, 403–413. https://doi.org/10.1016/j.procs.2024.04.040 (2024).
48. Reza, M. S., Amin, R., Yasmin, R., Kulsum, W. & Ruhi, S. Improving diabetes disease patients classification using stacking ensemble method with PIMA and local healthcare data. *PubMed Heliyon*. **10** (2), e24536. https://doi.org/10.1016/j.heliyon.2024.e24536 (2024).
49. Rashid, A. Diabetes dataset. *Mendeley Data*. **V1** https://doi.org/10.17632/wj9rwkp9c2.1 (2020).
50. Dong, F., Zhang, L. & Yang, G. Construction of a TF-miRNA-mRNA Regulatory Network for Diabetic Nephropathy. *Archivos Españoles De Urología*, **77**(1), 104–112. https://doi.org/10.56434/j.arch.esp.urol.20247701.14 (2024).
51. Dai, B. et al. The effect of Liuwei Dihuang decoction on PI3K/Akt signaling pathway in liver of type 2 diabetes mellitus(T2DM) rats with insulin resistance. *J. Ethnopharmacol.* **192**, 382–389. https://doi.org/10.1016/j.jep.2016.07.024 (2016).
52. Duan, W. et al. A TGF-β signaling-related lncRNA signature for prediction of glioma prognosis, immune microenvironment,and immunotherapy response. *CNS Neurosci. Ther.* **30**(4), e14489. https://doi.org/10.1111/cns.14489 (2024).
53. Cheng, X. et al. Quercetin: A promising therapy for diabetic encephalopathy through inhibition of hippocampal ferroptosis. *Phytomedicine,* **126**, 154887. https://doi.org/10.1016/j.phymed.2023.154887 (2023).
54. Liang, J. et al. The Regulation of Selenoproteins in Diabetes: A New Way to Treat Diabetes. *Curr. Pharmaceut. Des.* **30**(20), 1541–1547. https://doi.org/10.2174/0113816128302667240422110226 (2024).
55. Zhao, X., Zhang, Y., Yang, Y. & Pan, J. Diabetes-related avoidable hospitalisations and its relationship with primary healthcare resourcing in China: A cross-sectional study from Sichuan Province. *Health Soc. Care Community* **30**(4), e1143-e1156. https://doi.org/10.1111/hsc.13522 (2022).
56. Rustam, R. et al. Enhanced Detect. Diabetes Mellitus Using Novel Ensemble Feature. *Eng. Approach Mach. Learn. Model. Sci. Rep.*, **14**, Article 23274 (2024).

57. Faustin, U. M. & Zou, B. An improved homogeneous ensemble technique for early accurate detection of type 2 diabetes mellitus (T2DM). *Computation* **10**(7), 104 [Online]. Available: https://doi.org/10.3390/computation10070104. [Accessed: 13 Jan. 2025] (2022).
58. Reza, M. S., Amin, R., Yasmin, R., Kulsum, W. & Ruhi, S. Improving diabetes disease patients classification using stacking ensemble method with PIMA and local healthcare data. *Heliyon* **10**(2), e24536 (2024).
59. Saihood, Q. & Sonuç, E. A practical framework for early detection of diabetes using ensemble machine learning models. *Turkish J. Electr. Eng. Comput. Sci.* **31** (4), 722–738 (2023).
60. Daza, A., Ponce Sánchez, C. F., Apaza-Perez, G., Pinto, J. & Zavaleta Ramos, K. Stacking ensemble approach to diagnosing the disease of diabetes. *Inf. Med. Unlocked*, **44**, (2024).

Author contributions

Blessing Oluwatobi Olorunfemi : Conceptualization; Data curation; Formal analysis; Methodology; Writing - original draft; Software. Adewale Opeoluwa Ogunde : Investigation; Methodology; Writing - original draft; Writing - review & editing.Ahmad Almogren : Conceptualization; Investigation; Writing - review & editing; Methodology. Abidemi Emmanuel Adeniyi: Validation; Investigation; Writing - review & editing. Sunday Adeola Ajagbe: Visualization; Validation; Writing - review & editing.Salil Bharany: Project administration; Investigation; Methodology. Ayman Altameem: Funding Acquisition; Writing - review & editing; Software; Resources; Methodology. Ateeq Ur Rehman: Writing - review & editing; Software; Resources; Methodology. Asif Mehmood : Investigation; Methodology; Writing - original draft; Writing - review & editing.Habib Hamam: Writing - review & editing; Software; Resources; Methodology.

Funding

This work was supported by King Saud University, Riyadh, Saudi Arabia, through Researchers Supporting Project number RSP2025R498.

Declarations

Competing interests

The authors declare no competing interests.

Informed consent

Given that the study only used a publicly available dataset from Kaggle and did not involve direct interaction with human subjects, the concept of informed consent is not applicable. The study's dataset is freely available and does not include any personally identifiable information. Therefore, getting informed consent from participants was not required for this study.

Institutional Review Board Statement

As the study utilized a publicly available dataset sourced from the Kaggle, which primarily comprises anonymized data and does not involve human subjects directly, ethical review and approval were deemed unnecessary. Consequently, the study was conducted without the need for institutional review board (IRB) approval, as the data source inherently ensures compliance with ethical guidelines regarding data usage.

Additional information

Correspondence and requests for materials should be addressed to A.U.R. or A.M.