Cureus Journal of Computer Science

Open Access Original Article

# Hybrid Machine Learning Models for Accurate Type 2 Diabetes Mellitus Prediction Using a Stacking Classifier and a Meta-Model Approach

Md. Rashed [1] , Md. Imran Hossain [1] , Akif Mahdi [1] , Gulam Mustofa [1]

1. Department of Information and Communication Engineering, Pabna University of Science and Technology, Pabna, BGD

**Corresponding authors:** Md. Rashed , rashedulislam.ice.pust@gmail.com, Md. Imran Hossain, imran05ice@pust.ac.bd, Akif Mahdi, akif2100@gmail.com, Gulam Mustofa, gulammustofa.200615@s.pust.ac.bd

**Received** 01/27/2025
**Review began** 02/03/2025
**Review ended** 03/02/2025
**Published** 03/03/2025



**DOI:**
https://doi.org/10.7759/s44389-025-03135-0

## Abstract

Type 2 diabetes mellitus (T2DM) is a major global health issue that significantly reduces life expectancy and impairs overall quality of life. Early diagnosis of T2DM is critical, as it can help prevent or delay the development of associated complications. This study aims to assess the effectiveness of a stacked ensemble machine learning model, incorporating a meta-model approach, for the early detection of T2DM. In the first stage of modeling (Level 1), two base models - random forest (RF) and support vector machines (SVM) - are trained using distinct feature selection, tuning, and optimization strategies. In the second stage, ensemble methods, such as voting and stacking classifiers, are employed to combine the predictions from these base models. The final prediction is made by a meta-model, specifically a gradient boosting classifier (GBC), which classifies individuals as either positive or negative for T2DM. Further classification of positive instances is based on low-, moderate-, or high-risk levels. The system is designed for deployment to enable real-time assessment of diabetes risk. The Meta-Model (GBC), with an accuracy of 99.13%, a precision of 100%, specificity of 100%, and an F1 score of 99.25%, consistently outperforms the base models (SVM and RF) and ensemble models (voting and stacking classifiers), proving highly effective in the early detection of T2DM. This significant accuracy highlights the enormous promise of machine learning methods for assisting in the early identification of T2DM. The findings of this study will significantly enhance precision medicine screening for T2DM, empowering healthcare professionals to diagnose the condition early and ultimately improve patient outcomes.

**Categories:** Biotechnology and Computational Biology, Machine Learning (ML)
**Keywords:** type 2 diabetes mellitus (t2dm), ensemble machine learning, stacking classifier, meta model, gradient boosting classifier (gbc)

## Introduction

Diabetes mellitus (DM) is a chronic metabolic condition characterized by elevated blood glucose levels caused by inefficient insulin synthesis or use by the body. There are three types of diabetic conditions that may be identified (Figure *1*). In type 1 diabetes, also known as juvenile diabetes or insulin-dependent diabetes, the body’s defenses damage insulin-producing cells, resulting in the loss of insulin synthesis [1]. Type 2 diabetes (T2D) accounts for 90% of DM cases. Recently, there has been an annual increase in the prevalence of T2D [2]. Next, during pregnancy, usually in the second or third trimester, a type of diabetes known as gestational DM can develop. Insulin resistance and high blood glucose are its defining characteristics, which result from the body's inability to synthesis enough insulin to fulfill the needs of pregnancy. In this instance, 90-95% of all instances of diabetes are caused by T2DM. A number of related illnesses, including stroke, cancer, Alzheimer's disease, hypertension, and cardiovascular disease, can be made more likely by improper diabetes care, which can also impair glucose regulation. Consequently, improved results for those with diabetes are ensured by early identification of DM, or type 2 diabetes mellitus (T2DM), which is essential for accurately forecasting and controlling the disease. T2DM is mostly caused by insufficient creation of insulin to fulfill the body's needs. People with T2D are often diagnosed around the age of 40. The condition is frequently linked to aging, obesity, and genetic inheritance from parents [3]. Apart from obesity, T2D is highly correlated with age, gender, socioeconomic position, site of residence (rural or urban), drinking and smoking behaviors, and dietary patterns [4]. A few of these variables are changeable and have a significant influence on T2D treatment [3].

**How to cite this article**
Rashed M, Hossain M, Mahdi A, et al. (March 03, 2025) Hybrid Machine Learning Models for Accurate Type 2 Diabetes Mellitus Prediction Using a Stacking Classifier and a Meta-Model Approach. Cureus J Comput Sci 2 : es44389-025-03135-0. DOI https://doi.org/10.7759/s44389-025-03135-0


**FIGURE 1: Types of diabetes**

Worldwide, 537 million people are expected to suffer from diabetes [5]. Based on statistical data, 529 million individuals globally suffered from diabetes in 2021. It is projected that the population will reach 1.31 billion by 2050. Even if the proportion of diabetic patients varies somewhat between nations, diabetes is still one of the leading causes of mortality and disability worldwide. This is true irrespective of age groupings, gender factors, or national boundaries [6]. An estimated 760 billion dollars are spent each year on diabetes research [7]. Healthcare systems bear a massive and ever-growing financial burden from the expense of treating diabetes-related problems [8]. There is a considerable global increase in the number of diabetes individuals, which presents a major health concern. The top 15 nations with the highest DM rates are listed in Table *1* [9]. Many diabetes patients misjudge how terrible their health status is in the early stages of the disease [10].

| Country | Diabetes Rate (Age 20-79) |
|---|---|
| American Samoa | 20.3% |
| Egypt | 20.9% |
| French Polynesia | 25.2% |
| Guam | 19.1% |
| Kiribati | 22.1% |
| Kuwait | 24.9% |
| Malaysia | 19.0% |
| Marshall Islands | 23.0% |
| Mauritius | 22.6% |
| Nauru | 23.4% |
| New Caledonia | 23.4% |
| Pakistan | 30.8% |
| Qatar | 19.5% |
| Solomon Islands | 19.8% |
| Tuvalu | 20.3% |

**TABLE 1: The top 15 nations with the highest incidence of diabetes**

It is essential to develop methods for early identification and prediction of DM, as a high yearly death rate and several health complications arise from delayed diagnosis. The increasing prevalence of T2DM patients is regarded as a severe health problem [11]. It is vitally essential to predict T2DM in individuals of all ages. As a result, timely implementation of the necessary lifestyle modifications can aid in delaying the advancement of DM and the health issues that accompany it [12]. In the past year, data mining and machine learning have developed into trustworthy and valuable instruments in the medical field [13]. To increase the precision of illness prediction, machine learning algorithms were developed to automate the healthcare systems' operational model. The introduction of a distributed computing architecture based on Hadoop clusters facilitated the efficient processing and storage of enormous amounts of data in cloud environments. A novel machine learning method was implemented to predict diabetes on Hadoop-based clusters [14]. Numerous machine learning-based systems were utilized to classify and predict diabetic disease, including logistic regression (LR), feed-forward neural network, decision tree (DT), J48, random forest (RF), naive Bayes (NB), support vector machine (SVM), artificial neural network (ANN), k-nearest neighbor (KNN), and others [15]. To enhance efficiency and effectiveness in T2DM identification, we employ stacked ensemble machine learning in conjunction with Meta-Model. To improve their performance, the two fundamental models in the first level - RF and SVM - are trained using various feature selection strategies, hyperparameter adjustments, and optimization approaches. The second level involves combining the predictions from these basic models using ensemble approaches like voting and stacking classifiers. The final prediction is then produced by a Meta-Model, namely a gradient boosting classifier (GBC), which categorizes people as either positive or negative for T2DM. Based on their estimated risk, those who test positive are further divided into low-, moderate-, and high-risk categories.

The key contributions of this work are summarized as follows:

1. The model blends ensemble approaches (voting and stacking classifiers) with the best characteristics of many machine learning techniques, including SVM and RF. Through the use of many model's complementing properties, this hybrid method enhances accuracy and adaptability.

2. Using a GBC as the Meta-Model at the last prediction step significantly enhances predictive performance.

3. The model provides an additional risk level assessment (low, moderate, and high). This makes it possible for doctors to provide more targeted treatments by helping those categories patients according to how likely they are to have problems.

4. We implement the model for clinical application, facilitating quick and precise real-time diabetes risk assessment.

5. By incorporating predictions from different algorithms, healthcare providers can have more confidence in the results and explore which model's output contributed to the final decision.

6. Its modular design makes it simple to update and adjust as new algorithms or additional data become available.

## Literature survey

The early prediction of T2DM has become a critical area of research due to the increasing prevalence of the disease worldwide. Many studies have applied various machine learning algorithms to predict the onset of T2DM, with a particular focus on improving prediction accuracy and robustness. Recent advancements in hybrid models, which combine multiple classifiers in a stacking or ensemble approach, have shown promising results in diabetes prediction. These models leverage the strengths of different machine learning algorithms to achieve more reliable and precise predictions. The following Table *2* summarizes key works in the field of T2DM prediction, outlining the solution methods employed, the performance metrics achieved, and the limitations noted in each study.

| Ref. | Works | Method | Performance | Limitations |
|---|---|---|---|---|
| [16] | Application of Supervised ML for T2DM Prediction | Random Forest | Accuracy 93.8% | Requires more computational resources and complex interpretation. |
| [17] | Diabetes Detection and Classification | Random Forest with Kernel Entropy Component Analysis | Accuracy 96.75% | High computational cost due to kernel entropy; component analysis and random forest complexity. |
| [18] | Diabetes Prediction Using a Healthcare Framework | AdaBoost | Accuracy 80% | Moderate accuracy and generalization; may require more data and advanced feature engineering for improved performance. |
| [19] | Risk Prediction for Diabetic Retinopathy in Chinese T2DM Population | Multivariable Logistic Regression | Accuracy 79.6% | Low recall and F1 score indicate potential issues with imbalanced data or insufficient sensitivity. |
| [20] | Diabetes Detection in Bangladesh | Ensemble Technique | Accuracy 99.27% | Potential overfitting due to ensemble complexity; class imbalance issues despite SMOTE and ROS |
| [21] | T2DM Classification With Metal Exposure | Soft Voting Ensemble (XGBoost, Random Forest, LightGBM) | AUC improved from 0.792 to 0.847 with 12 baseline biomarkers + 3 metal variables | Potential bias due to dietary self-reporting; multicollinearity issues despite VIF adjustment |
| [22] | Early Diagnosis of T2D Using EHR and SB-SVM | SB-SVM | Accuracy 81.43% | Class imbalance issue handled, but generalization to other datasets may be limited |
| [23] | Diabetes Detection Using Artificial Neural Network (ANN) | ANN | Accuracy 80.79% | Requires a significant amount of data for training; may be sensitive to overfitting with small datasets |
| [24] | E-Diagnosis System for Diabetes Using ML Models | Naïve Bayes classifier, Random Forest classifier, J48 Decision Tree models | Accuracy 79.57% | Decision-making process can be difficult to interpret without model fine-tuning; may require additional features for better performance |
| [25] | Diabetes Prediction Using Deep Dense Layer Neural Network (DDLNN) | DDLNN | Accuracy 84.42% | Requires hyperparameter tuning and K-fold cross-validation; could benefit from more complex datasets or features |
| [26] | Enhancing Diabetes Prediction Using ML | Ensemble Soft Voting | Accuracy 97% | Class imbalance issue (Mendeley dataset); complexity of RDPVR method; further validation on diverse datasets needed |

| | | | | |
|---|---|---|---|---|
| [27] | Diabetes Prediction | Random Forest | Accuracy 83% | Limited to the selected features; only two classifiers evaluated |
| [28] | Diabetes Prediction Using Deep Learning | Convolutional Neural Network | Accuracy 92.31% | Requires a balanced dataset, computationally intensive, may not generalize well to imbalanced datasets |
| [29] | Diabetes Prediction Using Data Mining and ML Algorithms | Neural Network | Accuracy 88.6% | Neural network with multiple layers requires more computational resources, model performance may vary with different configurations |
| [30] | Deep Neural Network (DNN) Framework for Diabetes Classification | DNN with Stacked Autoencoders and Softmax Layer | Accuracy 86.26% | May require significant computational resources for training, and the risk of wrong predictions in medical diagnosis |
| [31] | Deep Learning for Predicting Diabetes Mellitus Complications | Deep Belief Network | Accuracy 81.25% | Limited to the dataset used; may not generalize well to other datasets or complications |
| [32] | Comparison of ML Algorithms for Type 2 Diabetes Prediction | DNN | Accuracy 99.5% | Results depend on dataset quality; noisy data may affect performance before preprocessing |
| [33] | Predictive Models for Chronic Kidney Disease in Type 2 Diabetes | Cox Proportional Hazards Model | Accuracy 87.4% | Limited generalizability to other populations; model performance could vary with different datasets |

**TABLE 2: Summary of existing studies**

T2DM: Type 2 Diabetes Mellitus, ML: Machine Learning, SB-SVM: Sparse Balanced Support Vector Machine, EHR: Electronic Health Record, SMOTE: Synthetic Minority Over-sampling Technique, ROS: Random Over-Sampling, RDPVR: Random Data Partitioning with Voting Rule, VIF: Variance Inflation Factor

## Materials And Methods

This study's methodology integrates ensemble-based Meta-Model techniques in an attempt to create a dependable machine learning model for diabetes diagnosis. The suggested strategy, which stacks classifiers with Meta-Model to improve diabetes analysis's accuracy and dependability, incorporates many machine learning techniques. The methodology for developing and assessing the Meta-Model is covered in depth in this section. The following explains the essential procedures that comprise this strategy:

### Dataset description

The Sylhet Diabetes Hospital in Sylhet, Bangladesh (SDHS-B) dataset, the publicly available dataset utilized in this work, is essential for forecasting diabetes outcomes and risk. It has 17 features, including both numerical and qualitative data, and is derived from 520 records at the Sylhet Diabetes Hospital, Sylhet, Bangladesh [34]. These characteristics cover a wide range of patient-related data, including joint thrush, polyuria, polydipsia, abrupt weight loss, weakness, polyphagia, blurred vision, itching, irritability, delayed healing, partial paresis, stiffness in the muscles, alopecia, and obesity. The distribution of gender and status among the sample population is depicted in the pie chart (Figure *2*). According to the gender distribution pie chart, 36.92% of the sample is female, and 63.08% of the sample is male. According to the status distribution pie chart, 61.54% of the participants have a positive status, whereas 38.46% have a negative status.

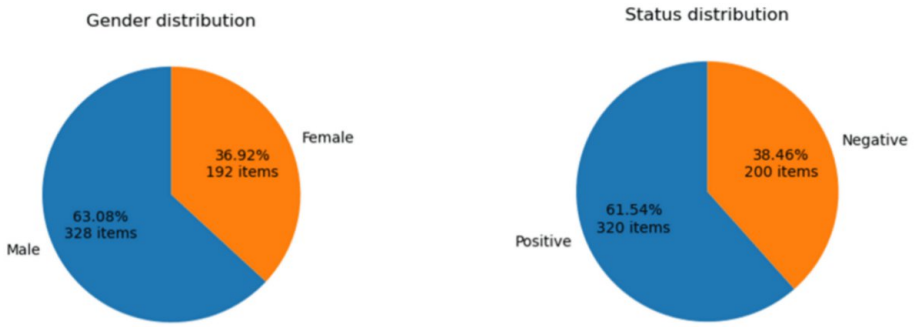


**FIGURE 2: Status and gender distribution pie chart**

The PIMA-I dataset [35] contains 768 patient records and eight features. Similarly, the Diabetes Dataset Frankfurt Hospital-Germany (DDFH-G) dataset, originating from Frankfurt Hospital, Germany [36], includes data from 2,000 cases, each with eight features. The features span various health metrics, such as the number of pregnancies and skin thickness. Notably, some features like glucose, insulin, blood pressure, BMI, and skin thickness include values of zero, which are unrealistic in practice. These zero values are treated as missing data and replaced by the mean of the corresponding feature column.

The target variable in the dataset specifies if a patient has been diagnosed with diabetes or not. Its extensive patient data and wide range of clinical characteristics offer a strong foundation for binary classification, which attempts to estimate the risk of diabetes precisely. This dataset was carefully chosen since it is extensive and relevant to diabetes, which will help the model generalize effectively and provide insightful results. A correlation heatmap, shown in Figure *3*, illustrates the connections between several variables. The correlation coefficient between two variables is represented by each square in the heatmap, with the colors representing the direction and intensity of the link. White squares denote no correlation, blue squares show negative correlations (variables move in opposing directions), and red squares indicate positive correlations (variables rise or decrease together).

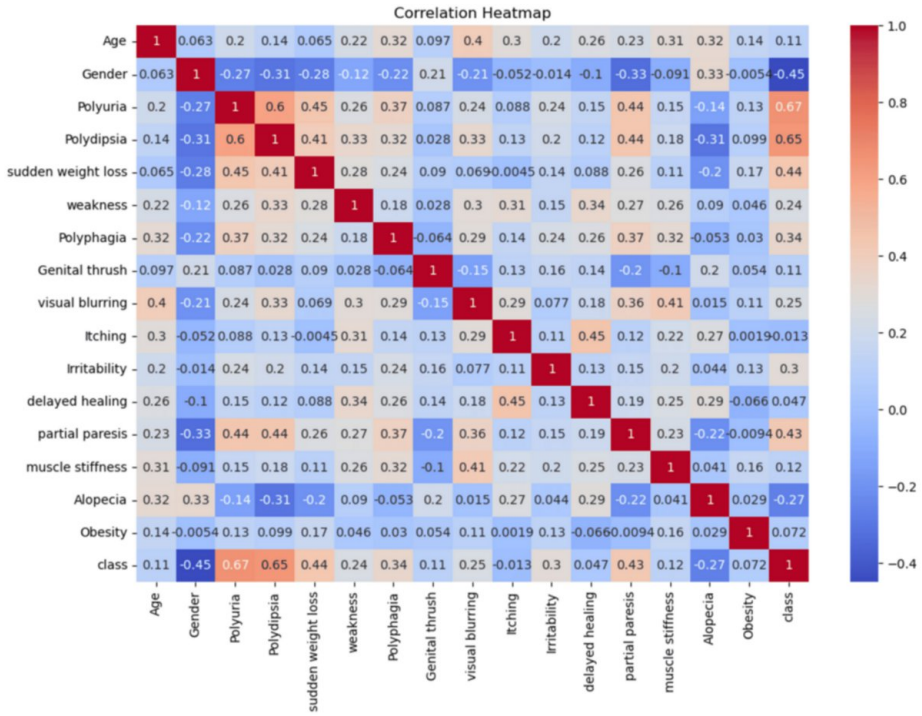


**FIGURE 3: Correlation heatmap of diabetes symptoms**

## Data preprocessing

An essential component of any machine learning pipeline is data preparation. To ensure that the learning algorithms can extract significant patterns and correlations from the data, raw data must be transformed

into a format appropriate for model training. Several important preprocessing steps were carried out in this diabetes prediction system, including the train-test split, normalization, and label encoding. Figure *4* presents the proposed workflow. A detailed explanation of these steps is as follows:

*Label Encoding*

Numerous categorical variables are included in the dataset, including gender (e.g., male/female), family history of diabetes (e.g., yes/no), and other medical conditions (e.g., polyuria, polydipsia, sudden weight loss, weakness, polyphagia, genital thrush, visual blurring, itching, irritability, delayed healing, partial paresis, muscle stiffness, alopecia, and obesity). Typically, machine learning models operate on numerical data. However, because most machine learning algorithms rely on mathematical operations like addition, subtraction, or comparison of numerical values, categorical variables convey qualitative input that cannot be employed directly. This was addressed by using label encoding. Using a process called label encoding, each category in a set of categorical variables is given a distinct integer value. The two categories (male and female), for example, might be encoded as follows: Male = 0 and Female = 1, in the case of gender. Likewise, for the presence of diabetes in the family: Yes = 1, No = 0. Because it enables algorithms like SVM, RF, and neural networks to handle categorical data and consider it as a part of their feature space, this transformation is essential.

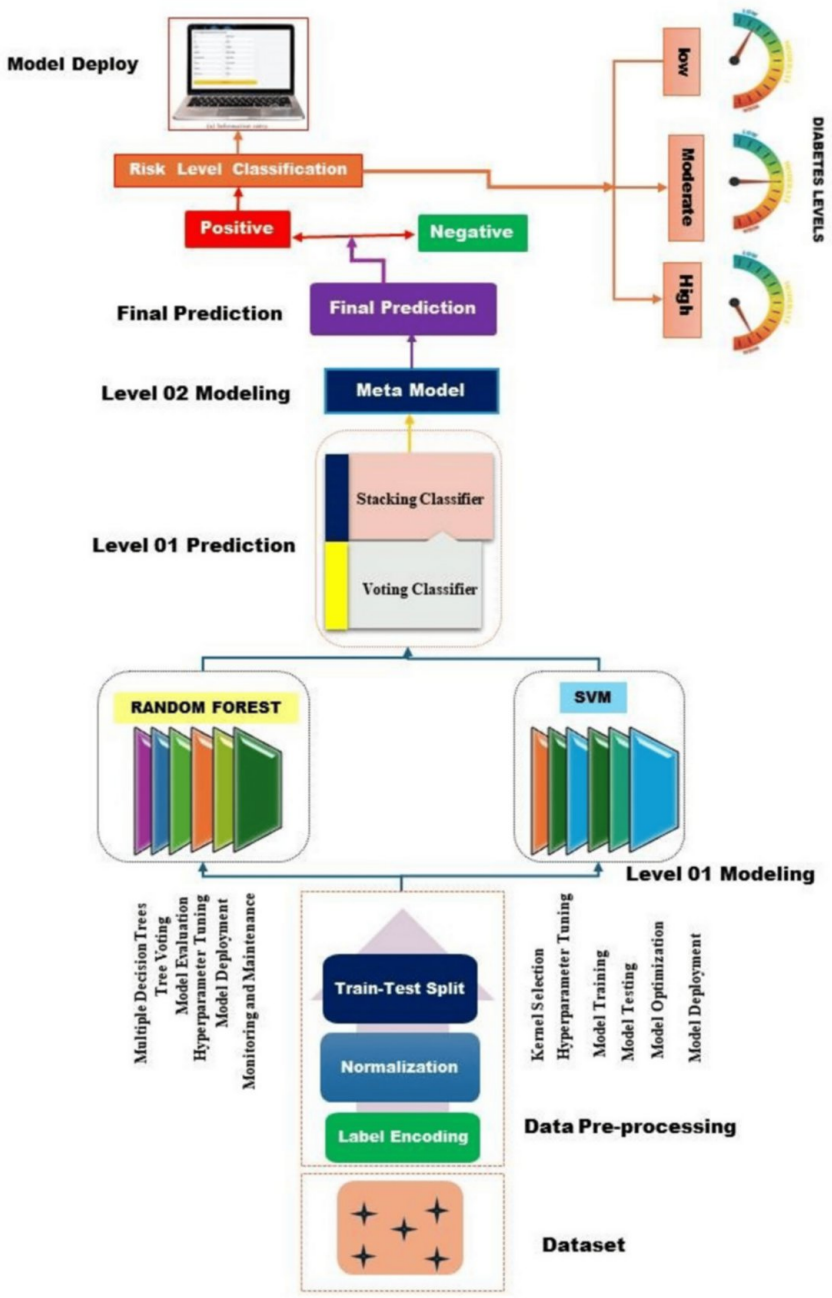


**FIGURE 4: Proposed Meta-Model-based framework**

SVM: Support Vector Machine

*Normalization*

Numerical information like the patient's age is included in the dataset. The min-max scaling approach was used to normalize the numerical variables, especially the age attribute. Using the following formula, normalization converts the data into a new range, usually between 0 and 1, where min (X) and max (X) are the feature's lowest and highest values in the dataset, and X represents the feature's initial value. For instance, the age feature's raw values would be converted into values between 0 and 1 by normalization if its minimum and maximum values were 16 and 90, respectively. Because the age variable has a wider numerical range than other data, such as gender, which is stored as either 0 or 1, this guarantees that it does not unduly affect the model. Because distance-based computations and weight optimizations can be sensitive to the size of the features, normalization is especially crucial when utilizing algorithms like SVM, RF, and neural networks.

$$Xnorm = \frac{X - min(X)}{max(x) - min(x))}$$

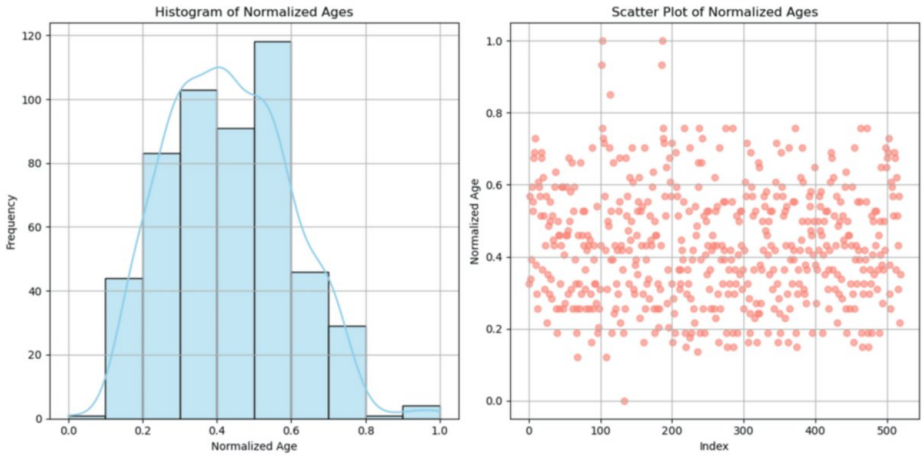


**FIGURE 5: Histogram and scatter plot of normalized ages**

Two plots, a scatter plot and a histogram, show the distribution of normalized ages in a dataset in Figure 5. An enhanced visualization is provided by the overlay density curve on the left-hand histogram, which displays the frequency of ages inside certain bins. The individual data points are shown in the scatter plot on the right, which makes it possible to examine the distribution in greater detail. Normalized age is represented by the x-axis in both plots, frequency is represented by the y-axis of the histogram, and normalized age is represented by the y-axis of the scatter plot. Plots such as this offer information about the dataset's age distribution, including its central tendency, dispersion, and any outliers.

*Train-Test Split*

The dataset was divided into two subsets: a training set that contained 75% of the data and a testing set that contained the remaining 25% in order to assess the model's performance and guarantee that it could generalize to new data. The machine learning models were trained on the training set to find patterns and connections between the input data and the target variable. In order to determine how effectively the models would function in actual situations, their performance on previously unknown data was assessed after training on the testing set. In order to avoid overfitting, make sure the model learns generalizable patterns rather than just memorizing the training data, and simulate the model's performance on fresh, unknown patient cases, the train-test split must be implemented.

## Base models

The idea of a base model is fundamental to machine learning and forms the basis for prediction tasks. Usually, a base model consists of a single machine learning algorithm that runs on its own and makes predictions using the input features that are supplied. We use two popular foundation models in this work: RF and SVM. The SVM approach with a radial basis function (RBF) kernel was used to address classification difficulties. SVM operates exceptionally well in high-dimensional fields by determining the optimal hyperplane that maximizes the margin between distinct classes - in this case, differentiating between those with and without diabetes. This was a very useful feature since the RBF kernel can handle non-linear decision boundaries, which is necessary to capture the complex patterns in the data. The model's performance was enhanced by optimizing hyperparameters such as the gamma coefficient, which determines the impact of a single training example, and the penalty parameter CCC, which balances the goals of maximizing margin and minimizing classification error, using grid search with cross-validation. By ensuring that the model struck the optimal balance between generalization and accuracy, our method successfully decreased the risk of overfitting while preserving strong predictive performance. Finding the optimal hyperplane that maximizes the margin between classes is one of the ways the SVM approach prevents overfitting in high-dimensional feature fields.

In contrast, the RF approach is an ensemble learning technique that creates a large number of DTs during training and determines the mode of each tree's classification forecast. By using the advantages of several trees, this model reduces overfitting and increases prediction accuracy. RF is a great option for complicated datasets since it excels at tolerating missing values and preserving accuracy for a large number of features. By using these two models as our foundational models, we are able to identify a variety of patterns in the data and offer a thorough categorization strategy that makes use of the

advantages of each technique.

The Level-1 modeling procedure and combination of base model outputs, where the preprocessed data is fed to both SVM and RF models, are shown in Figure 6. Every model produces predictions on its own, and performance is assessed using metrics like accuracy, precision, sensitivity, specificity, and confusion matrices. Use a stacking classifier to combine the prediction-1 (P1) and prediction-2 (P2).

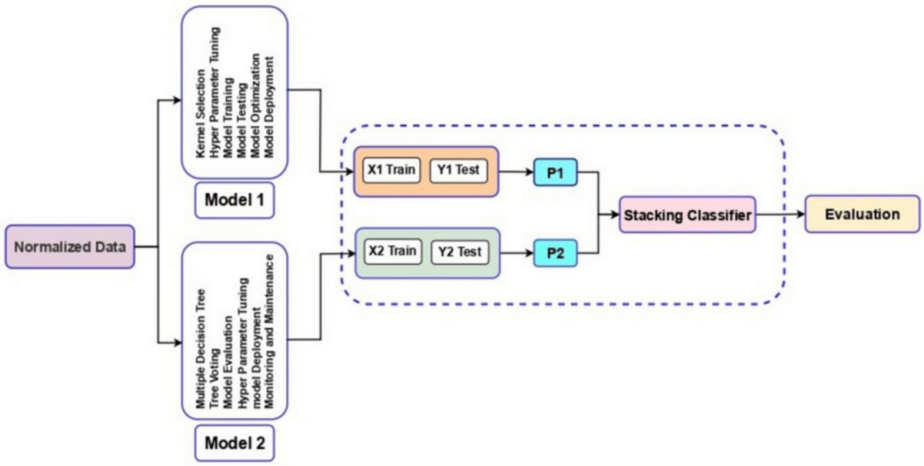


**FIGURE 6: Base model training and evaluation workflow with stacking classifier**

## Voting classifier to combine output of base models

The voting classifier is an effective ensemble technique that improves overall classification performance by combining the predictions of several base models. We use a soft voting technique in this methodology, which computes a final probability for each class based on the anticipated probabilities from each base model. This method is especially useful as it lets the classifier take advantage of the complementing features of each base model independently, which might increase resilience and accuracy. By combining the results of RF and SVM, the voting classifier lessens the possibility that any one model's biases may affect the outcome. The weighted average of the estimated probabilities from the basic models is how the voting classifier works. More dependable models are able to exert a larger effect on the result since each base model makes contributions based on its own level of prediction confidence. The combined result can offer a more balanced picture of the likelihood of class membership, for example, if the RF predicts 0.6 and the SVM predicts 0.8 for the positive class. When predictions are combined, performance measures like recall, F1 score, and accuracy may all be enhanced. This is especially true for unbalanced datasets where one class may predominate. All things considered, the voting classifier serves as an intelligent middleman, combining data from several sources to arrive at a better judgment.

## Stacking classifier to combine output of base model

Another ensemble method that improves prediction accuracy is the stacking classifier, which combines many base models and makes use of each one's unique advantages. The stacking classifier has a two-level design, in contrast to the voting classifier, which accumulates the predictions of each of its components directly. First, several base models are trained separately. Then, based on these base-level outputs, a Meta-Model is fed with the predictions of the base models to learn how to produce final predictions. Because of its hierarchical structure, the Meta-Model is able to represent intricate relationships and patterns that may not be apparent from the base models alone. A more sophisticated model, such as a gradient boosting classifier, may be the final estimator to which we feed the outputs of the SVM and RF. Through training, the Meta-Model gains insight into the advantages and disadvantages of the base models, which helps it fine-tune its decision-making process. By enabling the model to adaptively balance the contributions of the base models according to their performance in the validation dataset, this technique improves the predictive power of the entire model. Consequently, the stacking classifier can produce increased resilience against overfitting, higher accuracy, and better generalization to previously unknown data. The stacking classifier diagram used to aggregate the output of the basic model is shown in Figure 6. That would be ready for the Meta-Model's input.

## Final prediction using meta-model

The final Meta-Model in our hybrid framework plays a crucial role in refining predictions by aggregating insights from the base models - SVM and RF. Unlike traditional ensemble methods that average or vote on

base model outputs, our Meta-Model undergoes a structured learning process using the probability outputs from the base models as new features. This allows it to discern patterns, dependencies, and misclassifications that individual models may overlook. In our implementation, the GBC serves as the core of the hybrid model due to its sequential learning approach, where weak models are iteratively improved by minimizing residual errors. GBC builds an ensemble of decision trees, where each tree corrects the mistakes of the previous ones by focusing more on misclassified instances. This adaptive boosting mechanism enables the Meta-Model to capture complex, non-linear relationships in the transformed feature space while reducing overfitting. By learning from the errors of base models, GBC enhances generalization, leading to improved accuracy, precision, recall, and specificity. Additionally, GBC employs a loss function optimization process that balances bias and variance, making it particularly effective in managing diverse feature distributions and handling class imbalances. The ability of GBC to iteratively refine decision boundaries ensures that even subtle distinctions in data patterns are leveraged for more precise classification. This structured and adaptive learning process enables the final Meta-Model to provide a robust, reliable, and highly accurate classification output, significantly enhancing overall model performance in complex predictive tasks. Figure 7 shows the research in action.


**FIGURE 7: Meta-Model training and final prediction workflow**

## Performance metrics

The performance of the Meta-Model was evaluated using a broad variety of metrics. The many viewpoints that each of these indicators offers guarantee a thorough assessment of the model's prediction ability along several dimensions.

The following key terms are used to compute these metrics: TP: True Positives - Cases correctly identified as positive. TN: True Negatives - Cases correctly identified as negative. FP: False Positives - Cases incorrectly identified as positive. FN: False Negatives - Cases incorrectly identified as negative.

Accuracy: The percentage of correctly categorized cases relative to all instances is known as accuracy. Accuracy is helpful as a general measure, but in imbalanced datasets - where one class predominates over the other - it can be deceptive.

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

Precision: Precision is the percentage of genuine positives among all cases that were anticipated to be positive. When FP result in significant costs, it is especially crucial.

$$Precision = \frac{TP}{FP + TP}$$

Sensitivity (Recall): Recall, also known as sensitivity, quantifies the percentage of TP that were accurately detected. In medical diagnostics, where FN (missing positive instances) might be crucial, it is indispensable.

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

Specificity: Specificity is a crucial metric for minimizing FP since it quantifies the percentage of TN that were accurately recognized.

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

F1-Score: In situations when recall and accuracy are equally important, the F1 score - which is the harmonic mean of these two metrics - offers a fair assessment.

$$F - 1Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)}$$

Matthews correlation coefficient (MCC): Four categories (TP, TN, FP, and FN) in the confusion matrix are considered by the reliable MCC measure. For imbalanced datasets, it is quite beneficial.

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

ROC Curve: The receiver operating characteristic (ROC) curve provides a visual representation of the model's ability to differentiate across classes. It displays the genuine positive rate (sensitivity) against the FP rate (1 - specificity) at different threshold settings.

AUC Value: The region AUC, or Area Under the ROC Curve, offers a single scalar value that summarizes the performance of the model at each threshold level. Excellent model performance is indicated by an AUC value closer to 1, whereas random guessing is indicated by an AUC value near 0.5.

## Results

In Table *3*, the evaluation of the five classification models reveals significant differences in their performance across key metrics. The base model achieved an accuracy of 94.61%, while the voting and stacking classifiers both reached 98.46%, indicating a strong performance, yet the Meta-Model (GBC) excelled with an accuracy of 99.13%. In terms of precision, the base model scored 93.50%, with the voting and stacking classifiers improving to 98.64%, but the Meta-Model achieved a perfect precision of 100%. For sensitivity, the base model was at 97.29%, with voting and stacking classifiers at 98.64%, while the Meta-Model scored 98.52%. Specificity showed a similar trend, with the base model at 91.07%, voting and stacking classifiers at 98.21%, and the Meta-Model achieving a flawless 100%. The F1 score reflected this superiority, with the base model at 95.36%, voting and stacking classifiers at 98.64%, and the Meta-Model at 99.25%. Lastly, MCC displayed the Meta-Model's strength with a score of 98.22%, significantly outperforming the base model at 89.04%. Overall, the Meta-Model stands out with the highest values across most metrics, affirming its effectiveness as the proposed solution for classification tasks.

| Evaluation Indexes | Base Model | | Voting Classifier | Stacking Classifier | Meta-Model (GBC) |
|---|---|---|---|---|---|
| | SVM | RF | | | |
| Accuracy | 94.61 | 98.46 | 98.46 | 97.43 | 99.13 |
| Precision | 93.50 | 98.64 | 98.64 | 97.87 | 100 |
| Sensitivity | 97.29 | 98.64 | 98.64 | 97.87 | 98.52 |
| Specificity | 91.07 | 98.21 | 98.21 | 96.77 | 100 |
| F1 Score | 95.36 | 98.64 | 98.64 | 97.87 | 99.25 |
| MCC | 89.04 | 96.86 | 96.86 | 94.64 | 98.22 |

**TABLE 3: Performance evaluation of the proposed base model, voting classifier, stacking classifier, and Meta-Model using the SDHS-B dataset**

MCC: Matthews Correlation Coefficient, SVM: Support Vector Machine, RF: Random Forest, GBC: Gradient Boosting Classifier, SDHS-B: Sylhet Diabetes Hospital in Sylhet, Bangladesh

Confusion matrices were created to show the classification results and assess each model's performance further. In comparison to the RF model, the SVM model's confusion matrix in Figure *8* displays 51 TP, 5 FP, 2 FN, and 72 TN. This suggests that the SVM model's accuracy in recognizing non-diabetic patients is somewhat lower. With just one FP (+) and one FN (-), the RF and voting classifier models showed comparable results, demonstrating their potent predictive power. With two FP and two FN, the stacking classifier exhibited somewhat higher misclassification rates but was still quite accurate. With just one FN and no FP, the Meta-Model (GBC) performed the best, demonstrating its greater capacity to distinguish between patients with and without diabetes. These illustrations further demonstrate how ensemble methods - specifically, the Meta-Model - achieve higher accuracy, as shown in Figure *8*.

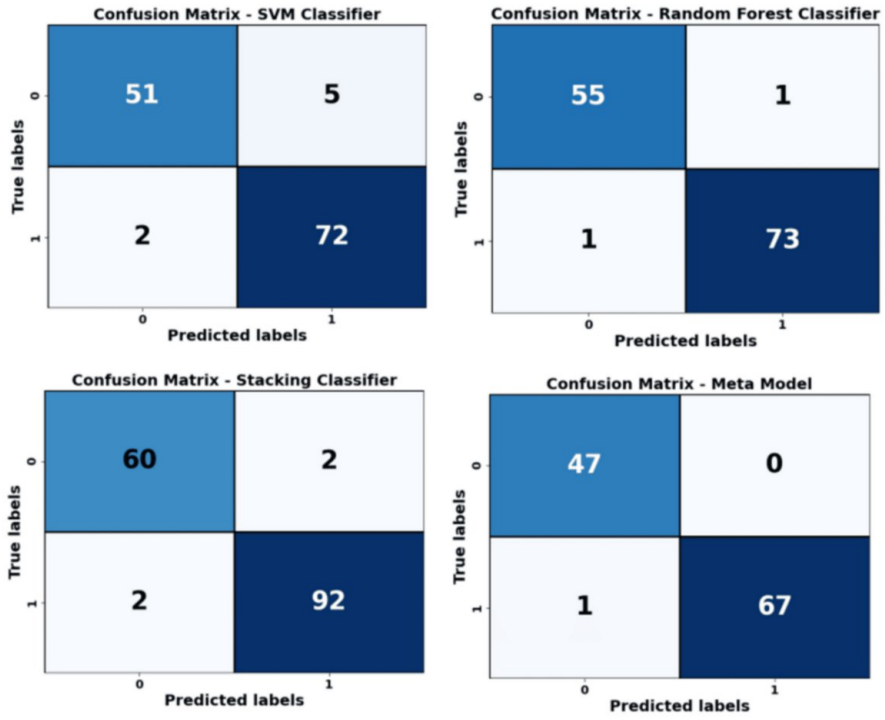


**FIGURE 8: Confusion matrix for SVM, random forest, stacking classifier, and Meta-Model (GBC) using SDHS-B dataset**

SVM: Support Vector Machine, GBC: Gradient Boosting Classifier, SDHS-B: Sylhet Diabetes Hospital in Sylhet, Bangladesh

The evaluation of classification models on both the DDFH-G and PIMA-I datasets showcases varying performance across multiple metrics, illustrating the strengths of each model in Tables *4* and 5. For the DDFH-G dataset, the Meta-Model (GBC) achieved the highest accuracy of 99.10%, closely followed by the voting classifier at 99.08%. Precision scores were impressive across the board, with the base model and both the voting and stacking classifiers attaining perfect scores of 100%, while the Meta-Model scored 98.63%. Sensitivity for the DDFH-G dataset saw the base model at 91.03% and the Meta-Model at 98.62%, indicating robust performance in identifying positive instances. Specificity remained high for all models, with the base model at 100% and the Meta-Model at 99.33%. The F1 score and MCC also highlighted the Meta-Model's effectiveness, scoring 98.62% and 97.94%, respectively.

In contrast, the evaluation of the PIMA-I dataset revealed that the voting classifier led with an accuracy of 99%, while the Meta-Model slightly outperformed it with an accuracy of 99.10%. While precision was perfect for the base, voting, and stacking classifiers at 100%, the Meta-Model recorded a strong precision of 98.63%. Sensitivity metrics were notable, with the base model at 91.41% and the GBC at 98.62%. Similar to the DDFH-G results, specificity was uniformly high, reaching 100% for the base and voting classifiers and 99.32% for the Meta-Model. The F1 score for the base model was 95.51%, while the Meta-Model excelled with 98.62%, and the MCC scores indicated strong overall performance, with the GBC achieving 97.95%. Collectively, these evaluations underscore the Meta-Model's consistent superiority across both datasets, emphasizing its effectiveness and reliability for classification tasks.

| Evaluation Indexes | Base Model | | Voting Classifier | Stacking Classifier | Meta-Model (GBC) |
|---|---|---|---|---|---|
| | SVM | RF | | | |
| Accuracy | 97.05 | 99.08 | 98.64 | 99.09 | 99.10 |
| Precision | 100 | 98.62 | 100 | 98.62 | 98.63 |
| Sensitivity | 91.03 | 98.62 | 95.86 | 98.62 | 98.62 |
| Specificity | 100 | 99.32 | 100 | 99.32 | 99.33 |
| F1 Score | 95.31 | 98.62 | 97.89 | 98.62 | 98.62 |
| MCC | 93.38 | 97.94 | 96.93 | 97.94 | 97.94 |

**TABLE 4: Performance evaluation of the proposed base model, voting classifier, stacking classifier, and Meta-Model using the DDFH-G dataset**

MCC: Matthews Correlation Coefficient, SVM: Support Vector Machine, RF: Random Forest, GBC: Gradient Boosting Classifier, DDFH-G: Diabetes Dataset Frankfurt Hospital-Germany

| Evaluation Indexes | Base Model | | Voting Classifier | Stacking Classifier | Meta-Model (GBC) |
|---|---|---|---|---|---|
| | SVM | RF | | | |
| Accuracy | 97.25 | 99 | 97.75 | 99.09 | 99.10 |
| Precision | 100 | 100 | 100 | 98.62 | 98.63 |
| Sensitivity | 91.41 | 96.88 | 92.97 | 98.62 | 98.62 |
| Specificity | 100 | 100 | 100 | 99.32 | 99.32 |
| F1 Score | 95.51 | 98.41 | 96.36 | 98.62 | 98.62 |
| MCC | 93.37 | 97.71 | 94.86 | 97.94 | 97.95 |

**TABLE 5: Performance evaluation of the proposed base model, voting classifier, stacking classifier, and Meta-Model using the PIMA-I dataset**

MCC: Matthews Correlation Coefficient, SVM: Support Vector Machine, RF: Random Forest, GBC: Gradient Boosting Classifier, PIMA-I: PIMA Indian

The visualization charts of these measures are shown in Figures *9*-*11*, for three datasets, which offer clear insights into the model's efficacy and highlight the validity of the Meta-Model for the diagnosis of diabetes. As can be seen by comparing the Meta-Model's evaluation metrics to those of the voting classifier, stacking classifier, and the individual base models (SVM and RF), the Meta-Model performs the best and most consistently out of all the models using several datasets.

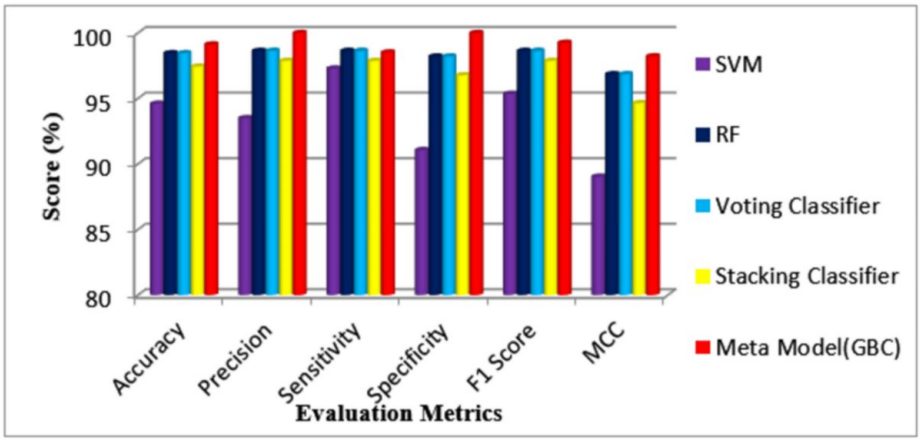


**FIGURE 9: Performance evaluation visualization of the proposed base model, voting classifier, stacking classifier, and Meta-Model using the SDHS-B dataset**

MCC: Matthews Correlation Coefficient, SVM: Support Vector Machine, RF: Random Forest, GBC: Gradient Boosting Classifier, SDHS-B: Sylhet Diabetes Hospital in Sylhet, Bangladesh

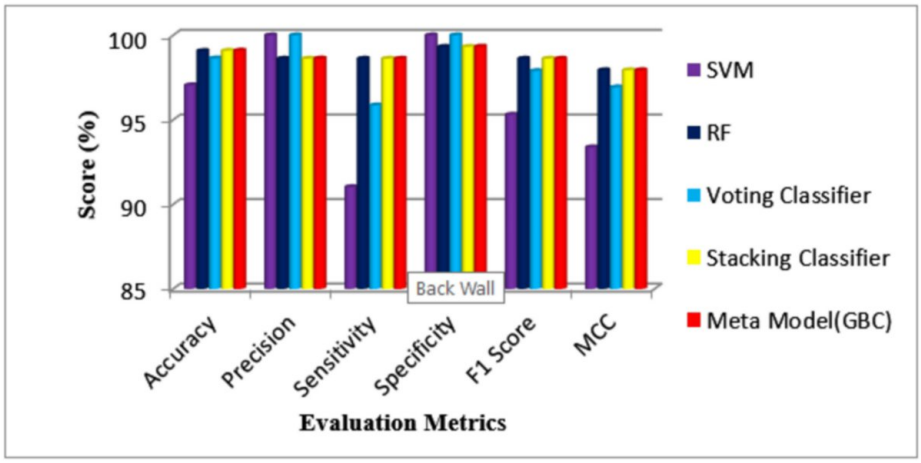


**FIGURE 10: Performance evaluation visualization of the proposed base model, voting classifier, stacking classifier, and Meta-Model using the DDFH-G dataset**

MCC: Matthews Correlation Coefficient, SVM: Support Vector Machine, RF: Random Forest, GBC: Gradient Boosting Classifier, DDFH-G: Diabetes Dataset Frankfurt Hospital-Germany

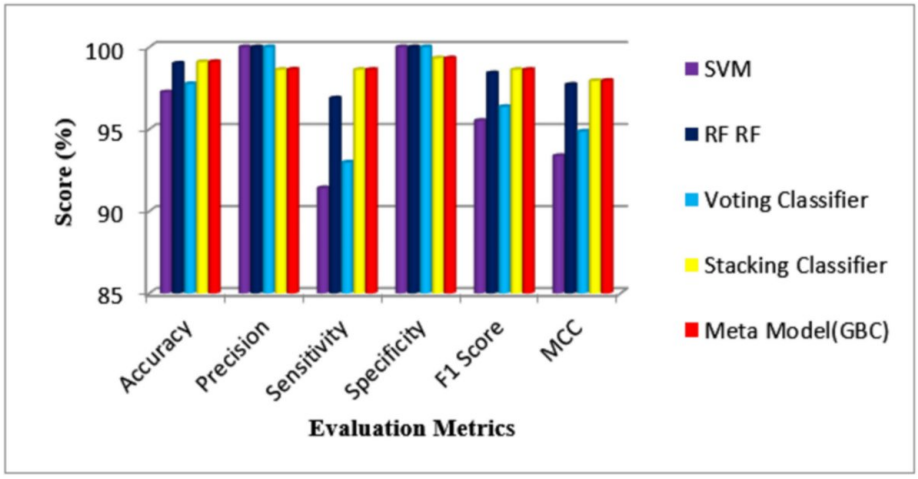


**FIGURE 11: Performance evaluation visualization of the proposed base model, voting classifier, stacking classifier, and Meta-Model using the PIMA-I dataset**

MCC: Matthews Correlation Coefficient, SVM: Support Vector Machine, RF: Random Forest, GBC: Gradient Boosting Classifier, PIMA-I: PIMA Indian

Plotting the true-positive rate (TPR) against the false-positive rate (FPR), the ROC curves for four models - SVM, voting classifier, stacking classifier, and Meta-Model - are displayed in Figure *12*. These curves each represent the classification performance of these models. Strong performance is indicated by the orange SVM curve, which initially shows a high TPR with few FP; however, performance gradually decreases as the number of FP rises. By virtue of its curve being closer to the top-left corner and its ability to combine many models through soft voting, the voting classifier outperforms the SVM. With a nearly flat ROC curve at the top, indicating an incredibly low FPR and nearly flawless classification, the purple stacking classifier performs virtually flawlessly. The Meta-Model is the last model to reach almost perfect performance; its ROC curve hugs the top-left corner, indicating near-flawless or flawless classification. The stacking classifier and Meta-Model perform the best out of all the models; the Meta-Model probably achieves an AUC near to 1, indicating a notable increase in predicting accuracy when employing ensemble approaches as opposed to more straightforward classifiers like SVM. The Meta-Model is the study's best-performing model for classification, demonstrating the value of ensemble techniques in your diabetes risk prediction framework.

The SVM, voting classifier, stacking classifier, and Meta-Model precision-recall (PR) curves in Figure *13* show how each model strikes a balance between accuracy and recall for diabetes risk prediction. The SVM model has great precision for the majority of recall values. Still, precision gradually decreases as recall levels rise, suggesting a trade-off between increased identification of genuine positives and FP. The voting classifier, which gains from mixing many classifiers, performs almost flawlessly with continuous excellent precision and recall. The purple-colored stacking classifier has a practically perfect PR curve, demonstrating its remarkable ability to classify with few false positives and flawless precision across all recall settings. In a similar vein, the Meta-Model exhibits almost perfect performance, with precision holding steady at 1.0 for the majority of recall values, albeit slightly declining at the highest recall points. All things considered, the ensemble models - in particular, the Meta-Model and stacking classifier - perform better than the individual classifiers because they strike the ideal balance between precision and recall, which makes them extremely useful for precise diabetes risk assessment.

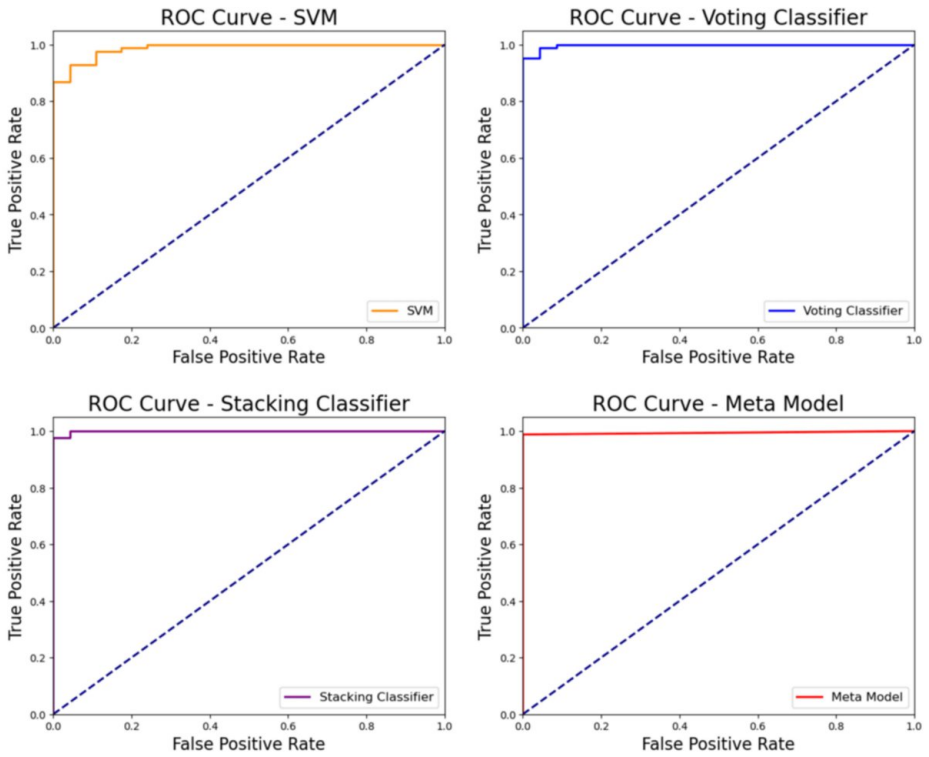


**FIGURE 12: ROC curve for SVM, voting classifier, stacking classifier, and Meta-Model using the SDHS-B dataset**

ROC: Receiver Operating Characteristic, SVM: Support Vector Machine, SDHS-B: Sylhet Diabetes Hospital in Sylhet, Bangladesh

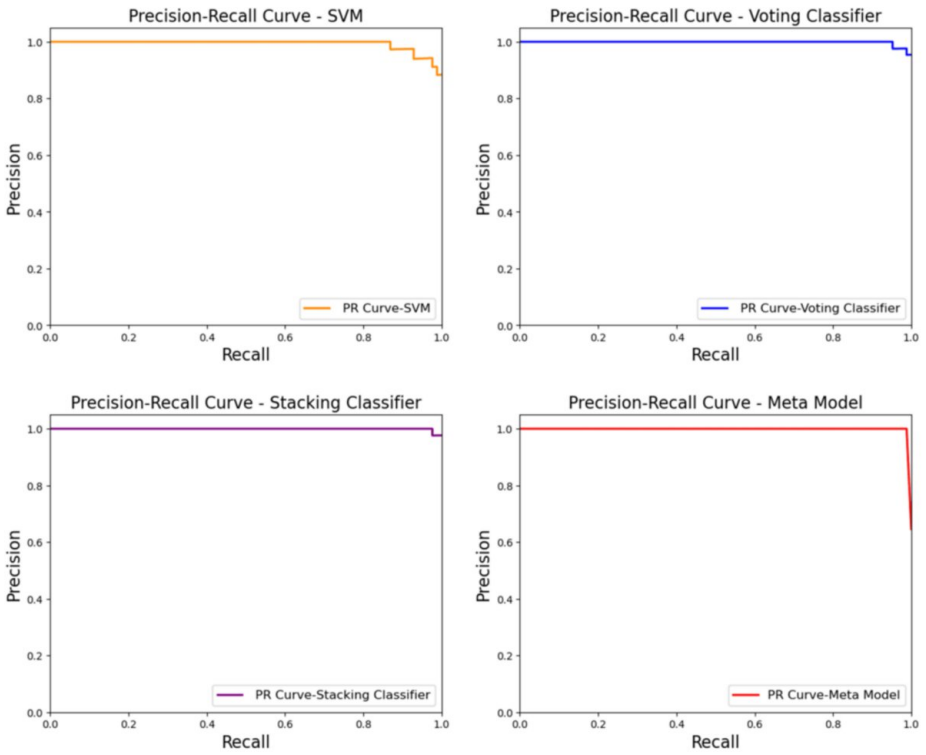


**FIGURE 13: Precision-recall curve for SVM, voting classifier, stacking classifier, and Meta-Model using the SDHS-B dataset**

SVM: Support Vector Machine, SDHS-B: Sylhet Diabetes Hospital in Sylhet, Bangladesh

To grasp the confidence of model predictions, it is essential to examine the calibration curves for the SVM, voting classifier, stacking classifier, and Meta-Model in Figure *14*. These curves show how well each model's projected probabilities match the actual results. The diagonal "perfectly calibrated" line is significantly deviated from by the SVM curve, which exhibits a high degree of fluctuation. This suggests that the model may be overconfident in certain probability ranges, causing the predicted probabilities not to match the true likelihood of positive outcomes closely. While it shows better alignment at extreme probability values, the voting classifier also has some inconsistencies. Its curve displays high peaks and deviations, indicating locations where the predicted probabilities are not well calibrated. Although it performs well in classification trials, the stacking classifier (purple) shows a similar pattern with significant deviations from the perfectly calibrated line, indicating space for improvement in probability prediction. The Meta-Model, on the other hand, shows an extremely flawless calibration curve that closely resembles the diagonal line in every probability range. This implies that the Meta-Model is the most trustworthy model for estimating the likelihood of diabetes risk because it not only performs well in terms of categorization but also produces extremely reliable predicted probabilities. Overall, these curves show that the Meta-Model is both highly accurate and well calibrated, whereas ensemble methods such as the voting and stacking classifiers demonstrate strong classification performance but may need further calibration to increase the reliability of their probability estimates.

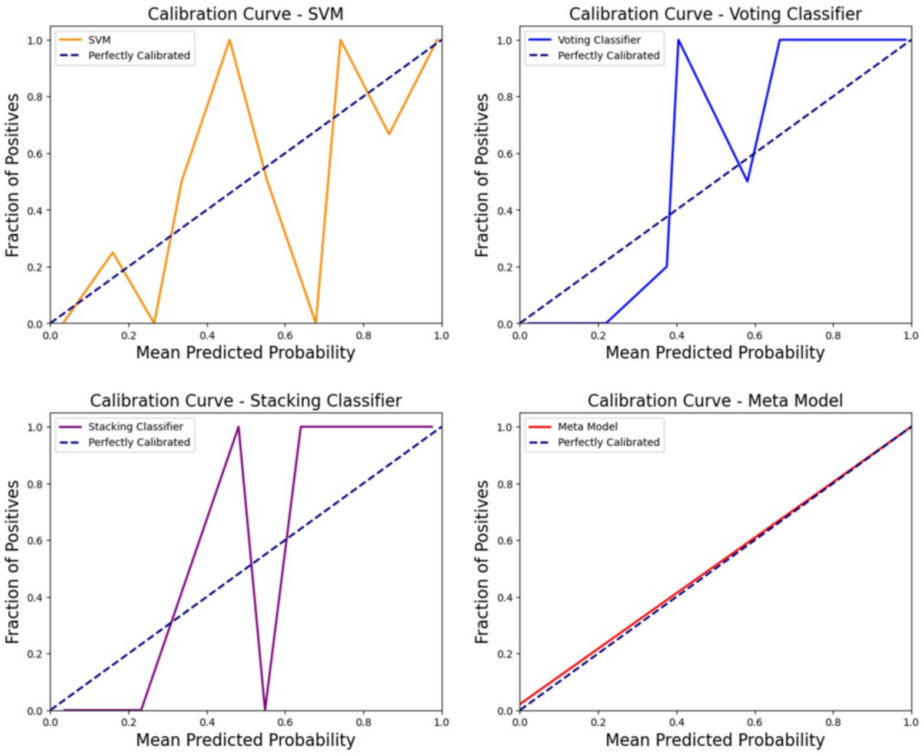


**FIGURE 14: Calibration curve for SVM, voting classifier, stacking classifier, and Meta-Model using the SDHS-B dataset**

SVM: Support Vector Machine, SDHS-B: Sylhet Diabetes Hospital in Sylhet, Bangladesh

In order to measure generalization performance as the training size rises, the learning curves for four models - SVM, voting classifier, stacking classifier, and Meta-Model - evaluated exclusively on the test data are shown in Figure *15*. Because of the short training size, all models initially show lower accuracy and increased variability; however, when additional data is added, the accuracy increases and stabilizes. The voting classifier performs marginally better, stabilizing at almost 97% accuracy, while the SVM steadily improves, reaching about 94% accuracy. Ensemble techniques such as the Meta-Model and stacking classifier typically outperform the others, with the Meta-Model exhibiting the best performance at around 99% accuracy and the stacking classifier obtaining approximately 98%. Confidence intervals are shown by the shaded areas surrounding the curves; as training size increases, these intervals go smaller, indicating lower variance and higher model confidence. When compared to the individual SVM model, ensemble models exhibit better test accuracy and stability overall, highlighting the benefits of merging many classifiers for reliable generalization.

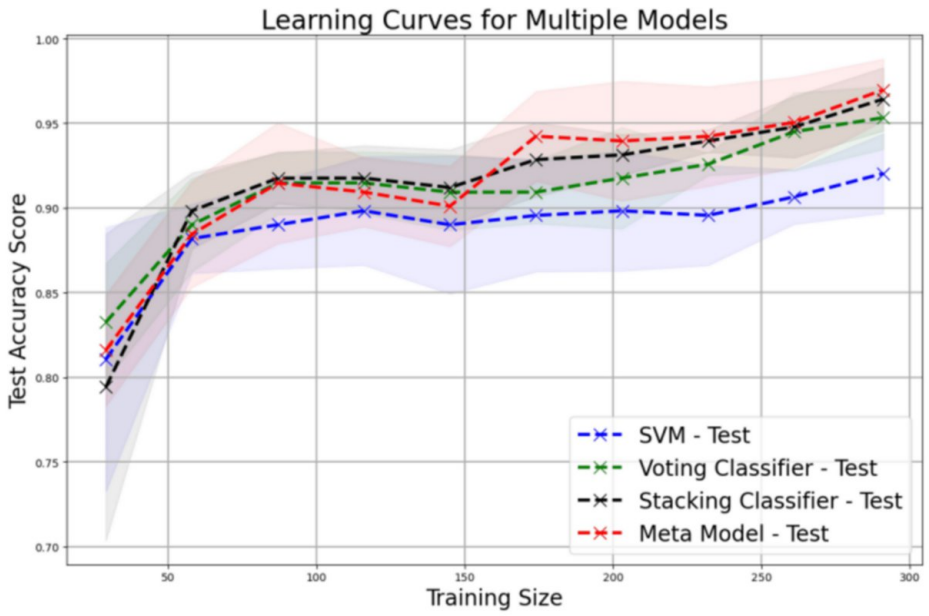


**FIGURE 15: Learning curve for SVM, voting classifier, stacking classifier, and Meta-Model using the SDHS-B dataset**

SVM: Support Vector Machine, SDHS-B: Sylhet Diabetes Hospital in Sylhet, Bangladesh

| Ref. | Dataset | Model | Accuracy |
|---|---|---|---|
| [3] | SDHS-B | RF | 98% |
| [17] | NHANES | Feature selection (LR) and Classifier (RF) | 94.25% |
| [18] | DHIS database public hospitals | RF | 93.8% |
| [20] | PIMA-I | LR | 80.2% |
| [26] | PIMA-I | Naive Bayes | 79.13% |
| [27] | PIMA-I | Deep Dense Layer Neural Network (DDLNN) | 84.42% |
| [28] | PIMA-I | Soft Voting (Logistic Regression, Naive Bayes, and RF) | 97% |
| [29] | PIMA-I | SVM | 81.4% |
| [30] | PIMA-I | CNN | 92.31% |
| [37] | DDFH-G | ML | 82% |
| [38] | PIMA-I | K-Means++ | 98% |
| [39] | PIMA-I, DDFH-G, IDPD-I | ANN, LSTM, CNN | 98.81%, 97.23%, 94.81% |
| **Proposed Model** | **SDHS-B** | **Meta-Model (GBC)** | **99.13%** |
| | **PIMA-I** | **Meta-Model (GBC)** | **99.10%** |
| | **DDFH-G** | **Meta-Model (GBC)** | **99.10%** |

**TABLE 6: Comparing the proposed Meta-Model with state-of-the-art approaches**

SDHS-B: Sylhet Diabetes Hospital in Sylhet, Bangladesh, NHANES: National Health and Nutrition Examination Survey, DHIS: District Health Information System, PIMA-I: PIMA Indian, DDFH-G: Diabetes Dataset Frankfurt Hospital-Germany, IDPD-I: Iraqi Diabetes Patient Dataset, LR: Logistic Regression, RF: Random Forest, SVM: Support Vector Machine, CNN: Convolutional Neural Networks, ML: Machine Learning, ANN: Artificial Neural Networks, LSTM: Long Short-Term Memory, GBC: Gradient Boosting Classifier

In Table 6, the performance comparison of various models across different datasets highlights the effectiveness of the proposed Meta-Model (GBC) in predicting T2DM. Studies such as [3], which used RF on the SDHS-B dataset, achieved 98% accuracy, demonstrating strong predictive performance. Other approaches, like LR and RF on the NHANES (National Health and Nutrition Examination Survey) dataset [17], yielded 94.25%, and RF applied to the District Health Information System database [18] reached 93.8%. Methods using the PIMA-I dataset, such as LR [20] and NB [26], resulted in lower accuracy, at 80.2% and 79.13%, respectively, while more advanced techniques like DDLNN [27] and CNN [30] improved accuracy to 84.42% and 92.31%, respectively. However, ensemble methods like soft voting [28], incorporating LR, NB, and RF, significantly boosted accuracy to 97%. Similarly, SVM [29] and K-means++ [35] on the PIMA-I dataset yielded 81.4% and 98% accuracy, respectively. The use of ANN, LSTM, and CNN on datasets like PIMA-I, DDFH-G, and Iraqi Diabetes Patient Dataset [39] showed high performance, with accuracies of 98.81%, 97.23%, and 94.81%, respectively.

## Discussion

The proposed Meta-Model, based on GBC, demonstrated exceptional performance across multiple datasets, establishing its superiority over traditional and machine learning approaches from previous research. On the SDHS-B dataset, the Meta-Model achieved an outstanding accuracy of 99.13%, significantly surpassing the performance of earlier models applied to the same data. Furthermore, its consistent accuracy of 99.10% on the PIMA-I and DDFH-G datasets highlights its robustness and adaptability to varied data environments. This consistency across diverse datasets underscores the Meta-Model's capacity to generalize effectively, ensuring reliable predictions beyond individual datasets. Such generalization is critical in real-world scenarios where data variability can challenge model reliability. The superior performance of the Meta-Model is a testament to its advanced design, leveraging the strengths of ensemble learning to optimize prediction accuracy. These findings reinforce its potential as a reliable, comprehensive solution for diabetes risk prediction, offering a significant advancement in early diagnosis and preventive healthcare strategies.

## Conclusions

Our research underscores the transformative role of machine learning in the early diagnosis of T2DM. By employing an advanced stacking ensemble framework, we achieved an impressive accuracy of 99.13%, demonstrating the effectiveness of predictive modeling in healthcare. Early identification is crucial for preventing long-term complications, and our study highlights how machine learning can contribute to the development of accessible, cost-effective, and efficient diagnostic tools, benefiting a wider population.

Despite the promising results, further research is necessary to enhance the model's generalizability and clinical applicability. Future work should focus on validating the model across larger, more diverse datasets to ensure its robustness across different demographics and medical conditions. Additionally, incorporating key physiological variables such as body size, height, and BMI could refine predictive accuracy by identifying subtle patterns related to diabetes onset. Another vital direction is the integration of longitudinal and time-series data to track patient health trends over time. This could improve the model's predictive capabilities by recognizing progressive changes indicative of early-stage diabetes. Moreover, exploring deep learning-based meta-models and hybrid architectures may further optimize the detection process, leading to more accurate and adaptable diagnostic solutions.

## Additional Information

### Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

**Acquisition, analysis, or interpretation of data:** Akif Mahdi, Md. Rashed , Md. Imran Hossain

**Critical review of the manuscript for important intellectual content:** Akif Mahdi, Md. Imran Hossain

**Supervision:** Akif Mahdi, Md. Imran Hossain

**Concept and design:** Md. Rashed , Md. Imran Hossain, Gulam Mustofa

**Drafting of the manuscript:** Md. Rashed , Gulam Mustofa

### Disclosures

**Human subjects:** All authors have confirmed that this study did not involve human participants or tissue.
**Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue.
**Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the

following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

## References


1. Hasan MK, Alam MA, Das D, Hossain E, Hasan M: Diabetes prediction using ensembling of different machine learning classifiers. IEEE Access. 2020, 8:76516-76531. 10.1109/access.2020.2989857
2. Saeedi P, Petersohn I, Salpea P, et al.: Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Research and Clinical Practice. 2019, 157:107843. 10.1016/j.diabres.2019.107843
3. Gowthami S, Venkata Siva Reddy R, Ahmed MR: Exploring the effectiveness of machine learning algorithms for early detection of type-2 diabetes mellitus. Measurement: Sensors. 2024, 31:100983. 10.1016/j.measen.2023.100983
4. Deepthi Y, Kalyan KP, Vyas M, Radhika K, Babu DK, Krishna Rao NV: Disease prediction based on symptoms using machine learning. Energy Systems, Drives and Automations. Lecture Notes in Electrical Engineering. Sikander A, Acharjee D, Chanda C, Mondal P, Verma P (ed): Springer, Singapore; 2020. 664:561-569. 10.1007/978-981-15-5089-8_55
5. Kumar A, Gangwar R, Zargar AA, Kumar R, Sharma A: Prevalence of diabetes in India: A review of IDF Diabetes Atlas 10th edition. Current Diabetes Reviews. 2024, 20:e130423215752. 10.2174/1573399819666230413094200
6. GBD 2021 Diabetes Collaborators: Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: A systematic analysis for the Global Burden of Disease Study 2021. The Lancet. 2023, 402:203-234. 10.1016/S0140-6736(23)01301-6
7. Williams R, Karuranga S, Malanda B, et al.: Global and regional estimates and projections of diabetes-related health expenditure: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Research and Clinical Practice. 2020, 162:108072. 10.1016/j.diabres.2020.108072
8. Parker ED, Lin J, Mahoney T: Economic costs of diabetes in the U.S. in 2022. Diabetes Care. 2023, 47:26-43. 10.2337/dci23-0085
9. Moro AR: Mapped: Diabetes rates by country in 2021. Visual Capitalist. (2023). Accessed: September 26, 2024: https://www.visualcapitalist.com/cp/diabetes-rates-by-country/.
10. Jaiswal V, Negi A, Pal T: A review on current advances in machine learning based diabetes prediction. Primary Care Diabetes. 2021, 15:435-443. 10.1016/j.pcd.2021.02.005
11. Shiva Prasad BVV, Gupta S, Borah N, Dineshkumar R, Lautre HK, Mouleswararao B: Predicting diabetes with multivariate analysis an innovative KNN-based classifier approach. Preventive Medicine. 2023, 174:107619. 10.1016/j.ypmed.2023.107619
12. Chaki J, Thillai Ganesh S, Cidham SK, Ananda Theertan S: Machine learning and artificial intelligence based diabetes mellitus detection and self-management: A systematic review. Journal of King Saud University - Computer and Information Sciences. 2022, 34:3204-3225. 10.1016/j.jksuci.2020.06.013
13. Zia A, Aziz M, Popa I, Khan SA, Hamedani AF, Asif AR: Artificial intelligence-based medical data mining. Journal of Personalized Medicine. 2022, 12:1359. 10.3390/jpm12091359
14. Sarwar MA, Kamal N, Hamid W, Shah MA: Prediction of diabetes using machine learning algorithms in healthcare. 2018 24th International Conference on Automation and Computing (ICAC). 2018, 1-6. 10.23919/iconac.2018.8748992
15. Maniruzzaman M, Rahman MJ, Ahammed B, Abedin MM: Classification and prediction of diabetes disease using machine learning paradigm. Health Information Science and Systems. 2020, 8:7. 10.1007/s13755-019-0095-z
16. Ebrahim OA, Derbew G: Application of supervised machine learning algorithms for classification and prediction of type-2 diabetes disease status in Afar regional state, Northeastern Ethiopia 2021. Scientific Reports. 2023, 13:7779. 10.1038/s41598-023-34906-1
17. Thotad PN, Bharamagoudar GR, Anami BS: Diabetes disease detection and classification on Indian demographic and health survey data using machine learning methods. Diabetes & Metabolic Syndrome: Clinical Research & Reviews. 2023, 17:102690. 10.1016/j.dsx.2022.102690
18. AlZu'bi S, Elbes M, Mughaid A, Bdair N, Abualigah L, Forestiero A, Zitar RA: Diabetes monitoring system in smart health cities based on big data intelligence. Future Internet. 2023, 15:85. 10.3390/fi15020085
19. Pan H, Sun J, Luo X, Ai H, Zeng J, Shi R, Zhang A: A risk prediction model for type 2 diabetes mellitus complicated with retinopathy based on machine learning and its application in health management. Frontiers in Medicine. 2023, 10:1136653. 10.3389/fmed.2023.1136653
20. Uddin MJ, Ahamad MM, Hoque MN, et al.: A comparison of machine learning techniques for the detection of type-2 diabetes mellitus: Experiences from Bangladesh. Information. 2023, 14:376. 10.3390/info14070376
21. Zhao M, Wan J, Qin W, Huang X, Chen G, Zhao X: A machine learning-based diagnosis modelling of type 2 diabetes mellitus with environmental metal exposure. Computer Methods and Programs in Biomedicine. 2023, 235:107537. 10.1016/j.cmpb.2023.107537
22. Bernardini M, Romeo L, Misericordia P, Frontoni E: Discovering the type 2 diabetes in electronic health records using the sparse balanced support vector machine. IEEE Journal of Biomedical and Health Informatics. 2020, 24:235-246. 10.1109/jbhi.2019.2899218
23. Pyne A, Chakraborty B: Artificial neural network based approach to diabetes prediction using pima indians diabetes dataset. 2023 International Conference on Control, Automation and Diagnosis (ICCAD). 2023, 01-06. 10.1109/ICCAD57653.2023.10152382
24. Chang V, Bailey J, Xu QA, Sun Z: Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms. Neural Computing and Applications. 2022, 35:16157-16173. 10.1007/s00521-022-07049-z
25. Gupta N, Kaushik B, Imam Rahmani MK, Lashari SA: Performance evaluation of deep dense layer neural network for diabetes prediction. Computers, Materials & Continua. 2023, 76:347-366. 10.32604/cmc.2023.038864

26. Alhalaseh R, AL-Mashhadany DAG, Abbadi M: The effect of feature selection on diabetes prediction using machine learning. 2023 IEEE Symposium on Computers and Communications (ISCC). 2023, 1-7. 10.1109/ISCC58397.2023.10218243
27. Sivaranjani S, Ananya S, Aravinth J, Karthika R: Diabetes prediction using machine learning algorithms with feature selection and dimensionality reduction. 021 7th International Conference on Advanced Computing and Communication Systems (ICACCS). 2021, 141-146. 10.1109/ICACCS51430.2021.9441935
28. García-Ordás MT, Benavides C, Benítez-Andrades JA, Alaiz-Moretón H, García-Rodríguez I: Diabetes detection using deep learning techniques with oversampling and feature augmentation. Computer Methods and Programs in Biomedicine. 2021, 202:105968. 10.1016/j.cmpb.2021.105968
29. Khanam JJ, Foo SY: A comparison of machine learning algorithms for diabetes prediction. ICT Express. 2021, 7:432-439. 10.1016/j.icte.2021.02.004
30. Kannadasan K, Edla DR, Kuppili V: Type 2 diabetes data classification using stacked autoencoders in deep neural networks. Clinical Epidemiology and Global Health. 2019, 7:530-535. 10.1016/j.cegh.2018.12.004
31. Shahin OR, Alshammari HH, Alzahrani AA, Alkhiri H, Taloba AI: A robust deep neural network framework for the detection of diabetes. Alexandria Engineering Journal. 2023, 74:715-724. 10.1016/j.aej.2023.05.072
32. Daanouni O, Cherradi B, Tmiri A: Predicting diabetes diseases using mixed data and supervised machine learning algorithms. Proceedings of the 4th International Conference on Smart City Applications. 2019, 10.1145/3368756.3369072
33. Sim R, Chong CW, Loganadan NK, Adam NL, Hussein Z, Lee SWH: Comparison of a chronic kidney disease predictive model for type 2 diabetes mellitus in Malaysia using Cox regression versus machine learning approach. Clinical Kidney Journal. 2022, 16:549-559. 10.1093/ckj/sfac252
34. Early stage diabetes risk prediction dataset. https://www.kaggle.com/datasets/ishandutta/early-stage-diabetes-risk-prediction-dataset.
35. Pima Indians diabetes database. https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database.
36. Dataset of diabetes, taken from the hospital Frankfurt, Germany. (2023). https://www.kaggle.com/datasets/johndasilva/diabetes.
37. Abdulhadi N, Al-Mousa A: Diabetes detection using machine learning classification methods. 2021 International Conference on Information Technology (ICIT). 2021, 350-354. 10.1109/ICIT52682.2021.9491788
38. Zhou H, Xin Y, Li S: A diabetes prediction model based on Boruta feature selection and ensemble learning. BMC Bioinformatics. 2023, 24:224 . 10.1186/s12859-023-05300-5
39. Al Reshan MS, Amin S, Zeb MA, Sulaiman A, Alshahrani H, Shaikh A: An innovative ensemble deep learning clinical decision support system for diabetes prediction. IEEE Access. 2024, 12:106193-106210. 10.1109/access.2024.3436641