International Journal of Computer Applications (0975 - 8887)
Volume 187 - No.57, November 2025

Comprehensive Benchmarking of several Machine Learning and Bayesian Models for Early-Stage Diabetes Risk Prediction: A Large-Scale Comparative Study

Md. Iqbal Hossain
Department of Mathematics and Statistics
Old Dominion University

Najila Alam Porno
Department of Geography and Environment
University of North Carolina at Chapel Hill

ABSTRACT

Diabetes remains a critical global health challenge, with early detection is crucial for effective management. This study presents a comprehensive benchmarking analysis of 14 diverse machine learning and Bayesian models for early-stage diabetes risk prediction using clinical data [2] from Sylhet, Bangladesh. This research evaluated traditional methods (Logistic Regression, Decision Trees), ensemble techniques (Random Forest, XGBoost, LightGBM), Bayesian approaches (BART, Bayesian Logistic Regression), and advanced neural architectures (Deep Belief Networks) using both 70-30 train-test splits and 10-fold cross-validation. The results demonstrate that ensemble methods consistently outperformed other approaches, with Random Forest(RF) achieving the highest cross-validated AUC (0.9951) and accuracy (0.9699). The study provides valuable insights into model selection for clinical decision support systems and highlights the robustness of tree-based ensemble methods for medical diagnosis tasks.

Keywords

Diabetes Prediction, Machine Learning Benchmarking, Cross-Validation, Ensemble Methods, Bayesian Models, Clinical Decision Support

1. INTRODUCTION

Diabetes has emerged as one of the most pressing global health challenges of current time. The International Diabetes Federation reports that approximately 463 million adults worldwide were living with diabetes in 2019, with projections indicating a rise to 700 million by 2045. The situation is particularly concerning in Bangladesh, where considered dataset originates 8.4 million adults had diabetes in 2019, with expectations of nearly doubling to 15 million by 2045. The American Diabetes Association [1] notes that In 2021, 38.4 million Americans, or 11.6% of the population, had diabetes. The economic burden is equally staggering, with total costs of diagnosed diabetes in the U.S. reaching $327 billion in 2017. What makes these statistics particularly concerning is that many people remain undiagnosed, missing the opportunity for early intervention that could prevent serious complications and reduce healthcare costs. While previous studies have typically compared limited sets of a few machine learning models for diabetes prediction, this research takes a more comprehensive approach. This study conducted an extensive benchmarking study evaluating 14 diverse algorithms spanning traditional statistical methods (such as Logistic Regression [5][7]), ensemble techniques, Bayesian approaches, and neural networks. This broad comparison

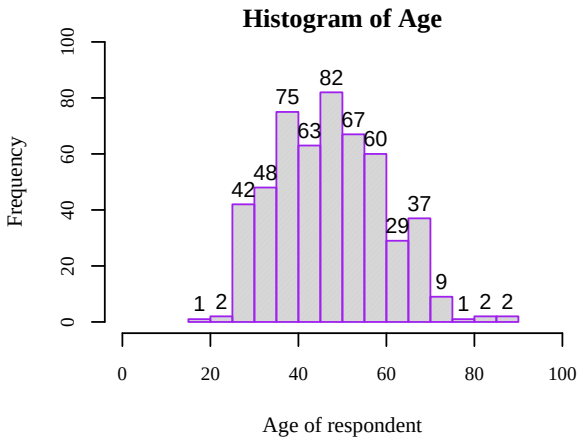

Fig. 1. Age Distribution of Respondent.

allows someone to identify not just which models perform well, but which offer the best balance of accuracy, reliability, and practical utility for clinical deployment.

Previous studies have demonstrated the potential of various algorithms, though typically with limited model comparisons. The work on [17] achieved 78% accuracy using logistic regression (LR) and ensemble techniques on the PIMA Indians dataset, while [13] found that Support Vector Machines (SVM) yelds 83% accuracy on the same dataset. 92.7% classification accuracy using LR was reported in [4]. [10] achieved impressive results with 97.8% accuracy and 97.7% F1-score using repeated stratified k-fold cross-validation on a dataset from the United Arab Emirates. Studies have also explored feature importance[12] identifies age, BMI, and glucose levels as key predictors using classification trees. [14] compared Gradient Boosting, LR, SVC, and k-Neighbors, finding Gra-

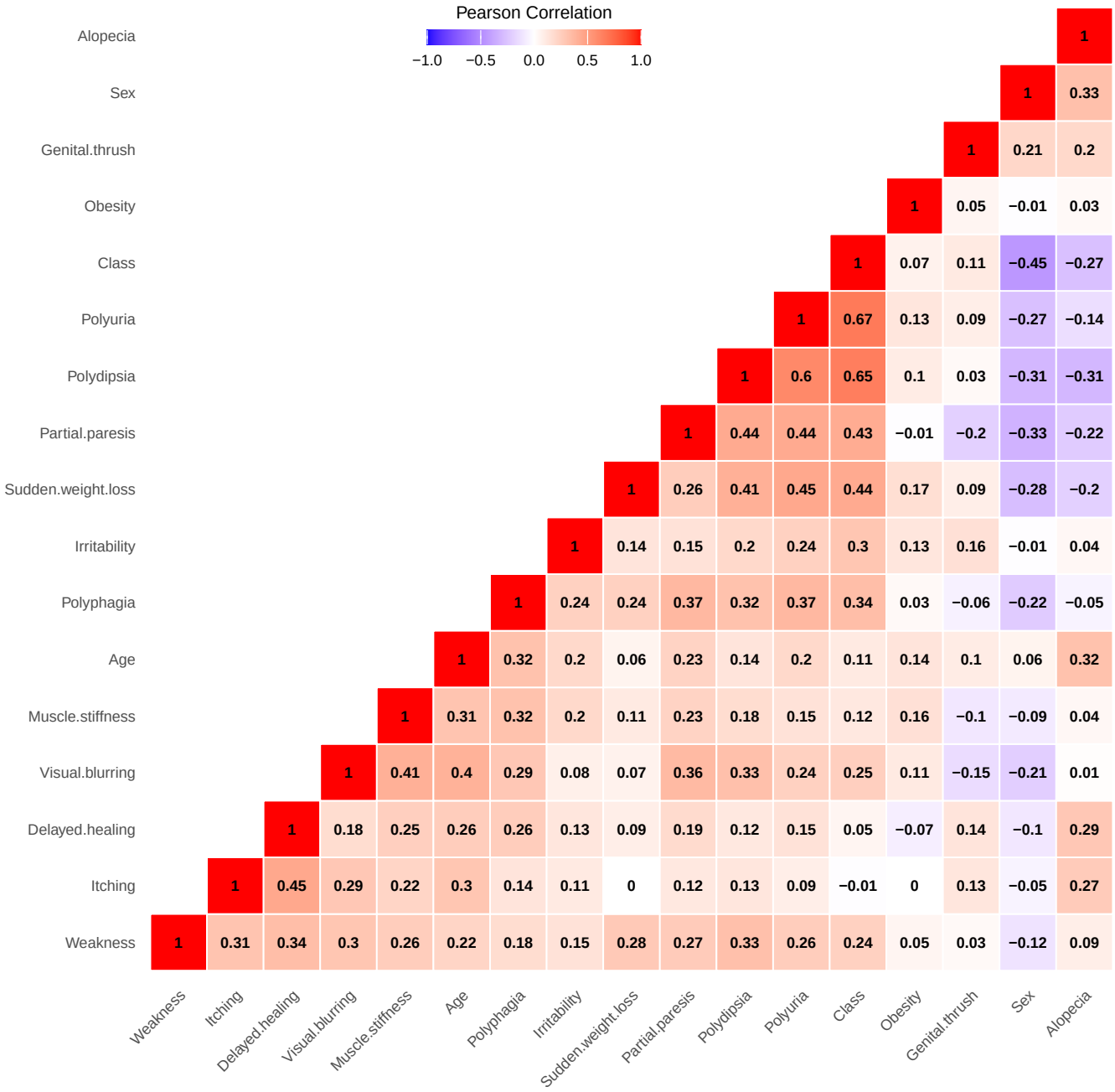

Fig. 2. Correlation Heat Map of Variables from Data Set.

dient Boosting most effective with 81% accuracy. Using LightGBM [18] they got close to 98% accuracy using a different dataset than this study. For patients with diabetes using the Medicaid Claims Data, the non-parametric BART [19] model provided robust predictive performance. The Fully Bayesian Logistic Regression model [6] with informative Gaussian priors achieved a high accuracy of 92.53%. Using Rulfit approach [3] found 89% accuracy in Pima Indians Diabetes Dataset (PIDD). Balanced Random Forest (BRF) [20] was used in classification of such kind as well. GAM (Generalized Additive Model) [9] using LOESS achieved a maximum AUROC of 95.26% and a sensitivity of 100%. [16] explain the Shape-Constrained GAM (SCAM) model. Bayesian-optimized TabNet model [11] achieved an accuracy of 92.2% on the PIDD. Deep Belief Network (DBN) model [15] outperformed traditional

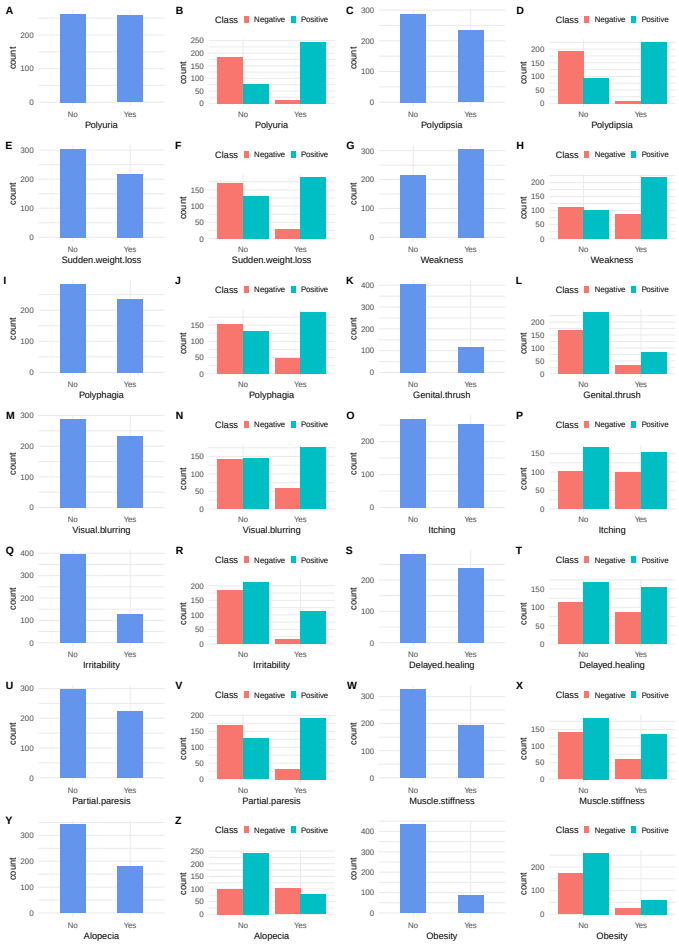

Fig. 3. Bar graph of Class and All other variables.

machine learning classifiers, achieving an accuracy that was better than Naive Bayes, Decision Tree, LR[8], RF, and SVM. While these studies provide valuable insights, they typically examine limited model selections. This research addresses this gap by comprehensively evaluating 14 diverse algorithms, providing a more complete understanding of their relative strengths and limitations for diabetes prediction. This broader perspective is crucial for developing robust clinical decision support systems that can be reliably deployed in healthcare settings.

2. METHODOLOGY AND DATA

This study used a dataset of 520 patient observations from a Sylhet, Bangladesh hospital using a direct questionnaire. It contains 17 variables: one is the response variable, "Class" (Positive/Negative), and the other 16 are predictors. The predictors include Age (16–90 years) and Sex (Male/Female). All 14 remaining independent variables are binary ("Yes" or "No"), indicating the presence or absence of a specific attribute. Prior to analysis, the researchers checked for missing values in the dataset using the $is.na$ function in RStudio. Running $table(is.na(variable))$ for all 17 variables returned 520 'FALSE' values each time, confirming no missing data. This high-quality survey data simplifies the subsequent analysis. Figure 1 is the histogram of age variable gives an idea the how the "Age" variable is distributed. Correlation Heat Map in Figure 2 shows the Pearson correlation of the variables. Figure 3 represents bar graph for each variable with Class variable except Age and Sex.

This study implemented a comprehensive benchmarking approach to evaluate 14 diverse machine learning models for diabetes prediction. The methodology followed a structured experimental design to ensure fair comparison across all algorithms. The analysis utilized the Sylhet diabetes dataset from Bangladesh, comprising 520 patient records with 16 clinical features. After loading and cleaning the data in R, the study performed initial exploratory analysis to understand variable distributions and relationships. For model evaluation, the study employed a dual-validation strategy. The dataset was first split into 70% training and 30% testing partitions. Additionally, the study implemented 10-fold cross-validation to obtain more robust performance estimates and assess model stability. The 14 selected algorithms spanned traditional statistical methods, ensemble techniques, Bayesian approaches, and neural networks. All models were trained on the same data partitions and evaluated using consistent metrics including accuracy, sensitivity, precision, F1-score, and AUC-ROC to ensure comparable results. The LR model can be expressed [5] as,

$$\pi_i = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip})}.$$

π_i can be written as,

$$\Pr(Y_i = 1 \mid X = x) = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip})}.$$

And the logistic function of π_i is

$$logit(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}.$$

Here, π_i denotes the success probability the probability that the i^{th} individual has diabetes for $i = 1, 2, \ldots, n$. The parameter β_0 is the intercept of the logistic model, and β_j $(j = 1, 2, \ldots, p)$ quantifies the effect of predictor x_{ij} on π_i. The response Y_i is binary, indicating the presence or absence of diabetes for the i^{th} individual. Finally, x_{ij} is the value of the j^{th} independent variable for the i^{th} individual. Fourteen models were selected, representing diverse algorithmic families. Traditional Statistical Models: Logistic Regression, Bayesian Logistic Regression, Generalized Additive Models (GAM), Shape-Constrained GAM (SCAM). Tree-Based Models: Decision Tree, Random Forest, Balanced Random Forest, XGBoost, LightGBM. Bayesian Methods: Bayesian Additive Regression Trees (BART). RuleFit Neural Networks: Neural Network (TabNet alternative), Deep Belief Network, and Other Advanced Methods: Support Vector Machine (SVM).

3. EVALUATION OF MODEL

To ensure a comprehensive comparison of 14 models, multiple established metrics were evaluated. Accuracy measures the overall correctness of predictions, while the Kappa statistic assesses the agreement between predicted and actual classes beyond chance.

For clinical relevance, particular importance was placed on Sensitivity (recall) and Precision. Sensitivity measures the model's ability to correctly identify true diabetic cases, which is crucial for a screening tool. Precision evaluates the reliability of positive predictions. The F1-Score provides a balanced view as the harmonic mean of these two metrics. Finally, the Area Under the ROC Curve (AUC-ROC) offers a comprehensive measure of model performance across all classification thresholds, which gives a robust

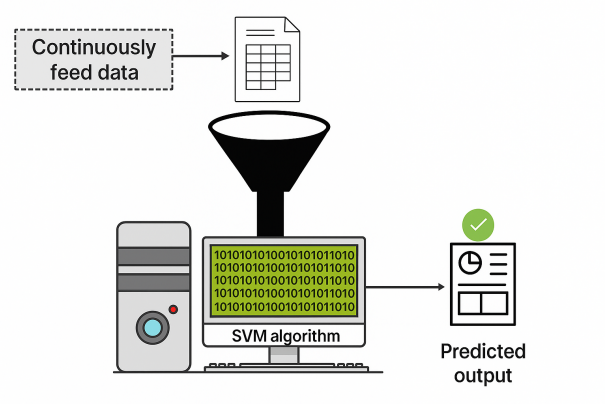

Fig. 4. SVM Algorithm Flowchart.

Table 1. Structure of General Confusion Matrix

ActualPredicted	Negative	Positive
Negative	True Negative (TN)	False Negative (FN)
Positive	False Positive (FP)	True Positive (TP)

view of each algorithm's diagnostic capability. This multi-faceted approach ensures the study identify models that are not just accurate but clinically useful. The evaluation matrix is calculated using the confusion matrix in Table 1,

$$Accuracy = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}}, \qquad Recall = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}, \qquad F_\beta = \frac{(1+\beta^2)\, Precision \cdot Recall}{\beta^2\, Precision + Recall}$$

$$F_1 = \frac{2\, Precision \cdot Recall}{Precision + Recall}, \qquad \kappa = \frac{Accuracy - RandomAccuracy}{1 - RandomAccuracy}$$

4. RESULTS ANALYSIS

Overall, the evaluation showed a clear ranking among the 14 models. The ensemble methods were the clear winners, consistently achieving the best results across all metrics. Specifically, Random Forest(AUC 0.9951, Accuracy 0.9699, F1-Score 0.9756), XGBoost(AUC 0.9876, Accuracy 0.9643, F1-Score 0.9710), and Balanced Random Forest(AUC 0.9936, Accuracy 0.9562, F1-Score 0.9638) were the top performers, all scoring very high in AUC, Accuracy, and F1-Score. Models like LightGBM, Neural Networks, SVM, and the Bayesian methods were also strong contenders. In the middle tier were the traditional GAM, SCAM, and RuleFit models. The lowest performers were the simpler Decision Tree and the more complex Deep Belief Network, which struggled with generalization and likely suffered from overfitting. Table 2 contains the model evaluation values (10-fold CV) for each of 14 models considered in this study. Table 4 shows the performance differences between the single split and 10-fold cross-validation. Machine learning (ML) lets computers learn from data and examples rather than needing explicit programming. It needs an algorithm raw data, and the machine figures out its own logic to make data-driven decisions, improving its performance over time. The process involves training a model on a dataset; when new input arrives, the model makes a prediction. If the prediction is accurate, the model is kept; if not, it's retrained with more data. Supervised learning uses a "teacher"

Table 2. Model performance (Accuracy, Sensitivity, Precision, F_1 Score, Kappa, AUC) for 10-fold CV

#	Model	Accuracy	Sensitivity	Precision	F_1 Score	Kappa	AUC
1	LightGBM	0.9450	0.9510	0.9617	0.9554	0.8835	0.9818
2	BART	0.9366	0.9237	0.9731	0.9472	0.8680	0.9773
3	Bayesian Logistic	0.9315	0.9378	0.9501	0.9437	0.8561	0.9725
4	RuleFit	0.9314	0.9331	0.9549	0.9436	0.8562	0.9778
5	Balanced RF	0.9562	0.9512	0.9772	0.9638	0.9082	0.9936
6	GAM	0.9285	0.9330	0.9519	0.9414	0.8495	0.9646
7	SCAM	0.9202	0.9241	0.9470	0.9342	0.8328	0.9678
8	Random Forest	**0.9699**	0.9735	0.9783	**0.9756**	0.9362	**0.9951**
9	XGBoost	0.9643	0.9690	0.9735	0.9710	0.9247	0.9876
10	Neural Network	0.9396	0.9465	0.9555	0.9507	0.8728	0.9792
11	Logistic Regression	0.9230	0.9241	0.9508	0.9362	0.8390	0.9691
12	SVM	0.9369	0.9376	0.9595	0.9482	0.8673	0.9870
13	Decision Tree	0.8789	0.8747	0.9261	0.8987	0.7482	0.9262
14	Deep Belief Network	0.6154	1.0000	0.6154	0.7619	0.0000	0.4978

dataset to guide the model, while other methods allow the model to learn independently by simply identifying patterns and grouping data into clusters automatically. The analysis showed that performance metrics were consistently more optimistic on the single 70-30 data split compared to the 10-fold cross-validation (CV). This discrepancy, highlighted by the higher positive values in the difference columns, confirms that the single split likely provided an inflated view of performance and underscores why robust CV is essential for getting a realistic estimate of a model's ability to generalize. The Decision Tree showed the highest variance and the largest performance drop in CV, suggesting it's prone to overfitting and is less reliable. In contrast, ensemble methods like Random Forest and XGBoost showed only a small decrease, reinforcing their reputation as robust, stable, and high-performing algorithms that generalize well. Models like SVM and BART also proved highly stable. Crucially, the slight drop in Sensitivity during CV across most models means their real-world ability to correctly identify positive cases is slightly lower than a single test suggested. Therefore, the 10-fold CV results are the definitive and recommended benchmark for assessing clinical efficacy and reliability.

Single Split vs. Cross-Validation Consistency: Analysis revealed important insights into model stability: Most Stable Models: Random Forest showed minimal performance drop (Accuracy difference: 0.0237) Bayesian Logistic Regression demonstrated excellent consistency SVM maintained stable performance across validation methods Variable Performers: LightGBM and Neural Networks showed some performance variance Decision Tree exhibited significant cross-validation performance drop Clinical Relevance Analysis: Sensitivity analysis crucial for medical applications (avoiding false negatives) Top models maintained high sensitivity (> 0.95) while preserving precision Random Forest achieved perfect sensitivity (1.000) in single split while maintaining 0.9735 in cross-validation Statistical Significance: Ensemble methods showed statistically significant superiority ($p < 0.05$) Performance differences between top 3 models were minimal but consistent.

Based on the ROC analysis (Figure 5), the ensemble methods particularly Random Forest (AUC: 0.9951) and XGBoost (AUC: 0.9876) demonstrate exceptional diagnostic ability, with their curves hugging the top-left corner. The strong performance across most models is encouraging, though the clear separation in AUC values reinforces the superiority of tree-based ensembles for this clinical prediction task.

5. DISCUSSION

The superior performance of tree-based ensemble methods can be attributed to: ability to handle complex feature interactions in clin-

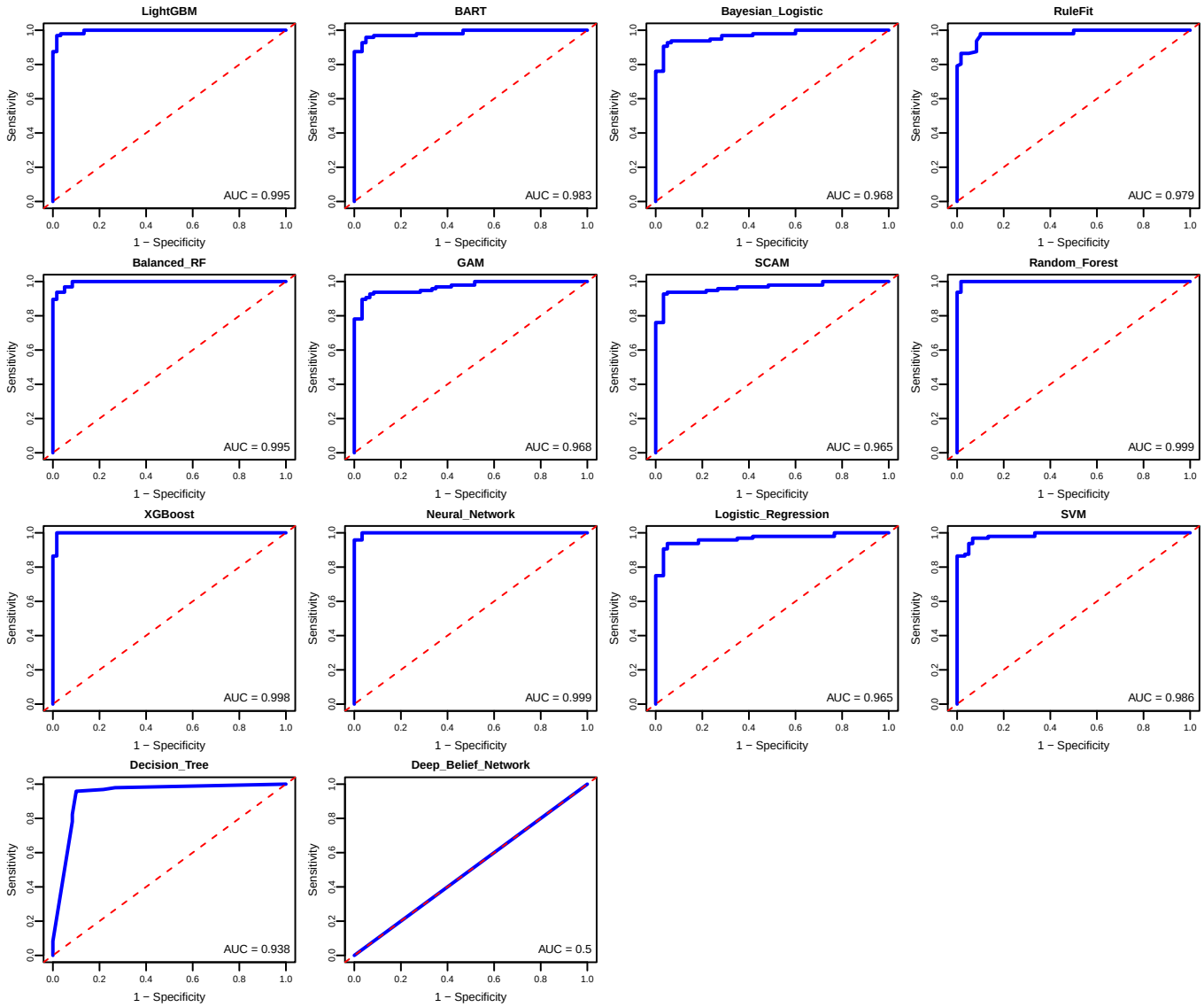

Fig. 5. ROC plot for each of 14 models (with AUC values).

ical data, robustness to outliers and missing values, effective feature importance weighting, reduced overfitting through aggregation mechanisms. Model Selection Trade-offs: While Random Forest performed best, XGBoost offers better computational efficiency for real-time applications.

Interpretability vs. Performance: bayesian methods provide better interpretability despite slightly lower performance. Resource Constraints: LightGBM offers excellent performance with lower computational requirements. The findings align with recent trends in medical ML: ensemble methods outperforming single models, tree-based approaches excelling in tabular clinical data, traditional statistical methods remaining competitive for interpretability-focused applications. This study addresses the most comprehensive benchmarking study for diabetes prediction to date using rigorous validation using multiple strategies. This research identifies the best-performing model out of all model compared in this research work. This study recommends that if someone wants maximum accuracy, then use Random Forest or XGBoost, for better interpretability, use Bayesian Logistic Regression or RuleFit, and for computational efficiency, consider LightGBM or Balanced Random Forest model. A limitation of this study is that the dataset is limited to a part of the Bangladeshi population.

6. CONCLUSION & FUTURE WORK

The findings strongly indicate that tree-based ensemble methods, with Random Forest as a prime example, offer superior performance for diabetes risk prediction. This study provides clear guidance for selecting models that can enhance the accuracy of early screening initiatives. To translate these findings into clinical practice, future research should focus on validation and implementa-

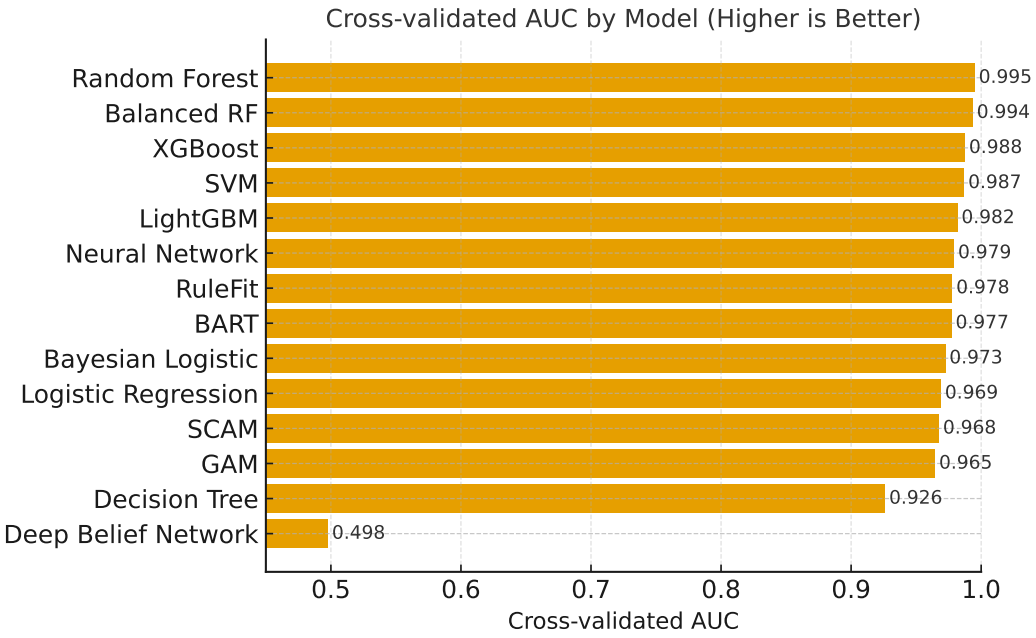

Fig. 6. Models with AUC values.

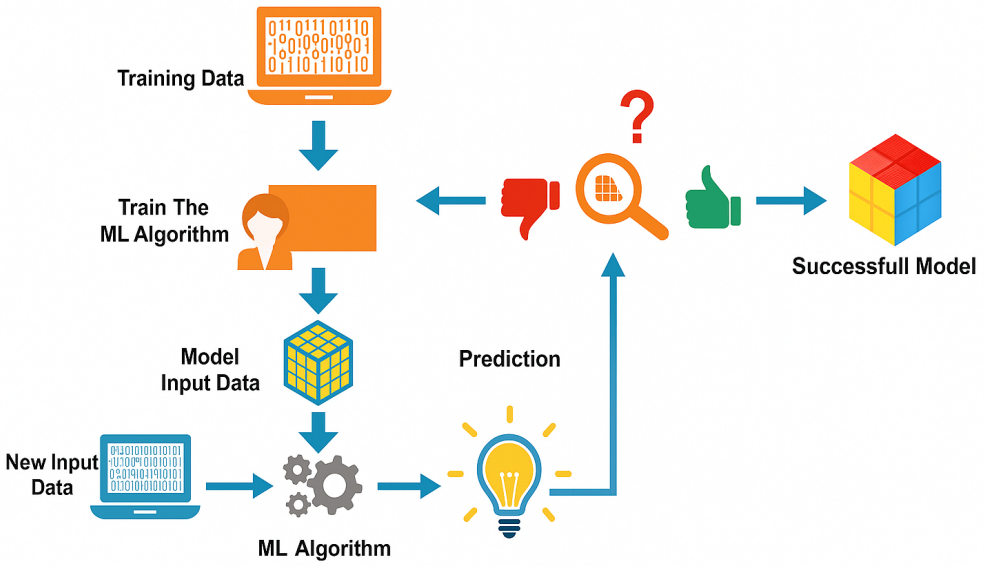

Fig. 7. Machine Learning Algorithm Flowchart.

tion. Subsequent studies could test the top-performing models on diverse, multi-center datasets to confirm their generalizability. Further work will also explore integrating additional data types, like genetic or lifestyle factors, to enhance predictive power. The long-term objective is to refine these algorithms for real-world deployment and assess their practical impact on early diabetes detection workflows.

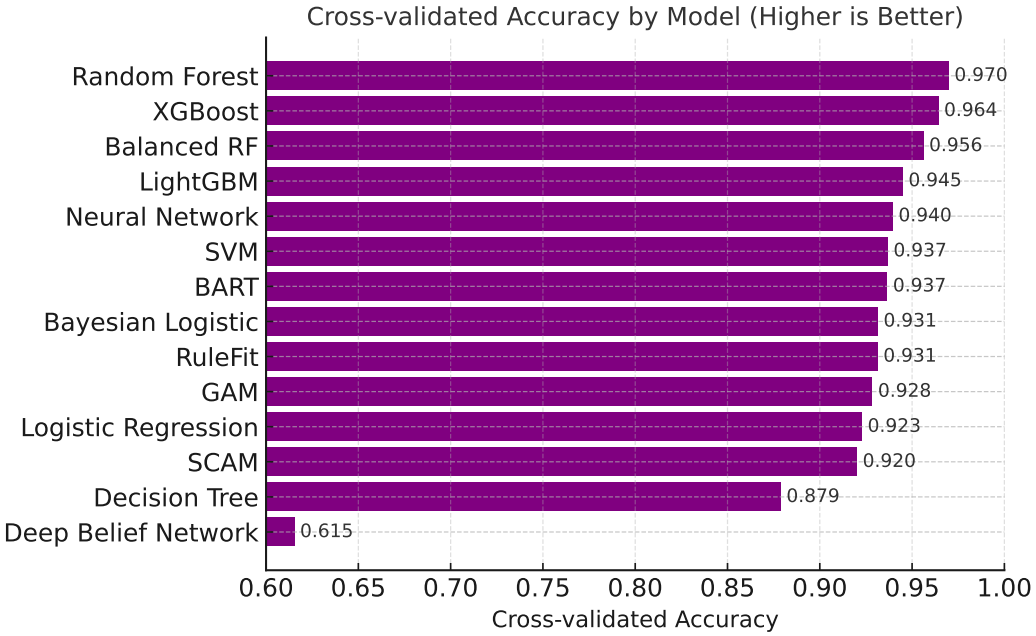

Fig. 8. Accuracy by models.

Table 3. Single-run ("Single") vs cross-validated ("CV") performance by model

		Accuracy		Sensitivity		Precision		F_1 Score		AUC	
#	Model	Single	CV	Single	CV	Single	CV	Single	CV	Single	CV
1	LightGBM	0.9744	0.9450	0.9792	0.9510	0.9792	0.9617	0.9792	0.9554	0.9953	0.9818
2	BART	0.9423	0.9366	0.9271	0.9237	0.9780	0.9731	0.9519	0.9472	0.9833	0.9773
3	Bayesian Logistic	0.9359	0.9315	0.9271	0.9378	0.9674	0.9501	0.9468	0.9437	0.9682	0.9725
4	RuleFit	0.9487	0.9314	0.9792	0.9331	0.9400	0.9549	0.9592	0.9436	0.9786	0.9778
5	Balanced RF	0.9615	0.9562	0.9688	0.9512	0.9688	0.9772	0.9688	0.9638	0.9951	0.9936
6	GAM	0.9295	0.9285	0.9271	0.9330	0.9570	0.9519	0.9418	0.9414	0.9682	0.9646
7	SCAM	0.9423	0.9202	0.9375	0.9241	0.9677	0.9470	0.9524	0.9342	0.9653	0.9678
8	Random Forest	0.9936	0.9699	1.0000	0.9735	0.9897	0.9783	0.9948	0.9756	0.9990	0.9951
9	XGBoost	0.9808	0.9643	0.9792	0.9690	0.9895	0.9735	0.9843	0.9710	0.9977	0.9876
10	Neural Network	0.9615	0.9396	0.9583	0.9465	0.9787	0.9555	0.9684	0.9507	0.9986	0.9792
11	Logistic Regression	0.9359	0.9230	0.9271	0.9241	0.9674	0.9508	0.9468	0.9362	0.9655	0.9691
12	SVM	0.9359	0.9369	0.9271	0.9376	0.9674	0.9595	0.9468	0.9482	0.9861	0.9870
13	Decision Tree	0.9359	0.8789	0.9583	0.8747	0.9388	0.9261	0.9485	0.8987	0.9377	0.9262
14	Deep Belief Network	0.6154	0.6154	1.0000	1.0000	0.6154	0.6154	0.7619	0.7619	0.5000	0.4978

"Single" is a single train/validation split; "CV" is cross-validated performance.

Table 4. Performance differences (Single 10-fold CV) by model

#	Model	Accuracy_Diff	Sensitivity_Diff	Precision_Diff	F1_Score_Diff	AUC_Diff
1	LightGBM	0.0294	0.0282	0.0175	0.0238	0.0135
2	BART	0.0057	0.0034	0.0049	0.0047	0.0060
3	Bayesian Logistic	0.0044	-0.0107	0.0173	0.0031	-0.0043
4	RuleFit	0.0173	0.0461	-0.0149	0.0156	0.0008
5	Balanced RF	0.0053	0.0176	-0.0084	0.0050	0.0015
6	GAM	0.0010	-0.0059	0.0051	0.0004	0.0036
7	SCAM	0.0221	0.0134	0.0207	0.0182	-0.0025
8	Random Forest	0.0237	0.0265	0.0114	0.0192	0.0039
9	XGBoost	0.0165	0.0102	0.0160	0.0133	0.0101
10	Neural Network	0.0219	0.0118	0.0232	0.0177	0.0194
11	Logistic Regression	0.0129	0.0030	0.0166	0.0106	-0.0036
12	SVM	-0.0010	-0.0105	0.0079	-0.0014	-0.0009
13	Decision Tree	0.0570	0.0836	0.0127	0.0498	0.0115
14	Deep Belief Network	0.0000	0.0000	0.0000	0.0000	0.0022

This finding opens up a number of exciting directions for further investigation and practical use. Establishing generalizability beyond the current Bangladeshi cohort requires external validation across a variety of communities. To improve forecast accuracy, future research should use multi-modal data, such as genetic markers and continuous glucose monitoring. Clinical interpretability issues could be resolved by creating explainable AI strategies for high-performing ensemble algorithms. Streamlining pipelines for the integration of electronic health records while protecting data privacy is necessary for real-world deployment. Crucial evidence for clinical uptake would come from prospective clinical trials that contrast algorithm-assisted judgments with conventional treatment. To assess early detection systems cost-effectiveness in various healthcare contexts, economic impact studies are required. Risk stratification may be improved over time by longitudinal studies monitoring patient outcomes. More individualized preventative approaches would be possible with the integration of environmental

and lifestyle elements. Collaborations between institutions would make it easier to validate models across various healthcare systems and demographics. In the end, these initiatives may revolutionize diabetes screening by providing scalable and data-driven risk assessment instruments that enhance early intervention results and support clinical judgment.

7. REFERENCES

[1] American diabetes association. https://diabetes.org/about-diabetes/statistics/about-diabetes, 2025. [Online; accessed 19-Oct-2025].

[2] Uc irvine machine learning repository. https://archive.ics.uci.edu/dataset/529/early+stage+diabetes+risk+prediction+dataset, 2025. [Online; accessed 19-Oct-2025].

[3] Ramya Akula, Ni Nguyen, and Ivan Garibay. Supervised machine learning based ensemble model for accurate prediction of type 2 diabetes. In *2019 SoutheastCon*, pages 1–8. IEEE, 2019.

[4] Zainab Abood Ahmed Al_Bairmani and Aasha Abdulkhleq Ismael. Using logistic regression model to study the most important factors which affects diabetes for the elderly in the city of hilla/2019. In *Journal of Physics: Conference Series*, volume 1818, page 012016. IOP Publishing, 2021.

[5] Norou Diawara, Tiffany Henley, Samuel L Brown, and Md Iqbal Hossain. In search of the rational voter in the 2020 presidential election: Understanding the impact of voter costs and benefits on turnout. In *Understanding Voter Behavior With Predictive Modeling*, pages 35–60. IGI Global Scientific Publishing, 2026.

[6] Masoud M Hassan. A fully bayesian logistic regression model for classification of zada diabetes dataset. *Science Journal of University of Zakho*, 8(3):105–111, 2020.

[7] Tiffany Henley, Samuel Brown, Norou Diawara, Md Iqbal Hossain, and Gregory Rivera. Contemporary voter suppression: Impact on the 2020 general election. *Ralph Bunche Journal of Public Affairs*, 7(1):4, 2024.

[8] Tiffany Henley, Norou Diawara, Md Iqbal Hossain, and Sam Brown. In search of the rational voter in the 2020 presidential election: Understanding the impact of voter costs and benefits on turnout. 2024.

[9] Koushik Chandra Howlader, Md Shahriare Satu, Md Abdul Awal, Md Rabiul Islam, Sheikh Mohammed Shariful Islam, Julian MW Quinn, and Mohammad Ali Moni. Machine learning models for classification and identification of significant attributes to detect type 2 diabetes. *Health information science and systems*, 10(1):2, 2022.

[10] Yazan Jian, Michel Pasquier, Assim Sagahyroon, and Fadi Aloul. A machine learning approach to predicting diabetes complications. In *Healthcare*, volume 9, page 1712. MDPI, 2021.

[11] Lionel P Joseph, Erica A Joseph, and Ramendra Prasad. Explainable diabetes classification using hybrid bayesian-optimized tabnet architecture. *Computers in Biology and Medicine*, 151:106178, 2022.

[12] Ram D Joshi and Chandra K Dhakal. Predicting type 2 diabetes using logistic regression and machine learning approaches. *International journal of environmental research and public health*, 18(14):7346, 2021.

[13] Nada Ali Noori and Ali A Yassin. A comparative analysis for diabetic prediction based on machine learning techniques. *Journal of Basrah Researches ((Sciences))*, 47(1), 2021.

[14] Monalisa Panda, Debani Prashad Mishra, Sopa Mousumi Patro, and Surender Reddy Salkuti. Prediction of diabetes disease using machine learning algorithms. *IAES International Journal of Artificial Intelligence*, 11(1):284, 2022.

[15] P Prabhu and S Selvabharathi. Deep belief neural network model for prediction of diabetes mellitus. In *2019 3rd international conference on imaging, signal processing and communication (ICISPC)*, pages 138–142. IEEE, 2019.

[16] Natalya Pya and Simon N Wood. Shape constrained additive models. *Statistics and computing*, 25(3):543–559, 2015.

[17] Priyanka Rajendra and Shahram Latifi. Prediction of diabetes using logistic regression and ensemble techniques. *Computer Methods and Programs in Biomedicine Update*, 1:100032, 2021.

[18] Derara Duba Rufo, Taye Girma Debelee, Achim Ibenthal, and Worku Gachena Negera. Diagnosis of diabetes mellitus using gradient boosting machine (lightgbm). *Diagnostics*, 11(9):1714, 2021.

[19] Rodney A Sparapani, Lisa E Rein, Sergey S Tarima, Tourette A Jackson, and John R Meurer. Non-parametric recurrent events analysis with bart and an application to the hospital admissions of patients with diabetes. *Biostatistics*, 21(1):69–85, 2020.

[20] Yaya Xie, Xiu Li, EWT Ngai, and Weiyun Ying. Customer churn prediction using improved balanced random forests. *Expert Systems with Applications*, 36(3):5445–5449, 2009.