FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 14, June, 2026, pp 345 – 352
DOI: https://doi.org/10.33003/fjs-2026-1014-5650

A Comparative Evaluation of Machine Learning Algorithms for Diabetes Risk Prediction

Tosin Comfort Olayinka

Department of Computing, College of Science and Computing, Wellspring University, Benin City, Edo State, Nigeria.

*Corresponding authors' email: tosin.olayinka@wellspringuniversity.edu.ng Phone: +2348030789007
https://orcid.org/0000-0001-5987-7390

Received: 31 Jul. 2026
Accepted: 17 Aug. 2026
Published: 27 Aug. 2026

ABSTRACT
Diabetes mellitus is a chronic metabolic disorder whose global prevalence continues to rise, creating an urgent need for scalable, low-cost tools for early risk identification. This study evaluates the effectiveness of five machine learning algorithms (Random Forest, XGBoost, Support Vector Machine [SVM], CatBoost, and TabNet) for predicting diabetes risk from routinely available clinical and lifestyle variables. Using the Pima Indians Diabetes dataset, a preprocessing pipeline was applied that included median imputation of physiologically implausible zero values, standardized (Z-score) feature scaling, Boruta-based feature selection, and class-imbalance handling through class-weight adjustment and the Synthetic Minority Oversampling Technique (SMOTE). Models were trained on an 80/20 train-test split and assessed using accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (ROC-AUC). Random Forest achieved the strongest overall performance (accuracy 0.753; F1-score 0.689; ROC-AUC 0.810), followed by CatBoost, SVM, and XGBoost, whereas TabNet performed worst with very low recall for the diabetic class. The best-performing model (Random Forest) was deployed in a lightweight Flask web application that returns a probability-based diabetes risk assessment, categorising each prediction as low, moderate, or high risk together with a tailored recommendation. The findings confirm that ensemble tree-based methods, particularly Random Forest, provide a reliable, interpretable, and deployable basis for diabetes risk screening, especially in resource-constrained settings. Key limitations include dataset homogeneity, residual class imbalance, and limited feature coverage.

Keywords: Class; Diabetes; Machine Learning; Prediction; Random Forest

INTRODUCTION

Diabetes mellitus is a chronic metabolic disorder characterized by persistently elevated blood glucose levels arising from insufficient insulin production, ineffective insulin utilization, or both. Diabetes Mellitus is a long-term illness that causes the body's improper digestion of carbohydrates, which raises blood glucose levels in humans as a result of reduced insulin production (Okikiola et al., 2023). Its global prevalence has risen sharply over recent decades; the World Health Organization reports that the number of people living with diabetes increased from 200 million in 1990 to 830 million in 2022, with prevalence rising most rapidly in low- and middle-income countries (World Health Organization, 2024). Traditional diagnostic methods such as fasting blood glucose and oral glucose tolerance tests, although accurate, are resource-intensive, require clinical settings and repeated monitoring, and often fail to identify pre-diabetic conditions or predict the likelihood of future onset (Mahesh, 2020).

Machine learning, a subset of artificial intelligence that enables systems to learn patterns from data without explicit programming, has emerged as a promising alternative for predictive modelling in healthcare (Mahesh, 2020; Jordan & Mitchell, 2015). Algorithms such as decision trees, support vector machines, random forests, and neural networks have demonstrated strong performance in identifying at-risk individuals from routinely collected health parameters (Khanam & Foo, 2021; Bavkar & Shinde, 2021). Their capacity to handle large datasets, model non-linear relationships, and integrate diverse data sources makes them well suited to early diabetes detection (Theerthagiri et al., 2022; Tasin et al., 2023).

Despite this promise, several challenges persist. Healthcare datasets are frequently imbalanced, with non-diabetic cases greatly outnumbering diabetic cases, which biases model outcomes (Chowdhury et al., 2023). Feature selection, model interpretability, and generalization to diverse populations remain open problems, and recent systematic reviews confirm that predictive accuracy on benchmark datasets has yet to be consolidated across studies (Kumar et al., 2023; Kiran et al., 2025). Much of the existing literature also evaluates individual algorithms in isolation and rarely translates trained models into accessible tools. This study addresses those gaps by systematically comparing five (5) algorithms under a common, imbalance-aware pipeline and deploying the best performer as a usable, web-based decision-support tool.

Related Works

Machine learning for diabetes prediction has attracted substantial research attention because of its potential for early diagnosis and intervention. Recent systematic and bibliometric reviews document rapid growth in this field over the past three decades and confirm that no single algorithm dominates across datasets, motivating rigorous comparative evaluation (Kiran et al., 2025; Ayoade et al., 2025). This study is organized to move from the general to the specific: it first introduces the concepts, types, and techniques of machine learning that are directly relevant to this study; it then reviews how these techniques have been applied in prior diabetes-prediction research; and it finally examines model

interpretability, deployment approaches, and the research gaps that motivate the present work.

Machine Learning: Concepts, Types, and Techniques

Machine learning (ML) is a subfield of artificial intelligence concerned with building systems that improve automatically through experience, learning statistical patterns from data rather than following explicitly programmed rules (Jordan & Mitchell, 2015; Mahesh, 2020). A learning task is typically framed as estimating a function that maps a set of input features to an output of interest, with predictive performance improving as the model is exposed to more representative data. ML methods are commonly grouped into four broad paradigms (Sarker, 2021): supervised learning, which trains on labelled input-output pairs for classification (predicting discrete categories) and regression (predicting continuous values); unsupervised learning, which discovers hidden structure in unlabelled data through clustering or dimensionality reduction; semi-supervised learning, which combines limited labelled data with abundant unlabelled data; and reinforcement learning, which optimises sequential decisions through reward feedback, with emerging applications in adaptive insulin-dosing strategies (Sarker, 2021; Bi et al., 2019). Deep learning, based on multi-layer neural networks, spans these paradigms and can model highly complex, non-linear relationships when sufficient data are available.

Machine learning plays a vital part in diabetes research in recognizing disorders at an early stage (Rufai et al; 2024). Diabetes risk prediction from structured clinical records is predominantly a supervised binary-classification problem, in which a model learns to distinguish diabetic from non-diabetic individuals using labelled examples. The present study therefore focuses on five supervised classifiers that represent complementary and widely used technique families. Decision-tree ensembles such as the Random Forest aggregate many decorrelated trees through bagging to reduce variance and overfitting while providing interpretable feature-importance estimates (Breiman, 2001). Kernel-based methods, exemplified by the Support Vector Machine, construct a maximum-margin separating hyperplane and, through kernel functions, model non-linearly separable data effectively (Cortes & Vapnik, 1995). Gradient-boosting frameworks build additive ensembles of weak learners in a stage-wise manner: XGBoost provides a scalable, regularized implementation that handles missing data and complex interactions (Chen & Guestrin, 2016), while CatBoost introduces ordered boosting and a principled treatment of categorical features to reduce prediction shift and overfitting (Prokhorenkova et al., 2018). Finally, attention-based deep tabular architectures such as TabNet combine deep learning with built-in, sequential-attention feature selection to improve interpretability on structured data (Arik & Pfister, 2021).

Machine Learning Applications in Diabetes Prediction

A large body of prior work has applied these techniques to diabetes prediction, most frequently using the Pima Indians Diabetes dataset as a benchmark. Studies employing single classifiers consistently report that tree-based and margin-based methods yield reliable, interpretable predictions, with accuracies typically plateauing in the 75-82% range (Khanam & Foo, 2021; Mujumdar & Vaidehi, 2019; Al-Sideiri et al., 2019). Support vector machines, in particular, have been reported to outperform simpler classifiers such as logistic regression on this dataset (Bavkar & Shinde, 2021). Neural-network and deep-learning approaches have also been explored and can achieve competitive precision, recall, and F1-scores, but at the cost of higher computational demand and reduced interpretability; crucially, comparative analyses find that deep models frequently fail to surpass gradient-boosted trees on small, tabular datasets such as Pima, where limited sample sizes constrain representation learning (Jithendra et al., 2023; Theerthagiri et al., 2022; Ayoade et al., 2025).

Ensemble and gradient-boosting methods have generally advanced the state of the art. XGBoost has been applied directly to diabetes risk prediction with strong results (Wardhani & Akbar, 2022), while stacked and hybrid ensembles that combine heterogeneous base learners with meta-learners, sometimes coupled with evolutionary feature selection, have reported further accuracy gains (Tan et al., 2022; Kumar et al., 2023). A broad set of comparative studies confirms that ensembles typically match or exceed individual classifiers, while emphasizing that relative performance is highly dataset-dependent (Bothra, 2021; Mani Nageshwar & Jayapradha, 2022; Almutairi & Abbott, 2023; Rakshith et al., 2022). Two preprocessing concerns recur throughout this literature. First, because diabetic cases are typically outnumbered by non-diabetic cases, class imbalance biases models toward the majority class; the Synthetic Minority Oversampling Technique (SMOTE) generates synthetic minority-class samples by interpolation and is the most widely used remedy (Chawla et al., 2002), improving minority-class recall in several diabetes studies (Kivrak et al., 2024; Chowdhury et al., 2023), although recent work cautions that on small or overlapping medical datasets SMOTE may generate clinically implausible samples and should be validated carefully (Gholampour, 2024). Second, feature selection is used to reduce dimensionality and improve generalization; the Boruta algorithm, a random-forest wrapper that compares real features against randomized shadow features, is a widely adopted all-relevant selection method (Kursa & Rudnicki, 2010). Collectively, recent reviews conclude that imbalance-aware, ensemble-based pipelines are among the most effective and reproducible strategies for diabetes prediction (Kiran et al., 2025; Tasin et al., 2023).

Interpretability, Deployment, and Research Gaps

As machine learning models move toward clinical use, interpretability and accessibility have become as important as predictive accuracy. Model-agnostic explanation frameworks such as SHAP (SHapley Additive exPlanations) attribute a prediction to individual features on a principled, game-theoretic basis (Lundberg & Lee, 2017) and are increasingly paired with diabetes classifiers to expose the influence of variables such as glucose, BMI, and age (Tasin et al., 2023). In parallel, the integration of continuous, real-time data from wearable devices is an active frontier: systematic reviews report that ensemble tree-based models are the most common choice for wearable-based blood-glucose forecasting, though standardization and clinical validation remain open challenges (Ahmed et al., 2023; Kiran et al., 2025). Despite these advances, comparatively few studies translate their trained models into accessible, end-user tools; most report offline metrics without providing a usable interface, which limits real-world clinical and public-health impact. This motivates the web-based deployment adopted in the present study, in which the best-performing model is exposed through a lightweight application that returns an interpretable, probability-based risk category.

Across this body of work, four recurring gaps emerge: (i) reliance on homogeneous datasets, such as the female-only Pima cohort, that limit generalizability; (ii) limited

integration of real-time data from wearable devices; (iii) insufficient attention to model interpretability and accessible deployment in clinical settings; and (iv) inconsistent handling of class imbalance, with over-reliance on resampling techniques whose clinical validity is not always established (Chowdhury et al., 2023; Bi et al., 2019; Gholampour, 2024; Kiran et al., 2025). The present study responds to the first, third, and fourth gaps by applying a consistent, imbalance-aware pipeline across multiple algorithms, by prioritizing an interpretable model, and by deploying that model as a usable web-based screening tool.

MATERIALS AND METHODS

Overview and System Design

The system is an end-to-end machine learning pipeline for diabetes risk prediction. It covers data preprocessing, model training and evaluation, selection and storage of the best model, and web-based deployment for categorical risk assessment shown in Figure 1.

Figure 1. System Architecture of the Diabetes Risk Prediction Application

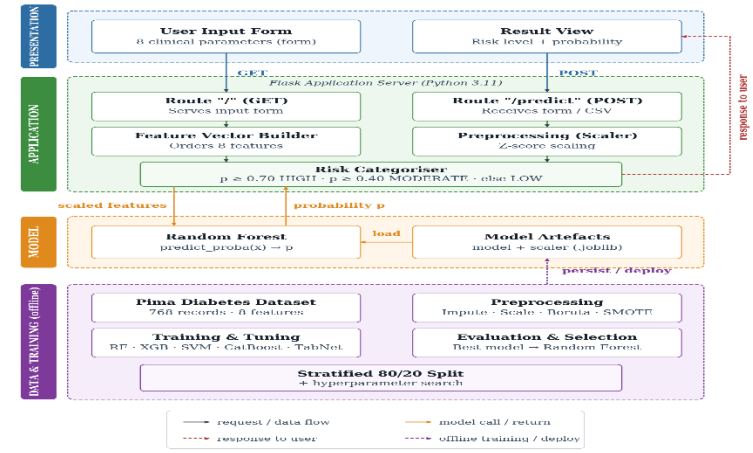

Figure 1: System architecture of the diabetes risk prediction application.

Dataset

The study used the Pima Indians Diabetes Dataset (Smith et al., 1988), comprising 768 records, eight numerical predictors/features (Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, and Age), and a binary outcome indicating diabetic (1) or non-diabetic (0) status. Participants were Pima Indian women aged 21 years and above.

Data Preprocessing

Implausible zero values in Glucose, Blood Pressure, Skin Thickness, Insulin, and BMI were replaced with missing values and median-imputed. Features were standardized using StandardScaler, while Boruta was applied for feature selection (Kursa & Rudnicki, 2010). Class imbalance was addressed using class weighting and SMOTE (Chawla et al., 2002). Glucose-to-BMI ratio and Age × Glucose were added as interaction.

Model Selection and Training

The dataset was divided into stratified training and testing sets using an 80:20 ratio. Grid or random search was used for hyperparameter tuning. The evaluated models were Random Forest, XGBoost, Support Vector Machine, CatBoost, and TabNet.

Evaluation Metrics

Performance was evaluated using accuracy, precision, recall and F1-score. Recall and F1-score were emphasized because of class imbalance and the need to minimize missed diabetic cases.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$\text{Precision} = TP / (TP + FP) \quad (2)$$

$$\text{Recall} = TP / (TP + FN) \quad (3)$$

$$F1 = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

Implementation

The system was developed as a Flask-based web application in Python 3.11. A pre-trained model and StandardScaler were loaded at startup, while the /predict endpoint processed user inputs or CSV uploads to generate diabetes risk probabilities. Risk levels were categorized as High (≥0.70), Moderate (≥0.40), or Low (<0.40), with results and recommendations displayed through the web interface. The application uses Flask, Pandas, NumPy, scikit-learn, CatBoost, PyTorch-TabNet, imbalanced-learn, and joblib, and runs on a standard personal computer without requiring an external database**.**

RESULTS AND DISCUSSION

Experimental Setup

All five (5) models were trained on the SMOTE-balanced, standardized training set and evaluated on the held-out 20% test set comprising 154 records (99 non-diabetic and 55 diabetic instances). The exploratory feature distributions used to inform preprocessing are shown in Figure 2, while the complete set of performance results for all models is consolidated in Table 1.

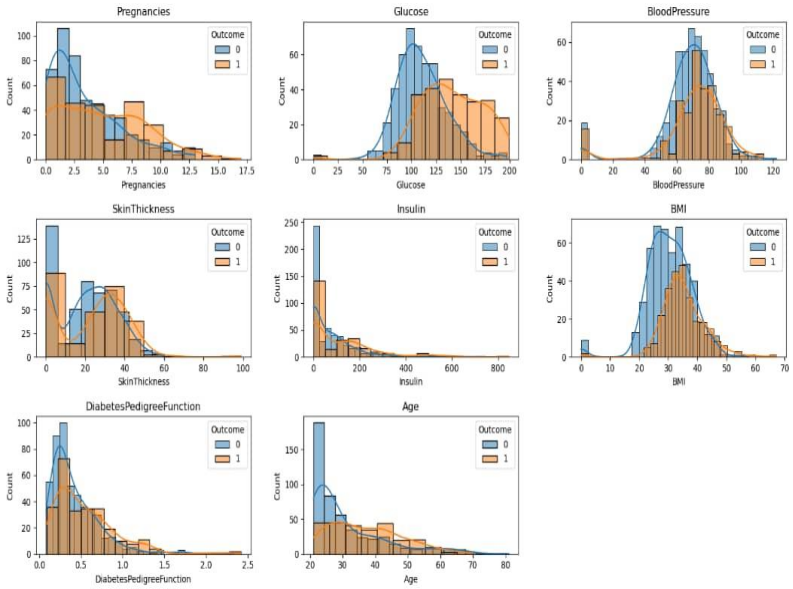

Figure 2: Feature-distribution histograms by outcome class after preprocessing

Comparative Results

Table 1 presents a comprehensive comparison of the five models across accuracy, precision, recall, F1-score, and ROC-AUC. The precision, recall, and F1-score are reported for the diabetic (positive) class, which is the clinically important target in a screening setting. Random Forest achieved the highest values on the aggregate metrics (accuracy 0.753; F1-score 0.689; ROC-AUC 0.810), combining strong recall for diabetic patients (0.76) with the best precision among the well-performing models (0.63). CatBoost ranked second (accuracy 0.714; ROC-AUC 0.806) with the second-highest diabetic recall (0.75). SVM and XGBoost followed closely on accuracy (0.701 and 0.688), but XGBoost achieved this with lower precision (0.55), implying more false-positive alarms, whereas SVM showed comparatively lower recall (0.67), implying more missed diabetic cases. TabNet performed worst by a wide margin: its diabetic-class recall of only 0.15 and ROC-AUC of 0.465 (below the 0.5 chance level) indicate that it failed to identify the great majority of diabetic patients and is therefore unsuitable for screening.

Table 1: Comprehensive Performance Comparison of the Evaluated Models on the Test Set (Precision, Recall, and F1-Score Are for the Diabetic/Positive Class)

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Random Forest	0.753	0.63	0.76	0.689	0.810
CatBoost	0.714	0.58	0.75	0.651	0.806
SVM	0.701	0.57	0.67	0.617	0.781
XGBoost	0.688	0.55	0.73	0.625	0.771
TabNet	0.604	0.36	0.15	0.208	0.465

Note. Accuracy, F1-score, and ROC-AUC are the aggregate metrics reported by the model-comparison results; precision and recall are the diabetic-class values taken from the corresponding classification reports. Higher values indicate better performance for all metrics; the best value in each column is achieved by Random Forest except for recall, where Random Forest (0.76) marginally leads CatBoost (0.75).

Discussion

The comparative results show that Random Forest provided the most appropriate balance of discrimination and case detection in the present experiment. Its accuracy of 0.753, F1-score of 0.689, and ROC-AUC of 0.810 were the highest among the five **(5)** evaluated models, while its diabetic-class recall of 0.76 marginally exceeded CatBoost (0.75) and XGBoost (0.73). This pattern supports the broader conclusion that ensemble methods are strong candidates for diabetes prediction on structured clinical data, but it also demonstrates that the best-performing ensemble can depend on the preprocessing, validation design, and optimization strategy used in a particular study. The Random Forest accuracy obtained here (75.3%) is consistent with the moderate performance range reported for conventional machine-learning models on the Pima Indians Diabetes dataset. Khanam and Foo (2021), using the same 768-record benchmark, found that logistic regression and SVM performed well among seven machine-learning algorithms, while their two-hidden-layer neural network achieved 88.6% accuracy. The present SVM accuracy of 70.1% is lower than the best result in that study, whereas Random Forest performed better than SVM within the current common pipeline. The difference should not be interpreted as a contradiction because Khanam and Foo used k-fold cross-validation and a distinct neural-network configuration, while the present metrics were obtained from one held-out 20% test set after median imputation, feature scaling, Boruta selection, feature engineering, and SMOTE. Thus, the comparison indicates that both model choice and experimental protocol materially influence the reported performance. The present findings also differ from studies in which boosting was superior. Tasin et al. (2023) reported that XGBoost combined with ADASYN produced 81% accuracy, an F1-score of 0.81,

and an AUC of 0.84 after combining the Pima data with 203 additional records from female participants in Bangladesh. In the present study, XGBoost achieved lower values of 68.8% accuracy, 0.625 F1-score, and 0.771 ROC-AUC, while Random Forest was the leading model. The higher XGBoost performance reported by Tasin et al. may reflect their enlarged data source, ADASYN-based balancing, mutual-information feature selection, and different validation configuration. Nevertheless, the two studies agree that imbalance treatment and ensemble learning are important, and both demonstrate the practical value of translating a trained classifier into an accessible web or mobile prediction interface.

A still higher benchmark was reported by Ganie et al. (2023), who compared five boosting algorithms on the Pima dataset and obtained 92.85% accuracy with Gradient Boosting after upsampling, normalization, feature selection, hyperparameter tuning, and k-fold cross-validation. That result exceeds the 75.3% accuracy of the present Random Forest. However, direct ranking across the two studies is not methodologically sound because k-fold cross-validation averages performance over repeated partitions, whereas the present study reports a single held-out split, and the preprocessing and candidate algorithms also differ. This contrast strengthens, rather than weakens, the interpretation of the current results: the reported values are specific to the defined test set and pipeline and should be validated with repeated stratified k-fold cross-validation before any claim of superiority beyond this experiment. The model-level trade-offs are clinically relevant. Random Forest's recall of 0.76 means that it detected more diabetic cases than the other models in this test, while its precision of 0.63 was also the best among the four competitive models. CatBoost was a close alternative, with ROC-AUC of 0.806 and recall of 0.75, suggesting that the performance gap between the two tree ensembles is small and may not persist under resampling. SVM produced fewer false-positive predictions than XGBoost only to a limited extent, but its lower recall of 0.67 implies more missed diabetic cases. Conversely, XGBoost's recall of 0.73 was achieved with precision of 0.55, indicating a larger follow-up burden from false-positive alerts. In screening, where a false negative can delay confirmatory testing, the higher recall of Random Forest and CatBoost is preferable, although threshold selection should ultimately be based on clinical costs rather than the default classification threshold alone. TabNet's results require particular caution. Its recall of 0.15 and ROC-AUC of 0.465 show that it did not learn a useful decision boundary for the diabetic class in this experiment. This outcome is compatible with evidence that deep-learning models do not consistently outperform tree-based methods on small tabular datasets and that performance depends strongly on dataset characteristics and tuning (Ayoade et al., 2025). With only 768 observations, the training partition may have been inadequate for stable optimization of the attention-based architecture. However, the result does not establish that TabNet is intrinsically unsuitable for diabetes prediction; it establishes only that, under the present dataset, preprocessing, and tuning configuration, it was substantially inferior to the four conventional models.

Overall, the comparison with previous authors places the present findings within, rather than above, the existing evidence. The Random Forest result is credible for a small, imbalanced benchmark and offers a favourable recall-precision balance, but it is lower than several studies that used cross-validation, additional observations, alternative resampling, or more extensive optimization. Random Forest was therefore selected for deployment because it was the best model under the study's controlled pipeline, not because it can be claimed to be universally superior. External validation on demographically diverse cohorts, probability calibration, decision-curve analysis, and repeated stratified cross-validation are necessary before the model's screening benefit or generalizability can be established.

Deployment and Web-Based User Interface

The best-performing model (Random Forest) was integrated into a lightweight Flask web application that provides users with a probability-based diabetes risk assessment, categorising each prediction as low, moderate, or high risk. The deployment follows the system architecture in Figure 1 and the client-server flow. As shown in Figure 3, the input screen presents a simple HTML form in which the user enters the eight clinical and lifestyle parameters (Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, and Age), or alternatively uploads a patient CSV file for the same fields. On submission, the request is sent by HTTP POST to the /predict route, where the server reconstructs the feature vector in the model's expected order, applies the fitted StandardScaler, and calls the model to obtain the probability of diabetes.

Figure 3: Web-based user interface: input form for entering patient clinical and lifestyle parameters

The predicted probability is then mapped to an interpretable risk category using the fixed thresholds defined as: probabilities of at least 0.70 are reported as High Risk; at least 0.40 as Moderate Risk; and otherwise Low Risk. As illustrated in Figure 4, the result screen displays the predicted risk level, the underlying probability, a colour-coded

probability bar, and a tailored recommendation. For example, a probability of 28% is reported as Low Risk with the advice to 'maintain healthy habits and regular checkups'; 44.5% is reported as Moderate Risk advising lifestyle changes and monitoring; and 93% is reported as High Risk advising the user to 'consult a healthcare professional immediately'. By translating raw model output into a clear, colour-coded risk category and actionable guidance, the application makes the model accessible to non-technical users and demonstrates the feasibility of integrating it into mobile or telemedicine platforms, while remaining, by design, a screening aid rather than a diagnostic replacement.

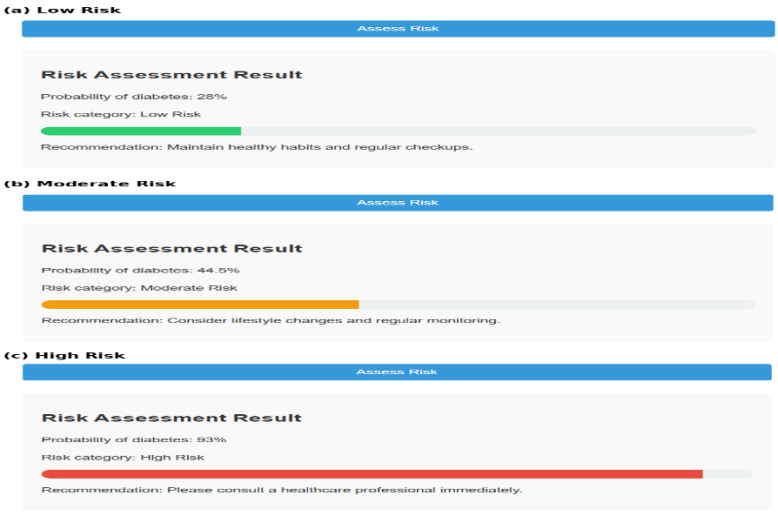

Figure 4: Web-based user interface: probability-based risk outputs categorised as (a) low, (b) moderate, and (c) high risk, each with a colour-coded probability bar and recommendation

Limitations

Several limitations qualify these findings. First, the Pima Indians Diabetes dataset is demographically homogeneous (adult females of a single heritage), which limits generalizability to broader populations. Second, despite SMOTE and class weighting, residual class imbalance may still influence results, and the clinical validity of synthetic oversampling on small datasets remains debated (Gholampour, 2024). Third, the feature set omits potentially informative genetic, environmental, and lifestyle variables. Fourth, the reported metrics derive from a single train-test split; k-fold cross-validation would provide more stable estimates. Finally, the deployment lacks external clinical validation and regulatory assessment, so the tool should be regarded as a screening aid rather than a diagnostic replacement.

CONCLUSION

This study developed and evaluated a machine learning system for diabetes risk prediction using clinical and lifestyle data from the Pima Indians Diabetes dataset. Under a common imbalance-aware preprocessing pipeline, five (5) algorithms were compared across accuracy, precision, recall, F1-score, and ROC-AUC, and Random Forest delivered the best overall performance (accuracy 0.753; F1-score 0.689; ROC-AUC 0.810), followed by CatBoost, SVM, and XGBoost, while TabNet performed poorly on the small dataset. The best model was deployed in a lightweight Flask web application that returns an interpretable low, moderate, or high-risk category with a supporting probability and recommendation, demonstrating a practical, scalable, and low-cost approach to early diabetes screening. Future work should validate the models on larger and more demographically diverse datasets; incorporate real-time data from wearable devices such as continuous glucose monitors; apply explainable AI techniques such as SHAP to further improve transparency; adopt k-fold cross-validation for more stable and rigorous evaluation; and pursue clinical and regulatory validation prior to real-world deployment. Periodic retraining and performance monitoring are also recommended to maintain accuracy as new data become available.

REFERENCES

Ahmed, A., Aziz, S., Abd-alrazaq, A., Farooq, F., Househ, M., & Sheikh, J. (2023). The effectiveness of wearable devices using artificial intelligence for blood glucose level forecasting or prediction: Systematic review. Journal of Medical Internet Research, 25, e40259. https://doi.org/10.2196/40259

Al-Sideiri, A., Cob, B. C., & Drus, S. B. M. (2019). Machine learning algorithms for diabetes prediction. https://doi.org/10.1145/3388218.3388231

Almutairi, E. S., & Abbott, M. F. (2023). Machine learning methods for diabetes prevalence classification in Saudi Arabia. Modelling, 4(1), 37-55. https://doi.org/10.3390/modelling4010004

Arik, S. O., & Pfister, T. (2021). TabNet: Attentive interpretable tabular learning. Proceedings of the AAAI Conference on Artificial Intelligence, 35(8), 6679-6687. https://doi.org/10.1609/aaai.v35i8.16826

Ayoade, O. B., Shahrestani, S., & Ruan, C. (2025). Machine learning and deep learning approaches for predicting diabetes progression: A comparative analysis. Electronics, 14(13), 2583. https://doi.org/10.3390/electronics14132583

Bavkar, V. C., & Shinde, A. A. (2021). Machine learning algorithms for diabetes prediction and neural network method for blood glucose measurement. Indian Journal of Science

and Technology, 14(10), 869-880. https://doi.org/10.17485/IJST/v14i10.2187

Bi, Q., Goodman, K. E., Kaminsky, J., & Lessler, J. (2019). What is machine learning? A primer for the epidemiologist. American Journal of Epidemiology, 188(12), 2222-2239. https://doi.org/10.1093/aje/kwz189

Bothra, R. (2021). Diabetes prediction using machine learning algorithms. International Journal of Engineering Applied Sciences and Technology, 6(5).
Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321-357. https://doi.org/10.1613/jair.953

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794). https://doi.org/10.1145/2939672.2939785

Chowdhury, M. M., Ayon, R. S., & Hossain, M. S. (2023). An investigation of machine learning algorithms and data augmentation techniques for diabetes diagnosis using class imbalanced BRFSS dataset. Healthcare Analytics, 5, 100297. https://doi.org/10.1016/j.health.2023.100297

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273-297. https://doi.org/10.1007/BF00994018

Ganie, S. M., Pramanik, P. K. D., Malik, M. B., Mallik, S., & Qin, H. (2023). An ensemble learning approach for diabetes prediction using boosting techniques. Frontiers in Genetics, 14, 1252159. https://doi.org/10.3389/fgene.2023.1252159

Gholampour, S. (2024). Impact of nature of medical data on machine and deep learning for imbalanced datasets: Clinical validity of SMOTE is questionable. Machine Learning and Knowledge Extraction, 6(2), 827-841. https://doi.org/10.3390/make6020039

Jithendra, V., Sai Mohit, R. M., Kusuma, S., Madhusudhan, M., & Jagadeesh, B. (2023). Diabetes prediction using machine learning techniques. Journal of Artificial Intelligence and Capsule Networks, 5(2), 190-206. https://doi.org/10.36548/jaicn.2023.2.008

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260. https://doi.org/10.1126/science.aaa8415

Khanam, J. J., & Foo, S. Y. (2021). A comparison of machine learning algorithms for diabetes prediction. ICT Express, 7(4), 432-439. https://doi.org/10.1016/j.icte.2021.02.004

Kiran, M., Xie, Y., Anjum, N., Ball, G., Pierscionek, B., & Russell, D. (2025). Machine learning and artificial intelligence in type 2 diabetes prediction: A comprehensive 33-year bibliometric and literature analysis. Frontiers in Digital Health, 7, 1557467. https://doi.org/10.3389/fdgth.2025.1557467

Kivrak, M., Avci, U., Uzun, H., & Ardic, C. (2024). The impact of the SMOTE method on machine learning and ensemble learning performance results in addressing class imbalance in data used for predicting total testosterone deficiency in type 2 diabetes patients. Diagnostics, 14(23), 2634. https://doi.org/10.3390/diagnostics14232634

Kumar, S., Jaiswal, M., & Tiwari, H. (2023). Diabetes prediction using optimization techniques with machine learning algorithms. International Journal of Electronic Healthcare, 13(2), 158. https://doi.org/10.1504/IJEH.2023.130515

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. Journal of Statistical Software, 36(11), 1-13. https://doi.org/10.18637/jss.v036.i11

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems 30 (pp. 4765-4774).

Mahesh, B. (2020). Machine learning algorithms: A review. International Journal of Science and Research (IJSR), 9(1), 381-386. https://doi.org/10.21275/ART20203995

Mani Nageshwar, & Jayapradha, S. (2022). Diabetes prediction using machine learning algorithms. In International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 46-51). https://doi.org/10.1109/ICACCS54159.2022.9785073

Mujumdar, A., & Vaidehi, V. (2019). Diabetes prediction using machine learning algorithms. Procedia Computer Science, 165, 292-299. https://doi.org/10.1016/j.procs.2020.01.047

Okikiola, F. M., Adewale, O. S., & Obe, O. O. (2023). A diabetes prediction classifier model using Naive Bayes algorithm. FUDMA Journal of Sciences, 7(1), 253–260. https://doi.org/10.33003/fjs-2023-0701-1301

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. In Advances in Neural Information Processing Systems 31 (pp. 6638-6648).

Rakshith, R., Varun, R., & Jagadeesh, M. G. (2022). A comparative analysis of diabetes prediction models using machine learning algorithms (pp. 261-265).

Rufai, M. A., Abdullahi, M. B., Abisoye, O. A., & Ojerinde, O. A. (2024). Utilizing a fusion of machine learning techniques for diabetes mellitus subtypes classification and identification. FUDMA Journal of Sciences, 8(3), 331–343. https://doi.org/10.33003/fjs-2024-0803-2510

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. SN Computer Science, 2(3), 160. https://doi.org/10.1007/s42979-021-00592-x

Smith, J. W., Everhart, J. E., Dickson, W. C., Knowler, W. C., & Johannes, R. S. (1988). Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In Proceedings of the Symposium on Computer Applications and Medical Care (pp. 261-265). IEEE Computer Society Press.

Tan, Y., Chen, H., Zhang, J., Tang, R., & Liu, P. (2022). Early risk prediction of diabetes based on GA-stacking. Applied Sciences, 12(2), 632. https://doi.org/10.3390/app12020632

Tasin, I., Nabil, T. U., Islam, S., & Khan, R. (2023). Diabetes prediction using machine learning and explainable AI techniques. Healthcare Technology Letters, 10(1-2), 1-10. https://doi.org/10.1049/htl2.12039

Theerthagiri, P., Ruby, A. U., & Vidya, J. (2022). Diagnosis and classification of diabetes using machine learning algorithms. SN Computer Science, 4(1).

Wardhani, K. D. K., & Akbar, M. (2022). Diabetes risk prediction using extreme gradient boosting (XGBoost). Jurnal Online Informatika, 7(2), 244-250. https://doi.org/10.15575/join.v7i2.970

World Health Organization. (2024). Diabetes [Fact sheet]. https://www.who.int/news-room/fact-sheets/detail/diabetes