# Improving Diabetes Prediction Accuracy and Interpretability with SMOTE and SHAP

**Wajeeha Iftikhar**
Faculty of Computing, Riphah International University, Lahore, Pakistan
wajeehaiftikhar47@gmail.com

**Muhammad Yaseen**
Faculty of Computing, Riphah International University, Lahore, Pakistan
muhammad.yaseen@riphah.edu.pk

**Gohar Rahman**
Faculty of Computing and Informatics, University Malaysia Sabah (UMS), Kota Kinabalu, Sabah, Malaysia
gohar_315@ums.edu.my (corresponding author)

**Muhammad Asif Nauman**
Faculty of Computing, Riphah International University, Lahore, Pakistan
asif.nauman@riphah.edu.pk

**Umar Farooq Khattak**
Faculty of Artificial Intelligence and Frontier Technologies, UNITAR International University, Selangor, Malaysia
umar.farooq@unitar.my

*Received: 13 November 2025 | Revised: 25 December 2025 | Accepted: 12 January 2026*

## ABSTRACT

**Diabetes can cause a lot of serious health problems, and its early detection is very important. This study proposes a hybrid machine learning framework that enhances diabetes prediction accuracy and interpretability by combining the Synthetic Minority Over-Sampling Technique (SMOTE) and Shapley Additive Explanations (SHAP), examining five machine learning models—Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), and XGBoost—on two datasets. The results showed that RF and XGBoost achieved the highest predictive performance. SMOTE improved class balance and model robustness, while SHAP provided transparent explanations of key predictors such as glucose, BMI, and age. The proposed approach demonstrates that the combination of SMOTE and SHAP enhances both the reliability and the interpretability of models for practical diabetes prediction.**

***Keywords-diabetes prediction; machine learning; random forest; SMOTE; Explainable AI (XAI); SHAP***

## I. INTRODUCTION

Diabetes is a widespread health issue that affects millions of people around the world and puts patients at risk of numerous comorbidities, such as cardiovascular disease, renal failure, and neuropathy. Early detection and strict disease control are the most critical in maximizing clinical outcomes, as they prevent or delay the development of serious complications. Recent advances in the healthcare domain involve the integration of Machine Learning (ML) algorithms to effectively predict diabetes, thus allowing early detection and development of personalized treatment plans.

ML models have high predictive accuracy in the diagnosis of diabetes, but often cannot be interpreted. In practice, it is necessary in clinical settings to explain how and why a model makes a specific prediction to gain the confidence of clinical practitioners to receive actionable knowledge. Previous studies have demonstrated the efficacy of ML models [1-3], but the lack of explainability remains a significant problem for their use in clinical practice. This has led to a strong need to develop comprehensible models that balance predictive accuracy with explainability. The imbalance in classes in datasets, including the PIMA Indian Diabetes dataset, has been another problem that has challenged the prediction of diabetes through ML. Non-diabetic cases far outnumber the diabetic ones, leading to biased predictions that reduce the ability of the model to accurately predict cases that fall in the minority sample. Previous studies have highlighted the impact of class imbalance on predictive accuracy [4, 5], but methods such as the Synthetic Minority Over-Sampling Technique (SMOTE) can reduce this problem. SMOTE is an artificial sampling technique that creates samples in the minority class, thus improving the performance of the model and making its predictions more reliable.

This study aimed to improve diabetes prediction by addressing both interpretability and class imbalance. The proposed method uses SMOTE to tackle class imbalance and SHAP (Shapley Additive Explanations) to improve the interpretability of ML models. SHAP is an effective tool for explaining the decisions made by a model [3, 6], allowing clinicians to understand the impact of features on the predictions. Combining these methods can lead to models that perform better in terms of predictive accuracy and provide clear and practical information that medical professionals can rely on. This study focused on a comparative analysis of five ML models, Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), and XGBoost, in terms of predictive performance and interpretability.

## II. RELATED WORK

Diabetes prediction has gained significant attention, and numerous researchers have investigated ML algorithms for this purpose. In [7], ML was used to predict diabetes through patient data. This study highlighted the practical use of predictive models in clinical practice, underscoring the importance of early intervention and individual treatment strategies based on ML predictions. LR is a common technique in healthcare studies because of its simplicity and interpretation, but it becomes weak when there are nonlinear relationships between predictors, limiting its effectiveness in complex diagnostic tasks such as detecting diabetes [8]. Empirical data show that LR is not as practical due to its inability to perform well against more advanced algorithms, such as RF or XGBoost, especially when dealing with nonlinear data patterns [9]. Although it is a simple method, numerous studies employ LR as a comparative model.

RF performs better in diabetes prediction, as it is suitable for dealing with nonlinear and feature interactions. In [10], RF achieved 82% accuracy, outperforming other models such as LR. In [11], RF and XGBoost were used with SMOTE to deal with class imbalance and improve prediction performance. Integrating ML models with techniques that handle class imbalance and explainability methods has become essential in the prediction of diabetes. XGBoost is a gradient boosting algorithm that has become popular because it can operate with big data and possesses high predictive capabilities, especially when applied in healthcare settings. In [12], XGBoost was used with SMOTE to overcome the issue of class imbalance, incorporating Explainable AI (XAI) to improve model transparency. The results of this study show that these methods can help achieve more accurate and interpretable diabetes prediction models, which would be more effective for clinical decision-making. XGBoost and RF can offer a good balance between accuracy, interpretability, and performance [7, 11]. XGBoost outperformed other models in [13], with an Area Under the Curve (AUC) of 0.963. SVM is known to be strong in the classification of high-dimensional data, but it faces the problem of interpretability and hyperparameter optimization. Previous studies have shown that SVM can be valuable in predicting diabetes, though with lower accuracy compared to RF and XGBoost.

Class imbalance is a common issue in diabetes prediction data sets, where the non-diabetic cases outnumber the diabetic ones significantly. To address such an imbalance, SMOTE can balance datasets by creating samples belonging to minority classes. Empirical evaluations have demonstrated the promising effect of SMOTE on model performance by enabling a more balanced class distribution [11, 12]. In [14], a sequential data preprocessing pipeline employed SMOTE to reduce bias and enhance the strength of the diabetes prediction models.

New developments in the prediction of diabetes have highlighted the need to not only optimize model performance but also to make them interpretable so that they can be clinically adopted. In [15], an interpretable ML framework combined SMOTE to balance the dataset and SHAP to improve the model explainability. SMOTE with SHAP has proven to be a significant approach to enhancing prediction accuracy and transparency in a range of healthcare applications. In [16], SMOTE and SHAP were used to predict chronic kidney disease, showing that using both methods could improve the model performance and explainability. The results of this study are easily applicable to diabetes prediction, as they highlight the importance of such methods to simplify complicated models and make them more reliable and comprehensible for clinicians. This will not only enhance the accuracy of prediction models but also their reliability in clinical decision-making, aligning with the objectives of developing reliable and transparent diabetes prediction models.

XAI tools, including SHAP, Partial Dependency, and LIME, have been identified as useful in improving the explainability of a model [17]. Interpretability is an important aspect of ML applications in healthcare because clinicians need an idea of the determinants that shape predictive results. The effectiveness of SHAP in explaining the behavior of a model has been confirmed in several studies, e.g. [18], demonstrating the potential to gain the confidence of clinicians. In addition, ensemble methods such as stacking have also been used to increase the predictive power of ML models. In [19], SHAP was used to clarify the decision-making process of an ensemble model. In [20], a stacking ensemble of RF and XGBoost was applied to improve the accuracy of diabetes predictions, highlighting the importance of using SHAP to improve the interpretability of the model's decision-making process.

In summary, ML models such as LR, SVM, RF, and XGBoost have been applied to diabetes prediction, with RF and XGBoost demonstrating better accuracy and robustness. However, issues with interpretability have prompted the integration of SHAP and other XAI methods to clarify model decisions. In addition, techniques such as SMOTE have been widely used to address class imbalance, improving the reliability of diabetes prediction models. Combining these techniques has significantly improved the predictive ability and transparency of ML models in healthcare settings.

## III. RESEARCH METHODOLOGY

The research method of this study is structured into four main phases: Data Preprocessing, Model Training, Model Evaluation, and Explainability Analysis. Two publicly available datasets were used: the PIMA Indians Diabetes Dataset [21] and the Diabetes Prediction Dataset [22].

### A. *Dataset Description*

Table I describes the datasets used in this study.

TABLE I. DATASET DESCRIPTION

| Characteristic | PIMA Indians | Diabetes Prediction |
|---|---|---|
| Samples | 768 | 1000 |
| Features | 8 predictors + 1 target | 13 predictors + 1 target |
| Key features | Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, DiabetesPedigreeFunction, Age | Gender, AGE, Urea, Cr, HbA1c, Chol, TG, HDL, LDL, VLDL, BMI |
| Target variable | *Outcome* (0: Non-Diabetic, 1: Diabetic) | *CLASS* (N: Non-Diabetic, Y: Diabetic, P: Prediabetes) |
| Source | [21] | [22] |

### B. *Data Preprocessing*

Non-predictive parameters were removed from the clinical datasets. Categorical values (Gender: M/F) were encoded using Label Encoder. For the PIMA dataset [21], zeros in clinical features (Glucose, BMI, and Insulin) were handled using the median imputation technique, while the Diabetes Prediction Dataset [22] did not contain missing values. All numerical features were normalized using StandardScaler to ensure that each feature contributed equally during model training. Stratified random sampling was used to divide each dataset into an 80:20 train-test split. SMOTE was applied only to the training data after splitting to avoid any data leakage. This process generates synthetic samples for the minority class, ensuring balanced class distributions before model training.

### C. *Model Training*

Five ML models were trained and evaluated on both datasets to perform a comparative performance analysis. These include LR, which is an interpretable linear baseline model, SVM, a nonlinear classifier implementing the Radial Bias Function (RBF) kernel for handling complex boundaries, DT classifier, a simple but powerful tree based model that is capable of capturing nonlinear relationships, RF classifier, an ensemble model including 100 DTs with unrestricted depth (*max_depth*=none) to improve predictive stability and reduce variance, and XGBoost, which involved training a gradient boosted tree algorithm with a learning rate of 0.1, 100 estimators and maximum depth of 6 to achieve high and robust results.

### D. *Model Evaluation*

Once the models were trained, their performance was evaluated using several metrics to ensure that the results they provide are reliable and accurate predictions [23]. The evaluation metrics used in this study are given below:

- Accuracy measures the overall correctness of the model. It can be defined as the ratio of cases correctly predicted (both true positives and true negatives) to the total number of instances.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{1}$$

where TP is True Positives, TN is True Negatives, FP is False Positives, and FN is False Negatives.

- Precision measures the accuracy of a model's positive predictions, calculated as the ratio of TP to the total number of positive instances. It is particularly important when the FP cases are high.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{2}$$

- Recall (Sensitivity) is the proportion of the observed positive values in the actual class that were correctly predicted, and is extremely important when the FN cases are high.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{3}$$

- F1-score is the harmonic mean of precision and recall, balancing both metrics.

$$\text{F1} - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

- AUC-ROC (Area under the Curve- Receiver Operating Characteristic) evaluates the model's ability to differentiate between the positive and negative classes. It represents the area under the ROC curve, which plots the True Positive Rate (TPR) against the False Positive Rate (FPR). AUC is a summary metric that provides an aggregate measure of a model's performance across all possible classification thresholds.

$$\text{TPR(Recall)} = \frac{\text{TP}}{\text{TP+FN}}\ ,\quad \text{FPR} = \frac{\text{FP}}{\text{FP+TN}} \tag{5}$$

### E. *Explainability Analysis*

SHAP was used to explain the contributions of global and local features. Local SHAP summary plots were generated for each dataset to identify the most influential predictors contributing to diabetes prediction across the whole population. Moreover, local SHAP waterfall plots were used to visualize the impact of features on patient predictions, showing how specific feature values impact the performance of the model for diabetic and non-diabetic classification. In both datasets, HbA1c, Glucose, and BMI consistently appeared as the most prominent predictors, showing their role in affecting diabetes outcomes and validating the clinical relevance of the model's interpretability.

### *F. Experimental Setup*

All experiments were performed using Python 3.10, scikit-learn 1.4.1, imbalanced learn 0.10.1, XGBoost 2.0.3, and SHAP 0.46.0 on a standard workstation (Intel i7, 16 GB RAM).

## IV. RESULTS AND DISCUSSION

This section describes the experimental results from the PIMA Dataset [21] and the Diabetes Prediction Dataset [22]. Each dataset was processed, balanced, and analyzed using the same ML workflow to maintain consistency. The results are discussed in terms of model performance, statistical validation, and interpretability.

Five ML models were trained and tested on the PIMA dataset after performing data preprocessing and balancing of classes using SMOTE. Table II summarizes their performance.

TABLE II. MODEL PERFORMANCE ON THE PIMA DATASET [21]

| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-score** | **AUC** |
|---|---|---|---|---|---|
| LR | 0.72 | 0.59 | 0.68 | 0.63 | 0.82 |
| SVM | 0.69 | 0.55 | 0.65 | 0.59 | 0.78 |
| DT | 0.68 | 0.55 | 0.52 | 0.53 | 0.64 |
| RF | 0.72 | 0.60 | 0.63 | 0.61 | 0.83 |
| XGBoost | **0.75** | **0.64** | **0.68** | **0.66** | **0.81** |

The XGBoost classifier was superior to other models, achieving the highest overall accuracy (75%) and performing well across all evaluation metrics. The results of RF were similar, confirming the strength of tree-based ensemble methods in the context of structured biomedical data. Their small improvements over the linear and kernel-based models highlight the importance of the nonlinear interaction of features in the diabetes prediction. Lastly, their performance range (68-75%) is in line with the results of similar works on similar datasets, pointing to the lack of overfitting or information leakage.

The Diabetes Prediction Dataset [22] was tested with the same models using SMOTE only on the training data. The data was more complex with three target classes, namely Non-Diabetic, Prediabetes, and Diabetic. Table III presents the results.

TABLE III. MODEL PERFORMANCE ON CLINICAL DATASET [22]

| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-score** | **AUC (OvR)** |
|---|---|---|---|---|---|
| LR | 0.85 | 0.83 | 0.82 | 0.82 | 0.88 |
| SVM | 0.87 | 0.85 | 0.84 | 0.84 | 0.90 |
| DT | 0.88 | 0.87 | 0.86 | 0.86 | 0.91 |
| RF | 0.92 | 0.91 | 0.90 | 0.90 | 0.94 |
| XGBoost | **0.94** | **0.93** | **0.91** | **0.92** | **0.95** |

XGBoost again outperformed the other models, achieving an accuracy of 94%, followed by RF with 92%. The ability of XGBoost and RF to model nonlinear relationships and complex feature interactions can be seen as the cause for their superior generalization capabilities. The accuracies of LR and SVM were somewhat less; however, they maintained good discriminative ability, as shown by their AUC (>0.85). The consistent AUC improvement across all models confirms the effectiveness of SMOTE in balancing class distributions without overfitting.

### *A. Statistical Validation*

A paired t-test was conducted on the RF and XGBoost classifiers on both datasets to determine whether the differences between the model performances were statistically significant. In the PIMA Dataset, the test resulted in $t = -0.277$ with a p-value of 0.783, which is not considered significantly different. The same pattern was replicated with the Clinical Predictions Dataset ($p > 0.05$), demonstrating that the two ensemble methods had the same behavior. These findings indicate that the differences in accuracy could be explained by random error and not by a significant model instability, which validates the same level of stability across datasets.

### *B. Confusion Matrix Analysis*

To gain a more detailed insight into the classification performance demonstrated by each model, confusion matrices were examined for the best-performing algorithms, namely, RF and XGBoost, on both datasets. For the PIMA Dataset, both RF and XGBoost correctly identified the majority of diabetic and non-diabetic instances, with XGBoost slightly reducing FN. In the Diabetes Prediction Dataset, the same pattern was observed, where the ensemble models achieved near-perfect classification for the predominant class and correctly distinguished pre-diabetic and diabetic categories with minimal misclassifications. These results confirm that the ensemble models are robust and capable of handling heterogeneous, imbalanced medical data while maintaining high discriminative accuracy. In addition, the confusion matrices revealed that the likelihood of misclassification was minimal, which undermined the fact that the class-balancing strategy, applied through SMOTE, was efficient and did not give rise to noise and bias in the training process.

### *C. Explainability Analysis*

To enhance interpretability, XGBoost, which was the best-performing classifier, was studied with SHAP. SHAP is a single model through which complex models can be explained by measuring the impact of every feature on the individual predictions and model behavior. This double-staged explainability can provide a global measure of feature significance and a local explanation of the effect of particular features on classification.

Each dataset was plotted by creating a global SHAP summary plot to establish the most powerful predictors in all cases (Figure 2). In the PIMA Indians Diabetes Dataset [21], the key contributors were Glucose, BMI, and Age, which supports the known medical information that an increased level of glucose in the body and increased body mass index level are the strong predictors of diabetes risk. On the other hand, HbA1c, BMI, and LDL were the most prominent features identified in the Diabetes Prediction Dataset [22], which demonstrates the central role of glycemic control and lipid metabolism in the diagnosis of clinical diabetes.

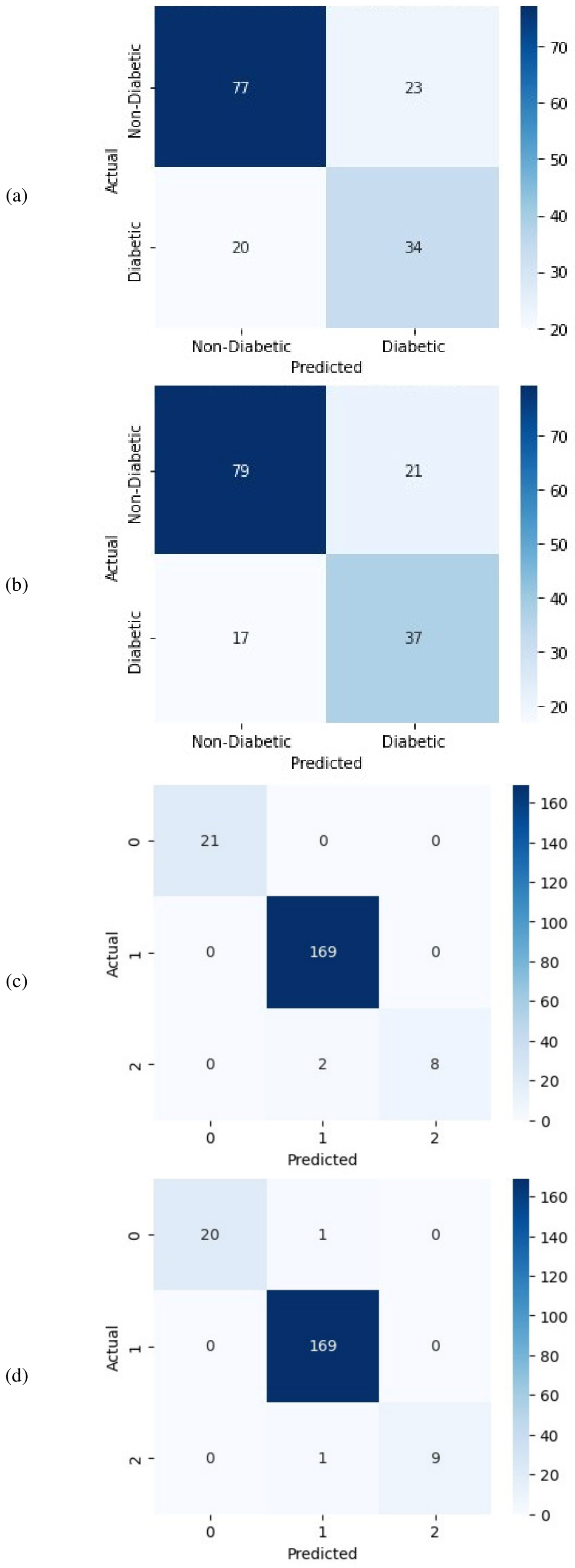

Fig. 1. Confusion matrices: (a) RF on [21], (b) XGBoost on [21], (c) RF on [22], (d) XGBoost on [22].

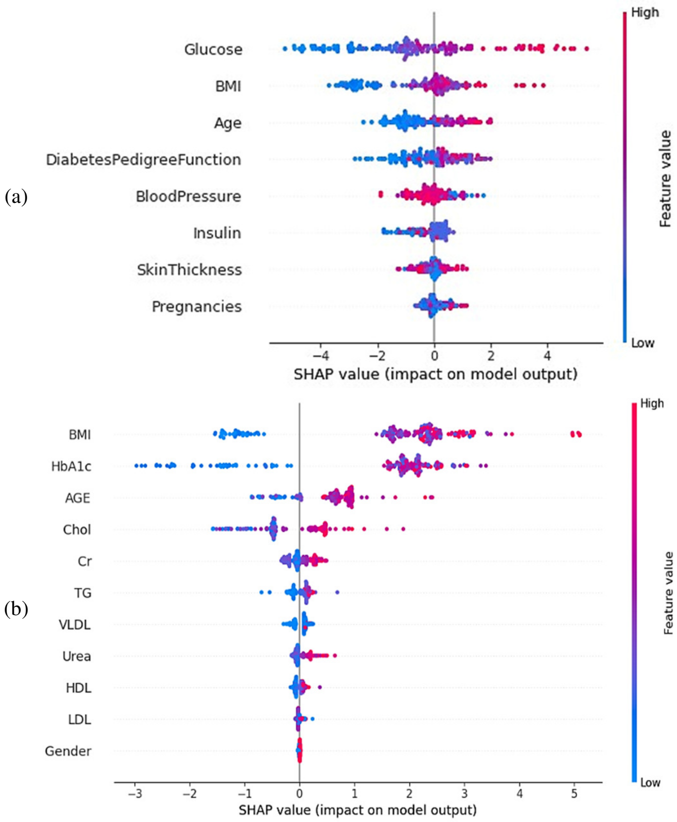

Fig. 2. Global SHAP summary plots for XGBoost on (a) [21], (b) [22].

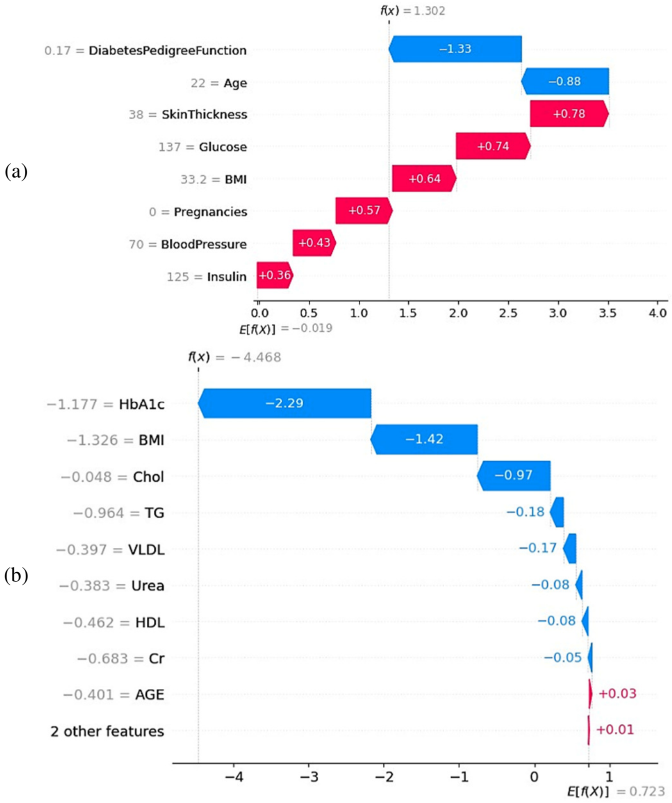

Fig. 3. SHAP waterfall plots for XGBoost on: (a) [21], (b) [22].

Local SHAP waterfall plots were then used to explain how one single feature value affected single-patient predictions (Figure 3). These graphs demonstrate that specific variables, e.g., high HbA1c or Glucose, increase the probability of a positive classification of diabetes, but other variables, e.g., normal BMI or HDL level, decrease it. Color-coded bars represent the direction and magnitude of the contribution of each feature, offering an intuitive representation of the decision-making process of the model. In general, SHAP analysis showed that the decision-making process of the XGBoost model can be clinically explained and follows the known medical logic. In both datasets, the metrics for glycemia and body mass prevailed, confirming the accuracy and transparency of the model. With added SHAP-based explainability, the study makes sure that a clinician is more likely to trust the model outputs, bridging the gap between ML predictions and clinical interpretability and enabling informed decision-making in the healthcare context.

## V. CONCLUSION AND FUTURE WORK

This study presented an ML framework that combines SMOTE and SHAP to improve both the predictions and the interpretability of diabetes classification models. The proposed framework was tested using two benchmark datasets [21, 22]. The findings indicate that ensemble-based models, such as XGBoost and RF, performed better in all measured metrics. In the PIMA Dataset [21], XGBoost achieved an accuracy of 75% and an AUC of 0.81, while in the Diabetes Prediction Dataset [22], its accuracy and AUC was 94% and 0.95, respectively.

A paired t-test reinforced the performance results of the ensemble models, indicating model stability and robustness. The application of SMOTE effectively addressed the class imbalance problem, which helped to enhance model generalization without overfitting. At the same time, the explainability provided through SHAP improved predictive transparency, as it was able to identify key risk factors, including HbA1c, glucose, and body mass index (BMI). This combined strategy not only enhanced the diagnostic accuracy but also enhanced the reliability and clinical explainability of the models, which is highly critical in the implementation of AI-based tools in healthcare.

### A. *Future Work*

Future research in this field should focus on improving the performance of models and developing practical, real-world applications. To begin with, it would be necessary to test the proposed framework on larger and more heterogeneous datasets, including Electronic Health Records (EHR) or inter-clinic clinical trials, to determine its effectiveness in various populations. Deep learning models, such as neural networks, can help improve predictive accuracy, especially in situations with large amounts of data. Moreover, multi-modal integration of information, such as medical images, genome profiles, and sensor data, could provide a better picture of diabetes risk factors and enhance predictive performance. Continuous monitoring and early diagnosis of diabetes can be conducted with the help of real-time prediction systems such as wearable health devices. Handling bias and enhancing fairness in the model would allow making relatively fair predictions in demographic subgroups, and cooperation with medical professionals would assist in converting the output of the model into actionable medical interventions. All these developments can significantly increase the predictive capabilities, explainability, and practical utility of ML models in healthcare.

## REFERENCES

[1] U. Allani, "Interactive Diabetes Risk Prediction Using Explainable Machine Learning: A Dash-Based Approach with SHAP, LIME, and Comorbidity Insights." arXiv, 2025, https://doi.org/10.48550/ARXIV.2505.05683.

[2] V. Bardia and E. Sophiya, "Diabetes Prediction Using Machine Learning Algorithm: A Comparative Analysis," in *2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Mar. 2024, pp. 1973–1979, https://doi.org/10.1109/ICACCS60874.2024.10717264.

[3] S. Mahmud, B. U. Islam, N. Haque Anik, and T. Ghosh, "Diabetes Prediction: A Comparative Analysis of Machine Learning Algorithms with SMOTE," in *2024 IEEE International Conference on Computing, Applications and Systems (COMPAS)*, Sept. 2024, pp. 1–6, https://doi.org/10.1109/COMPAS60761.2024.10796405.

[4] M. Abdelaoui, "Analysis of the diabetes dataset using a SMOTE machine learning approach," *Studies in Engineering and Exact Sciences*, vol. 5, no. 2, 2024, Art. no. e12076, https://doi.org/10.54021/seesv5n2-772.

[5] R. Taher, S. H. Basha, and A. Abdalla, "Improving Machine Learning Techniques with Imbalanced Data Treatment for Predicting Diabetes," in *Proceedings of the 9th International Conference on Advanced Intelligent Systems and Informatics 2023*, 2023, pp. 380–391, https://doi.org/10.1007/978-3-031-43247-7_34.

[6] M. Alghamdi, M. Al-Mallah, S. Keteyian, C. Brawner, J. Ehrman, and S. Sakr, "Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: The Henry Ford ExercIse Testing (FIT) project," *PLOS ONE*, vol. 12, no. 7, 2017, Art. no. e0179805, https://doi.org/10.1371/journal.pone.0179805.

[7] A. N. Okere, T. Li, C. Theran, E. Nyasani, and A. A. Ali, "Evaluation of factors predicting transition from prediabetes to diabetes among patients residing in underserved communities in the United States – A machine learning approach," *Computers in Biology and Medicine*, vol. 187, Mar. 2025, Art. no. 109824, https://doi.org/10.1016/j.compbiomed.2025.109824.

[8] R. Kaur, R. Kumar, S. Kaur, G. Singh, A. Kaur, and S. Singh, "Machine Learning for Diabetes Prediction: Performance Analysis Using Logistic Regression, Naïve Bayes, and Decision Tree Models," *Healthcraft Frontiers*, vol. 02, no. 04, pp. 169–187, Dec. 2024, https://doi.org/10.56578/hf020401.

[9] W. Li, Y. Peng, and K. Peng, "Diabetes prediction model based on GA-XGBoost and stacking ensemble algorithm," *PLOS ONE*, vol. 19, no. 9, 2024, Art. no. e0311222, https://doi.org/10.1371/journal.pone.0311222.

[10] N. Nagarjuna and D. L. Hn, "Predictive Modeling of Diabetes Mellitus Utilizing Machine Learning Techniques," *CVR Journal of Science and Technology*, vol. 26, no. 1, pp. 112–117, June 2024.

[11] H. El-Sofany, S. A. El-Seoud, O. H. Karam, Y. M. Abd El-Latif, and I. A. T. F. Taj-Eddin, "A Proposed Technique Using Machine Learning for the Prediction of Diabetes Disease through a Mobile App," *International Journal of Intelligent Systems*, vol. 2024, no. 1, 2024, Art. no. 6688934, https://doi.org/10.1155/2024/6688934.

[12] Z. Rafie, M. S. Talab, B. E. Z. Koor, A. Garavand, C. Salehnasab, and M. Ghaderzadeh, "Leveraging XGBoost and explainable AI for accurate prediction of type 2 diabetes," *BMC Public Health*, vol. 25, no. 1, Oct. 2025, Art. no. 3688, https://doi.org/10.1186/s12889-025-24953-w.

[13] A. Nkemdirim Okere, T. Li, C. Theran, E. Nyasani, and A. A. Ali, "Evaluation of factors predicting transition from prediabetes to diabetes among patients residing in underserved communities in the United States – A machine learning approach," *Computers in Biology and Medicine*, vol. 187, Mar. 2025, Art. no. 109824, https://doi.org/10.1016/j.compbiomed.2025.109824.

[14] A. M. AbdulAbbas, R. Alkanany, Y. A. K. Al-Nuaimi, and Z. M. A. Al-Hamdawee, "A Sequential Data Preprocessing Pipeline for Diabetes Prediction: A Data Leakage Prevention and Dual-Validation Approach," *Engineering, Technology & Applied Science Research*, vol. 15, no. 6, pp. 30059–30066, Dec. 2025, https://doi.org/10.48084/etasr.14155.

[15] P. Netayawijit, W. Chansanam, and K. Sorn-In, "Interpretable Machine Learning Framework for Diabetes Prediction: Integrating SMOTE Balancing with SHAP Explainability for Clinical Decision Support," *Healthcare*, vol. 13, no. 20, Oct. 2025, https://doi.org/10.3390/healthcare13202588.

[16] P. Gogoi and J. A. Valan, "Chronic kidney disease prediction using machine learning techniques: a comparative study of feature selection methods with SMOTE and SHAP," *Multiscale and Multidisciplinary Modeling, Experiments and Design*, vol. 8, no. 4, Mar. 2025, Art. no. 216, https://doi.org/10.1007/s41939-025-00806-2.

[17] Md. M. Islam, H. R. Rifat, Md. S. B. Shahid, A. Akhter, M. A. Uddin, and K. M. M. Uddin, "Explainable Machine Learning for Efficient Diabetes Prediction Using Hyperparameter Tuning, SHAP Analysis, Partial Dependency, and LIME," *Engineering Reports*, vol. 7, no. 1, 2025, Art. no. e13080, https://doi.org/10.1002/eng2.13080.

[18] M. Kutlu, T. B. Donmez, and C. Freeman, "Machine learning interpretability in diabetes risk assessment: a SHAP analysis," *Computers and Electronics in Medicine*, vol. 1, no. 1, pp. 34–44, July 2024, https://doi.org/10.69882/adba.cem.2024075.

[19] H. B. Kibria, M. Nahiduzzaman, M. O. F. Goni, M. Ahsan, and J. Haider, "An Ensemble Approach for the Prediction of Diabetes Mellitus Using a Soft Voting Classifier with an Explainable AI," *Sensors*, vol. 22, no. 19, Sept. 2022, https://doi.org/10.3390/s22197268.

[20] M. B. Almadhoun and M. A. Burhanuddin, "Optimizing Feature Selection and Machine Learning Algorithms for Early Detection of Prediabetes Risk: Comparative Study," *JMIR Bioinformatics and Biotechnology*, vol. 6, no. 1, July 2025, Art. no. e70621, https://doi.org/10.2196/70621.

[21] "Pima Indians Diabetes Database." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database.

[22] "Diabetes Prediction Dataset." Kaggle, [Online]. Available: https://www.kaggle.com/dataset s/marshalpatel3558/diabetes-prediction-dataset-legit-dataset.

[23] G. Yudheksha, V. Murugadoss, P. S. Reddy, T. Harshavardan, and S. Sriramulu, "A Machine Learning based Approach to Detect Early Stage Diabetes Prediction," in *2022 6th International Conference on Electronics, Communication and Aerospace Technology*, Dec. 2022, pp. 919–924, https://doi.org/10.1109/ICECA55336.2022.10009113.