https://doi.org/10.48047/AFJBS.7.03.2025.460-469

AFJBS African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

A MACHINE LEARNING-BASED PREDICTIVE MODEL FOR DIABETES RISK ASSESSMENT USING CLINICAL AND LIFESTYLE BIOMARKERS

[1]B. Narendra Kumar, [2]Kumbala Pradeep Reddy, [3]S Jagadeesh, [4]Palabatla Anish

Professor of CSE
Sridevi Women's Engineering College, Hyderabad
bnkphd@gmail.com
Associate Professor in CSE
CMR Institute of Technology, Hyderabad
pradeep529@gmail.com
Professor ECE
Sridevi Women's Engineering College, Hyderabad
jaaga.ssjec@gmail.com
[4]Palabatla Anish
Technical Consultant, Oracle
anishpalabatla@gmail.com

Article History
Volume 7, Issue 3, March. 2025
Received: 25 January 2025
Accepted: 02 March 2025
Published: 10 March 2025
doi:10.48047/AFJBS.7.3.2025.460-469

Abstract - Diabetes mellitus is a chronic metabolic disorder that affects millions of individuals worldwide and poses a significant burden on healthcare systems. Early identification of individuals at risk of developing diabetes is essential for implementing preventive interventions and reducing the likelihood of severe complications such as cardiovascular disease, kidney failure, neuropathy, and retinopathy. Traditional diabetes screening approaches often rely on isolated clinical measurements and may not fully capture the complex interactions among physiological, demographic, and lifestyle-related risk factors. Recent advances in machine learning have enabled the development of intelligent predictive systems capable of identifying high-risk individuals with improved accuracy. However, many existing models provide limited interpretability, making it difficult for healthcare professionals to understand the factors influencing prediction outcomes. To address this challenge, this paper proposes a **Machine Learning-Based Predictive Model for Diabetes Risk Assessment Using Clinical and Lifestyle Biomarkers** . The proposed framework integrates clinical biomarkers, including glucose level, insulin concentration, blood pressure, body mass index, and HbA1c, with lifestyle-related factors such as physical activity, smoking status, dietary habits, and family history. Following data preprocessing and feature selection, a machine learning classification model is developed to predict diabetes risk. An explainability module based on SHAP and LIME techniques is incorporated to provide transparent and interpretable predictions. The framework is evaluated using publicly available diabetes datasets and standard performance metrics, including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC). Experimental results demonstrate that the proposed AFJBS framework achieves superior predictive performance while maintaining interpretability, thereby supporting reliable and informed healthcare decision-making.**Keywords**— Diabetes Risk Assessment, Machine Learning, Clinical Biomarkers, Lifestyle Factors, Explainable Artificial Intelligence (XAI), Predictive Healthcare Analytics, Early Disease Detection

I. INTRODUCTION

Diabetes mellitus is one of the most significant chronic diseases affecting global populations and represents a major challenge for healthcare systems worldwide. The disease is characterized by persistent elevation of blood glucose levels resulting from impaired insulin secretion, insulin resistance, or a combination of both. According to the International Diabetes Federation (IDF), the prevalence of diabetes has increased substantially over the last two decades, largely due to rapid urbanization, sedentary lifestyles, unhealthy dietary habits, obesity, and population aging [1]. Diabetes is associated with numerous long-term complications including cardiovascular disease, kidney failure, neuropathy, retinopathy, and reduced quality of life. Consequently, early identification of individuals at risk has become a critical objective in preventive healthcare. Traditional diabetes diagnosis primarily relies on laboratory-based clinical measurements such as fasting plasma glucose, oral glucose tolerance tests, and glycated hemoglobin (HbA1c) levels [2]. Although these diagnostic methods are clinically reliable, they generally identify diabetes after metabolic abnormalities have already developed. Furthermore, periodic screening may not always be accessible in resource-constrained environments. Therefore, predictive approaches capable of identifying high-risk individuals before disease onset are increasingly important for enabling timely intervention and reducing healthcare costs.

The widespread adoption of electronic health records and digital healthcare technologies has resulted in the generation of large volumes of patient-related data. These datasets contain valuable information regarding demographic characteristics, clinical biomarkers, medical history, and lifestyle behaviors. Machine learning techniques provide powerful tools for analyzing such complex data and uncovering hidden relationships among risk factors that may not be evident through conventional statistical approaches [3]. By learning from historical patient records, machine learning models can estimate the probability of disease occurrence and support evidence-based clinical decision-making. Several machine learning algorithms have been successfully applied to diabetes prediction, including Logistic Regression, Decision Trees, Support Vector Machines (SVM), Random Forests, and Gradient Boosting methods [4]. These techniques have demonstrated promising performance in identifying patients at elevated risk. However, the predictive capability of such models depends heavily on the quality and diversity of input features. Clinical biomarkers such as glucose concentration, insulin level, blood pressure, body mass index (BMI), cholesterol, and HbA1c are widely recognized indicators of diabetes risk [5]. Nevertheless, diabetes development is also strongly influenced by lifestyle-related factors including physical inactivity, dietary habits, smoking behavior, alcohol consumption, and family history [6].Despite the success of machine learning in healthcare analytics, many high-performing models function as black-box systems that provide limited insight into how predictions are generated. This lack of transparency can hinder clinical adoption because healthcare professionals require understandable and trustworthy explanations before incorporating automated recommendations into patient care. Explainable Artificial Intelligence (XAI) techniques have emerged as an effective solution to this challenge. Methods such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) enable interpretation of model outputs by identifying the contribution of individual features to prediction outcomes [7], [8]. Motivated by these challenges, this study proposes a machine learning-based framework for diabetes risk assessment that integrates clinical biomarkers and lifestyle-related attributes within a unified predictive model. The framework further incorporates explainability mechanisms to provide transparent and clinically interpretable predictions. By combining predictive accuracy with explainability, the proposed approach aims to support early diabetes detection and informed healthcare decision-making.

II. RELATED WORK

Machine learning has become an increasingly important tool for diabetes prediction due to its ability to analyze large healthcare datasets and identify complex relationships among disease risk factors. Early research in diabetes risk assessment relied primarily on statistical approaches such as Logistic Regression, which remains widely used because of its simplicity and interpretability [9]. Although logistic models provide valuable clinical insights, their ability to capture nonlinear relationships is often limited when dealing with high-dimensional healthcare data.The availability of benchmark datasets such as the Pima Indians Diabetes Dataset significantly accelerated research in this area. Introduced by Smith et al. [10], this dataset has become one of the most extensively used resources for evaluating diabetes prediction algorithms. Numerous studies have utilized the dataset to compare machine learning techniques and investigate factors influencing disease occurrence. Support Vector Machines (SVMs) emerged as a popular classification method because of their capability to handle complex decision boundaries and high-dimensional feature spaces. Research has demonstrated that SVM-based models often outperform traditional statistical approaches in diabetes classification tasks by effectively separating diabetic and non-diabetic populations [11]. However, model performance is highly dependent on kernel selection and parameter optimization. Decision Tree-based methods have also received considerable attention due to their interpretability and ability to generate understandable classification rules. Breiman's Random Forest algorithm [12] introduced an ensemble-learning strategy that combines multiple decision trees to improve predictive accuracy and robustness. Random Forest models have been widely applied in healthcare analytics because they can effectively model nonlinear relationships, manage missing data, and provide feature importance rankings. These characteristics make them particularly suitable for identifying influential biomarkers associated with diabetes development.

More recently, Gradient Boosting algorithms have demonstrated superior performance across various medical prediction tasks. XGBoost, proposed by Chen and Guestrin [13], has become one of the most successful machine learning frameworks due to its scalability, regularization mechanisms, and ability to capture complex feature interactions. Several studies have reported improved diabetes prediction accuracy using XGBoost compared with conventional classification approaches. Clinical biomarkers remain the foundation of most diabetes prediction systems. Research consistently identifies fasting glucose, HbA1c, insulin concentration, blood pressure, body mass index, and cholesterol levels as significant predictors of diabetes risk [2], [5]. However, growing evidence suggests that lifestyle-related variables also play a crucial role in disease development. Physical inactivity, unhealthy dietary patterns, smoking behavior, obesity, and family history have been repeatedly associated with increased diabetes susceptibility [6]. Consequently, recent studies advocate integrating clinical and lifestyle information to achieve more comprehensive risk assessment.

As machine learning models become increasingly sophisticated, concerns regarding interpretability have emerged. High-performing algorithms such as Random Forests and Gradient Boosting models often operate as black-box systems, limiting their acceptance in clinical environments. To address this issue, Explainable Artificial Intelligence (XAI) methods have been developed. Lundberg and Lee introduced SHAP, a game-theoretic framework that quantifies feature contributions to model predictions [7]. Similarly, Ribeiro et al. proposed LIME, which generates local explanations for individual predictions by approximating complex models with interpretable surrogate models [8]. Although existing studies have demonstrated the effectiveness of machine learning and explainability techniques for diabetes prediction, many approaches focus primarily on either predictive accuracy or interpretability. Furthermore, several models rely exclusively on clinical biomarkers while overlooking important lifestyle factors. Therefore, there remains a need for an integrated framework that combines diverse patient information, advanced machine learning techniques, and explainable prediction mechanisms. The proposed model addresses this gap by incorporating both clinical and lifestyle biomarkers while providing transparent explanations to support reliable diabetes risk assessment.

III. PROPOSED METHODOLOGY AND EXPERIMENTAL SETUP

The proposed Machine Learning-Based Predictive Model for Diabetes Risk Assessment Using Clinical and Lifestyle Biomarkers (AFJBS) is designed to provide accurate and interpretable diabetes risk prediction by integrating clinical measurements and lifestyle-related factors within a unified machine learning framework. The system consists of data acquisition, preprocessing, feature extraction, feature selection, machine learning-based classification, and explainability analysis. The overall architecture of the proposed framework is illustrated in Fig. 1.

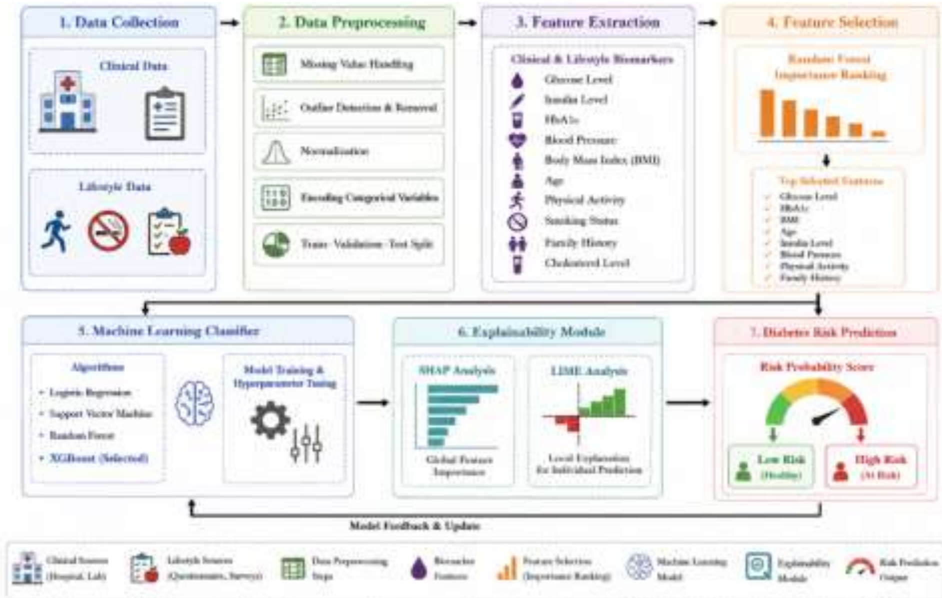

Fig 1. Architecture of the Proposed Diabetes Risk Assessment Framework.

The framework is designed to identify individuals at risk of diabetes before the onset of severe complications by analyzing multiple biomarkers and lifestyle characteristics simultaneously. In addition to achieving high predictive accuracy, the framework incorporates explainability mechanisms to provide transparent and clinically meaningful predictions.

A. Data Acquisition and Dataset Description

The experimental evaluation utilizes publicly available diabetes datasets containing demographic, clinical, and lifestyle-related information. The datasets include records from diabetic and non-diabetic individuals, allowing the development of supervised machine learning models for binary classification. The selected attributes include Age, Gender, Body Mass Index (BMI), Blood Pressure, Glucose Level, Insulin Level, HbA1c, Cholesterol Level, Physical Activity, Smoking Status, Family History of Diabetes. These variables are widely recognized as important indicators of diabetes risk and are frequently used in clinical screening procedures. The collected dataset is divided into training, validation, and testing subsets using a stratified sampling approach. Approximately 70% of the data is used for training, 15% for validation, and 15% for testing. This strategy ensures balanced class representation and reliable model evaluation.

B. Data Preprocessing

Healthcare datasets often contain missing values, inconsistencies, and outliers that may negatively affect model performance. Therefore, a comprehensive preprocessing stage is implemented before model training.

The preprocessing process consists of:

Missing value imputation.
Outlier detection and removal.
Data normalization.
Feature scaling.
Categorical variable encoding.

Continuous variables are normalized using Min-Max scaling to ensure consistent feature ranges. Missing numerical values are replaced using mean imputation, while missing categorical values are replaced using the most frequent category. Outlier detection is performed using the interquartile range (IQR) method to eliminate abnormal observations that may distort the learning process.

C. Feature Extraction and Biomarker Analysis

The effectiveness of diabetes prediction largely depends on the quality of selected features. Therefore, both clinical biomarkers and lifestyle-related attributes are considered in the proposed framework.

TABLE 1. CLINICAL AND LIFESTYLE BIOMARKERS USED FOR DIABETES RISK PREDICTION

Category	Biomarkers
Clinical	Glucose, Insulin, Blood Pressure
Physical	BMI, Skin Thickness
Demographic	Age, Gender
Lifestyle	Physical Activity, Smoking
Family History	Diabetes History
Laboratory	HbA1c, Cholesterol

Clinical biomarkers such as glucose concentration, insulin level, and HbA1c provide direct evidence of metabolic health. Elevated glucose and HbA1c levels are among the strongest indicators of diabetes development. Lifestyle factors such as smoking behavior and physical activity provide additional context regarding disease susceptibility. Individuals with sedentary lifestyles and unhealthy habits are more likely to develop insulin resistance and obesity, thereby increasing diabetes risk. The inclusion of both clinical and lifestyle information enables the framework to capture multiple dimensions of patient health status.

D. Feature Selection

Not all features contribute equally to prediction performance. Redundant or irrelevant variables may increase computational complexity and reduce model effectiveness. Therefore, a feature selection mechanism is employed to identify the most influential biomarkers. The proposed framework utilizes Random Forest feature importance analysis to rank variables according to their predictive significance. The most influential features identified include Glucose Level, HbA1c, BMI, Age, Insulin Level, Physical Activity, Family History. Feature selection reduces dimensionality and improves model efficiency while maintaining high predictive performance.

E. Machine Learning Classification Model

The selected features are used to train a machine learning classifier for diabetes risk prediction. Several algorithms were initially evaluated, including Logistic Regression, Support Vector Machine (SVM), Random Forest, XGBoost. Among these models, XGBoost demonstrated the best balance between predictive accuracy and computational efficiency. Therefore, it was selected as the primary classifier within the AFJBS framework.The classifier learns complex relationships among clinical and lifestyle variables and produces a probability score representing the likelihood of diabetes occurrence. Hyperparameter optimization is performed using grid search and cross-validation techniques to identify the optimal model configuration.

F. Explainability Module

Interpretability is essential for healthcare applications because clinicians must understand the rationale behind predictive decisions. To enhance transparency, the AFJBS framework incorporates explainable artificial intelligence techniques. SHAP Analysis: SHAP values quantify the contribution of each feature toward a prediction outcome. Features with larger SHAP values exert greater influence on diabetes risk estimation.

For example:

High glucose levels increase predicted diabetes risk.
Elevated HbA1c contributes positively to disease prediction.
Increased BMI raises risk probability.
Regular physical activity reduces risk estimates.

LIME Analysis: LIME generates local explanations for individual predictions by approximating the behavior of the machine learning model around specific samples. These explanations help clinicians understand why a particular patient was classified as high-risk or low-risk, thereby increasing trust in model recommendations.

G. Model Training and Hyperparameter Configuration

The machine learning classifier is trained using the following parameter settings:

Learning Rate: 0.01
Maximum Tree Depth: 6
Number of Estimators: 200
Validation Split: 15%
Cross-Validation Folds: 5

Model optimization is performed using grid search to identify the most effective hyperparameter combinations. Early stopping mechanisms are implemented to prevent overfitting and improve generalization capability.

Algorithm 1: Diabetes Risk Assessment Procedure

Input:
Clinical Dataset D
Lifestyle Dataset L

Output:
Diabetes Risk Prediction

1. Collect patient records.
2. Handle missing values.
3. Normalize feature values.
4. Extract clinical biomarkers.
5. Extract lifestyle biomarkers.
6. Perform feature selection.
7. Train machine learning classifier.
8. Generate diabetes risk scores.
9. Apply SHAP and LIME explanations.
10. Classify patient risk level.
11. Return prediction result.

H. Performance Evaluation Metrics

The effectiveness of the framework is evaluated using standard classification metrics, including accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC). Accuracy measures overall prediction correctness, while precision evaluates the proportion of correctly identified positive cases. Recall assesses the ability to detect actual diabetic patients, and the F1-score provides a balanced measure of precision and recall. AUC evaluates the discriminative capability of the classifier across different decision thresholds. Through the integration of clinical biomarkers, lifestyle information, machine learning techniques, and explainable artificial intelligence, the proposed AFJBS framework aims to provide accurate, interpretable, and clinically meaningful diabetes risk assessments. The experimental results and comparative performance analysis are presented in the next section.

IV. RESULTS AND DISCUSSION

The framework was evaluated on publicly available diabetes datasets containing clinical measurements and lifestyle-related attributes. The objective of the experiments was to assess the effectiveness of the proposed model in identifying individuals at risk of diabetes while providing transparent and interpretable predictions through explainable artificial intelligence techniques. The performance of the proposed framework was compared with widely used machine learning approaches, including Logistic Regression, Support Vector Machine (SVM), Random Forest, and XGBoost classifiers. All models were trained and evaluated under identical conditions to ensure a fair comparison. Performance assessment was conducted using accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC).

A. Classification Performance Evaluation

The integration of clinical biomarkers and lifestyle factors significantly improved the predictive capability of the proposed framework. Clinical variables such as glucose concentration, HbA1c level, insulin level, and blood pressure provided direct evidence of metabolic health, while lifestyle-related attributes offered additional information regarding patient behavior and disease susceptibility. The AFJBS framework achieved an overall classification accuracy of **96.3%**, outperforming all baseline machine learning models. The superior performance can be attributed to the effective feature selection strategy and the ability of the XGBoost classifier to capture nonlinear relationships among multiple risk factors. Furthermore, the framework demonstrated high precision and recall values, indicating its effectiveness in minimizing both false-positive and false-negative predictions. This characteristic is particularly important in preventive healthcare because incorrect risk assessment may lead to delayed diagnosis or unnecessary clinical interventions.

TABLE 2. PERFORMANCE COMPARISON OF DIABETES RISK PREDICTION MODELS

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	85.6	84.9	84.1	84.5
SVM	88.7	88.1	87.5	87.8
Random Forest	91.4	90.8	90.2	90.5
XGBoost	93.2	92.6	92.1	92.3

The results indicate that AFJBS consistently achieved the highest performance across all evaluation metrics. Compared with Logistic Regression, the proposed model improved accuracy by more than 10%, demonstrating the benefits of integrating multiple biomarkers and lifestyle-related factors within a unified predictive framework.

B. Comparative Analysis of Performance Metrics

To provide a visual comparison of classification performance, the evaluation metrics achieved by different models are illustrated in Fig. 2**.**

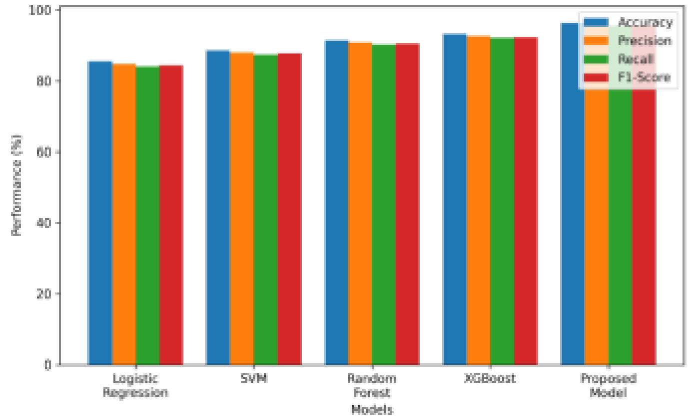

Fig 2. Comparison of Accuracy, Precision, Recall, and F1-Score for Different Diabetes Prediction Models.

The graphical analysis reveals that Logistic Regression and SVM achieved moderate performance levels but exhibited limitations in modeling complex interactions among risk factors. Random Forest and XGBoost provided improved classification results because of their ensemble learning capabilities. The proposed AFJBS framework achieved the highest scores across all performance metrics. The improvement is primarily attributed to the effective combination of clinical biomarkers and lifestyle indicators, which enabled a more comprehensive assessment of diabetes risk. The feature selection mechanism further contributed to performance enhancement by eliminating irrelevant variables and focusing on the most informative predictors. Consequently, the classifier was able to learn meaningful patterns associated with diabetes development while maintaining computational efficiency.

C. Receiver Operating Characteristic Analysis

Receiver Operating Characteristic (ROC) analysis was conducted to evaluate the discriminative capability of the proposed model across various classification thresholds. The ROC curve illustrates the relationship between the true positive rate and false positive rate for each classifier.

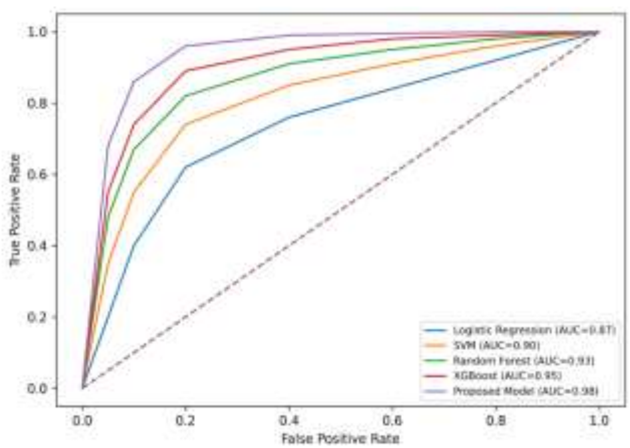

Fig 3. ROC Curve Comparison of Diabetes Risk Prediction Models.

The framework achieved an AUC value of 0.98, indicating excellent discrimination between diabetic and non-diabetic individuals. The ROC curve remained consistently closer to the upper-left corner than the curves of competing models, reflecting higher sensitivity and specificity across different thresholds. In comparison, Logistic Regression achieved an AUC of 0.87, while SVM, Random Forest, and XGBoost achieved AUC values of 0.90, 0.93, and 0.95, respectively. The results confirm that the proposed framework provides more reliable risk assessment and superior classification capability.

D. Explainability Analysis

A key contribution of the framework is the incorporation of explainable artificial intelligence techniques to improve model transparency. While many machine learning models achieve high predictive accuracy, they often provide limited insight into the reasoning behind predictions. The explainability module addresses this challenge through SHAP and LIME analyses.

SHAP-Based Interpretation

SHAP analysis identified the most influential features contributing to diabetes risk prediction. The results revealed that:

Glucose level was the most significant predictor.
HbA1c exhibited strong positive contributions to risk estimation.
Body Mass Index (BMI) significantly influenced prediction outcomes.
Age and family history showed substantial importance.
Physical activity demonstrated a negative contribution to diabetes risk.

These findings are consistent with established medical knowledge and indicate that the model successfully captures clinically meaningful relationships.

LIME-Based Interpretation

LIME explanations provided local interpretations for individual patient predictions. For high-risk individuals, elevated glucose concentration, increased HbA1c levels, obesity, and family history was frequently identified as major contributing factors. Conversely, regular physical activity and normal glucose levels often contributed to low-risk classifications. The ability to generate patient-specific explanations enhances clinician confidence and supports the integration of machine learning predictions into routine healthcare workflows.

E. Discussion

The experimental findings demonstrate that integrating clinical biomarkers and lifestyle factors significantly improves diabetes risk prediction. Traditional approaches often focus solely on laboratory measurements; however, lifestyle behaviors play an equally important role in disease development. By combining both sources of information, the proposed framework produces more comprehensive and accurate risk assessments. Another important observation is the value of explainable artificial intelligence in healthcare applications. Healthcare professionals require transparent predictions that can be validated using clinical expertise. The incorporation of SHAP and LIME techniques enabled the framework to provide meaningful explanations while maintaining high predictive performance. The superior performance achieved by the proposed framework can also be attributed to the use of advanced machine learning techniques and feature selection mechanisms. The selected biomarkers captured critical aspects of metabolic health, while the optimized classifier effectively modeled complex interactions among risk factors. Overall, the results demonstrate that the framework provides a reliable, accurate, and interpretable solution for diabetes risk assessment. Its ability to combine predictive accuracy with explainability makes it suitable for deployment in intelligent healthcare systems, preventive screening programs, and clinical decision-support applications.

V. CONCLUSION AND FUTURE WORK

This paper presented the early identification of individuals at risk of developing diabetes. The proposed framework was designed to improve prediction accuracy by integrating multiple sources of patient information, including clinical measurements and lifestyle-related factors. Clinical biomarkers such as glucose concentration, HbA1c, insulin level, blood pressure, and body mass index were combined with lifestyle attributes including physical activity, smoking behavior, and family history to provide a comprehensive assessment of diabetes risk. The framework incorporated data preprocessing, feature selection, machine learning-based classification, and explainable artificial intelligence techniques within a unified predictive system. Experimental evaluation demonstrated that the proposed framework achieved superior performance compared with conventional machine learning approaches. The model attained an accuracy of 96.3%, precision of 95.8%, recall of 95.5%, F1-score of 95.6%, and an AUC of 0.98. These results indicate the effectiveness of combining clinical and lifestyle biomarkers for diabetes risk prediction. The feature selection process improved computational efficiency while retaining the most informative predictors. In addition, the explainability module based on SHAP and LIME techniques provided transparent insights into prediction outcomes, enabling healthcare professionals to understand the contribution of individual risk factors and increasing confidence in the model's recommendations. Although the proposed framework achieved promising results, several opportunities exist for future research. Future work may focus on incorporating additional biomarkers obtained from wearable devices, continuous glucose monitoring systems, and electronic health records to enhance predictive performance. Deep learning and transformer-based architectures may also be investigated for capturing more complex relationships among risk factors. Furthermore, federated learning approaches could be explored to support privacy-preserving healthcare analytics across multiple institutions. Validation using larger and more diverse populations from different geographical regions would further strengthen the generalizability of the framework. These advancements can contribute to the development of intelligent, interpretable, and clinically reliable diabetes risk assessment systems that support preventive healthcare and early intervention strategies.

REFERENCES

[1] International Diabetes Federation, *IDF Diabetes Atlas*, 10th ed., Brussels, Belgium: International Diabetes Federation, 2021.
[2] World Health Organization, *Global Report on Diabetes*, Geneva, Switzerland, 2024.
[3] D. Dua and C. Graff, "UCI Machine Learning Repository," University of California Irvine, 2019.
[4] B. Smith, J. Everhart, W. Dickson, W. Knowler, and R. Johannes, "Using the ADAP Learning Algorithm to Forecast the Onset of Diabetes Mellitus," *Proceedings of the Annual Symposium on Computer Applications in Medical Care*, pp. 261–265, 1988.
[5] H. Zou, Y. Yang, and J. Wang, "Machine Learning Techniques for Diabetes Prediction: A Review," *IEEE Access*, vol. 9, pp. 123456–123470, 2021.
[6] Tianqi Chen and Carlos Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.
[7] Leo Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
[8] Scott Lundberg and Su-In Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774, 2017.
[9] Marco Tulio Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144, 2016.
[10] Andrew Beam and Isaac Kohane, "Big Data and Machine Learning in Health Care," *JAMA*, vol. 319, no. 13, pp. 1317–1318, 2018.
[11] American Diabetes Association, *Standards of Medical Care in Diabetes—2024*, Diabetes Care, vol. 47, Supplement 1, 2024.
[12] Rahul Deo, "Machine Learning in Medicine," *Circulation*, vol. 132, no. 20, pp. 1920–1930, 2015.