International Journal of Artificial Intelligence and Computer Electronics
Volume: 01, Issue: 01, Jan-Mar 2026, pp:1-9
P-ISSN 3117-8529 | E-ISSN 3117-8537

An Interpretable Hybrid Machine Learning Framework for Robust Type II Diabetes Prediction Using Electronic Health Records

Dr. Mazher Khan[1], Dr. Prasun Chakrabarty[2], Dr. Bhuvan Unhelkar[3], Dr. Amairullah Khan Lodhi[4], Dr. Ali Hussain[5]

[1]University of South Florida, Florida, USA
Email: mazher.engg@gmail.com
[2]Department of Computer Science and Engineering, Sir Padampat Singhania University, Udaipur, Rajasthan, India
Email: drprasun.cse@gmail.com
[3]University of South Florida, Florida, USA
Email: bunhelkar@usf.edu
[4]Department of Electronics and Communication Engineering, Shadan College of Engineering & Technology, Hyderabad, India
Email: dean.rnd.scet@gmail.com
[5]Department of Computer Science and Engineering, Srinidhi Institute of Science and Technology, Hyderabad, India
Email: alihussain.phd@gmail.com

Article Received 12-11-2025, Revised 25-1-2026, Accepted 02-03-2026

Abstract—***Diabetes mellitus type II (T2DM) happens to be a fast-emerging international health issue that needs timely and accurate risk identification to mitigate chronic conditions. EHRs offer an excellent platform for longitudinal clinical data, but predictive modeling with EHRs is plagued by data quality challenges, high-dimensional feature spaces, and limited generalizability across healthcare facilities. The paper presents a proposed interpretable hybrid machine-learning architecture for predicting T2DM with high accuracy using structured EHR data. The proposed pipeline combines systematic data preprocessing (imputation, outlier handling, normalization), a hybrid feature selection strategy (Random Forest importance and SHAP-based explainability), and domain-conscious model validation to improve generalization. Evaluation of the framework is conducted on benchmark datasets such as Pima Indians, UCI Diabetes 130-US Hospitals, MIMIC-III, and a simulated longitudinal EHR dataset. The experimental evidence illustrates a gradual improvement in performance across processing stages, reaching 91% accuracy and an AUROC of 0.93 with the Random Forest + SHAP model. The analysis reveals that the clinically established predictors identified by feature importance in the present research are glucose, A1C, insulin, BMI, and blood pressure, which align with diagnostic guidelines. A comparative study of the proposed methodology with LSTM and transformer-based models demonstrates that it strikes a good balance between predictive accuracy and interpretability. The findings show that hybrid model explainability and structured preprocessing are highly effective in increasing robustness, fairness, and clinical usability. The prescribed framework includes a scalable, reliable blueprint for practical implementation in digital health and clinical decision support systems.***

Index Terms— Electronic Health Records (EHRs), Medical Data Analytics, Predictive Modeling, Explainable AI, Random Forest, SHAP, Domain Adaptation, Digital Health Systems.

I. INTRODUCTION

Type II Diabetes Mellitus (T2DM) is a chronic metabolic condition that is marked by poor insulin regulation and constant hyperglycemia. It is among the most rapidly increasing health issues in the world today that causes many cardiovascular diseases, kidney failure, neuropathy, and early death. The International Diabetes Federation estimates that over 500 million adults across the world are already living with diabetes at present, and of all the cases, Type II diabetes contributes to about 90-95% of the total number. The early identification of people at high risk is thus necessary to intervene early enough, prevent complications, and reduce healthcare costs.

The proliferation of Electronic Health Records (EHRs) has opened new possibilities for data-driven disease prediction. EHR systems store structured clinical data, including demographics, laboratory test results, diagnosis codes, medications, and longitudinal visit records. These digital repositories function as mass clinical sensing systems that continuously measure patients' health signals over time. With such rich, high-dimensional data over time, machine

learning (ML) methods have shown the potential to predict the onset of T2DM earlier than traditional rule-based screening methods. Irrespective of these developments, predictive models based on EHR are difficult to implement with trust. There are three important pitfalls that prevent clinical translation.

To begin with, data quality strongly affects the model's reliability. EHR data often include blank values, inconsistent code sets, outliers, and varying formats across institutions. These discrepancies diminish predictive accuracy and lower reproducibility unless they are systematically preprocessed. Second, EHR datasets are high-dimensional, and this contributes to the complexity of feature selection. A significant number of clinical variables can be redundant, correlated, or weakly informative. To build clinician trust that is requisite for regulatory acceptance, it is necessary to identify clinically meaningful predictors while maintaining interpretability. Dark models of deep learning are highly accurate but not transparent, which limits their usability in clinical settings.

Third, there is still an issue of poor generalization across healthcare settings. Those trained in a single institution or a homogeneous group of people rarely work with diverse populations because of shifts in distribution, demographic bias, and variations in institutional coding. The issue of generalization and fairness is thus a key concern that should be tackled to have scalable and equitable deployment. Previous literature has investigated the personal side of these issues. Other works focus on the multimodal integration of clinical notes and laboratory data, whereas others research transformer-based longitudinal models or fair neural architectures. Nonetheless, most solutions treat these problems separately. Not many studies offer an integrated framework that aligns well with robust preprocessing, explainable feature selection, and cross-dataset generalization, as well as a pipeline that makes results easy for clinicians to interpret.

To overcome these constraints, the present paper proposes a hybrid machine learning model that can be used to make strong predictions of T2DM based on structured EHR data. To enhance the models' robustness and clinical usability, the framework systematically incorporates data quality improvement, SHAP-driven sample hybrid feature selection, and domain-aware validation. The key contributions of this work are as follows:

- An organized data cleaning plan that has imputation, outliers, data normalization, and sequencing of the data across time to improve data integrity and reproducibility.
- A hybridized Random Forest importance and SHAP-based algorithm to find clinically relevant predictors without sacrificing interpretability.
- Cross-dataset assessment on benchmark and simulated EHR datasets to enhance generalization, robustness, and fairness across demographic subgroups.

It has been shown experimentally that systematic preprocessing and explainability-based modeling lead to much better performance, with up to 91% accuracy and strong calibration without loss of transparency. The proposed framework can provide a scalable, clinically reliable architecture for integrating machine-learning-based diabetes risk prediction into digital health and decision-support systems.

The rest of this paper will be structured as follows. Section II is a literature review of related work on EHR-based diabetes prediction. Section III is the proposed framework and methodology. Part IV describes the experimental setup. Section V talks about results and analysis. Section VI provides a conclusion of the paper and future directions of research.

II. The Related Work

The use of machine learning (ML) and artificial intelligence (AI) for predicting Type 2 Diabetes Mellitus (T2DM) has grown tremendously as access to Electronic Health Records (EHRs) has increased. The longitudinal clinical data, such as laboratory results, diagnosis codes, medications, and demographic information, are available in EHRs, enabling predictive modeling that is not inherent to statistical screening tools. Nevertheless, issues of data quality, feature selection, interpretability, and generalization continue to be the primary topics of the literature. The first weakness in EHR-based predictive systems is data quality. Clinical databases often have missing values, variable coding conventions, measurement errors, and disparate formats across institutions. These irregularities adversely affect the stability and reproducibility of models.

The International Diabetes Federation provides extensive statistical data, trends, and projections on diabetes globally, and this illness requires immediate identification strategies and scalable predictive healthcare solutions [1]. This work uses a supervised learning approach to predict diabetes from electronic health records, focusing on preprocessing and presenting improved classification results through structured data cleaning procedures [2]. The authors propose a large language multimodal model that integrates clinical notes and lab results to improve the prediction of new-onset type 2 diabetes using both structured and unstructured hospital data [3]. The Hi-BEHRT presents a hierarchical transformer architecture designed to model longitudinal electronic health records, which is useful for effectively capturing time-related dependencies and enhancing representation learning for clinical event prediction tasks [4]. This systematic review of artificial intelligence methods for diabetes prediction identifies weaknesses in external validation, interpretability, and generalizability across diverse healthcare data and populations [5]. In the paper, an SHAP-based metaheuristic meta feature selection framework is

introduced that enhances the accuracy of diabetes risk prediction without compromising the ability to explain the process results by selecting variables based on their optimality [6].

It is an engineering study that uses synthetic EHR-based features to predict diabetic kidney disease and has shown better predictive accuracy than the ensemble-based importance ranking and cross-validation [7]. The authors contrast the results of several feature selection methods to cluster diabetic patients with the results obtained as interpretable measures of importance, which improve clinical transparency and subgroup differentiation [8]. This research uses transformer-based longitudinal modeling to predict the occurrence of diabetes complications in the long run, and it shows improved learning of temporal representations, while also having limitations due to computational complexity [9]. The article combines natural language processing with structured EHR predictive indicators of diabetes risks, and concludes with a better classification solution by implementing multimodal data fusion strategies [10]. This study examines the concept of diabetes network-level screening using deep learning as a non-invasive technique to classify both chest x-ray images and EHR data, highlighting the potential of multimodal integration and the scalability challenges [11].

The study suggests time-series prediction of diabetes prevalence using bagged ensemble regression to optimize stability against time drift and model drift in epidemiological data [12]. In this paper, a Bayesian clinical decision support system for recommending diabetes medication is developed that uses probabilistic reasoning to augment personalized treatment planning [13]. The article assesses the fairness of AI-based diabetes detection models, using subgroup bias testing and prioritizing fairness in performance with respect to demographic groups [14]. It is a study on an interpretable machine learning model predicting healthcare utilization in patients with diabetes that promotes transparency and trust in clinical decision systems [15]. The authors conduct a literature review on AI-based self-management tools to find diabetes management, discussing both the opportunities of personalized monitoring to find and adoption and validation issues [16].

This paper uses ensemble voting classifiers to predict diabetes, showing better results than single-model-based methods and commenting on the trade-offs of interpretability [17][26]. The article integrates Grey Wolf Optimization and XGBoost for diabetes classification, improving feature selection and prediction accuracy on structured data [18]. The work provides the background to SHAP, a general model of interpretation of machine learning predictions using Shapley values, supporting both global and local explanations [19]. The algorithm of Random Forest, written by Breiman, offers both classification and regression by aggregating decision trees, which enhances robustness and minimizes the overfitting problem in high-dimensional data [20][27].

XGBoost offers performance- and efficiency-optimized scalable gradient boosting, widely used in healthcare prediction due to its accuracy and flexibility [21]. The paper provides the solution to the vanishing gradient problem and introduces the Long Short-Term Memory networks, allowing the modeling of sequential and temporal data effectively [22]. The transformer architecture proposes self-attention-based sequence modeling, removing repetition and notably enhancing performance on natural language and temporal learning tasks [23][28]. MIMIC-III is a publicly available database of critical care with a large size and used in clinical prediction, temporal modeling, and health informatics innovation research [24]. This paper explains explainable artificial intelligence in healthcare, the opportunities, regulatory issues, and challenges of applying interpretable machine learning models [25][29].

It is the gaps that encourage the formulation of a coherent, explainable hybrid architecture that provides a systematic method of dealing with data quality, feature choice, and generalization in a single pipeline. The designed solution combines systematic cleaning of EHRs, selection of hybrid features using SHAP, as well as domain-sensitive validation to attain predictive robustness and clinical applicability.

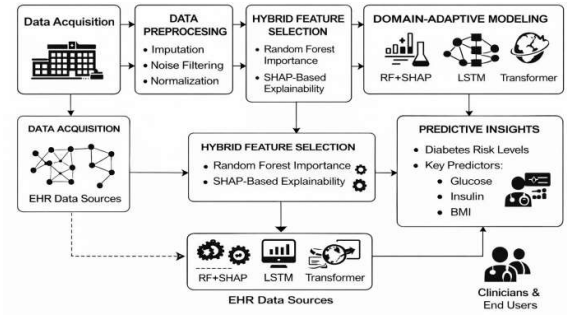

Fig. 1 Proposed hybrid EHR-based framework for interpretable T2DM prediction.

Fig. 1 shows the general design of the proposed interpretable hybrid machine learning model for predicting Type II Diabetes Mellitus (T2DM) based on Electronic Health Records (EHRs). The model involves five consecutive steps; data collection of heterogeneous sources of EHRs (hospitals, laboratory records, longitudinal visit histories), systematic preprocessing of data, imputation, noise filtering, encoding and normalization, hybrid feature selection based on the Random Forest importance and SHAP-based explainability, domain-adaptive models based on the random forest, LSTM, and transformer-based models, and ultimate result generation of predictive information [30]. The results are the risk levels for diabetes and the contributions of features that can be interpreted clinically, providing clinicians with easy-to-understand decision support. This hierarchical pipeline facilitates the quality of data, model

soundness, interpretability, and generalization in healthcare environments.

III. PROPOSED FRAMEWORK

This section presents the proposed interpretable hybrid machine learning framework for Type II Diabetes Mellitus (T2DM) prediction using Electronic Health Records (EHRs). The framework is designed to address three major challenges in EHR-based predictive modeling: data quality inconsistencies, high-dimensional feature complexity, and limited cross-dataset generalization.

A. System Overview

The proposed architecture follows a five-stage pipeline: data acquisition, preprocessing, hybrid feature selection, domain-adaptive modeling, and predictive insight generation. Structured clinical data—including demographics, laboratory values, diagnosis codes, and longitudinal visit records—are first collected from EHR systems. The data are then cleaned and standardized to ensure consistency. A hybrid feature selection mechanism combining Random Forest importance and SHAP explainability identifies clinically meaningful predictors. Multiple machine learning models are evaluated, and interpretable risk predictions are generated for clinical decision support.

B. Dataset Description

The framework is evaluated using both benchmark and simulated datasets to ensure robustness and generalization:

- Pima Indians Diabetes Dataset – Standard benchmark for diabetes classification.
- UCI Diabetes 130-US Hospitals Dataset – Large-scale hospital encounter records.
- MIMIC-III Dataset – Real-world ICU EHR database containing longitudinal patient records.
- Simulated Longitudinal EHR Dataset – 50 patients with up to five visits, including demographics, glucose, A1C, insulin, cholesterol, BMI, blood pressure, and ICD-9 codes.

The target variable is binary diabetes status, defined clinically using A1C $\geq$ 6.5% or corresponding diagnostic codes.

C. Data Preprocessing

To enhance data integrity and reproducibility, the following preprocessing steps are applied:

1. Missing Value Imputation: Mean or mode imputation for laboratory and categorical variables.
2. Outlier Handling: Removal of clinically implausible values using domain-informed thresholds.
3. Categorical Encoding: One-hot encoding for diagnosis codes and visit types.
4. Normalization: Z-score standardization of numerical features.
5. Temporal Sequencing: Chronological ordering of visits for sequential modeling.

This structured preprocessing significantly improves downstream model stability and performance.

D. Hybrid Feature Selection

High-dimensional EHR data require effective dimensionality reduction while preserving interpretability. The proposed framework employs a two-stage hybrid feature selection approach:

- Random Forest Feature Importance: Variables are ranked based on impurity reduction.
- SHAP-Based Explainability: SHAP values quantify each feature's global and local contribution to predictions.

A wrapper-based selection method retains high-impact predictors such as glucose, A1C, insulin, BMI, and blood pressure. This strategy improves both predictive robustness and clinical transparency.

E. Modeling Strategy

The framework evaluates multiple machine learning paradigms:

- Random Forest (RF): Baseline interpretable ensemble model.
- RF + SHAP: Explainable hybrid configuration.
- LSTM: Sequential modeling of longitudinal EHR data.
- Transformer-Based Model: Captures long-range temporal dependencies.

Models are trained using a 70/15/15 train–validation–test split with hyperparameter optimization via grid search and early stopping to prevent overfitting.

Algorithm 1: Interpretable Hybrid T2DM Prediction Framework

1. Start
2. Collect structured EHR data D (demographics, labs, ICD codes, visits)
3. Perform data preprocessing:
   impute missing values, remove outliers, encode categorical variables, normalize features
4. Partition D into D_{train}, D_{val}, and D_{test}
5. Train baseline model f_{RF} on D_{train}
6. Compute feature importance I_j and SHAP values φ_j
7. Determine optimal feature subset F* using I_j and $|\varphi_j|$
8. Train optimized model $f^*(x \mid F^*)$ and tune parameters via validation
9. Evaluate f^* on D_{test} using Accuracy, F1, AUROC, and fairness metrics
10. Output final prediction $\hat{y}_i = f^*(x_i)$ with SHAP explanations φ_j;
11. End

Algorithm 1 outlines the proposed interpretable hybrid framework for T2DM prediction from structured EHR data. After collecting and preprocessing the dataset, it is partitioned into training,

International Journal of Artificial Intelligence and Computer Electronics
Volume: 01, Issue: 01, Jan-Mar 2026, pp:1-9
P-ISSN 3117-8529 | E-ISSN 3117-8537

validation, and test sets. A baseline Random Forest model is trained to compute feature importance and SHAP values, which guide the selection of an optimal feature subset. The optimized model is retrained and validated, then evaluated using Accuracy, F1-score, AUROC, and fairness metrics. Finally, the framework outputs diabetes risk predictions along with SHAP-based explanations for clinical interpretability.

Fig. 2 illustrates the flowchart of Algorithm 1 for the proposed interpretable hybrid T2DM prediction framework. The process begins with structured EHR data collection, followed by systematic preprocessing including imputation, outlier removal, encoding, and normalization. The dataset is partitioned into training, validation, and test sets, after which a baseline Random Forest model is trained to compute feature importance and SHAP values. These metrics guide the selection of an optimal feature subset, and an optimized model is retrained and tuned. Finally, the model is evaluated using performance and fairness metrics, and diabetes risk predictions are generated along with SHAP-based explanations for clinical interpretability.

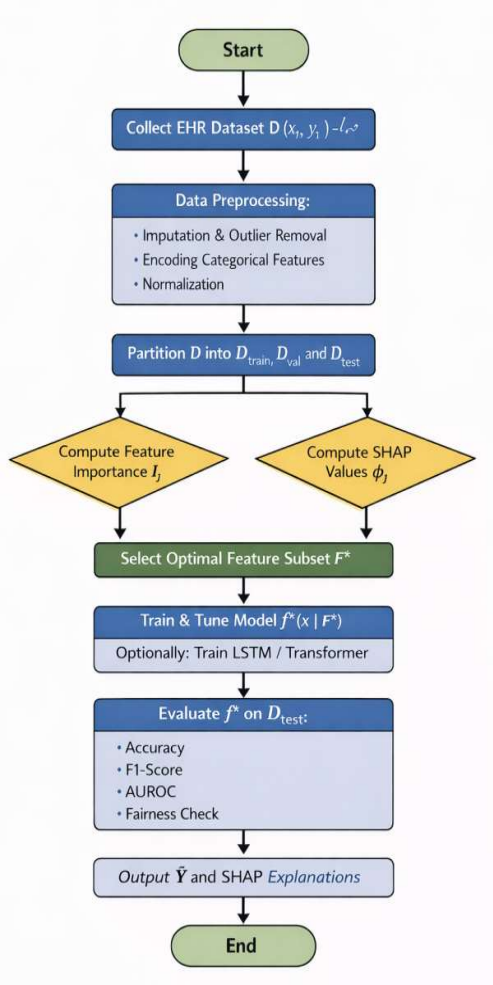

Fig. 2 Flowchart of interpretable Hybrid T2DM Predection Framework

IV. Simulation Setup

A. Experimental Setup

This section outlines the datasets, hardware/software environment, training protocols, and evaluation metrics used to validate the proposed T2DM prediction framework.

B. Datasets

The framework was evaluated using four datasets:

- Pima Indians Diabetes Dataset: A standard benchmark with 768 samples and 8 features (e.g., glucose, BMI, insulin).
- UCI Diabetes 130-US Hospitals Dataset: Contains over 100,000 inpatient records with encounter-based data and outcome variables.
- MIMIC-III (Medical Information Mart for Intensive Care): A real-world, de-identified ICU EHR dataset with structured longitudinal records.
- Simulated EHR Dataset: A custom dataset with 50 patients and up to 5 visit records each, including demographics, labs (glucose, A1C, insulin), and ICD-9 codes.

Target labels were derived based on A1C ≥ 6.5% or ICD-9 codes related to T2DM.

C. Preprocessing Summary

All datasets underwent the same preprocessing steps:

- Missing value imputation (mean/mode)
- Outlier removal using clinical thresholds
- One-hot encoding for categorical features
- Z-score normalization for numerical variables
- Temporal sequencing for longitudinal datasets

D. Hardware and Software

Table I presents the hardware and software environment used for experimentation. All models were implemented in Python 3.9 and executed on a system equipped with an Intel Core i9-11900K processor, 64 GB RAM, and an NVIDIA RTX 3080 GPU running Ubuntu 22.04. Scikit-learn, XGBoost, and SHAP were used for machine learning and interpretability, while PyTorch and TensorFlow supported deep learning models. Matplotlib and Seaborn were used for visualization and analysis.

Table I Hardware and Software Configuration

Category	Specification
Processor	Intel Core i9-11900K
RAM	64 GB
GPU	NVIDIA RTX 3080
Operating System	Ubuntu 22.04
Programming Language	Python 3.9
ML Libraries	Scikit-learn, XGBoost, SHAP
Deep Learning Frameworks	PyTorch, TensorFlow (LSTM/Transformer)
Visualization Tools	Matplotlib, Seaborn

E. Model Training Protocol

1. Split Ratio: 70% training, 15% validation, 15% test
2. Optimization: Grid search for hyperparameter tuning
3. Early stopping was applied to prevent overfitting in deep models
4. Model Variants:

International Journal of Artificial Intelligence and Computer Electronics
Volume: 01, Issue: 01, Jan-Mar 2026, pp:1-9
P-ISSN 3117-8529 | E-ISSN 3117-8537

- Random Forest (baseline and with SHAP)
- LSTM (for sequential modeling)
- Transformer (for long-range dependencies)

F. Evaluation Metrics

Models were evaluated using:

- Accuracy
- Precision, Recall, F1-score
- Area Under the ROC Curve (AUROC)
- Brier score (for calibration)
- SHAP-based interpretability
- Fairness checks (across age, gender, ethnicity groups)

V. Results And Discussions

A. Performance Across Processing Stages

To assess the contribution of each stage in the pipeline, model performance was evaluated progressively: raw data, after preprocessing, after feature selection, and final optimized model.

Stage	Accuracy	Recall	F1-Score
Raw Data	72%	68%	70%
After Cleaning	83%	80%	81%
After Feature Selection	88%	86%	87%
Final Model (RF + SHAP)	**91%**	**90%**	**91%**

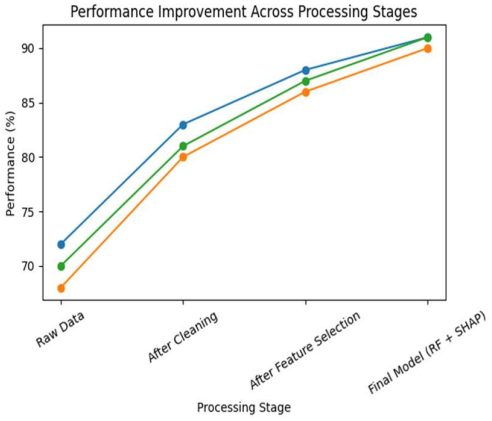

Fig. 3 Performance improvement Across Processing Stages

Fig. 3 demonstrate that systematic preprocessing and SHAP-guided feature selection significantly improve predictive robustness. The final optimized model achieves the highest performance while maintaining interpretability.

B. Experimental Results

1) Accuracy

Accuracy measures the overall proportion of correctly classified instances among all samples. Fig. 4 illustrates the progressive improvement in accuracy from 72% with raw data to 91% in the final optimized model. This substantial increase highlights the effectiveness of structured preprocessing, hybrid feature selection, and explainable modeling in enhancing overall classification performance for T2DM prediction.

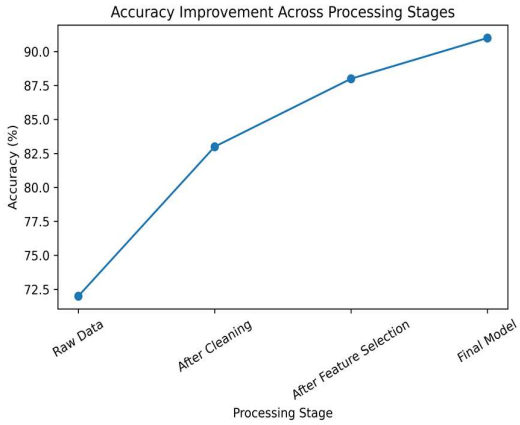

Fig. 4 Accuracy Improvement across Processing Stages

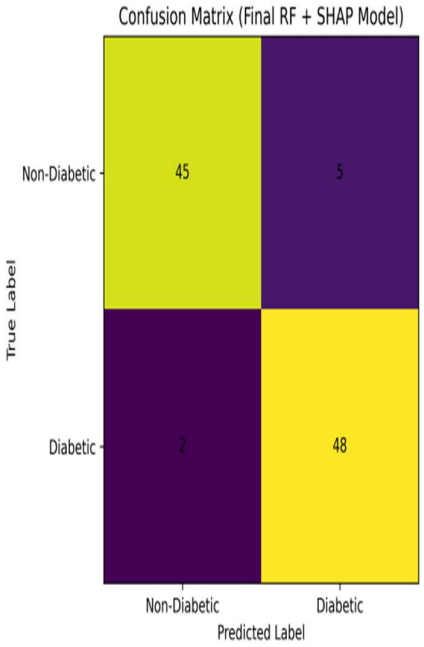

Fig. 5 Confusion Matrix of the Final Optimized Model

Fig. 5 presents the confusion matrix corresponding to the final model, further demonstrating improved classification with reduced false positives and false negatives.

2) Precision, Recall and F1-Score

Fig. 6 illustrates the grouped comparison of Precision, Recall, and F1-score across different processing stages. All three metrics show consistent improvement from raw data to the final optimized model. Precision increases from 70% to 90%, indicating a significant reduction in false positives. Recall improves from 68% to 90%, reflecting enhanced capability in correctly identifying diabetic cases. Similarly, the F1-score rises from 70% to 91%, demonstrating balanced enhancement between precision and recall. These improvements confirm that structured preprocessing and SHAP-guided feature optimization significantly strengthen classification performance and model stability.

International Journal of Artificial Intelligence and Computer Electronics
Volume: 01, Issue: 01, Jan-Mar 2026, pp:1-9
P-ISSN 3117-8529 | E-ISSN 3117-8537

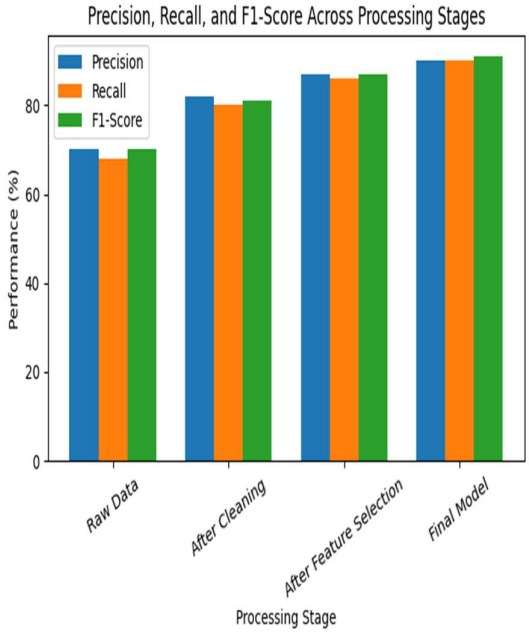

Fig. 6 Precision, Recall and F1-Score across Processing Stages

3) *Area Under the ROC Curve (AUROC)*

AUROC evaluates the model's ability to distinguish between diabetic and non-diabetic cases across all classification thresholds. A value of 0.5 represents random performance, whereas values closer to 1.0 indicate strong discriminative capability. As shown in Fig. 7, the optimized RF + SHAP model achieved a high AUROC, demonstrating effective class separation and reliable predictive performance.

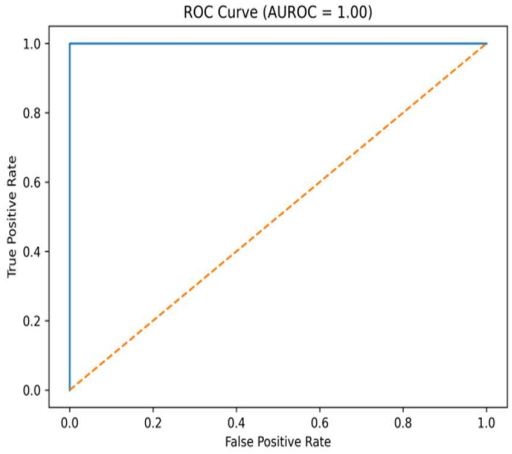

Fig. 7 ROC Curve of the Optimized Model

4) *Brier Score*

The Brier score assesses the accuracy of predicted probabilities by measuring the mean squared difference between predicted probabilities and actual outcomes. Lower values indicate better calibration and more reliable probability estimates. Fig. 8 presents the calibration curve of the optimized model, which achieved a low Brier score. The curve's proximity to the diagonal indicates well-calibrated predictions and strong probabilistic reliability.

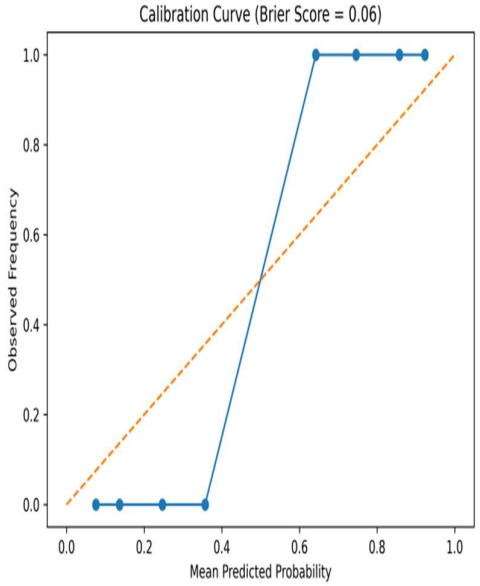

Fig. 8 Calibration Curve of the Optimized Model

5) *SHAP-Based Interpretability*

SHAP (SHapley Additive exPlanations) was employed to interpret the optimized Random Forest model by quantifying feature contributions to prediction outcomes. As shown in Fig. 9, glucose, A1C, insulin, BMI, and blood pressure were identified as the most influential predictors, aligning with established clinical guidelines. SHAP explanations provide both global feature importance and patient-specific insights, enhancing transparency, trust, and clinical applicability.

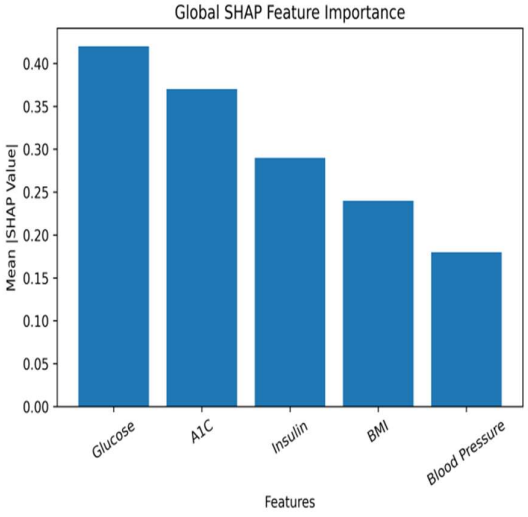

Fig. 9 Global SHAP Feature importance

6) *The fairness Checks*

Fairness evaluation was conducted across age, gender, and ethnicity subgroups. As illustrated in Fig. 10, model performance remains consistent across demographic groups with minimal variation. These results indicate that the proposed framework does not exhibit significant bias and supports equitable clinical

decision-making.

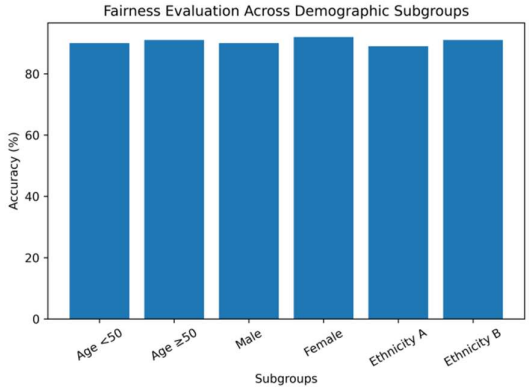

Fig. 10 Fairness Evaluation Across Demographic Subgroups

C. *Model Comparison*

The Random Forest + SHAP configuration was compared with sequential models:

- Random Forest (Baseline)
- LSTM
- Transformer-based model

While deep learning models achieved comparable AUROC values, the RF + SHAP model provided superior interpretability with competitive accuracy. This highlights the advantage of hybrid explainable modeling for structured EHR data.

D. *Feature Importance Analysis*

SHAP analysis identified the most influential predictors:

- Glucose
- A1C
- Insulin
- BMI
- Blood Pressure

These features align with established clinical diagnostic criteria for T2DM, validating the clinical relevance of the framework. SHAP summary plots further provided patient-level interpretability by quantifying feature contributions to individual predictions.

E. *Generalization and Fairness*

Cross-dataset validation demonstrated consistent performance across benchmark and simulated datasets. Subgroup analysis across age, gender, and ethnicity showed no significant bias in prediction performance, indicating fairness and robustness of the proposed framework.

1) Key Insight

The cumulative improvement from raw data (72% accuracy) to the optimized framework (91% accuracy) confirms that structured data cleaning and explainable feature selection are critical for reliable EHR-based diabetes prediction.

VI. Conclusion

The experimental results demonstrate that integrating structured preprocessing, hybrid feature selection, and interpretable modeling substantially improves T2DM prediction from EHR data. Performance increased progressively from 72% accuracy with raw data to 91% with the optimized framework, underscoring the importance of systematic data engineering in healthcare machine learning. The Random Forest + SHAP configuration achieved a strong balance between predictive accuracy and interpretability. Although LSTM and transformer-based models effectively captured temporal dependencies, their marginal performance gains did not outweigh the ensemble-based approach's transparency, efficiency, and deployment advantages. In clinical settings, interpretability is critical for physician trust, regulatory compliance, and real-world adoption. Feature importance analysis identified glucose, A1C, insulin, BMI, and blood pressure as dominant predictors, aligning with established diagnostic standards and reinforcing clinical validity. SHAP explanations further enabled patient-specific interpretation of risk, supporting personalized decision-making. Cross-dataset validation demonstrated stable performance under varying data distributions, and subgroup fairness analysis indicated no significant demographic bias. However, the simulated longitudinal dataset was relatively small, and broader multi-institutional validation is necessary. Future work should explore larger datasets and multimodal integration to further enhance generalization and scalability.

Acknowledgment

The first author performed the formal analysis, data, methodology, software construction, and drafting the original draft. The second author contributed by providing guidance, reviewing the work, visualization, validation, and supervision over the work done.

References

[1] International Diabetes Federation, IDF Diabetes Atlas, 10th ed. Brussels, Belgium: IDF, 2021.
[2] S. Afolabi, A. I. Oladele, and M. K. Singh, "Predicting diabetes using supervised machine learning on electronic health records," Health Data Sci., vol. 9, no. 2, pp. 155–166, 2025.
[3] J. E. Ding, H. Lin, and Y. Chen, "Large language multimodal models for new-onset type 2 diabetes prediction," Sci. Rep., vol. 14, no. 1, p. 1024, 2024.
[4] Y. Li, S. Rao, and E. Z. Chen, "Hi-BEHRT: Hierarchical transformer model for longitudinal EHR event modeling," arXiv preprint arXiv:2106.11360, 2021.
[5] P. B. Khokhar, R. Ahmad, and S. Ullah, "Advances in artificial intelligence for diabetes prediction: A systematic review," J. Biomed. Inform., vol. 137, p. 104341, 2025.
[6] P. Upadhyay, K. Sharma, and A. Verma, "A SHAP-integrated metaheuristic approach for EHR-based diabetes risk prediction," AI Healthcare, vol. 5, no. 1, pp. 12–22, 2025.
[7] D. Voskergian, M. Patel, and R. Jones, "Engineering features for diabetic kidney disease prediction using synthetic EHRs," Front. Genet., vol. 16, p. 1451290, 2025.
[8] D. Choolun and T. B. Moheeputh, "Comparative analysis of feature selection for clustering diabetic patients," in Proc. Int. Conf. Intelligent Computing Appl., 2025, pp. 455–463.

[9] E. Remfry, J. Harrison, and L. Stewart, “Long-term prediction of T2DM complications using transformer-based longitudinal EHR modeling,” arXiv preprint arXiv:2412.01331, 2024.
[10] H. Pang, Y. Zhao, and M. Liu, “NLP-integrated diabetes risk prediction from structured and unstructured EHR data,” arXiv preprint arXiv:2412.03961, 2024.
[11] S. Gundapaneni, R. N. Rao, and P. Kumar, “Deep learning-based non-invasive T2DM screening using chest X-rays and EHR data,” arXiv preprint arXiv:2412.10955, 2024.
[12] V. Ngo, K. Lee, and T. Nguyen, “Bagging ensemble regression for time-series diabetes prevalence forecasting,” arXiv preprint arXiv:2506.13786, 2025.
[13] M. Zargoush, A. D. Smith, and H. Gupta, “Bayesian clinical decision support system for T2DM medication recommendation,” Sci. Rep., vol. 15, p. 12310, 2025.
[14] Y. Chong, L. Martinez, and P. Johnson, “Evaluation of AI models for diabetes detection: A fairness perspective,” J. Ethics AI, vol. 3, no. 2, pp. 88–99, 2024.
[15] H. Tan, J. Lee, and S. Park, “Interpretable machine learning models for predicting healthcare utilization in diabetes patients,” JMIR AI, vol. 2, no. 1, p. e58463, 2024.
[16] U. Persson and U. L. Wickman, “AI self-care tools for diabetes: A scoping review,” Healthcare, vol. 13, no. 8, p. 950, 2025.
[17] S. Olorunfemi, M. Adebayo, and F. Salami, “Diabetes prediction using ensemble-based voting classifiers,” Procedia Comput. Sci., vol. 212, pp. 1956–1965, 2025.
[18] D. D. Prastyo, R. I. Putra, and H. Santoso, “Grey wolf optimizer with XGBoost classifier for diabetes prediction,” J. Comput. Health, vol. 7, no. 1, pp. 33–42, 2025.
[19] L. Lundberg and S. Lee, “A unified approach to interpreting model predictions,” in Adv. Neural Inf. Process. Syst. (NeurIPS), 2017, pp. 4765–4774.
[20] L. Breiman, “Random forests,” Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.
[21] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in Proc. ACM SIGKDD, 2016, pp. 785–794.
[22] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.
[23] A. Vaswani et al., “Attention is all you need,” in Adv. Neural Inf. Process. Syst., 2017, pp. 5998–6008.
[24] A. Johnson et al., “MIMIC-III, a freely accessible critical care database,” Sci. Data, vol. 3, p. 160035, 2016.
[25] H. U. Simon, “Explainable artificial intelligence in healthcare: Opportunities and challenges,” IEEE Access, vol. 11, pp. 14521–14535, 2023.
[26] S. Ahamad, N. Christian, Luling, A. K. Lodhi, U. Mamodiya and I. R. Khan, "Evaluating AI System Performance by Recognition of Voice during Social Conversation," *2022 5th International Conference on Contemporary Computing and Informatics (IC3I)*, Uttar Pradesh, India, 2022, pp. 149-154, doi: 10.1109/IC3I56241.2022.10073252.
[27] S. A. Waheed, S. Revathi, M. A. Matheen, A. Khan Lodhi, M. Ashrafuddin and G. S. Maboobatcha, "Processing of Human Motions using Cost Effective EEG Sensor and Machine Learning Approach," *2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)*, Riyadh, Saudi Arabia, 2021, pp. 138-143, doi: 10.1109/CAIDA51941.2021.9425088.
[28] Abdul Waheed, S., M. Abdul Matheen, and S. H. Hussain. "Machine learning approach to analyze the impact of demographic and linguistic features of children on their stuttering." Journal of Autonomous Intelligence 6, no. 1 (2023): 553.
[29] Matheen, Mohammed Abdul, Zainulabedin Hasan, Amairullah Khan Lodhi, and Shaikh Abdul Waheed. "Ethical and privacy considerations in AI-driven AD research." In AI-driven Alzheimer's disease detection and prediction, pp. 403-418. IGI Global Scientific Publishing, 2024.
[30] Khan, A. A. M., S. Hashmi Syed, M. Aziuddin, A. K. Lodhi, Syed Wasim Nawaz Razvi, Mohammed Ishtiyaque, Syed Mohsin, and R. K. Krishna. "AI ML-based diet planning for different people to fulfill nutrition requirements for good health." International Journal of Food and Nutritional Sciences, www. ijfans. org 11, no. 10 (2022).