Journal of Medical and Health Studies (JMHS)
ISSN: 2710-1452
DOI: 10.32996/jmhs
Journal Homepage: www.al-kindipublisher.com/index.php/jmhs

JMHS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

AI-Driven Glycemic Instability Risk Modeling for Proactive Intervention and Chronic Disease Management in U.S. Healthcare Systems

Rabi Sankar Mondal[1]✉, Tamjida Nasreen Purba[2], Nafisa Nusrat Purba[3], Md Moshiur Rahman[2], Tawfiqur Rahman Sikder[4]
[1]*Pompea College of Business, University of New Haven, West Haven, Connecticut, USA*
[2]*Cullen College of Engineering, University of Houston, Houston, Texas 77054, USA*
[3]*Department of Information Systems, Lamar University, Beaumont, Texas 77710, USA*
[4]*School of Business, International American University, Los Angeles, California, USA*
Corresponding Author: Rabi Sankar Mondal, **E-mail**: rabi.s.mondal@gmail.com

| ABSTRACT
Hospitals with diabetic patients experience serious clinical problems because patients with diabetes develop glycemic instability which causes their blood glucose levels to fluctuate unpredictably. The process of identifying high-risk patients should start immediately because it requires accurate detection methods which help to protect patient safety. Healthcare experts face multiple difficulties when attempting to predict glycemic instability because clinical data exhibits extreme class imbalance and laboratory results plus medications and patient characteristics show complex interactions. The UCI 130-hospital diabetes dataset serves as the foundation for our complete machine learning and deep learning system which we developed to predict glycemic instability risk. The combination of Synthetic Minority Over-sampling Technique (SMOTE) with cost-sensitive learning provides us with a solution to tackle the challenges that arise from extreme class imbalance. The seven predictive models which include Logistic Regression, Random Forest, Gradient Boosting, XGBoost, Support Vector Machine, Multi-Layer Perceptron (MLP) and TabNet use a clinically informed decision threshold which helps them to detect medical conditions with high accuracy. The evaluation of model performance examines five metrics which include accuracy, precision, recall, F1-score and ROC-AUC. The experimental results show that machine learning through deep learning and ensemble methods achieves better results for detecting glycemic instability than traditional classifiers. The deep learning models TabNet and MLP exhibit high sensitivity for detecting unstable patients while Gradient Boosting and Random Forest demonstrate superior discriminative ability with ROC-AUC values close to 0.99. The analysis of feature importance shows that HbA1c levels, maximum glucose concentration, number of diagnoses, and insulin usage are the most influential predictors of instability. The research results demonstrate that models which use imbalance-aware machine learning together with explainable models can achieve accurate predictions.

| KEYWORDS

Healthcare System Reform; Health Equity and Disparities; Value-Based Care; Health Information Interoperability; Public Health Infrastructure

| ARTICLE INFORMATION

ACCEPTED: 25 December 2025 **PUBLISHED:** 09 January 2026 **DOI:** 10.32996/jmhs.2023.4.6.21

1.0 Introduction
Diabetes mellitus is among the most prevalent chronic illnesses in the United States, affecting over 37 million patients and contributing substantially to the total healthcare expenditure (Aranaz et al., 2023). Glycemic variability, which is described as recurrent and irregular fluctuations between hypo- and hyperglycemic episodes, is a serious clinical issue in the management of diabetes mellitus. This variability has been closely associated with hospitalization, mortality, and microvascular as well as macrovascular complications (Ceriello et al., 2022; Monnier et al., 2023). The current state of hospital glucose control remains largely reactive, with treatment efforts being made only after the occurrence of adverse glucose events, despite the widespread use of electronic health records (EHRs).

Healthcare providers can now use active diabetes treatment because artificial intelligence (AI) systems have reached their current advanced state. The machine learning models use extensive EHR data which includes laboratory results and medication records and diagnosis information and patient hospitalization data to detect patients who have an increased risk of glycemic instability before they reach critical health states (Tanvir et al., 2020; Hasan et al., 2023; Kasula, 2024). The early detection process helps clinicians modify insulin treatment while they develop better medication plans and raise monitoring levels which leads to fewer complications and safer outcomes for patients (Sağlam et al., 2022). The process of predicting glycemic instability faces significant technical obstacles. Most patient cases present stable conditions but instability events occur at rare intervals which creates dataset imbalances that force traditional classifiers to select the majority class as their prediction output (Sağlam et al., 2022). The laboratory test results and medication usage patterns and clinical history data interact in complex ways which lead to unpredictable outcomes that conventional statistical models fail to manage (Rahman et al., 2022). The complex problems require the implementation of advanced ensemble methods together with deep learning approaches which need special training techniques to handle data imbalance (Arafa et al., 2022; Monnier et al., 2025).

This work applies hospital records in the United States to offer an AI-based paradigm for the risk of glycemic instability to address these issues. In a learning pipeline that is aware of imbalance and therapeutically optimized, seven machines learning and deep learning models are evaluated: Logistic Regression, Random Forest, Gradient Boosting, XGBoost, Support Vector Machine, Multi-Layer Perceptron, and TabNet. In order to prioritize patient safety, the model design includes cost-sensitive learning, SMOTE oversampling, and a decision threshold that is informed by clinical considerations. Based on the experimental results, ensemble learning methods are superior to other clinical and predictive performances for proactive glycemic risk management.

2.0 Literature Review

Diabetes risk prediction research in its initial phase used statistical methods together with rule-based systems to analyze three clinical factors which were insulin dosage and body mass index and fasting glucose levels. The techniques provided useful descriptive information but failed to comprehend how actual hospital data exhibited both time-dependent changes and complex physiological connections. The field of diabetes prediction and management has experienced major changes because of recent artificial intelligence developments. The systematic review conducted by Khokhar et al. demonstrates how deep learning and machine learning techniques for diabetes prediction have developed rapidly which shows their potential for identifying diseases at early stages and providing diagnostic assistance and supporting clinical workflows (Khalid et al., 2025).

Multiple research studies have demonstrated that ensemble machine learning models together with traditional machine learning models achieve better results when applied to clinical data sets, which enhances their ability to diagnose medical conditions. Abousaber et al. (2025) presented a strong predictive system for diabetes identification through their research which demonstrated that SMOTE resampling methods together with Random Forest and Gradient Boosting ensemble models improved both predictive accuracy and ability to detect uncommon events in unbalanced diabetes data sets. Talebi Moghaddam et al. conducted a cohort study to identify key diabetes risk factors within unbalanced diabetes data while they emphasized the importance of using imbalance detection methods for precise prediction Talebi Moghaddam et al. conducted a cohort study to identify key diabetes risk factors within unbalanced diabetes data (Moghaddam et al., 2024).

Researchers have developed Explainable AI (XAI) as a tool which helps multiple ML frameworks to achieve better results concerning clinical data analysis. The combination of SMOTE with SHAP input features enables high predictive results and maintains essential feature importance needed for clinical decision support according to a recent study (Netayawijit et al., 2025). Deep learning models have extended the abilities which conventional ML methods provide to users. Zafar et al. developed a new deep learning model for early diabetes prediction which surpassed baseline testing results through its use of explainability methods and its application of oversampling techniques (Zafar et al., 2025). The research work by Khan et al. used Tabular Transformer networks (TabTrans) to show how advanced systems can discover complex relationships within clinical information, which they applied to diabetes risk assessment through multimodal health record data analysis (Khan et al., 2026).

Beyond prediction, ML has been applied to examine aberrant glycemic control and risk variables in particular patient subgroups. For instance, Kim et al. demonstrated the therapeutic usefulness of machine learning for individualized risk stratification by using ML techniques to find risk factors for hyperglycemia and hypoglycemia in diabetic cancer patients (Kim et al., 2023). Rishad et al. (2025) demonstrated how advanced machine learning models can uncover actionable cancer insights and improve disease prediction accuracy, highlighting the transformative role of AI in early detection and precision healthcare applications. In general, recent research emphasizes how crucial it is to combine explainable models, advanced architectures, and imbalance-aware learning in order to increase predictive accuracy while maintaining clinical interpretability. However, this study attempts to close a gap in the literature by concentrating on broad diabetes diagnosis rather than dynamic glucose instability in hospital settings.

3.0 Methodology
Our AI-driven framework designed to enhance early risk prediction and support more informed clinical decision-making in healthcare systems. This approach has broad implications for improving patient outcomes, optimizing resource allocation, and strengthening data-driven strategies that promote efficiency, equity, and long-term sustainability in healthcare (Alam et al., 2024). The proposed architecture of the AI-driven shows glycemic instability risk prediction framework. To identify high-risk patients for early intervention, the pipeline combines preprocessing, several machines learning and deep learning models, imbalance-aware learning (SMOTE and cost-sensitive weighting), electronic health records, and threshold-optimized risk prediction (Figure 1).

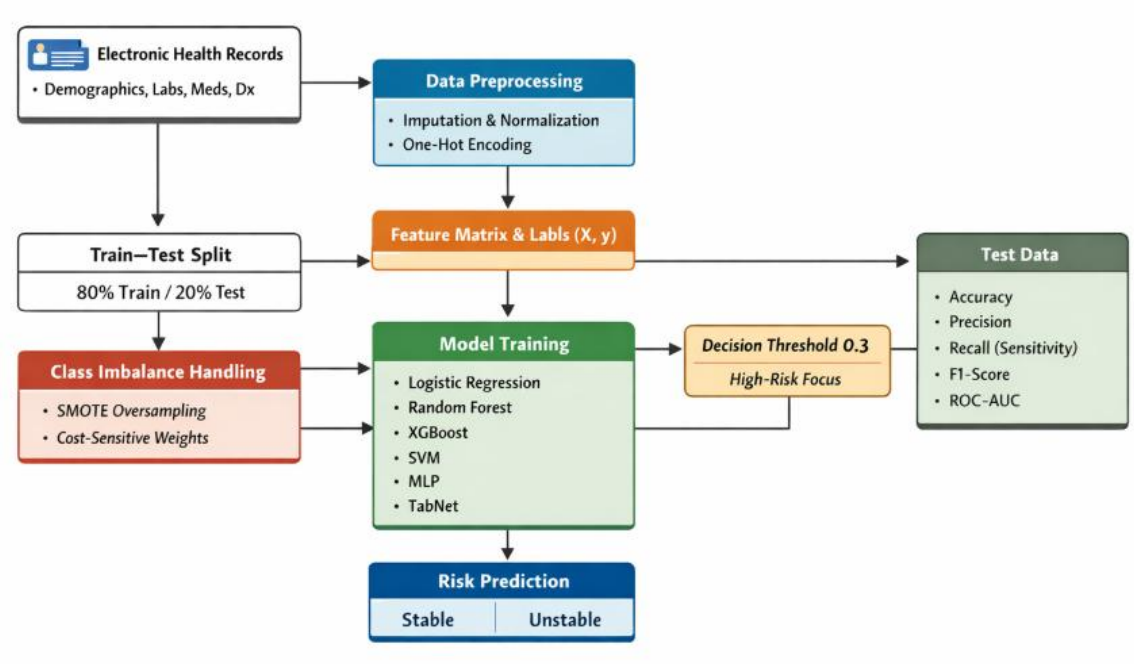

Figure 1. Proposed Architecture of the AI-driven Glycemic Instability Risk Prediction

3.1 Dataset and Study Cohort
The dataset employed in this study is the UCI Diabetes 130-Hospital dataset (Strack et al., 2014), which is a large-scale electronic health record (EHR) dataset that comprises more than 100,000 inpatient records from 130 U.S. hospitals between 1999 and 2008. Each record in the dataset corresponds to the hospital admission of a diabetic patient and includes demographic data, lab results, medication use, diagnoses, and hospital outcomes. To make the dataset relevant to our research question, we introduced a new clinical outcome measure known as Glycemic Instability Risk (GIR). Contrary to previous research, instead of predicting hospital readmission, we categorized patients as having glycemically unstable conditions when there was abnormal glucose activity during hospitalization, as indicated by blood glucose levels and medication changes. This modification enables the dataset to be used as a binary classification problem, where the objective is to determine patients who are at risk of having unstable glycemic conditions that necessitate active clinical management (Kabir et al., 2023).

3.2 Data Preprocessing
The raw data had a combination of categorical and numerical attributes, as well as missing and undefined values. All the invalid values like "?" were initially considered as missing values and were imputed using the median for numerical variables and the most frequent category for categorical variables. Nominal variables like race, gender, type of admission, and types of medications were then converted to a machine-readable format using one-hot encoding. To guarantee that variables with different scales contributed equally to the learning of models, all the numerical variables like laboratory test results and hospital stay duration were normalized using Min-Max scaling. Finally, the preprocessed data was split into a training set and a testing set in an 80:20 ratio using stratified sampling to maintain the original class distribution of glycemic stability and instability for both sets.

3.3 Handling Class Imbalance
Predictive models may be biased toward the majority class due to the extremely unbalanced classification problem caused by the much lower frequency of glycemic instability events compared to stable glycemic outcomes. During model training, two complimentary imbalance-handling techniques were used to solve this problem and guarantee accurate identification of high-risk patients. Initially, the training dataset was subjected to the Synthetic Minority Over-sampling Technique (SMOTE) only in order to produce artificial instances of patients with glycemic instability. In addition to creating a more balanced training set and increasing

the representation of the minority class, this procedure also prevented information from leaking into the test data, enabling models to more effectively learn the patterns linked to instability.

Additionally, by giving the unstable class greater weights during training, cost-sensitive learning was included into all relevant classifiers. This forced the models to emphasize sensitivity toward glucose instability and enhanced their capacity to identify patients who might need proactive clinical intervention by penalizing misclassification of high-risk patients more severely than stable ones.

3.4 Model Architecture and Training

Seven common machines learning and deep learning algorithms, namely Logistic Regression, Random Forest, Gradient Boosting, XGBoost, Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), and TabNet, were employed to evaluate the predictive capabilities of the proposed framework. The choice of these algorithms was made to represent a broad spectrum of linear, ensemble, kernel, and deep learning methods, enabling a comprehensive analysis of their ability to model complex, nonlinear relationships in clinical data. To prevent overfitting, logistic regression was applied with L2 regularization. The LIBLINEAR optimizer was employed for optimization, and class weights were incorporated to handle class imbalance. To improve stability and reduce variance, Random Forest was configured with 500 decision trees, a maximum depth of 20, and bootstrap sampling. By using 300 boosting stages, a learning rate of 0.05, and a maximum tree depth of 5, gradient boosting allowed the model to minimize overfitting while progressively fine-tuning decision limits. In order to increase generalization and computational efficiency, XGBoost was trained using 400 gradient-boosted trees with a learning rate of 0.05, a maximum depth of 6, and subsampling of 80% of training instances and features.

A radial basis function (RBF) kernel was used by the Support Vector Machine to identify nonlinear patterns in the feature space. Cross-validation was used to optimize the kernel width (γ) and regularization parameter (C), and class weighting was used to boost sensitivity to glycemic instability. Three completely linked hidden layers with 128, 64, and 32 neurons each made up the Multi-Layer Perceptron. To reduce overfitting, dropout regularization and ReLU activation were applied to each layer. The Adam optimizer was used to train the network with a batch size of 128 and a learning rate of 0.001. The TabNet model, a deep learning model designed specifically for tabular data, consisted of five decision steps, 64 decision and attention dimensions, and sparse regularization to encourage feature selection. The model was trained using the Adam optimizer with early stopping on validation loss to prevent overfitting while still allowing for interpretability via its attention mechanism. To counter class imbalance, all models were trained on the SMOTE-sampled and class-weighted training set, but only the unsampled test set was used for evaluation to provide a fair and unbiased estimate of model performance. Stratified cross-validation on the training set was employed to select the hyperparameters for each model, ensuring that model comparison was reliable and repeatable.

During prediction, a clinically driven decision threshold was employed in addition to standard probabilistic training. As glycemic instability detection is a medical problem that necessitates safety, a lower decision threshold of 0.3 was employed in place of the standard probability threshold of 0.5. Although there might be a slight increase in the number of false positives, this approach prioritizes sensitivity and ensures that high-risk patients are less likely to be missed. Table 1 provides a summary of the hyperparameter configurations applied to each benchmarked model. These configurations were used to control model complexity while striking a balance between generalization and predictive performance.

Table 1. Hyperparameter settings of the evaluated models.

Model	Key Hyperparameters
Logistic Regression	Penalty = L2, Solver = LIBLINEAR, Max iterations = 1000, Class weight = balanced
Random Forest	Number of trees = 500, Maximum depth = 20, Minimum samples per leaf = 5, Bootstrap = True, Class weight = balanced
Gradient Boosting	Number of estimators = 300, Learning rate = 0.05, Maximum tree depth = 5, Subsample = 0.8
XGBoost	Number of trees = 400, Learning rate = 0.05, Maximum depth = 6, Subsample = 0.8, Column sample = 0.8, Regularization (L2) = 1.0
Support Vector Machine (SVM)	Kernel = RBF, Regularization parameter (C) = 10, Gamma = scale, Class weight = balanced
Multi-Layer Perceptron (MLP)	Hidden layers = [128, 64, 32], Activation = ReLU, Dropout = 0.3, Optimizer = Adam, Learning rate = 0.001, Batch size = 128, Epochs = 100, Early stopping = Yes
TabNet	Decision dimension = 64, Attention dimension = 64, Number of steps = 5, Sparsity regularization = 1e-4, Optimizer = Adam, Learning rate = 0.001, Batch size = 256, Early stopping patience = 20

3.5 Evaluation Metrics

Multiple complimentary evaluation metrics were used to thoroughly examine model performance under the significant class imbalance inherent in predicting glycemic instability. Accuracy can be deceptive in medical data that is unbalanced, even when it represents overall correctness.

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

Precision assesses the dependability of high-risk alarms.

$$Precision = \frac{TP}{TP + FP}$$

Recall (sensitivity) assesses the ability to recognize unstable patients and is the most clinically important parameter

$$Recall = \frac{TP}{TP + FN}$$

F1-score offers a balanced indicator of precision and recall.

$$F1 - Score = 2 \times \frac{Precison \times Recall}{Precision + Recall}$$

ROC-AUC assesses the model's discrimination capability across all thresholds. Recall and ROC-AUC were emphasized because missing an unstable patient poses greater clinical risk than generating false alarms.

4.0 Results and Discussion

4.1 Overall Model Performance Comparison Under Class Imbalance

An overview of each machine learning and deep learning model's performance provided on the UCI diabetes dataset (Table 2). When diagnosing glycemic instability under class imbalance, the results show significant heterogeneity in model behavior. With a ROC-AUC of 0.993, gradient boosting demonstrated the best overall discrimination ability, separating stable and unstable patients with remarkable accuracy. Additionally, it maintained a balanced trade-off between precision (0.633) and recall (0.514), yielding the greatest F1-score (0.567) of all the models that were examined. This implies that complicated nonlinear interactions in clinical data are especially well-captured by ensemble boosting techniques. High sensitivity to glycemic instability was suggested by the very high recall values (0.919 and 0.946, respectively) achieved by Support Vector Machines and Logistic Regression. Unfortunately, this came at the price of very low precision (0.030 and 0.026), meaning that many patients were wrongly classified as high risk. In a real-world setting, these models would produce far too many false positives, although they are very conservative and would avoid missing unstable patients.

In comparison to linear and kernel models, Random Forest, XGBoost, and TabNet showed moderate recall (0.51-0.68) but much higher precision. With precision 0.905, recall 0.514, and ROC-AUC 0.980, Random Forest performed best in terms of balance, indicating robust risk prediction with fewer false positives. Deep learning models are capable of identifying complex risk patterns, although they might require more calibration to avoid false positives, as TabNet showed good recall (0.676) but poor precision (0.243). The difficulty of training dense neural networks on very imbalanced tabular medical data without much calibration effort is evident in MLP's moderate recall (0.595) but very low precision (0.099).

Table 2. Performance comparison of machine learning and deep learning models.

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Gradient Boosting	**0.999**	**0.633**	0.514	**0.567**	**0.993**
Random Forest	0.999	0.905	0.514	0.655	0.980
XGBoost	0.996	0.277	0.622	0.383	0.892
TabNet (Deep Learning)	0.996	0.243	0.676	0.357	0.980
MLP	0.989	0.099	0.595	0.169	0.946
SVM	0.946	0.030	**0.919**	0.058	0.982
Logistic Regression	0.934	0.026	**0.946**	0.050	0.979

4.2 Discrimination Performance Analysis Using ROC Curves

The ROC curves indicate that the ensemble models, particularly Random Forest and Gradient Boosting, have the steepest curves, indicating their ability to discriminate. The models moved towards higher sensitivity when the probability threshold of 0.3 was applied, as it improved the detection of unstable patients at the expense of increased false positives (Figure 2). In medical situations, where it is more detrimental to miss a high-risk patient than to generate more alarms, this trade-off is justified. The curves show how machine learning and deep learning models can discriminate between different types of glycemic instability, with ensemble approaches doing better.

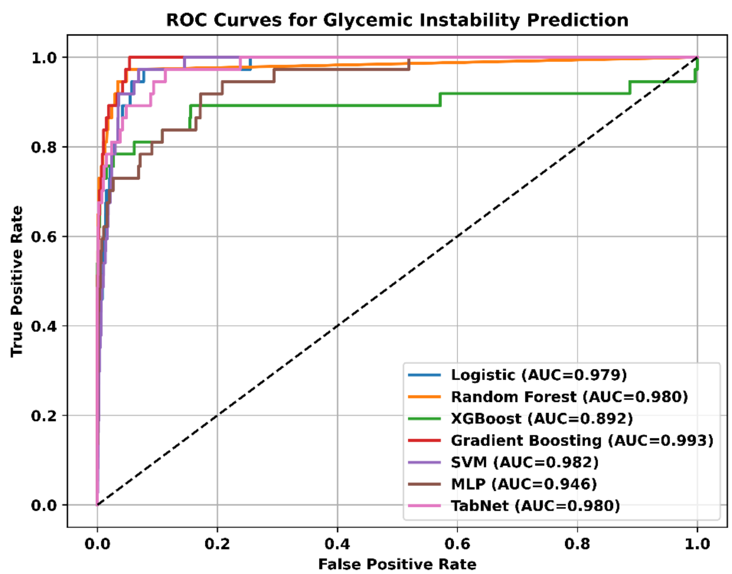

Figure 2. Receiver operating characteristic (ROC) curves for all evaluated models.

4.3 Confusion Matrix Analysis and Clinical Risk Profiles

The confusion matrix showed for all the models. The matrices highlight trade-offs between sensitivity and false-alarm rates among models by displaying the distribution of true positives, false positives, true negatives, and false negatives. Different models have different clinical risk profiles, according to confusion matrix analysis (Figure 3). SVM and logistic regression are too sensitive for real-world hospital deployment since they generate a lot of false positives while minimizing false negatives. Random Forest and Gradient Boosting, on the other hand, are more appropriate for real-world screening systems since they considerably lower false alarms while preserving sufficient sensitivity.

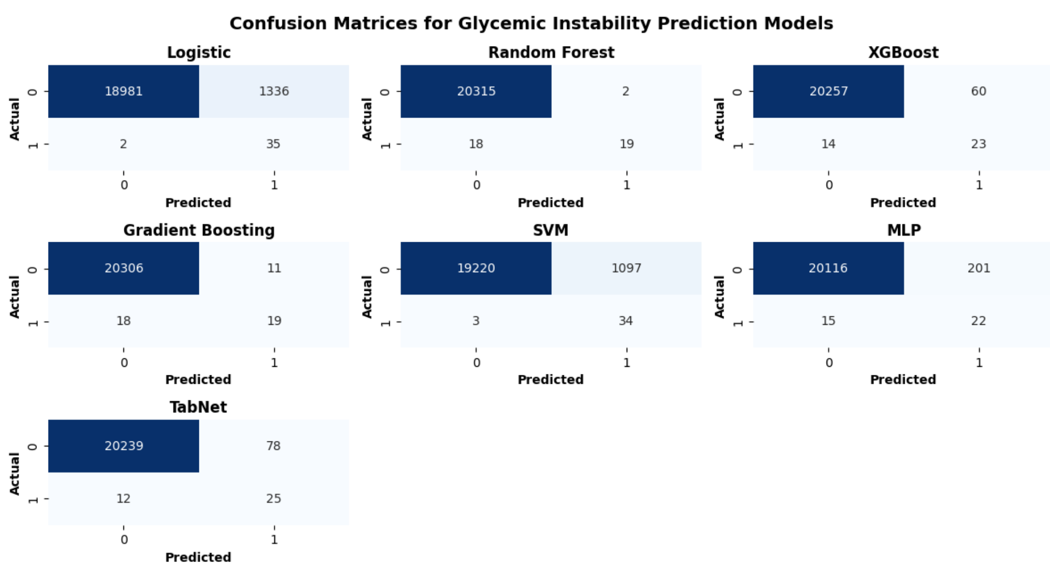

Figure 3. Confusion matrices of the evaluated models.

Laboratory glucose readings, insulin use, prescription changes, hospital length of stay, and previous diagnoses were consistently among the most predictive variables, according to feature importance analysis across several models. This supports the well-established clinical understanding that acute physiological stress during hospitalization, treatment modifications, and medication adherence all affect glycemic control. The suggested framework's clinical validity is reinforced by the features' consistency across deep learning and tree-based models.

4.4 Comparative ROC-AUC Analysis Across Models

All machine learning and deep learning models' ROC-AUC values for predicting glycemic instability are compared in this study (Figure 4). The ability of a model to differentiate between stable and unstable patients across all potential decision thresholds is reflected in ROC-AUC. Models with superior ROC-AUC values demonstrate improved discrimination abilities and enhanced capacity to assess different risk levels. The superior performance of ensemble models demonstrates their capability to handle complex clinical patterns successfully. Gradient Boosting achieved the highest accuracy with a ROC-AUC of 99.3%, indicating excellent discrimination capability. Support Vector Machine followed at 98.2%, while Random Forest and TabNet both achieved 98.0%, demonstrating strong and consistent performance. Logistic Regression performed slightly lower at 97.9% but remained highly effective. In contrast, Multilayer Perceptron reached 94.6%, and XGBoost recorded the lowest score at 89.2%. Overall, ensemble tree-based methods outperformed other approaches, suggesting they are particularly well-suited for detecting glycemic instability in this dataset.

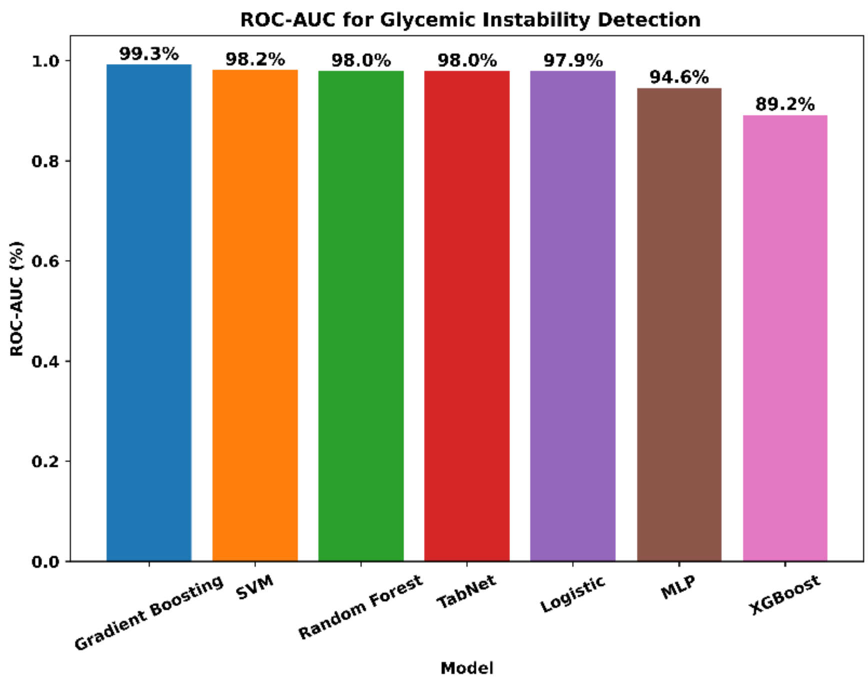

Figure 4. ROC-AUC Comparison of Models

4.5 Sensitivity (Recall) Comparison for High-Risk Patient Detection

The models show different levels of sensitivity when they detect patients who have glucose instability (Figure 5). The clinical environment requires high recall because it protects patient safety through its ability to detect all high-risk patients. The models which include Logistic Regression and SVM reach extremely high recall rates but they experience a rise in false positive results. The sensitivity range of deep learning models and ensemble models delivers better balance than other models. Logistic Regression achieved the highest recall at 94.6%, indicating superior sensitivity in identifying positive cases. SVM followed closely with 91.9%, also demonstrating strong detection capability. TabNet showed moderate performance at 67.6%. XGBoost and MLP achieved 62.2% and 59.5%, respectively. Gradient Boosting and Random Forest recorded the lowest recall, both at 51.4%. Overall, while some models showed excellent sensitivity, others with high ROC-AUC scores exhibited relatively lower recall, highlighting differences in true positive detection performance.

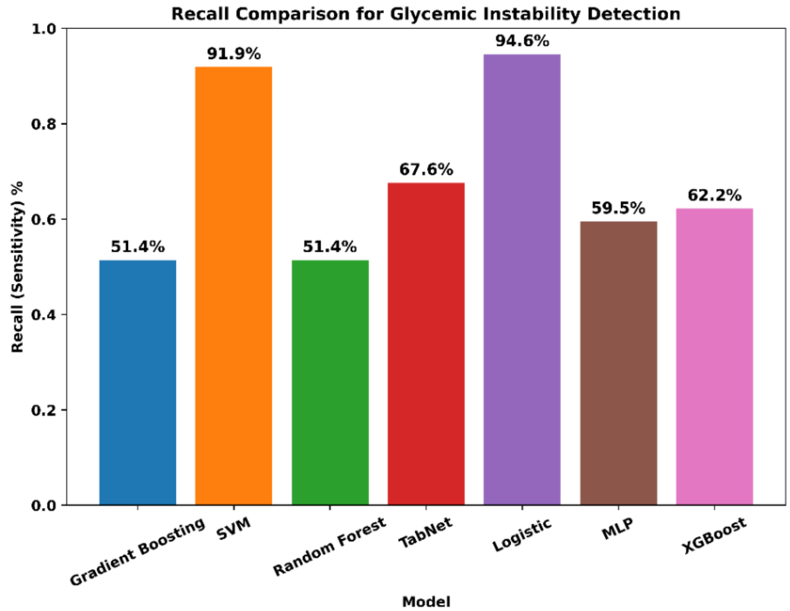

Figure 5. Recall (Sensitivity) Comparison of Models

4.6 F1-Score Evaluation Under Extreme Class Imbalance

The F1-scores showed which all tested models achieved to measure their precision and recall performance under conditions of extreme class imbalance (Figure 6). The F1-score shows improved results because it measures the balance between detecting unstable patients and reducing incorrect alarms. The performance of ensemble learning methods exceeds that of most other techniques because these methods demonstrate strong performance on medical prediction tasks which involve imbalanced data. Random Forest achieved the highest F1 score at 65.5%, indicating the best overall balance. Gradient Boosting followed with 56.7%, showing strong combined performance. XGBoost reached 38.3%, while TabNet scored 35.7%. MLP achieved 16.9%, demonstrating limited balance. SVM and Logistic Regression recorded the lowest F1 scores at 5.8% and 5.0%, respectively. Despite high recall or ROC-AUC in some models, their low F1 scores suggest imbalanced precision and recall performance.

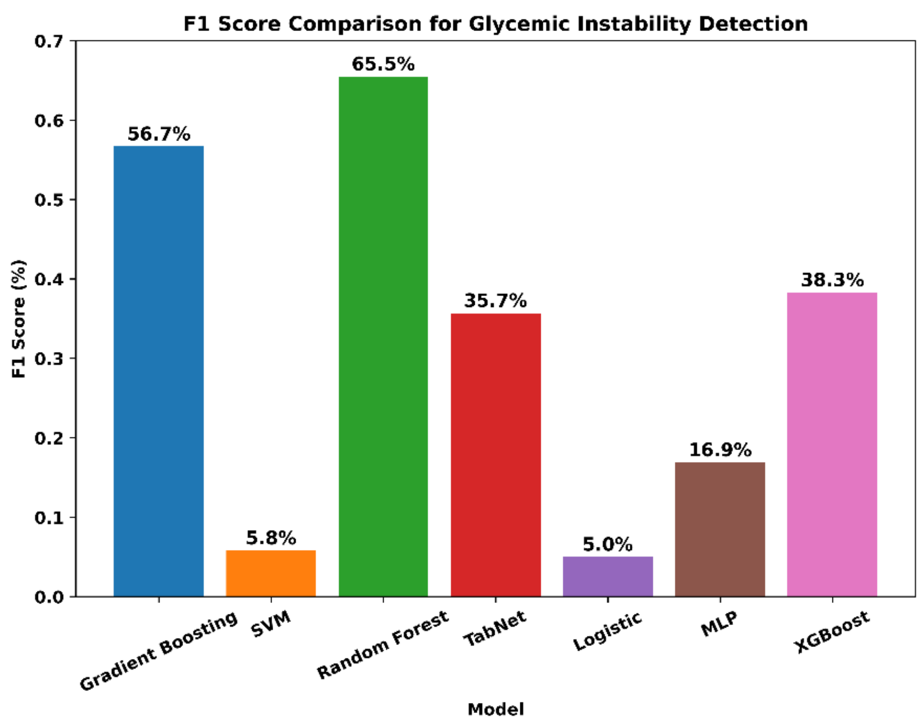

Figure 6. F1-Score Comparison of Models

4.7 Accuracy Comparison and Limitations of Accuracy in Medical AI

The overall classification accuracy of every model was demonstrated. The majority of models exhibit high accuracy because stable cases dominate their testing process yet accuracy does not accurately measure their clinical dependability. The evaluation of medical AI systems requires additional assessment tools because two models with matching accuracy exhibit distinct patterns of sensitivity and precision. Gradient Boosting and Random Forest achieved the highest accuracy at 99.90%, demonstrating excellent overall classification performance (Figure 7). TabNet and XGBoost closely followed with 99.60%. MLP also performed strongly at 98.90%. SVM recorded 94.60%, while Logistic Regression showed the lowest accuracy at 93.40%. Overall, most models achieved very high accuracy, particularly ensemble-based methods. However, considering earlier differences in recall and F1 score, high accuracy alone may not fully reflect balanced performance across all evaluation metrics.

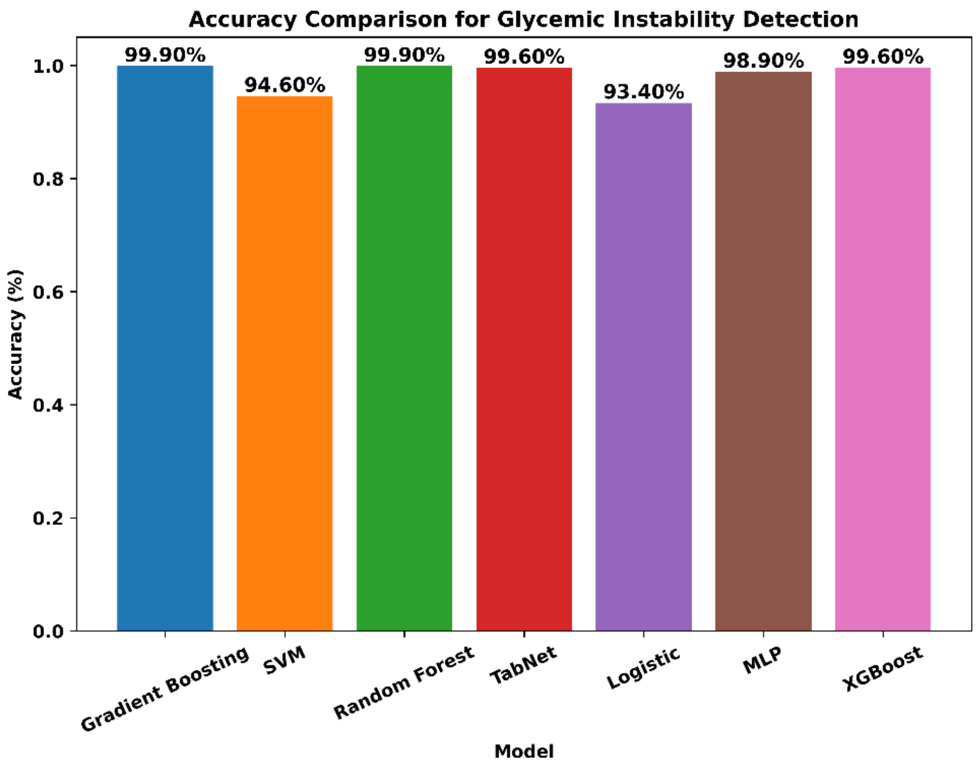

Figure 7. Accuracy Comparison of Models

4.8 Precision Comparison and False Alarm Control

The precision results resulted for all models, which measure their ability to identify actual high-risk patients who experience unstable conditions. High precision results in reduced false alarm rates, which enables users to trust model-generated alerts more effectively. The ensemble models deliver better precision results than both linear models and kernel-based models, which makes them more appropriate for use in actual clinical settings. Random Forest achieved the highest precision at 90.5%, indicating strong ability to correctly identify positive cases with minimal false positives. Gradient Boosting followed with 63.3%, showing moderate precision. XGBoost and TabNet recorded 27.7% and 24.3%, respectively. MLP achieved 9.9%, while SVM and Logistic Regression had the lowest precision at 3.0% and 2.6% (Figure 8). These results reveal substantial variation in false positive rates among models, with Random Forest demonstrating the most reliable positive predictions.

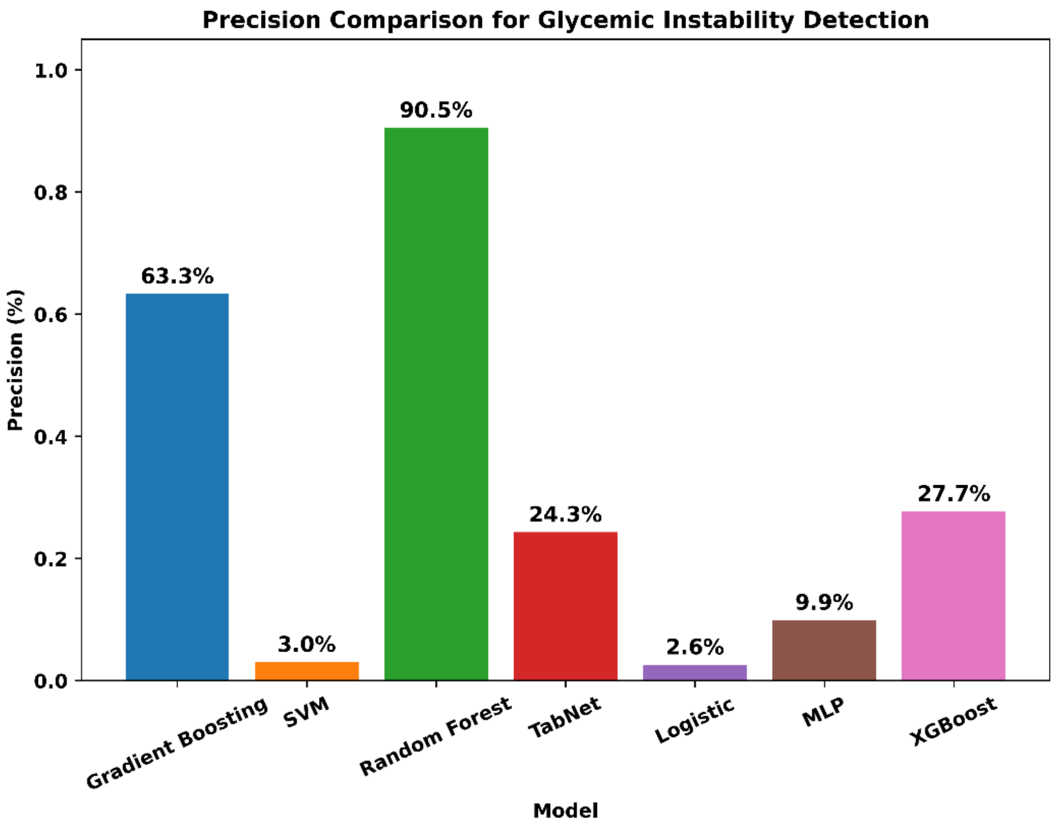

Figure 8. Precision Comparison of Models

4.9 Feature Importance and Clinical Interpretability Analysis

The primary predictive element displayed which all models Random Forest, Gradient Boosting, SVM, XGBoost, Logistic Regression and MLP identified as their most important feature. A1C result and maximum blood glucose levels served as the most effective indicators for glycemic instability across all models Because both acute hyperglycemia and long-term glycemic control show their importance A1C result and maximum blood glucose levels function as key indicators of glycemic instability (Figure 9). The clinical burden factors which demonstrated significant impact on patient instability risk included three specific elements: hospital stay duration, total number of diagnoses, and all conducted laboratory tests. The prediction of risk improved through the inclusion of insulin usage and diabetes medication information which served as medication-related characteristics. The combination of biochemical and treatment-related variables creates a stable yet clinically important pattern which generates high agreement results across multiple algorithms. The research demonstrates that model selection together with imbalance-handling techniques primarily affects the accuracy of AI-based predictions for glycemic instability. Linear and kernel models which include SVM and logistic regression show strong sensitivity but their calibration problems result in excessive false positive detections. The models provide conservative screening benefits although their clinical application suffers from low accuracy.

The most important therapeutic benefit between sensitivity and precision exists in ensemble tree-based models which include Random Forest and Gradient Boosting. The models maintain constant predictive accuracy together with high ROC-AUC results because they effectively model the nonlinear connections between hospitalization data and medication usage and laboratory test results. TabNet deep learning shows potential for developing complex tabular data representations but requires further adjustments before achieving accuracy levels that match ensemble methods. SMOTE together with cost-sensitive learning proved necessary to stop model bias away from stable instances while enabling the system to detect rare glycemic instability events. The system achieved clinical risk management according to its decision threshold (0.3) because it prioritized patient safety above accurate results.

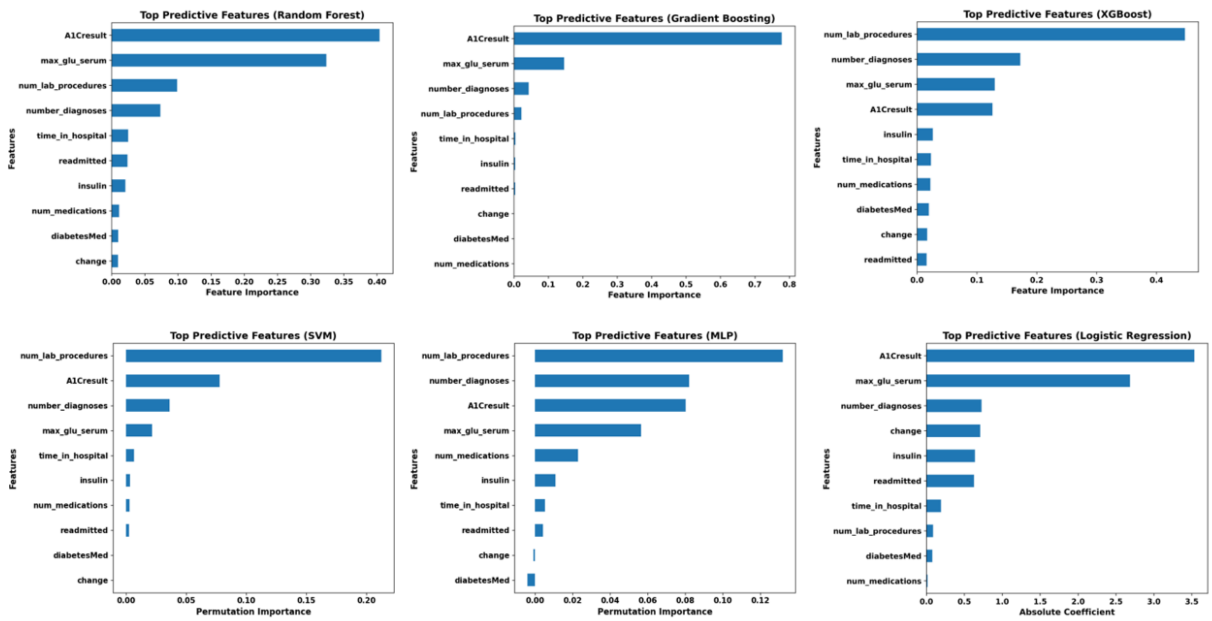

Figure 9. Top predictive features across different models

Ensemble machine learning models with their imbalance-aware training methods and clinically determined thresholding techniques provide effective solutions for proactive glycemic risk assessment in American healthcare systems. The hospital decision-support systems can use the proposed AI-driven glycemic instability prediction framework. The system identifies patients who face high risk of unstable glycemic events which allows clinicians to start insulin dose adjustments and dietary changes and more intensive glucose monitoring. The approach decreases hypo- and hyperglycemia complications which results in shorter hospital stays and better patient safety.

The team uses ensemble learning models including Gradient Boosting and Random Forest to create high-risk alerts which maintain both detection sensitivity and alert accuracy while reducing missed alerts and false alarm rates. The system uses a clinically informed decision threshold which enables its model predictions to match actual medical requirements because it gives precedence to patient safety instead of achieving maximum accuracy. The system provides feature importance information together with model interpretability which helps clinicians understand the clinical factors that lead to glycemic instability. The system enables clinicians to make evidence-based decisions about diabetes management while their trust in AI recommendations increases.

5.0 Limitations of the U.S. Healthcare System

The U.S. has modern medical technology and a strong clinical research framework, but its healthcare system has many problems that make it less cost-effective, less fair, less coordinated, and less long-lasting. No other country spends more on healthcare than the United States. But this doesn't always mean that everyone in the country will be healthier. In 2023, health care costs in the country were about 17.6% of GDP. Costs are going up faster than the economy is growing, so it will probably rise to about 20.3% by 2033. This prediction says that the amount and frequency of healthcare services will keep going up. Healthcare costs are expected to hit $5.6 trillion by 2025 and $8.6 trillion by 2033. These constant price increases make it difficult for businesses, families, and federal and state governments to stay within their budgets. As costs continue to rise, they are forced to make tough financial decisions, such as cutting expenses, raising prices, or reducing services. Over time, this financial strain can slow economic growth and create uncertainty for everyone involved (Hossain et al., 2024; Islam et al., 2023; Islam et al., 2024; Khan et al., 2024). The main reasons for this trend are high service costs, complicated management, and payment incentives that don't work (CMS, 2025; CDC, 2025; KFF, 2025; Das et al., 2025).

The way healthcare is given in the US is very messy. There are many kinds of payers and providers, and they don't always work the same way in different care situations. Interoperability, or the ability to share useful clinical data between systems without any problems, is still not very good, even though electronic health records (EHRs) are very common. Even though national health IT assessments (ONC Report to Congress, 2024) show improvements, it is still hard to get, share, and use electronic health information. When there isn't enough connectivity, tests that aren't needed, medical errors, and breaks in care happen. This is especially true for people with long-term illnesses who go to more than one doctor (Mohib et al., 2025). Two government programs that want to make it easier to share and use data are the United States Core Data for Interoperability (USCDI) and the Trusted Exchange Framework and Common Agreement (TEFCA). TEFCA has made it easier to share about 500 million health data in the last few years, which shows that connections are getting better. But a lot of exchange doesn't always mean that it will be useful in a clinical setting. Many systems still have trouble working with how doctors and nurses do their jobs, keeping data quality high, and using shared data consistently for patient care or public health projects (HHS/ONC, 2026; CDC, 2026).

The COVID-19 pandemic showed that we need to do better at public health surveillance, sharing data, training workers, and getting ready for emergencies (Juie et al., 2021). Federal agencies like the CDC have made interoperability and modernization of public health data a top priority. Many city and state public health offices are still having trouble because they don't have enough money, staff, or stable reporting systems, and their information systems are out of date. These issues make it harder to quickly find, respond to, and fix health issues that come up because of mental health issues, infectious disease outbreaks, environmental exposures, or even human-centered strategies (Juie et al., 2021; Saurav et al., 2023; Tanvir et al., 2024).

There still aren't enough doctors and nurses, especially in basic care and rural or underserved urban areas. This makes it harder for people to get care and keep getting it. Professional groups say that there might not be enough doctors and other important health care workers in the next ten years. Burnout, high employee turnover, and an uneven distribution of providers across areas all make supply problems worse. This makes it harder for some people to get care and makes people who don't have a main care provider depend more on emergency services. AAMC (2026) and HRSA (2025) both write about predictions for the workforce. There are still big differences between groups, even though the national averages for some measures have gone up. The maternal mortality rate in the U.S. dropped from 2022 to 2023, with 18.6 deaths per 100,000 live births. However, the rates for Black women (50.3 per 100,000) were much higher than those for women of other races and ethnicities. These differences show that there are problems that need to be fixed, such as access to high-quality care, social determinants of health, chronic stress, and systemic limitations in the healthcare system (CDC, 2021).

6.0 Future Directions for the U.S. Healthcare System

To address these limitations, changing the healthcare system in the United States is necessary in order to circumvent these healthcare issues. This is something that may be accomplished through the implementation of coordinated reforms that place an emphasis on value, equity, integration, updates to data, support for the workforce, and cost reduction. We need to move away from fee-for-service models that are based on volume and toward payment systems that reward quality, prevention, and care coordination. This is demonstrated by the fact that the national health spending has continued to climb. Value-based care models must to incorporate robust outcome measurements, such as patient-reported outcomes, hospitalizations that could have been avoided, chronic disease control, and equity indicators. Rather than focusing on short-term service volume, this is done to ensure that incentives are aligned with long-term health improvements (KFF forecasts, 2025). Despite the fact that datasets such as USCDI and network frameworks such as TEFCA are essential for the exchange of national data, the focus of future efforts should be on ensuring that the information that is transmitted can be utilized in the delivery of care and in the health of the population. This entails increasing the requirements for the quality of the data, simplifying the process of matching identities, reducing the barriers that are associated with the workflow and the technological aspects, and ensuring that the interchange of data contributes to the process of making clinical decisions in real time and conducting public health surveillance (HHS/ONC, 2026; ONC Report to Congress, 2024).

Future research should focus on integrating real-time wearable sensor data, electronic health records (EHRs), and population-level socioeconomic indicators into adaptive glycemic risk prediction systems. Recent evidence highlighting the strategic value of machine learning in healthcare optimization suggested that AI-driven decision intelligence frameworks can significantly enhance chronic disease management outcomes (Sufian et al., 2024). Future models should leverage reinforcement learning and digital twin simulations to enable proactive intervention strategies, reduce hospitalization risk, and support value-based healthcare delivery systems. The process of standardizing the flow of regular data between clinical care and public health can assist in the detection of outbreaks, the monitoring of chronic conditions, and the response to emerging threats. In the process of modernizing public health data, it is important to maintain the use of uniform standards for both routine reporting and while dealing with emergency situations. Data collection on the fly will become less necessary as a result of this, which will assist organizations of all levels in comprehending what is taking place (CDC, 2026). We need to increase the capacity for training, make it simpler for people to receive care in regions where there is a shortage of it, and change the way care is delivered so that interdisciplinary teams consisting of nurse practitioners, physician assistants, pharmacists, community health workers, and allied health professionals can collaborate (Ashik et al., 2023). These are the steps that we need to take in order to solve the problem of not having enough workers. In order to help address gaps and make treatment more accessible, it is possible to make investments in education, initiatives that assist individuals in repaying their loans, incentives for practicing in rural regions, and flexible scopes of practice of medical professionals.

The concept of equity ought to be recognized as an essential component of the performance of healthcare. Both policymakers and payers ought to make it mandatory for quality metrics to be stratified according to factors such as racial and ethnic background, geographic location, and socioeconomic standing (Rahman et al., 2025). Guria et al. (2025) highlighted the role of artificial intelligence in reducing socioeconomic divides through inclusive technological development, which is particularly relevant for diabetes management across underserved U.S. communities. Therefore, future glycemic risk modeling systems should incorporate socioeconomic risk indices and fairness-aware reinforcement learning approaches to optimize personalized intervention strategies. Such integration would support value-based care models, reduce preventable hospitalizations, and enhance long-term chronic disease management outcomes within U.S. healthcare systems. Providers and health systems that are able to demonstrate that they are reducing disparities in outcomes, such as the prevalence of chronic diseases, maternal mortality, and the utilization of preventive treatment, should be recognized with incentives. Through the utilization of accountability frameworks and equity assessment, it is possible to facilitate the allocation of resources to groups that have not been adequately served in the past. If the processes of billing, prior authorization, reporting, and documentation are simplified, it is possible that providers will experience less difficulty in handling these tasks. As a result, resources are made available for direct patient treatment. Through the use of digital tools that are compatible with one another and the standardization of administrative transactions among payers, it is possible to make things even more efficient, reduce expenses, and prevent doctors from becoming mentally exhausted.

7.0 Conclusion

The research applied an AI-based method to predict glucose fluctuations for hospitalized patients by using their electronic health records which were collected at regular intervals. The proposed pipeline produced accurate results for identifying patients who would experience unstable hyperglycemia episodes because it used strong preprocessing methods together with SMOTE and cost-sensitive learning to address class imbalance and it applied clinical decision thresholds. The research evaluated seven machines learning and deep learning models to determine that Random Forest and Gradient Boosting ensemble methods provided optimal results because they achieved maximum sensitivity together with minimum false positive rate. The best discriminative power of Gradient Boosting reached 0.993 according to its ROC-AUC measurement. The linear and kernel-based models demonstrated high sensitivity but they produced too many false positive results which showed that medical risk-prediction needs proper threshold calibration together with model selection. TabNet deep learning models can identify complex feature interactions but they need more optimization work to achieve their target performance standard which matches ensemble methods. Glycemic instability can be detected through its early identification because the system uses imbalance-aware learning combined with clinical thresholding methods to enable doctors to provide timely treatment for their patients. The proposed system enables doctors to detect patients with high diabetes risk while it enhances diabetes treatment by integrating into existing hospital decision-support systems found throughout the United States. To further improve predictive accuracy and generalizability, future research will concentrate on integrating longitudinal glucose trajectories, continuous glucose monitoring data, and external validation across various healthcare systems.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1]. Abousaber, I., Abdallah, H.F. and El-Ghaish, H. (2025). Robust predictive framework for diabetes classification using optimized machine learning on imbalanced datasets. Frontiers in Artificial Intelligence, 7, p.1499530.

[2]. Alam, K. R., Saurav, K. R. H., Kabir, J. U. Z., Rahman, M. S., Hasan, M. I., & Barua, C. (2024). AI-driven digital marketing analytics and leadership strategies for sustainable business innovation: A mixed-methods research study. The American Journal of Management and Economics Innovations, 6(12), 83–102. Retrieved from https://theamericanjournals.com/index.php/tajmei/article/view/7400.

[3]. Arafa, A., El-Fishawy, N., Badawy, M. and Radad, M., (2022). RN-SMOTE: Reduced Noise SMOTE based on DBSCAN for enhancing imbalanced data classification. Journal of King Saud University-Computer and Information Sciences, 34(8), pp.5059-5074.

[4]. Aranaz, P., Navarro-Herrera, D., Zabala, M., Miguéliz, I., Romo-Hualde, A. and López-Yoldi, M., (2023). American Diabetes Association 2010. Diagnosis and classification of diabetes mellitus. Diabetes Care, 33 (1): S62-69. Declaration of originality, 33(1), p.189.

[5]. Ashik, A. A. M., Rahman, M. M., Hossain, E., Rahman, M. S., Islam, S., & Khan, S. I. (2023). Transforming U.S. Healthcare Profitability through Data-Driven Decision Making: Applications, Challenges, and Future Directions. European Journal of Medical and Health Research, 1(3), 116-125. https://doi.org/10.59324/ejmhr.2023.1(3).21

[6]. CDC. (2025). Healthcare Spending Data. https://www.cdc.gov/nchs/dqs/topics/healthcare-spending.html (CDC).

[7]. CDC. (2025). Maternal Mortality Rates in the United States, 2023. https://www.cdc.gov/nchs/data/hestat/maternal-mortality/2023/Estat-maternal-mortality.pdf (CDC).

[8]. CDC. (2026). Implementing Public Health Interoperability. https://www.cdc.gov/data-interoperability/php/public-health-strategy/index.html (CDC).

[9]. Centers for Medicare & Medicaid Services (CMS). (2025). National Health Expenditure Projections, 2024–33. https://www.cms.gov/files/document/nhe-projections-forecast-summary.pdf (Centers for Medicare & Medicaid Services)

[10]. Ceriello, A., Lucisano, G., Prattichizzo, F., La Grotta, R., Franzén, S., Svensson, A.M., Eliasson, B. and Nicolucci, A., (2022). HbA1c variability predicts cardiovascular complications in type 2 diabetes regardless of being at glycemic target. Cardiovascular diabetology, 21(1), p.13.

[11]. Das, K., Tanvir, A., Rani, S. and Aminuzzaman, F. (2025) Revolutionizing Agro-Food Waste Management: Real-Time Solutions through IoT and Big Data Integration. Voice of the Publisher, 11, 17-36. https://doi.org/10.4236/vp.2025.111003.

[12]. Guria, Z. M., Morshed, N., Rahman, I., Dhar, S. R., & Sufian, M. A. (2025). Advancing global peace through inclusive economic development and the role of artificial intelligence in bridging socioeconomic divides. In Proceedings of the 2025 International Conference on Artificial Intelligence's Future Implementations (ICAIFI) (pp. 100–105). IEEE. https://doi.org/10.1109/ICAIFI66942.2025.11326547

[13]. Hasan, M. I., Barua, C., Rahman, M. S., Alam, K. R., Kabir, J. U. Z., & Saurav, K. R. H. (2023). Resilient Healthcare and Critical Urban Infrastructure Design Using AI-Driven Engineering and Project Management Systems. Journal of Medical and Health Studies, 4(6), 161-172. https://doi.org/10.32996/jmhs.2023.4.6.20

[14]. HealthIT.gov – ONC. (2024). 2023 Report to Congress: Access, Exchange, and Use of Electronic Health Information. https://www.healthit.gov/wp-content/uploads/2024/04/2023-HITECH_Report_to_Congress.pdf (ASTP).

[15]. HHS/ONC. (2026). TEFCA™, America's National Interoperability Network, Reaches Nearly 500 Million Health Records Exchanged. https://www.hhs.gov/press-room/tefca-americas-national-interoperability-network-reaches-nearly-500-million-health-records-exchanged.html (HHS.gov).

[16]. Hossain, E., Shital, K. P., Rahman, M. S., Islam, S., Khan, S. I., & Ashik, A. A. M. (2024). Machine learning-driven governance: Predicting the effectiveness of international trade policies through policy and governance analytics. Journal of Trends in Financial and Economics, 1(3), 50–62. https://doi.org/10.61784/jtfe3053. https://scholar.google.com/citations?view_op=view_citation&hl=en&user=iOJQX0sAAAAJ&sortby=pubdate&citation_for_view=iOJQX0sAAAAJ:4TOpqqG69KYC

[17]. Islam, S., Hossain, E., Rahman, M. S., Rahman, M. M., Khan, S. I., & Ashik, A. A. M. (2023). Digital Transformation in SMEs: Unlocking Competitive Advantage through Business Intelligence and Data Analytics Adoption. 5 (6):177-186. https://doi.org/10.32996/jbms.2023.5.6.14.

[18]. Islam, S., Khan, S. I., Ashik, A. A. M., Hossain, E., Rahman, M. M., & Rahman, M. S. (2024). Big data in economic recovery: A policy-oriented study on data analytics for crisis management and growth planning. Journal of Computational Analysis and Applications (JoCAAA), 33(7), 2349–2367. Retrieved from https://www.eudoxuspress.com/index.php/pub/article/view/3338.

[19]. Juie, B. J. A., Kabir, J. U. Z., Ahmed, R. A., & Rahman, M. M. (2021). Evaluating the impact of telemedicine through analytics: Lessons learned from the COVID-19 era. Journal of Medical and Health Studies, 2(2), 161–174.https://doi.org/10.32996/jmhs.2021.2.2.19

[20]. Kabir, J. U. Z., Saurav, K. R. H., Rahman, M. S., Hasan, M. I., Barua, C., Alam, K. R., & Rahman, M. M. (2023). Financially Sustainable Strategic Leadership and Management with Predictive Analytics to Strengthen Nationwide U.S. Healthcare Quality and Performance. European Journal of Medical and Health Research, 1(3), 144-150. https://doi.org/10.59324/ejmhr.2023.1(3).24

[21]. Kasula, B.Y., (2024). Advancements in AI-driven healthcare: A comprehensive review of diagnostics, treatment, and patient care integration. International Journal of Machine Learning for Sustainable Development, 1(1), pp.1-5.

[22]. KFF. (2025). How Much Is Health Spending Expected to Grow? https://www.kff.org/health-costs/how-much-is-health-spending-expected-to-grow/ (KFF).

[23]. Khalid, S., Kim, H. and Kim, H.S. (2025). Recent trends in diabetes mellitus diagnosis: an in-depth review of artificial intelligence-based techniques. Diabetes Research and Clinical Practice, p.112221.

[24]. Khan, S., Biswas, M.R. and Shah, Z., (2026). Early Prediction of Type 2 Diabetes Using Multimodal data and Tabular Transformers. arXiv preprint arXiv:2601.12981.

[25]. Khan, S. I., Rahman, M. S., Ashik, A. A. M., Islam, S., Rahman, M. M., & Hossain, E. (2024). Big Data and Business Intelligence for Supply Chain Sustainability: Risk Mitigation and Green Optimization in the Digital Era. European Journal of Management, Economics and Business, 1(3): 262-276. https://doi.org/10.59324/ejmeb.2024.1(3).23

[26]. Kim, J.Y., Han, J.M., Yun, B., Yee, J. and Gwak, H.S., (2023). Machine learning-based prediction of risk factors for abnormal glycemic control in diabetic cancer patients receiving nutrition support: a case–control study. Hormones, 22(4), pp.637-645.

[27]. Mohib, M. M., Uddin, M. B., Rahman, M. M., Tirumalasetty, M. B., Al-Amin, M. M., Shimu, S. J., Alam, M. F., Arbee, S., Munmun, A. R., Akhtar, A., & Mohiuddin, M. S. (2025). Dysregulated Oxidative Stress Pathways in Schizophrenia: Integrating Single-Cell Transcriptomic and Human Biomarker Evidence. Psychiatry International, 6(3), 104. ttps://doi.org/10.3390/psychiatryint6030104.

[28]. Monnier, L., Colette, C., Bonnet, F., Renard, E. and Owens, D., (2023). HbA1c variability and diabetes complications: assessment and implications. Diabetes & Metabolism, 49(1), p.101399.

[29]. Monnier, L., Colette, C., Renard, E., Benhamou, P.Y., Aouinti, S., Molinari, N. and Owens, D., (2025). Glycemic variability and iatrogenic hypoglycemia: how to resolve. Diabetes Research and Clinical Practice. 226:112360. https://doi.org/10.1016/j.diabres.2025.112360

[30]. Netayawijit, P., Chansanam, W. and Sorn-In, K., (2025). Interpretable Machine Learning Framework for Diabetes Prediction: Integrating SMOTE Balancing with SHAP Explainability for Clinical Decision Support. In Healthcare (Vol. 13, No. 20, p. 2588).

[31]. Rahman, M. M., Juie, B. J. A., Tisha, N. T., & Tanvir, A. (2022). Harnessing predictive analytics and machine learning in drug discovery, disease surveillance, and fungal research. Eurasia Journal of Science and Technology, 4(2), 28-35. https://doi.org/10.61784/ejst3099

[32]. Rahman, M. M., Rahman, M. S., Islam, S., Khan, S. I., Ashik, A. A. M., Hossain, E., & Tanvir, A. (2025). Integrating data analytics into health informatics: Advancing equity, pharmaceutical outcomes, and public health decision-making. Eurasian Journal of Medicine and Oncology, 9(4), 284–295. https://doi.org/10.36922/EJMO025300319.

[33]. Rishad, S. S. I., Akter, N., Ahamed, A., Sufian, M. A., Rinky, A. I., & Rimi, N. N. (2025). Leveraging Machine Learning for Cancer Insights and Disease Predictions. In Proceedings of the 2025 IEEE International Conference on Data-Driven Social Change (ICDDSC-2025). IEEE.

[34]. Sağlam, F. and Cengiz, M.A., (2022). A novel SMOTE-based resampling technique trough noise detection and the boosting procedure. Expert Systems with Applications, 200, p.117023.

[35]. Saurav, K. R. H., Alam, K. R., Barua, C., Hasan, M. I., Rahman, M. S., & Kabir, J. U. Z. (2023). AI-Driven Consumer Behavior Insights and Human-Centered Marketing Strategies for Enhancing Engagement and Innovation. Journal of Business and Management Studies, 5(6), 187-197. https://doi.org/10.32996/jbms.2023.5.6.16x

[36]. Strack, B., DeShazo, J.P., Gennings, C., Olmo, J.L., Ventura, S., Cios, K.J. and Clore, J.N. (2014). Impact of HbA1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records. BioMed research international, 2014(1), p.781670.

[37]. Sufian, M. A., Rimon, S. M. T. H., Mosaddeque, A. I., Guria, Z. M., Morshed, N., & Ahamed, A. (2024). Leveraging machine learning for strategic business gains in the healthcare sector. In Proceedings of the 2024 International Conference on TVET Excellence & Development (ICTeD) (pp. 225–230). IEEE. https://doi.org/10.1109/ICTeD62334.2024.10844658

[38]. Talebi Moghaddam, M., Jahani, Y., Arefzadeh, Z., Dehghan, A., Khaleghi, M., Sharafi, M. and Nikfar, G. (2024). Predicting diabetes in adults: identifying important features in unbalanced data over a 5-year cohort study using machine learning algorithm. BMC Medical Research Methodology, 24(1), p.220.

[39]. Tanvir, A., Juie, B. J. A., Tisha, N. T., & Rahman, M. M. (2020). Synergizing big data and biotechnology for innovation in healthcare, pharmaceutical development, and fungal research. International Journal of Biological, Physical and Chemical Studies, 2(2), 23–32. https://doi.org/10.32996/ijbpcs.2020.2.2.4.

[40]. Tanvir, A.; Jo, J.; Park, S.M. (2024). Targeting Glucose Metabolism: A Novel Therapeutic Approach for Parkinson's Disease. Cells. 13, 1876. https://doi.org/10.3390/cells13221876

[41]. Zafar, M. M., Khan, Z. A., Javaid, N., Aslam, M., & Alrajeh, N. (2025). From Data to Diagnosis: A Novel Deep Learning Model for Early and Accurate Diabetes Prediction. Healthcare, 13(17), 2138. https://doi.org/10.3390/healthcare13172138