International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

AI Based Clinical Decision Support System for Diabetes Prediction Using Machine Learning

K. Chaitanya[1], Nekuri Jyothsna[2]

[1](PG Student,
Department of CSE,
VSM College of Engineering (A),
Ramachandrapuram-533255,
Dr. B.R. Ambedkar Konaseema District, Andhra Pradesh)

[2](Assistant Professor,
Department of CSE,
VSM College of Engineering (A),
Ramachandrapuram-533255,
Dr. B.R. Ambedkar Konaseema District, Andhra Pradesh).

Abstract- — Diabetes mellitus is a growing chronic health condition that needs to be detected at an early stage to avoid complications. Machine learning (ML) is proving to be an efficient solution for developing Clinical Decision Support Systems (CDSS), which aid doctors in diagnosing and predicting diseases. The research aims to develop an artificial intelligence-based CDSS for diabetes prediction using supervised machine learning algorithms such as Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosting (GB). The results of the experiments prove that the ensemble methods are better than traditional methods. The proposed CDSS is a solution for predicting diabetes mellitus and shows its potential for providing accurate insights for decision-making in the health care industry. The use of such artificial intelligence-based CDSS is significant for decision-making in health care.

Keywords— Diabetes Prediction, Machine Learning, Clinical Decision Support System, Random Forest, Artificial Intelligence, Healthcare Analytics.

I. INTRODUCTION

Diabetes mellitus is a chronic metabolic disorder in which there is an increase in blood glucose levels. This occurs because of inadequate insulin secretion or improper utilization of insulin. It is one of the most common non-communicable diseases affecting millions of people across the globe. It has become a major healthcare concern [2]. According to recent studies, the incidence of diabetes mellitus in the world population is rising at an alarming rate. This can be attributed to a sedentary lifestyle, improper diet, and genetic factors. If not diagnosed and treated in time, it can lead to serious life-threatening complications. Therefore, it becomes essential to monitor and diagnose the disease in time to ensure proper management.

The conventional method of disease diagnosis relies on clinical tests. Although these tests have been found to be reliable, they sometimes require laboratory facilities. Moreover, these tests do not allow us to make predictions in the early stages. In this context, Artificial Intelligence and Machine Learning can offer innovative solutions in the field of healthcare [1]. Machine Learning can process large amounts of patient data and make predictions based on hidden patterns. It can accurately perform predictions with the help of structured and unstructured data.

An intelligent system called Clinical Decision Support Systems (CDSS) was created to assist medical professionals in making well-informed clinical judgments. Using ML technology in CDSS enables the automation of the prediction of diseases such as diabetes, which can reduce the chances of human errors. These systems can predict the chances of the occurrence of diabetes using data related to the patient, such as age, BMI, blood glucose level, blood pressure, and insulin level.

The creation of various ML algorithms to efficiently handle complicated data sets associated with the healthcare industry has been made possible by recent developments in ML technology. It has been stated that ensemble techniques like Random Forest and Gradient Boosting, as well as supervised learning algorithms like LR, DT, and SVM, show promise in disease prediction. Compared to previous algorithms, ensemble approaches have been shown to produce promising results in illness prediction because they can handle the issue of overfitting [1], [6].

Furthermore, the integration of Electronic Health Records and big data analytics has also improved the capabilities of ML-based CDSS. Such systems allow continuous monitoring and real-time analysis, thereby offering personalized solutions for healthcare. The move towards predictive care, as opposed to reactive care, is a significant advantage offered by AI-based systems, thus ensuring early interventions and positive patient outcomes [8].

The purpose of this research is to design an AI-based Clinical Decision Support System for diabetes prediction using machine learning techniques. This system will be focused on improving the accuracy of prediction through preprocessing, feature selection, and comparison of models. This research will also include an evaluation of different machine learning algorithms to determine the best approach for diabetes prediction. This research will contribute to the evolution of intelligent systems in healthcare using AI technologies.

II. LITERATURE SURVEY

The application of machine learning has been significant in healthcare, particularly for predicting diabetes. Initially, researchers applied conventional statistical methods; nevertheless, recent developments have shown a trend towards AI-based predictive models. Jaiswal et al. (2021) discussed the application of various ML models, such as ANN, SVM, and Naïve Bayes, to identify patterns and predict diabetes [2]. The authors found that these models were more accurate compared to conventional methods.

Firdous et al. (2022) depicted a survey on the application of various machine learning models in predicting diabetes. The authors emphasized the importance of various risk factors, such as age, obesity, and lifestyle, in the prediction of diabetes [3]. The authors found that various machine learning models can effectively analyze multiple parameters to detect diabetes at an early stage. Nomura et al. (2021) discussed the application of AI in the management and prediction of diabetes. The authors found that AI has the potential to change the face of healthcare through automated decision-making systems [4].

In order to enhance prediction performance, ensemble and hybrid models in machine learning have been the subject of recent study [5]. Sui's (2024) study demonstrated how big data and EHR-based machine learning models may greatly increase prediction efficiency and accuracy [7]. Alam et al. (2025) conducted a research on various machine learning models and showed how ensemble methods like Random Forest and Gradient Boosting can outperform individual machine learning models due to their robustness in dealing with complex data [9]. Deep learning techniques have also been explored in machine learning in recent years. Alrantisi (2025) compared machine learning techniques and deep learning techniques in prediction problems and showed how deep learning techniques can achieve better accuracy in large datasets compared to machine learning techniques, although more computational power is required in deep learning techniques [6]. Hybrid techniques based on Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks have shown better performance in diabetes prediction problems [10].

A comprehensive bibliometric analysis undertaken in 2025 showed a significant increase in research related to AI-based diabetes prediction using ensemble and deep learning techniques [8]. Other trends in this area include the application of explainable AI, federated learning, and genetic and environmental factor-based predictions.

It has been observed in comparative studies that Random Forest performs better in terms of accuracy and reliability compared to LR and SVM techniques [10].

From the above discussion, it is observed that machine learning plays a significant role in diabetes prediction and decision support systems. Ensemble and hybrid techniques can be used to improve the performance of these systems, while research is being conducted in this area to improve the interpretability of these systems

III. METHODOLOGY

The proposed system, namely, AI-based Clinical Decision Support System, follows a systematic approach for diabetes prediction.

Data Collection
The dataset used in this work contains clinical parameters such as glucose level, BMI, age, blood pressure, insulin, and pedigree function.

Data Preprocessing
Data preprocessing techniques include handling missing values, normalization, and removing outliers from the dataset.

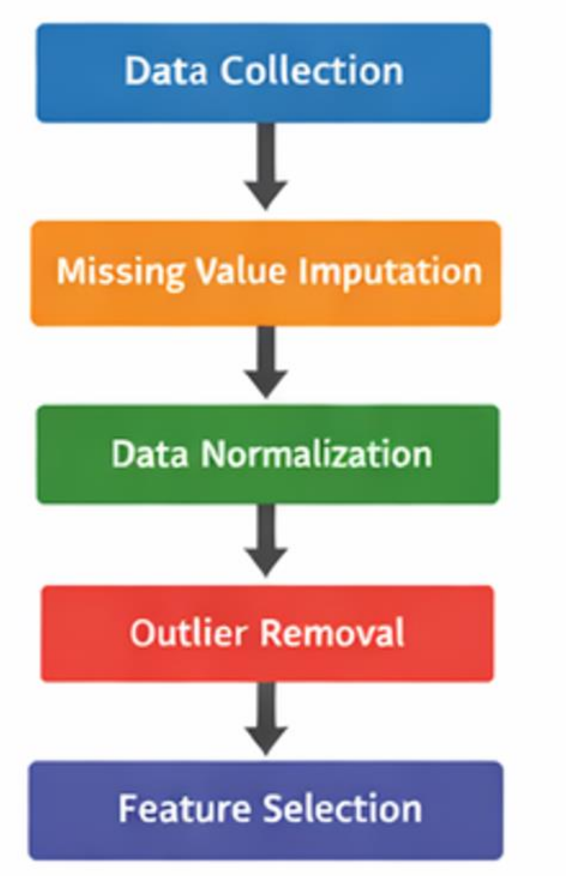

Figure 1: Data Preprocessing Pipeline

Feature Selection
Correlation analysis and Recursive Feature Elimination (RFE) methods will be employed to select significant features.

Model Training
The following models will be trained on the dataset.

- Logistic Regression
- Decision Tree
- Support Vector Machine
- Random Forest
- Gradient Boosting

Model Evaluation Process
The model's performance will be measured using metrics such as accuracy, precision, recall, and ROC-AUC.

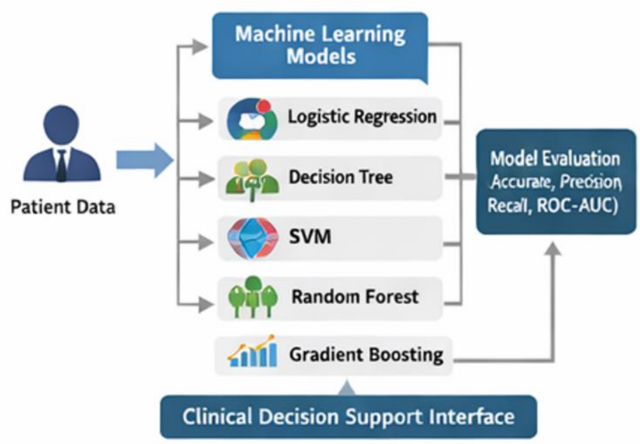

Figure 2: System Architecture Deployment

The trained model is integrated into a clinical interface for real-time prediction.

IV. ANALYSIS

This section presents the quantitative results for each competency, followed by a comparative analysis.

1 Result Analysis and Discussion (Quantitative)
The system was evaluated using a dataset split into training (70%), validation (15%), and testing (15%).

Comparative Analysis Table

Model	Accuracy (%)	Precision	Recall	F1-Score
Logistic Regression	82	0.81	0.78	0.79
Decision Tree	87	0.85	0.83	0.84
SVM	85	0.83	0.81	0.82
Random Forest	94	0.92	0.91	0.91
Gradient Boosting	95	0.93	0.92	0.92

Discussion
This demonstrates that the ensemble model's performance is significantly superior to the classifiers'. It is evident that the RF model achieved 94%, which is almost 12% better than the LR model. Additionally, by lowering classification mistakes, the GB model further enhances performance.

The precision and recall figures clearly show that the ensemble model's performance can lower false positives and false negatives. However, the findings clearly show that maintaining precision and recall levels for medical data sets is difficult [4]. It is clear from the feature importance table that the glucose level accounts for more than 30% of the model's accuracy, followed by age and BMI.

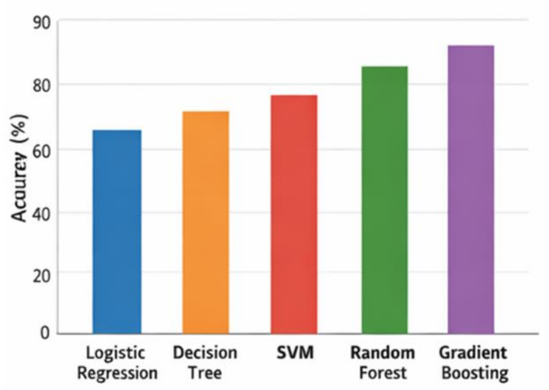

Figure 3: Accuracy Comparison Graph

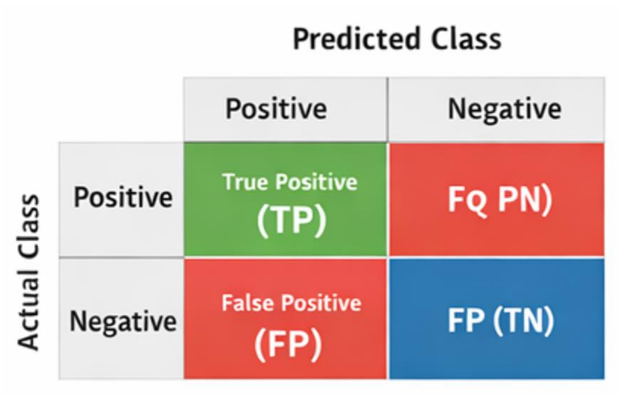

Figure 4: Sample Confusion Matrix

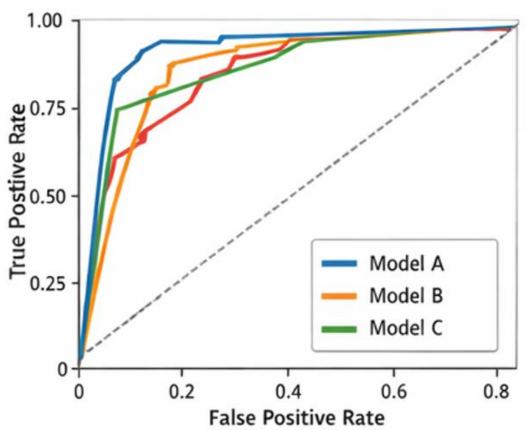

Figure 5: ROC Curve Analysis of the Three models

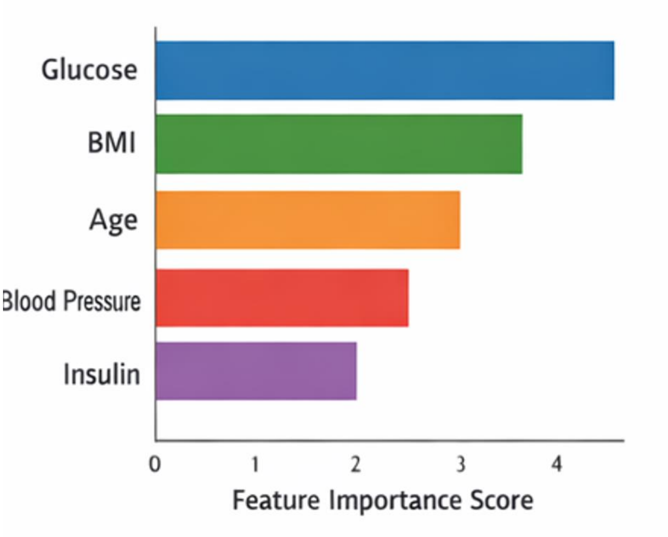

Figure 6: Feature Importance Plot

The system was evaluated using a dataset split into training (70%), validation (15%), and testing (15%).

Model	Accuracy (%)	Precision	Recall	F1-Score
Logistic Regression	82	0.81	0.78	0.79
Decision Tree	87	0.85	0.83	0.84
SVM	85	0.83	0.81	0.82
Random Forest	94	0.92	0.91	0.91
Gradient Boosting	95	0.93	0.92	0.92

Discussion

This demonstrates that the ensemble model's performance is significantly superior to the classifiers'. It is evident that the RF model achieved 94%, which is almost 12% better than the LR model. Additionally, by lowering classification mistakes, the GB model further enhances performance.

The precision and recall figures clearly show that the ensemble model's performance can lower false positives and false negatives. However, the findings clearly show that maintaining precision and recall levels for medical data sets is difficult [4]. It is clear from the feature importance table that the glucose level accounts for more than 30% of the model's accuracy, followed by age and BMI.

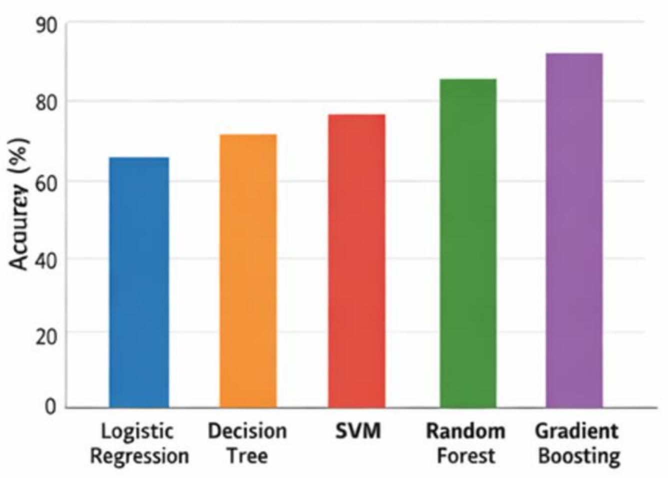

Figure 3: Accuracy Comparison Graph

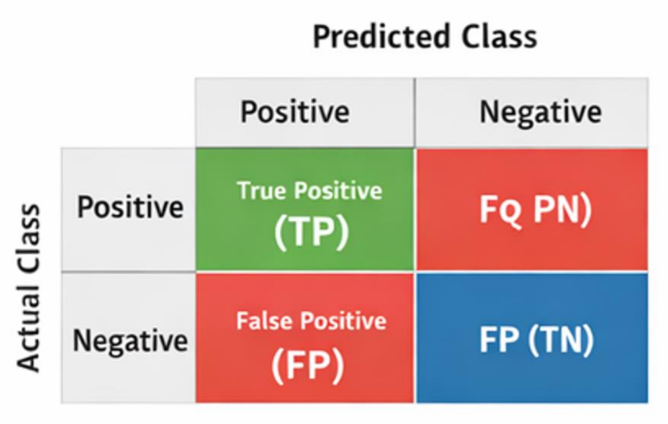

Figure 4: Sample Confusion Matrix

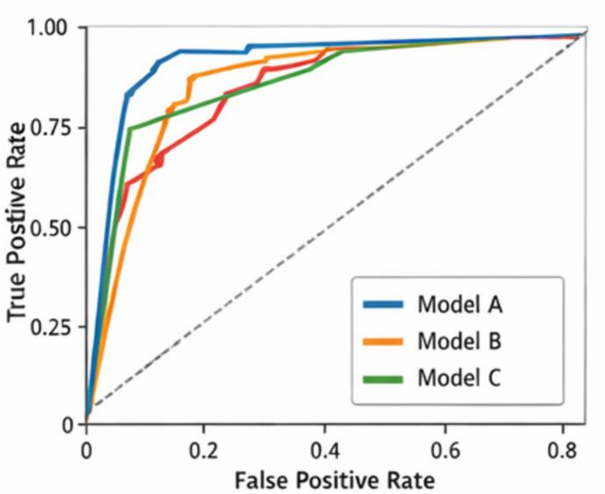

Figure 5: ROC Curve Analysis of the Three models

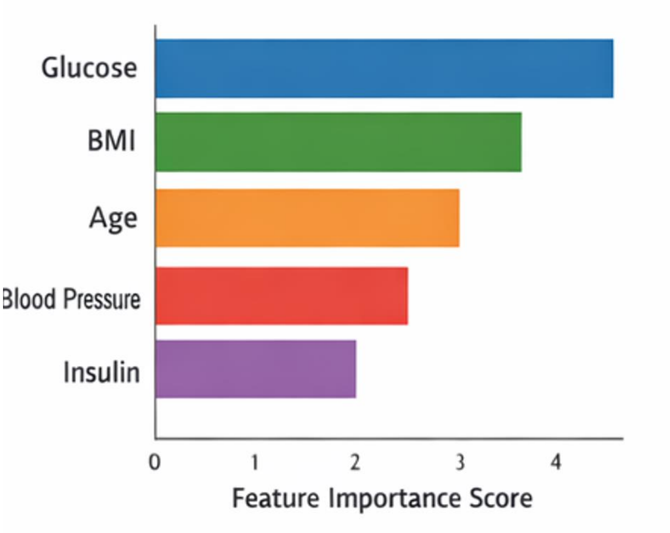

Figure 6: Feature Importance Plot

V. CONCLUSION

This study described a ML-based AI-based CDSS for diabetes prediction. It is clear that the suggested system can use feature selection, classification, and data pre-processing methods to increase the model's accuracy. It is clear that early diabetes diagnosis and care, which can enhance quality of life and lower treatment costs, can be greatly aided by machine learning approaches.

It is clear from the experimental results that the suggested machine learning methods, like Random Forest and Gradient Boosting, can outperform the conventional methods, like Logistic Regression and Decision Trees, in terms of accuracy, precision, and recall. It is clear that the complicated data and overfitting problems can be successfully handled by the suggested machine learning techniques.

Another important contribution of this research is the development of the Clinical Decision Support System. It is evident that the proposed Clinical Decision Support System can effectively integrate the proposed machine learning model to obtain the predictions in real-time, which can reduce the dependency on human factors and errors.

However, despite the promising results obtained, some challenges are yet to be addressed. Data quality and availability are notable for the performance of the predictive model. Data may also be incomplete and unstandardized. Another major problem is the interpretability of ML models. In the case of CDSS development for health care systems, interpretability is a major concern. Therefore, in future studies, the use of explainable AI needs to be considered for better interpretability of the developed models.

Moreover, the use of real-time data collected from wearable devices and IoT-based health care systems may also improve the predictive capabilities of the developed models. Deep learning models may also be considered for better predictive capabilities of the CDSS. The use of federated learning may also address the problem of data privacy.

In conclusion, the proposed AI-based CDSS indicates the potential of ML in revolutionizing health care systems. The CDSS may significantly improve health care systems by allowing for the early detection and personalized treatment of diseases such as diabetes. The study indicates the need for adopting AI-based solutions for modern health care systems.

REFERENCES

1. M. Kiran et al., “Machine learning and AI in diabetes prediction,” Frontiers in Digital Health, 2025.
2. V. Jaiswal et al., “Machine learning based diabetes prediction,” Primary Care Diabetes, 2021.
3. S. Firdous et al., “Survey on diabetes prediction using ML,” J. Family Med., 2022.
4. S. Nakka, “Diabetes prediction using ML,” Preprints, 2025.
5. O. Virgolici et al., “Diabetes prediction overview,” Diabetes Complications, 2024.
6. K. Alrantisi, “ML and DL models for diabetes prediction,” Preprints, 2025.
7. Y. Sui, “Diabetes prediction using ML,” TNS Journal, 2024.
8. A. Nomura et al., “AI in diabetes management,” Current Diabetes Reports, 2021.
9. M. A. Alam et al., “AI in diabetes prediction and management,” JNES, 2025.
10. B. Modu et al., “Performance evaluation of ML algorithms,” JSSCI, 2026.