International Journal of Advanced Engineering and Management Research

Vol. 11, No. 02; 2026

ISSN: 2456-3676

AI Smart Health Risk Assessment

Nenavath Chander[1], P.Siddarda2, P.Shiva[3], T.Aryan[4]
[1,2,3,4]Malla Reddy University, Cyber Security Department, Hyderabad, Telangana, India 500100

doi.org/10.51505/ijaemr.2026.11231 URL: http://dx.doi.org/10.51505/ijaemr.2026.11231

Received: Mar 25, 2026 Accepted: Mar 30, 2026 Online Published: Apr 23, 2026

Abstract
AI Smart Health Risk Assessment is a system designed to facilitate the early detection of diabetes and other metabolic disorders using machine learning techniques. Traditional diagnostic approaches typically rely on clinical consultations and laboratory testing, which can be time-consuming and may not be easily accessible to all individuals. To address these limitations, the proposed system employs the Extreme Gradient Boosting (XGBoost) algorithm in combination with a rule-based Clinical Interpretation Engine to provide rapid and reliable health risk predictions. The system analyzes seven key health parameters: age, body mass index (BMI), systolic and diastolic blood pressure, fasting glucose, glycated hemoglobin (HbA1c), and total cholesterol. Based on these inputs, it generates a probability score and classifies the risk into three categories: low, moderate, or high. In addition to prediction, the system offers interpretability by explaining the contribution of each health parameter to the overall risk. It evaluates glucose levels, HbA1c values, blood pressure categories, indicators of metabolic syndrome, and cardiovascular risk factors. Based on this analysis, personalized recommendations are provided, enabling users to make informed decisions regarding medical consultation. All assessment results are stored in a MongoDB database with timestamps, allowing users to monitor their health trends over time through visual representations. Furthermore, the system includes additional functionalities such as BMI calculation, result comparison, medical recommendations, and report generation. Overall, the proposed system demonstrates the potential of artificial intelligence in enhancing preventive healthcare by delivering faster, more accessible, and interpretable risk assessment solutions.

Keywords: Diabetes Prediction, Artificial Intelligence, XGBoost, Machine Learning, Health Risk Assessment, Clinical Decision Support, FastAPI, MongoDB, Metabolic Syndrome.

1. Introduction

The integration of machine learning in healthcare is creating new opportunities for early disease detection and continuous health monitoring. Modern health risk assessment systems aim to identify potential health issues at an early stage, enabling individuals to take preventive measures before conditions become severe. Chronic conditions such as diabetes, hypertension, and cardiovascular diseases are increasing globally, yet many individuals remain undiagnosed due to

Vol. 11, No. 02; 2026

ISSN: 2456-3676

limited access to healthcare services. Traditional diagnostic methods often involve multiple steps, including laboratory testing, medical consultations, and follow-up visits, making the process time-consuming and costly. This limits their effectiveness for regular and widespread health screening. A major limitation of conventional approaches is their reactive nature.

Individuals typically seek medical attention only after experiencing noticeable symptoms, missing the opportunity for early intervention. Additionally, patients often lack a clear understanding of how different clinical measurements contribute to their overall health, making it difficult to make informed decisions. Machine learning offers a more proactive and efficient alternative. These systems analyze large volumes of clinical data to identify patterns and predict disease risk with high accuracy. By evaluating multiple health indicators simultaneously and generating probability-based predictions, machine learning models provide more reliable and data-driven insights compared to traditional methods. When deployed through web-based platforms, these tools enable users to access instant health assessments without relying on conventional healthcare infrastructure. This approach improves accessibility, reduces dependency on clinical visits, and supports preventive healthcare practices for a broader population.

II. Related Work

Recent studies have explored the application of machine learning techniques for predicting diabetes risk and assessing overall health conditions. Although traditional diagnostic methods remain reliable, they typically involve multiple steps, making them time-consuming and costly[7]. As a result, regular health screening becomes less accessible to a broader population. Several researchers have demonstrated that machine learning models can effectively analyze clinical datasets and identify relationships between various health parameters and medical outcomes. In particular, ensemble learning methods such as gradient boosting have shown superior performance compared to conventional classification approaches. However, many of these studies primarily emphasize model accuracy, often overlooking usability and accessibility for end users. A study conducted by Kavakiotis et al. reviewed multiple machine learning techniques and concluded that ensemble models are highly effective in diabetes prediction[8]. While their findings highlight performance improvements, the study does not address the integration of such models into user-friendly systems capable of delivering real-time results. Furthermore, research on Extreme Gradient Boosting (XGBoost) has demonstrated its capability to efficiently handle structured medical data and produce accurate predictions[9]. Despite these advantages, most implementations remain limited to experimental environments and are not incorporated into fully functional health monitoring systems for practical use.

Recent advancements have also introduced web-based platforms that integrate machine learning models, enabling users to input health data and receive predictions online[10]. Although these systems improve accessibility, they often lack interpretability, making it difficult for users to understand how individual health parameters influence the predicted outcomes. Additionally, some studies have investigated the integration of predictive models with rule-based clinical

Vol. 11, No. 02; 2026

ISSN: 2456-3676

systems to enhance interpretability and transparency. These hybrid approaches provide more meaningful insights; however, their adoption in real-world healthcare applications remains limited[11].In summary, existing research demonstrates strong predictive capabilities but reveals significant gaps in interpretability, personalization, and real-time usability. The proposed system aims to address these limitations by integrating machine learning-based prediction, clinical interpretation, and a user-friendly interface into a unified platform[12].

III. Proposed Model

The proposed AI Smart Health Risk Assessment model is designed to provide a rapid and user-friendly approach for evaluating the risk of diabetes and metabolic disorders using machine learning techniques. Unlike conventional diagnostic methods that depend on laboratory testing and expert analysis, the proposed system enables users to obtain risk predictions through an interactive web interface. The model considers seven key health parameters, namely age, body mass index (BMI), systolic and diastolic blood pressure, glucose level, glycated hemoglobin (HbA1c), and total cholesterol, to estimate the likelihood of developing diabetes. These features are processed using a trained Extreme Gradient Boosting (XGBoost) classifier, which generates a probability score ranging from 0 to 1. Based on this score, users are categorized into three risk levels: low, moderate, and high, thereby ensuring ease of interpretation. To enhance transparency and interpretability, the system incorporates a rule-based Clinical Interpretation Engine.

This component evaluates each input parameter against established medical standards, enabling the system not only to predict risk but also to provide meaningful explanations regarding the contributing factors. The overall system architecture consists of a React-based frontend for user interaction, a Fast API backend for handling application logic and requests, the machine learning model for prediction, and a Mongo DB database for storing user records. This integrated architecture ensures efficient processing, accurate predictions, and scalable data management. This has a feature of downloading the report as pdf format so that the user can show their report directly to the doctor.

IV. Methodology & Implementation of AI Smart Health Riskassessment

The AI Smart Health Risk Assessment system utilizes a machine learning framework to accurately predict diabetes risk in real time based on clinical biomarker data. The platform provides a web-based interface where users can directly access a health input form, enter their seven clinical parameters, and receive a detailed risk assessment without the need for registration or login.Each prediction request undergoes validation using Pydantic to ensure data accuracy and consistency, and is securely processed by the FastAPI backend. The trained XGBoost model then generates a probability score indicating the likelihood of diabetes risk. Subsequently, the Clinical Interpretation Engine evaluates each biomarker based on established medical guidelines to provide meaningful and clinically relevant insights.

Vol. 11, No. 02; 2026

ISSN: 2456-3676

A. Data Preparation

The dataset utilized in this study was obtained from publicly available medical sources and includes records of both diabetic and non-diabetic patients. It encompasses a diverse range of patient profiles to support effective model training and generalization. Each record consists of seven clinical parameters along with a corresponding label indicating the presence or absence of diabetes. The dataset was carefully examined to ensure consistency and reliability of the information. Additionally, appropriate measures were taken to maintain a balanced distribution between diabetic and non-diabetic cases. This balance is essential for preventing model bias and ensuring accurate and fair predictions across different classes.

B. Data Preprocessing

The dataset was preprocessed to ensure accuracy and consistency prior to model training. Missing values and duplicate records were identified and removed to maintain data integrity. Outliers were detected and eliminated using statistical techniques, ensuring that only clinically realistic data was retained for training. This step helps improve the reliability and robustness of the model. The dataset was then divided into training and testing subsets using an 80:20 ratio. The split was performed in a manner that preserved class balance across both subsets. Furthermore, feature scaling was applied to normalize all input variables, ensuring that each feature contributes equally during the training process. This prevents any single parameter from disproportionately influencing the model's predictions.

C. Model Training

The Extreme Gradient Boosting (XGBoost) model was trained using the preprocessed dataset. Regularization techniques were applied to reduce over fitting and improve the model's generalization performance on unseen data. The trained model produces a probability score representing the likelihood of diabetes risk. Based on this score, users are classified into three categories: low, moderate, and high risk. This classification approach enhances interpretability and allows users to easily understand their health risk level.

D. Clinical Interpretation

A rule-based clinical system is incorporated to evaluate patient health data in accordance with established medical standards. The system analyzes key parameters such as blood glucose levels, glycated hemoglobin (HbA1c), and blood pressure categories. In addition to individual parameters, the system also assesses broader health conditions, including indicators of metabolic syndrome. This approach enables the model to provide meaningful clinical insights alongside predictive results.

E. System Deployment

The AI Smart Health Risk Assessment System was implemented as a complete web application using modern and efficient technologies to ensure smooth performance and wide accessibility.

The frontend was developed using React and served through the Vite development server, providing a fast and interactive user experience. The application can be accessed through any modern web browser via localhost on port 5173, without requiring software installation or user registration. The trained XGBoost model and the StandardScaler used for data preprocessing were saved as serialized pickle files. These components are loaded into memory during application startup, ensuring that subsequent prediction requests are processed instantly without delays associated with model loading.

F. Health Monitoring

All prediction records are automatically stored in MongoDB along with timestamps for each entry. This ensures that historical health data is preserved and can be accessed for long-term analysis. These stored records are visualized through six interactive trend charts on the Health Trends Dashboard. The charts display key health indicators such as Glucose, HbA1c, Blood Pressure, BMI, Risk Probability, and Multi-Risk Trends, allowing users to effectively monitor changes in their health over time.

V. System Implementation and Results

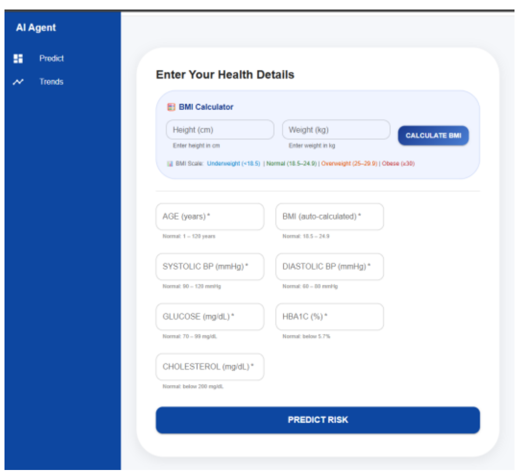

Figure 1: Interface

Figure 1 is the main user interface of the system, where users input their health parameters to assess their risk. The interface is designed to be intuitive and user-friendly, with clearly labeled fields for entering details such as age, body mass index (BMI), blood pressure, glucose levels, glycated hemoglobin (HbA1c), and cholesterol. Additionally, a built-in BMI calculator is provided to assist users in determining their BMI if it is not already known. Once all required inputs are entered, users can initiate the risk assessment process by clicking the “Predict Risk” button. This interface design ensures ease of use and accessibility, enabling users to quickly obtain health risk predictions without requiring technical expertise.

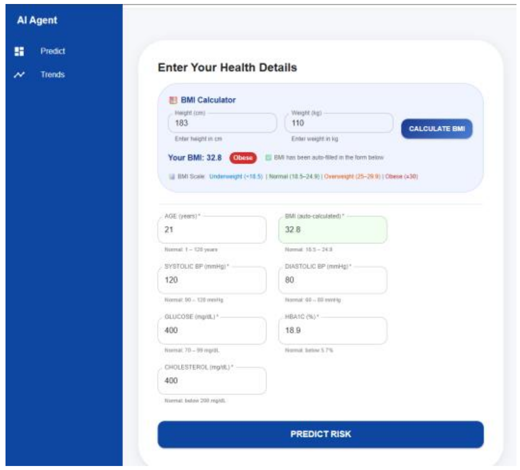

Figure 2: Entered values

Figure 2 presents the interface after the user has entered all the required health parameters. At this stage, all input fields are completed, and the BMI is automatically calculated where applicable. This step allows users to review and verify their entered information, ensuring accuracy before submission. Once the user confirms the inputs, the system processes the data to generate the risk prediction.

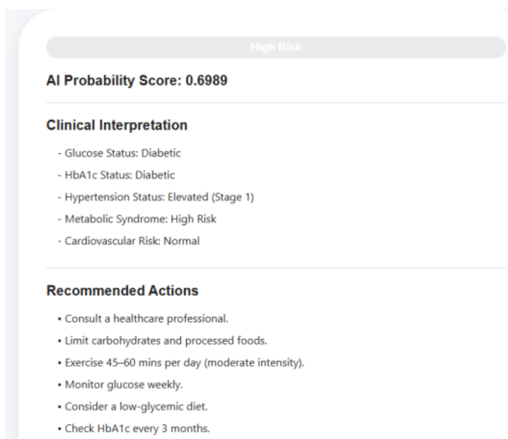

Figure 3: Prediction

In figure 3 we have prediction interface it gives the probability score, clinical interpretation and recommended actions according to the user's values

Vol. 11, No. 02; 2026

ISSN: 2456-3676

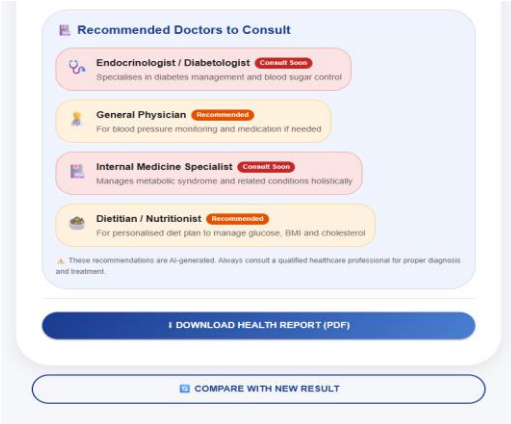

Figure 4: Recommended Actions

This Figure 4 recommended action page suggests the user to consult their respective doctors based on their values. It directly gives the doctor's list to visit . There is also an option to download the report. There is also feature called compare results . If the user had checked their values before some days and now again clicking on that option they can enter their present values and can compare. It shows the results of previous one's and present one's . So that the user can know his status compared to his previous status.

VI. Conclusion

The proposed system demonstrates how machine learning can significantly enhance the early detection of diabetes risk in an efficient and accessible manner. By integrating predictive models with clinical interpretation, the system provides insights that extend beyond simple classification outcomes. The platform reduces dependence on traditional diagnostic procedures by enabling users to obtain risk assessments using basic health parameters. This improves accessibility to health screening, particularly for individuals with limited access to medical facilities. Furthermore, the system allows users to monitor their health over time through data storage and visualization features. This capability supports better awareness and encourages timely preventive actions. Overall, the proposed solution presents a practical application of artificial intelligence in preventive healthcare, emphasizing usability, interpretability, and real-world effectiveness. The proposed system can be further enhanced by extending its capabilities to support the prediction of additional health conditions such as cardiovascular diseases, kidney disorders, and liver-related illnesses. This would transform the system into a comprehensive multi-disease prediction platform. Future developments may include the integration of mobile applications to improve accessibility and user engagement. Additionally, incorporating wearable device connectivity would enable automatic collection of health data, reducing manual input and allowing real-time health monitoring. The performance of the model can be further improved by training it on larger and more diverse datasets, thereby increasing its accuracy across different population groups. Moreover, the adoption of privacy-preserving techniques such as federated learning can enhance model performance while ensuring data security.

Vol. 11, No. 02; 2026

ISSN: 2456-3676

Incorporating user authentication mechanisms would enable the creation of personalized profiles, making the system more suitable for long-term use by both individuals and healthcare professionals.

References

Chen, T., and C. Guestrin. "XGBoost: A scalable tree boosting system." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016): 785-794.

Chen, T., T. He, M. Benesty, and V. Khotilovich. "XGBoost: Extreme gradient boosting." *R Package Version* (2015).

Chen, T., T. He, M. Benesty, and V. Khotilovich. "Package XGBoost." *R Version* (2019).

Mujumdar, A., and V. Vaidehi. "Diabetes prediction using machine learningalgorithms." *Procedia Computer Science* 165 (2019): 292-299.

Rani, K. J. "Diabetes prediction using machine learning." *Research in Computer Science, Engineering and Applications* (2020).

Hasan, M. K., M. A. Alam, D. Das, and E. Hossain. "Diabetes prediction using ensembling of different machine learning classifiers." *IEEE Access* 8 (2020): 76516-76531.

Singh, K., J. K. Rout, and H. Das. "Diabetes prediction using machine learning techniques." *Proceedings of International Conference on Intelligent Computing* (2019): 271-280.

Ramraj, S., N. Uzir, and R. Sunil. "Experimenting XGBoost algorithm for prediction and classification of different datasets." *International Journal of Control Theory and Applications* 9.40 (2016): 651-662.

Li, W., Y. Yin, X. Quan, and H. Zhang. "Gene expression value prediction based on XGBoost algorithm." *Frontiers in Genetics* 10 (2019): 1077.

Li, J., X. An, Q. Li, C. Wang, H. Yu, X. Zhou, and Y. Geng. "Application of XGBoost algorithm in the optimization of pollutant concentration." *Atmospheric Research* 276 (2022): 106238.

Asselman, A., M. Khaldi, and S. Aammou. "Enhancing the prediction of student performance based on the machine learning XGBoost algorithm." *Interactive Learning Environments* 31.6 (2023): 3360-3379.

Li, S., and X. Zhang. "Research on orthopedic auxiliary classification and prediction model based on XGBoost algorithm." *Neural Computing and Applications* 32 (2020): 1971-1979.

International Diabetes Federation, "IDF Diabetes Atlas, 10th ed.," Brussels, Belgium: IDF, 2021. [Online]. Available: https://diabetesatlas.org.