RESEARCH PAPER

AI-Powered Diabetes Prediction Using Electronic Health Records

P. K. Manoj Kumar[1*], P. Stella Mary[2], G. Jenitha[3], Jayasundar S[4], Monica Apte[5], Vengalapudi Appalakonda[6], M. Krishnaveni[7], Anil Kumar Dubey[8]

[1]Associate Professor, Department of School of Computational Intelligence, Joy University, Tirunelveli, Tamil Nadu, India, 627116. Email: pkmanojkumar@joyuniversity.edu.in

[2]Assistant Professor, Department of Computer Science, Sri Ramakrishna College for Arts & Science for Women, Coimbatore, Tamil Nadu, India, 641006. Email: stellacs@srcw.ac.in

[3]Associate Professor, Department of Mathematics, AMET deemed to be University, Chennai, Tamil Nadu, India, 603112. Email: jenitha.g@ametuniv.ac.in

[4]Professor & Head, Department of Computer Science and Engineering, A.K.T Memorial College of Engineering and Technology, Kallakurichi, Tamil Nadu, India, 606213. Email: chisundar123@gmail.com

[5]Associate Professor, Department of Computer Science, SPPU, Pune, Maharashtra, India, 411058. Email: monicaapte@gmail.com

[6]Assistant Professor, Department of Artificial Intelligence and Machine Learning, Aditya University, Surampalem, Andhra Pradesh, India, 533005. Email: appalakonda.v@adityauniversity.in

[7]Head & Assistant Professor, Department of Biochemistry, Nadar Saraswathi College of Arts and Science (Autonomous), Theni, Tamilnadu, India, 625531. Email: venibio87@gmail.com

[8]Associate Professor, Department of Applied Mathematics, Bhilai Institute of Technology Durg, Durg, Chhattisgarh, India, 491001. Email: anilkumardby70@gmail.com

***Corresponding Author:** Dr. P. K. Manoj Kumar, Associate Professor, Department of School of Computational Intelligence, Joy University, Tirunelveli, Tamil Nadu, India, 627116. Email: pkmanojkumar@joyuniversity.edu.in

Abstract

Diabetes mellitus is a chronic metabolic disorder that has emerged as a major global public health challenge, affecting millions worldwide. Early detection plays a crucial role in preventing complications such as cardiovascular diseases, neuropathy, nephropathy, and retinopathy. In recent years, Artificial Intelligence (AI) and Machine Learning (ML) have significantly transformed predictive healthcare, particularly through the utilization of Electronic Health Records (EHRs). This study explores the role of AI-driven models in predicting diabetes using structured and unstructured clinical data extracted from EHR systems. The research focuses on the integration of supervised learning algorithms, deep learning techniques, and data preprocessing methods to enhance predictive accuracy and clinical decision-making. EHR-based prediction systems provide a comprehensive dataset including patient demographics, laboratory results, medical history, and lifestyle factors. AI algorithms such as Random Forest, Support Vector Machines, Logistic Regression, and Neural Networks have demonstrated high accuracy in identifying individuals at risk of developing diabetes. Furthermore, advanced techniques like feature selection, class imbalance handling, and explainable AI contribute to model transparency and reliability.

This paper presents a detailed review of existing literature, discusses methodologies, analyzes case studies, and evaluates challenges such as data privacy, bias, and interpretability. The findings highlight that AI-powered EHR systems can significantly improve early diagnosis, reduce healthcare costs, and support personalized treatment strategies. However, ethical considerations and data governance remain critical for successful implementation.

Overall, this study emphasizes the transformative potential of AI in predictive healthcare and advocates for interdisciplinary collaboration to enhance diabetes prediction systems.

Keywords: *Artificial Intelligence, Machine Learning, Diabetes Prediction, Electronic Health Records, Deep Learning, Healthcare Analytics, Predictive Modeling, Clinical Decision Support*

How to cite this article: Kumar PKM, Mary PS, Jenitha G, Jayasundar S, Apte M, Appalakonda V, Krishnaveni M, Dubey AK. AI-Powered Diabetes Prediction using Electronic Health Records. Int J Drug Deliv Technol. 2026;16(12s): 87-93. DOI: 10.25258/ijddt.16.12s.10

I. Introduction

Diabetes mellitus has emerged as one of the most pressing global health challenges of the twenty-first century, affecting over 500 million individuals worldwide and contributing significantly to morbidity and mortality. The condition is characterized by chronic hyperglycemia resulting from defects in insulin secretion, insulin action, or both. Its long-term complications, including cardiovascular disease, kidney failure, neuropathy, and retinopathy, pose a substantial burden on healthcare systems and significantly diminish patients' quality of life. Early detection and timely intervention are therefore critical to mitigating these adverse outcomes. However, traditional diagnostic approaches, which rely primarily on periodic clinical screenings such as fasting glucose tests or HbA1c measurements, often fail to identify high-risk individuals before the onset of irreversible complications.

The rapid advancement of Artificial Intelligence (AI) and Machine Learning (ML) technologies has opened new avenues for predictive healthcare, enabling early identification of disease risk through data-driven approaches. AI refers to the simulation of human intelligence processes by machines, particularly computer systems, while ML represents a subset of AI that focuses on algorithms capable of learning patterns from data. These technologies are particularly well-suited for analyzing complex, high-dimensional healthcare datasets, where traditional statistical methods may fall short. As Singla notes, AI-driven healthcare systems have the capacity to "identify hidden patterns and correlations within clinical datasets that remain undetected through conventional diagnostic techniques" (Singla 2019).

Electronic Health Records (EHRs) serve as a cornerstone for AI-based predictive modeling. EHRs are digital repositories of patient information that include structured data such as laboratory test results, medication histories, and vital signs, as well as unstructured data such as physician notes and clinical narratives. The longitudinal nature of EHR data allows for comprehensive tracking of patient health over time, making it an invaluable resource for predictive analytics. According to Brisimi et al., EHR-based models enable the integration of diverse clinical variables, thereby enhancing the accuracy and robustness of disease prediction systems (Brisimi et al. 2018).

In the context of diabetes prediction, AI models leverage EHR data to identify risk factors such as age, body mass index (BMI), family history, blood pressure, and glucose levels. Machine learning algorithms, including Logistic Regression, Decision Trees, Random Forests, Support Vector Machines (SVM), and Neural Networks—have been widely applied to classify individuals as diabetic or non-diabetic based on these features. More recently, deep learning techniques, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have demonstrated superior performance by capturing complex nonlinear relationships within the data (Esmaeilyfard et al. 2025).

Another important development in this field is the integration of predictive models into Clinical Decision Support Systems (CDSS). These systems assist healthcare professionals by providing real-time insights and recommendations based on patient data. Tuppad et al. emphasize that AI-powered CDSS can significantly enhance diagnostic accuracy and reduce clinical workload by automating routine decision-making processes (Tuppad et al. 2022). Furthermore, the adoption of explainable AI techniques has addressed concerns regarding the interpretability of machine learning models, enabling clinicians to understand the rationale behind predictions.

Despite these advancements, several challenges remain. Data quality issues, such as missing values, inconsistencies, and noise, can adversely affect model performance. Additionally, concerns regarding data privacy and security are particularly relevant in the context of EHR systems, which contain sensitive patient information. Ethical considerations, including algorithmic bias and fairness, must also be addressed to ensure equitable healthcare outcomes. As Hennebelle et al. argue, the integration of secure data-sharing frameworks and ethical guidelines is essential for the responsible deployment of AI in healthcare (Hennebelle et al. 2022).

This study aims to explore the application of AI-powered diabetes prediction using EHR data by examining existing literature, analyzing real-world case studies, and discussing methodological and ethical considerations. By synthesizing current research and identifying key challenges and opportunities, this paper contributes to the growing body of knowledge on AI-driven predictive healthcare and highlights its potential to transform diabetes management.

II. Methodology

The methodology adopted in this study involves a comprehensive analytical framework integrating EHR data preprocessing, feature engineering, model selection, and evaluation. Initially, structured datasets are extracted from Electronic Health Records, including demographic attributes (age, gender), clinical parameters (glucose levels, BMI, blood pressure), and medical history. Data preprocessing techniques such as normalization, missing value imputation, and outlier detection are applied to ensure data quality. Feature selection methods, including mutual information and correlation analysis, are employed to identify the most relevant predictors of diabetes.

Subsequently, multiple machine learning algorithms, including Logistic Regression, Random Forest, Support Vector Machines, and Neural Networks, are implemented to develop predictive models. Advanced techniques such as SMOTE are utilized to address class imbalance issues. Model performance is evaluated using accuracy, precision, recall, F1-score, and ROC-AUC metrics. Additionally, explainable AI tools such as SHAP and LIME are incorporated to interpret model decisions. The overall framework ensures robustness, scalability, and clinical relevance in predicting diabetes risk.

III. Review of Literature

The application of AI in diabetes prediction has been widely explored across multiple studies. Early research emphasized statistical models, but recent advancements have shifted towards machine learning and deep learning techniques.

Khokhar et al. highlight that machine learning algorithms such as SVM, XGBoost, and Logistic Regression have shown high effectiveness in diabetes prediction (Khokhar et al. 2025). Similarly, Kiran and Xie provide a bibliometric analysis of 33 years of research, demonstrating the growing complexity and adoption of AI techniques in diabetes prediction (Kiran and Xie 2025).

Singla emphasizes the role of AI in EHR-based diagnosis, stating that predictive models can identify undiagnosed diabetes cases and improve early intervention (Singla 2019). Tasin et al. developed a machine learning model using the Pima Indian dataset and achieved significant accuracy through ensemble techniques (Tasin et al. 2022).

Larabi-Marie-Sainte discusses hybrid machine learning models, noting that combining multiple algorithms enhances predictive performance (Larabi-Marie-Sainte 2019). Tuppad et al. highlight the importance of clinical decision support systems in diabetes management (Tuppad et al. 2022).

Recent studies have explored deep learning approaches. Farnoosh et al. introduced an attention-based CNN model that improves prediction accuracy by addressing data imbalance (Farnoosh et al. 2025). Esmaeilyfard et al. found that deep learning models outperform traditional methods in predictive accuracy (Esmaeilyfard et al. 2025).

Research also focuses on EHR-based prediction. Brisimi et al. demonstrated the use of EHR data in predicting diabetes-related hospitalizations (Brisimi et al. 2018). Nasir and Malik proposed SACDNet, achieving high accuracy using EHR datasets (Nasir and Malik 2023).

Additionally, Bontha et al. emphasized the integration of lifestyle and clinical data for improved prediction (Bontha et al. 2025). Rahman et al. highlighted the importance of feature selection and data balancing techniques (Rahman et al. 2025).

Recent advancements include wearable technology integration, which enhances real-time data collection and predictive modeling (Fraser et al. 2025). Osonuga noted the importance of supervised learning in predictive healthcare systems (Osonuga 2025).

Overall, the literature indicates that AI-based diabetes prediction using EHR data is a rapidly evolving field with significant potential for improving healthcare outcomes.

Recent research by Choi et al. explores the use of recurrent neural networks (RNNs) for predicting chronic diseases, including diabetes, using longitudinal EHR data. The study demonstrates that temporal modeling significantly improves prediction accuracy by capturing patient health trajectories over time. The authors emphasize that sequential learning methods outperform static models when dealing with time-series healthcare data (Choi et al. 2016).

Similarly, Rajkomar et al. conducted a large-scale study using deep learning models trained on EHR data from multiple hospitals. Their findings indicate that deep neural networks can predict various clinical outcomes, including diabetes onset, with high precision. The study highlights the scalability of AI systems in handling large healthcare datasets and improving predictive analytics (Rajkomar et al. 2018).

In another study, Miotto et al. introduced the “Deep Patient” model, which uses unsupervised learning techniques to extract meaningful representations from EHR data. The model successfully predicts future disease risks, including diabetes, by identifying hidden patterns in patient records. This research underscores the importance of representation learning in healthcare analytics (Miotto et al. 2016).

Beam and Kohane provide a comprehensive overview of big data applications in healthcare, emphasizing the role of machine learning in disease prediction. They argue that AI-driven models can transform clinical practice by enabling early diagnosis and personalized treatment strategies. Their work also discusses the challenges of data integration and model interpretability (Beam and Kohane 2018).

Research by Shickel et al. focuses on deep learning approaches for EHR analysis, highlighting the potential of neural networks in extracting insights from complex clinical data. The authors note that deep learning models can automatically learn hierarchical features, reducing the need for manual feature engineering and improving prediction performance (Shickel et al. 2017).

Chen et al. examine the application of gradient boosting algorithms in diabetes prediction. Their study demonstrates that ensemble methods, such as XGBoost, outperform traditional machine learning techniques in terms of accuracy and robustness. The authors attribute this performance to the algorithm's ability to handle nonlinear relationships and missing data effectively (Chen et al. 2019).

Another important contribution comes from Zhang et al., who developed a hybrid model combining machine learning and statistical methods for diabetes risk prediction. The study highlights the benefits of integrating domain knowledge with data-driven approaches, resulting in improved predictive accuracy and clinical relevance (Zhang et al. 2020).

Kavakiotis et al. provide a systematic review of machine learning applications in diabetes research, covering classification, clustering, and regression techniques. The authors conclude that machine learning has significantly advanced diabetes diagnosis and management, particularly through the use of EHR data and wearable devices (Kavakiotis et al. 2017).

In a more recent study, Islam et al. explore the use of explainable AI models in diabetes prediction. Their research emphasizes the importance of interpretability in clinical settings, demonstrating how explainable models can provide insights into risk factors and support decision-making processes (Islam et al. 2021).

Finally, Ristevski and Chen discuss the integration of big data analytics and AI in healthcare, focusing on predictive modeling and decision support systems. They highlight the role of EHR data in enabling comprehensive analysis and improving disease prediction accuracy, while also addressing challenges related to data privacy and security (Ristevski and Chen 2018).

IV. Case Studies

Case Study 1: Large-Scale EHR-Based Deep Learning Model (United States)

A landmark study conducted across multiple healthcare institutions in the United States utilized large-scale EHR datasets to develop a deep learning-based diabetes prediction model. The dataset included millions of patient records encompassing demographic information, laboratory results, medication histories, and clinical diagnoses. The researchers implemented a deep neural network architecture capable of processing both structured and unstructured data. The model demonstrated superior predictive performance compared to traditional machine learning algorithms, achieving high sensitivity and specificity in identifying individuals at risk of developing diabetes. One of the key strengths of this approach was its ability to incorporate temporal data, allowing the model to analyze changes in patient health over time. This longitudinal analysis enabled early detection of diabetes risk even before clinical symptoms became apparent. Nasir and Malik highlight that such EHR-based deep learning models can significantly enhance early diagnosis and reduce hospital admissions by enabling proactive intervention strategies (Nasir and Malik 2023). Furthermore, the integration of this model into hospital information systems facilitated real-time clinical decision support, improving patient outcomes and optimizing resource allocation.

Case Study 2: Machine Learning-Based Prediction in Bangladesh Healthcare System

In a developing country context, Tasin et al. developed a machine learning-based diabetes prediction system using both public datasets (such as the Pima Indian Diabetes dataset) and locally collected clinical data from healthcare facilities in Bangladesh. The study aimed to create an accessible and cost-effective solution for early diabetes detection in resource-constrained environments.

The researchers employed multiple machine learning algorithms, including Decision Trees, Random Forests, and Support Vector Machines, to evaluate predictive performance. Feature selection techniques were applied to identify the most significant risk factors, such as glucose levels, BMI, age, and insulin levels. Additionally, the Synthetic Minority Oversampling Technique (SMOTE) was used to address class imbalance issues within the dataset. The final model achieved high accuracy and was deployed as a mobile application, enabling healthcare providers and patients to assess diabetes risk in real time. According to Tasin et al., this approach demonstrates the potential of AI-driven solutions to bridge gaps in healthcare accessibility and improve early diagnosis in low-resource settings (Tasin et al. 2022). The study also emphasized the importance of user-friendly interfaces and scalability in ensuring widespread adoption.

Case Study 3: Integration of AI in Clinical Decision Support Systems (Global Perspective)

Another significant case study involves the integration of AI-based diabetes prediction models into Clinical Decision Support Systems (CDSS) used in hospitals worldwide. These systems analyze EHR data to provide real-time recommendations to clinicians, aiding in diagnosis, treatment planning, and patient monitoring. Singla describes how AI-powered CDSS can identify high-risk patients by analyzing patterns in EHR data, including lifestyle factors, comorbidities, and laboratory results (Singla 2019). For example, patients with prediabetic conditions can be flagged for further testing and preventive interventions, thereby reducing the likelihood of disease progression.

Moreover, these systems incorporate explainable AI techniques, allowing clinicians to understand the factors contributing to each prediction. This transparency enhances trust and facilitates clinical adoption. The implementation of AI-driven CDSS has been associated with improved diagnostic accuracy, reduced medical errors, and enhanced patient outcomes. However, challenges such as data interoperability and system integration remain significant barriers to widespread implementation.

Case Study 4: Wearable Device and EHR Integration for Real-Time Prediction

Recent advancements have explored the integration of wearable health devices with EHR systems to enable real-time diabetes prediction. Wearable devices, such as continuous glucose monitors and fitness trackers, generate large volumes of real-time data, including physical activity, heart rate, and glucose levels. Fraser et al. developed a predictive model that combines wearable device data with EHR information to provide continuous risk assessment. The model utilizes machine learning algorithms to analyze both historical and real-time data, enabling dynamic prediction of diabetes risk (Fraser et al. 2025).

This approach represents a significant step toward personalized healthcare, as it allows for continuous monitoring and timely intervention. Patients can receive alerts and recommendations based on their current health status, promoting proactive disease management. However, issues related to data integration, privacy, and device accuracy must be addressed to ensure reliability and effectiveness.

V. Discussion

The application of AI in diabetes prediction using Electronic Health Records represents a transformative development in modern healthcare, offering significant improvements in early diagnosis, risk assessment, and personalized treatment. The integration of machine learning algorithms with EHR data enables healthcare systems to leverage vast amounts of clinical information, facilitating data-driven decision-making and enhancing patient outcomes. One of the primary advantages of AI-based prediction models is their ability to handle high-dimensional and heterogeneous data. EHR datasets typically include a wide range of variables, such as demographic information, laboratory results, medication histories, and lifestyle factors. Machine learning algorithms can process these diverse data types and identify complex relationships that are not easily detectable through traditional statistical methods. As Khokhar et al. note, advanced algorithms such as Random Forest and XGBoost have demonstrated high accuracy in predicting diabetes by capturing nonlinear interactions among variables (Khokhar et al. 2025).

Deep learning techniques have further advanced the field by enabling the analysis of unstructured data, such as clinical notes and imaging records. These models can automatically extract relevant features from raw data, reducing the need for manual feature engineering. Esmaeilyfard et al. emphasize that deep learning models outperform conventional approaches in terms of predictive accuracy and scalability, particularly when dealing with large datasets (Esmaeilyfard et al. 2025). However, the complexity of these models also raises concerns regarding interpretability and computational requirements.

Another critical aspect of AI-powered diabetes prediction is feature selection. Identifying the most relevant predictors is essential for improving model performance and reducing computational complexity. Studies have shown that variables such as glucose levels, BMI, age, and family history are among the most significant predictors of diabetes. Rahman et al. highlight the importance of feature selection techniques in enhancing model accuracy and reducing overfitting (Rahman et al. 2025). Additionally, addressing class imbalance through techniques such as SMOTE is crucial for ensuring reliable predictions, particularly in datasets with a disproportionate number of non-diabetic cases.

Explainability is a key requirement for the adoption of AI in healthcare. Clinicians must be able to understand and trust the predictions generated by machine learning models. Explainable AI techniques, such as SHAP and LIME, provide insights into the contribution of individual features to model predictions. Tasin et al. argue that these techniques enhance transparency and facilitate clinical decision-making by enabling healthcare professionals to interpret model outputs (Tasin et al. 2022). Despite the numerous advantages, several challenges must be addressed to ensure the successful implementation of AI-based diabetes prediction systems. Data quality remains a significant concern, as EHR datasets often contain missing values, inconsistencies, and noise. Robust data preprocessing techniques are essential for mitigating these issues and ensuring reliable model performance. Furthermore, data privacy and security are critical considerations, given the sensitive nature of healthcare data. Hennebelle et al. propose the use of blockchain technology and secure data-sharing frameworks to address these challenges and ensure data integrity (Hennebelle et al. 2022).

Ethical considerations, including algorithmic bias and fairness, also play a crucial role in the deployment of AI in healthcare. Bias in training data can lead to unequal predictions across different demographic groups, potentially exacerbating existing healthcare disparities. Therefore, it is essential to develop unbiased models and implement fairness-aware algorithms to ensure equitable outcomes. Another important challenge is the integration of AI systems into existing healthcare infrastructures. Interoperability issues between different EHR systems can hinder data sharing and limit the effectiveness of predictive models. Standardization of data formats and the adoption of interoperable systems are necessary to facilitate seamless integration. In addition to these challenges, the adoption of AI in healthcare requires collaboration between multiple stakeholders, including clinicians, data scientists, policymakers, and technology developers. Training healthcare professionals to effectively use AI tools is also essential for maximizing their potential benefits.

Overall, AI-powered diabetes prediction using EHR data holds immense promise for improving healthcare outcomes. By enabling early detection, personalized treatment, and efficient resource allocation, these systems can significantly enhance the quality of care. However, addressing technical, ethical, and organizational challenges is essential for realizing the full potential of AI in predictive healthcare.

VI. Conclusion

AI-powered diabetes prediction using Electronic Health Records has emerged as a powerful tool in modern healthcare. The integration of machine learning and deep learning techniques enables early detection, personalized treatment, and improved patient outcomes. This study demonstrates that EHR-based predictive models can significantly enhance clinical decision-making and reduce healthcare costs. Despite the advancements, challenges such as data quality, interpretability, and ethical concerns remain. Future research should focus on developing explainable and unbiased models, improving data integration, and ensuring privacy protection. Interdisciplinary collaboration between healthcare professionals, data scientists, and policymakers is essential to maximize the potential of AI in diabetes prediction. In conclusion, AI-driven predictive healthcare systems hold immense promise in combating diabetes and transforming global healthcare delivery.

Works Cited

1.Bontha, S. S., et al. *Predicting Risk and Complications of Diabetes*. 2025.
2.Brisimi, T. S., et al. "Predicting Diabetes Hospitalizations Using EHR." 2018.
3.Esmaeilyfard, R., et al. *Enhancing AI-driven Forecasting of Diabetes*. 2025.
4.Farnoosh, R., et al. *DiabetesXpertNet Model*. 2025.
5.Fraser, R. A., et al. *AI and Wearable Technology in Diabetes*. 2025.
6.Hennebelle, A., et al. *AI Blockchain Diabetes Prediction*. 2022.
7.Khokhar, P. B., et al. *AI for Diabetes Prediction Review*. 2025.
8.Kiran, M., and Y. Xie. *ML and AI in Diabetes Prediction*. 2025.
9.Larabi-Marie-Sainte, S. *Techniques for Diabetes Prediction*. 2019.
10.Nasir, T., and M. Malik. *SACDNet EHR Model*. 2023.
11.Osonuga, A. *Generative AI in Diabetes Prediction*. 2025.
12.Rahman, F., et al. *Feature Selection in Diabetes Prediction*. 2025.
13.Singla, R. *AI in Diabetes Care*. 2019.
14.Tasin, I., et al. *ML-based Diabetes Prediction*. 2022.

15.Tuppad, A., et al. *Clinical Decision Support Systems*. 2022.
16.Malik, M. B., et al. *Lifestyle-based Diabetes Prediction*. 2022.
17.Ismail, L., et al. *AI Methods in Diabetes Prediction*. 2022.
18.Qin, X., et al. *Lifestyle Data and Diabetes Prediction*. 2023.
19.Yuk, H., et al. *AI-based Diabetes Prediction in Korea*. 2023.
20.Aiello, M., et al. *AI Healthcare Systems*. 2019.
21.Pranto, T., et al. *Diabetes Prediction Models*. 2021.
22.Hennebelle, A., et al. *HealthEdge Framework*. 2023.
23.Khan, S., et al. *AI Risk Prediction Study*. 2025.
24.Lee, H., et al. *Diabetes Prediction in Older Adults*. 2025.
25.Ganie, S. M., et al. *Lifestyle Prediction Models*. 2022.