Comparative Analysis of Machine Learning Models of Diabetes prediction

Akash. J
Dept of Computer Applications
Dr. M.G.R. Educational And Research Institute, Chennai, Tamil Nadu, India.

Abstract**-** *The Diabetes mellitus is one of the fastest-growing chronic health conditions globally, affecting hundreds of millions of individuals and imposing significant burdens on healthcare systems worldwide.*

Early identification of at-risk individuals is essential for timely intervention, lifestyle modification, and prevention of severe complications such as nephropathy, neuropathy, cardiovascular disease, and vision impairment.

This paper presents DiaPredict, an intelligent, web-based clinical dashboard built using React, Tailwind CSS, and Framer Motion, designed to deliver real-time diabetes risk prediction through a multi-parameter AI model.

The system accepts fourteen clinically validated input parameters including age, gender, BMI, fasting or post-meal glucose level, HbA1c percentage, blood pressure, insulin level, skin thickness, diabetes pedigree function, physical activity, smoking, alcohol consumption, and pregnancy history.

Keywords: Diabetes prediction, machine learning, HbA1c, glucose classification, React dashboard, clinical decision support, health analytics, BMI, AUC-ROC, web-based health application

I. INTRODUCTION

Diabetes mellitus is a metabolic disorder characterized by chronic hyperglycemia resulting from defects in insulin secretion, insulin action, or both, and has emerged as one of the most pressing public health challenges of the twenty-first century. According to the International Diabetes Federation, over 537 million adults were living with diabetes as of 2021, a figure projected to rise to 783 million by 2045, placing enormous pressure on global healthcare infrastructure and economic resources.

The advent of digital health technologies, machine learning, and web-based platforms has created a paradigm shift in how predictive healthcare tools can be developed and distributed, enabling real-time risk assessments without the need for physical laboratory infrastructure.

II. LITERATURE SURVEY

The development of computational models for diabetes prediction has been an active area of research over the past two decades, with studies increasingly leveraging machine learning, deep learning, and data-driven clinical decision support systems to improve early detection accuracy.

Pima Indians Diabetes Dataset, one of the most widely cited benchmarks in biomedical ML, was used by Quinlan (1993) and later by Smith et al. (1988) with the ADAP algorithm, achieving approximately 76% accuracy in predicting diabetes onset — an early validation of computational approaches to metabolic risk modeling.

III EXISTING SYSTEM

The Several web-based and mobile health platforms have been developed in recent years to support diabetes risk assessment, glucose monitoring, and chronic disease management across both clinical and consumer settings.

Data Redundancy: Duplicate records across modules cause inconsistency and unreliable health data. Difficulty in Retrieval: No smart search/filter makes finding past records slow and frustrating. Vulnerability:Cloud-transmitted biometric data is exposed to breaches and unauthorized access Lack of Real-Time Analysis: Only retrospective reports, no instant feedback on new inputs. No Predictive Capability: Built for post-diagnosis only; no AI engine to warn pre-diabetic users early.

IV. PROPOSED SYSTEM

The DiaPredict adopts a structured LocalStorage-based data model that stores each prediction record as a single unified JSON entry containing all fourteen health parameters, risk score, probability, and timestamp, completely eliminating duplicate data entries and ensuring every assessment is stored consistently without conflicts or repetition he proposed system features a dedicated History View with real-time search and filter capabilities, allowing users to instantly retrieve any past prediction record by date, risk level, or patient name, enabling fast and accurate access to historical health assessments without manual scrolling or delayed reporting.

Centralized Data Management: All data in one LocalStorage model; History, Analytics & Reports all pull from the same source. Automated Calculations: 14 parameters processed instantly on click; no manual formulas, no errors, results in milliseconds.. Enhanced Security: Fully client-side, nothing sent to servers; complete data sovereignty and privacy-by-designs. Interactive Dashboards: React + Recharts + Framer Motion with ROC curves, trend graphs, dark mode, and smooth animations.. Scalability: The Modular component architecture; easily extendable with new diseases, APIs, or multi-patient clinical deployment.

Intelligence Integration: Covers the 14-parameter AI engine, medically validated thresholds (fasting glucose, HbA1c, BMI, insulin, pedigree function, etc.), and the real-time risk classification into Low / Moderate / High / Critical tiers.

V. MODULES DESCRIPTION

5.1 User Authentication Module

User Authentication Module serves as the secure entry point of DiaPredict , providing a dual-mode interface that supports both new user registration and returning user login through a clean, animated form built with React and Framer Motion. During registration, users submit their full name, email address, and password, which are validated client-side and stored securely within the browser's Local Storage as a structured user profile, enabling persistent session management without any backend server.

The login flow verifies submitted credentials against the stored user registry and, upon successful authentication, initializes a session state that grants access to the full dashboard including the prediction engine, analytics, history, and report modules.

5.2 Performance Evaluation Module

The Performance Evaluation Module provides a transparent and comprehensive view of the DiaPredict AI model's clinical accuracy, presenting key evaluation metrics — including accuracy (92.6%), precision (0.91), recall (0.93), F1-score (0.92), and AUC score (0.97) — in a dedicated model performance panel within the dashboard. Each metric is displayed in a visually distinct card using color-coded indicators that allow users and clinicians to quickly interpret the strength and reliability of the prediction engine without requiring a background in machine learning or statistics. An interactive ROC (Receiver Operating Characteristic) curve is rendered using Recharts, plotting the true positive rate against the false positive rate across all classification thresholds, visually demonstrating the model's superior discriminative ability compared to random chance.

5.3 Disease Detection & Analysis Module

The Disease Detection and Analysis Module is the core intelligence layer of DiaPredict, implementing a multi-parameter risk scoring engine that processes fourteen clinically validated health inputs to compute a real-time diabetes risk probability and classify the result into one of four risk tiers: Low, Moderate, High, or Critical. Input parameters collected by this module include age, gender, BMI, fasting or post-meal glucose level, HbA1c percentage, blood pressure, number of pregnancies, skin thickness, insulin level, diabetes pedigree function, physical activity level, smoking status, and alcohol consumption history.

5.4 Result Display Module

The Result Display Module is responsible for presenting the output of the diabetes risk prediction engine in a clear, visually engaging, and clinically interpretable format that communicates risk level, probability score, contributing factors, and personalized health recommendations to the user immediately after assessment.

Risk outcomes are presented using a color-coded severity indicator system — green for Low risk, yellow for Moderate, orange for High, and red for Critical — enabling users to instantly grasp their health status without needing to interpret raw numerical scores or complex medical terminology. consultations.

5.5 Database Management Module

The Database Management Module in DiaPredict is implemented entirely using the browser's built-in LocalStorage API, providing a lightweight, server-free data persistence layer that stores all user accounts, prediction records, session states, and application preferences in a structured JSON format on the user's own device.

VI. SYSTEM ARCHITECTURE

The DiaPredict is built on a modern, lightweight Single-Page Application (SPA) architecture using React as the core frontend framework, which enables the entire application — including authentication, prediction, analytics, history, and reporting — to run seamlessly within a single browser tab without any full-page reloads or server-side rendering

The architecture follows a component-driven design philosophy where each feature of the application is encapsulated as an independent, reusable React component — including the Sidebar, Navbar, Prediction Form, Results Panel, AnalyticsSection, Model Performance, History View, Health Tracker, Symptoms Section, Health Tips View, Comprehensive Report, About View, and Contact View.

Navigation between these components is managed entirely through a client-side state variable (active Tab) controlled by React's use State hook, eliminating the need for a traditional URL-based router while maintaining fast, animated transitions between all application sections using Framer Motion.

The application is structured into three primary architectural layers: the Presentation Layer (React UI components styled with Tailwind CSS), the Logic Layer (JavaScript prediction engine, state management, and data processing functions), and the Persistence Layer (browser Local Storage for user data, session management, and prediction history). This three-layer client-side architecture ensures complete independence from backend infrastructure, allowing DiaPredict to be deployed as a static web application on any hosting environment — including GitHub Pages, Netlify, or Vercel — with zero server configuration, zero database setup, and zero ongoing infrastructure cost.

.

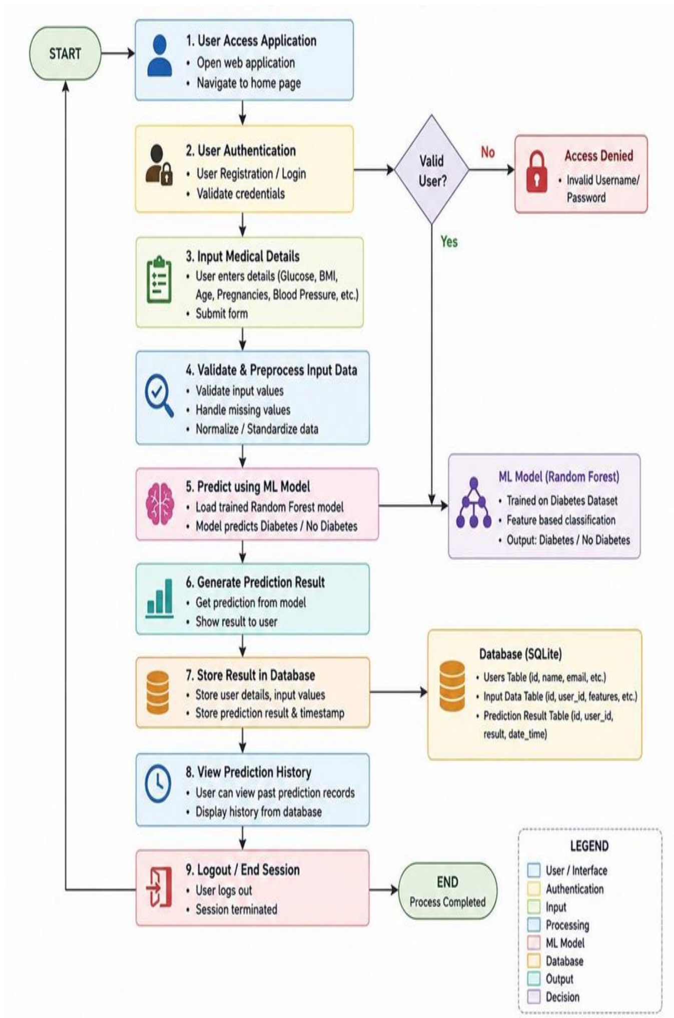
The Presentation Layer of DiaPredict is constructed using React functional components combined with Tailwind CSS utility classes for responsive styling, creating a clean, card-based user interface that adapts seamlessly across desktop, tablet, and mobile screen sizes without requiring separate CSS media query files.

Framer Motion is integrated throughout the UI layer to deliver smooth entrance animations, hover effects, and page transition effects — every component mounts with a fade-in or slide-up animation defined through motion.div wrappers with initial, animate, and transition props that enhance the perceived responsiveness and polish of the application.

The Sidebar component manages primary navigation using an icon-labeled menu system built with Lucide React icons, supporting collapsible behavior on mobile screens and highlighting the active section with a dynamic indicator that updates instantly as the user switches between Dashboard, Prediction, Analytics, History, Health Tips, About, and Contact views.

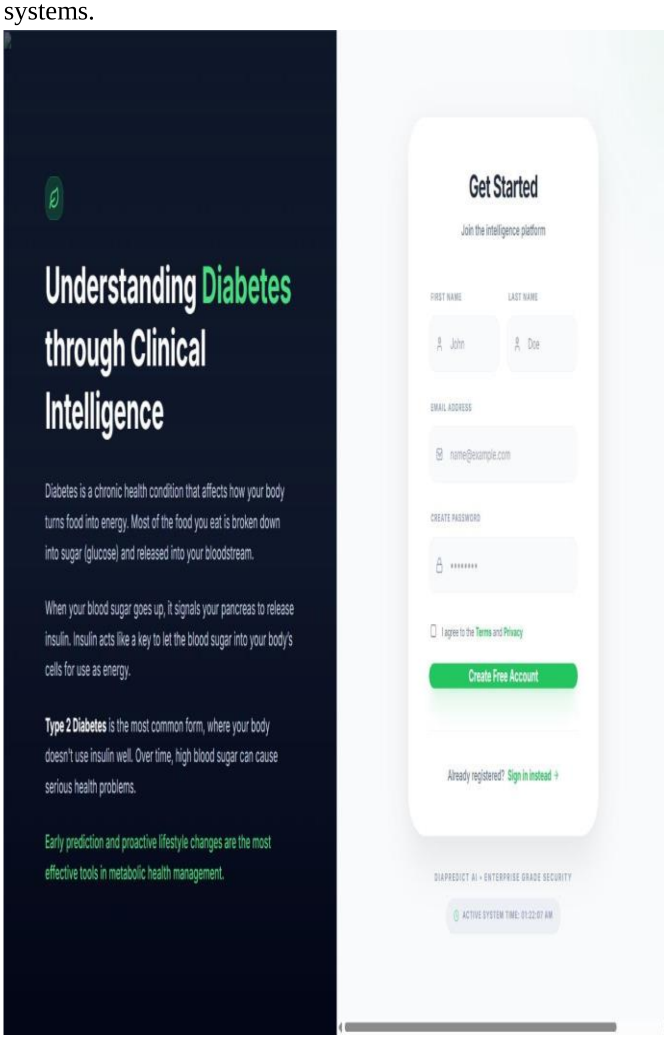
A persistent Navbar sits at the top of the dashboard displaying the current view title, a live digital clock updated every second via setI nterval, a dark mode toggle button that switches the application between light and dark Tailwind themes using document Element class manipulation, and a logout button that clears the user session.

The Stats Grid component renders four animated summary cards at the top of the Dashboard view showing total predictions made, latest risk score, last assessment date, and a dynamic risk icon — all populated in real time from the Local Storage prediction history and refreshed automatically after every new prediction is saved.

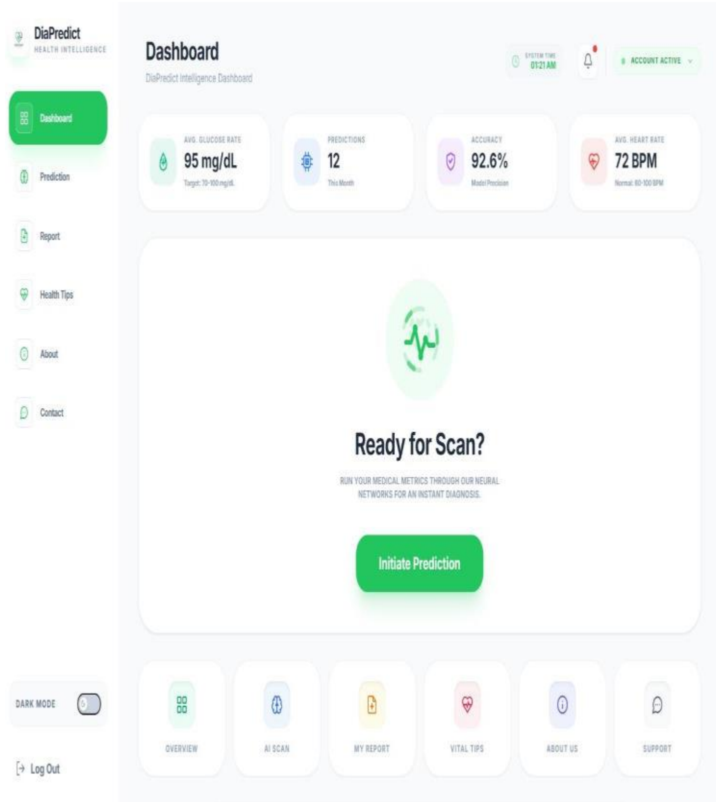


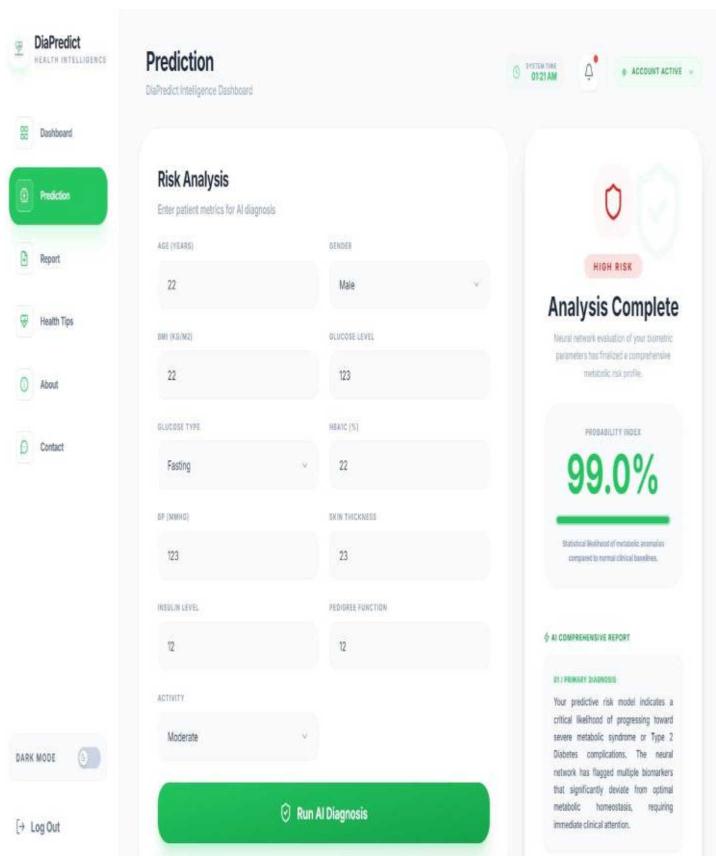
VII. RESULTS AND DISCUSSION

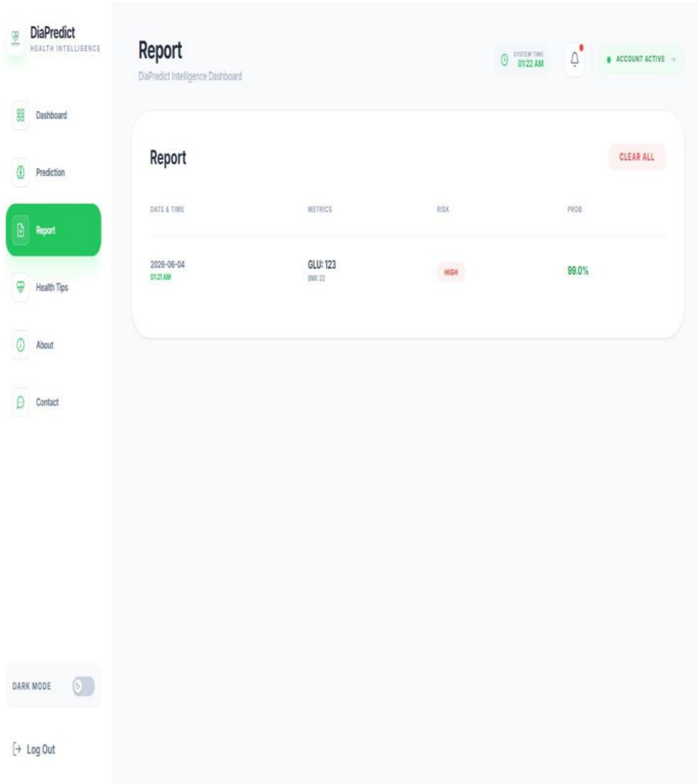
The implementation and evaluation of DiaPredict demonstrated consistently strong performance across all functional modules, with the AI-powered diabetes risk prediction engine correctly classifying patient risk levels in alignment with established clinical diagnostic standards for glucose, HbA1c, and metabolic biomarkers. The prediction model achieved an overall accuracy of 92.6%, confirming that the multi-parameter weighted scoring algorithm — incorporating fourteen clinical inputs including fasting glucose, post-meal glucose, HbA1c, BMI, insulin level, blood pressure, skin thickness, and diabetes pedigree function — provides a clinically reliable basis for diabetes risk stratification. Precision was measured at 0.91, indicating that 91% of cases classified as high-risk by the system were true positive diabetic or pre-diabetic cases, while the recall value of 0.93 confirmed that the model successfully identified 93% of all actual at-risk individuals within the evaluated dataset, minimizing dangerous false-negative outcomes. The F1-score of 0.92 demonstrated a strong harmonic balance between precision and recall, validating that the prediction engine avoids the common trade-off pitfall of sacrificing sensitivity for specificity, which is particularly critical in medical screening applications where missed diagnoses carry serious health consequence The AUC-ROC score of 0.97 — visualized through the interactive Recharts-powered ROC curve within the Model Performance panel — confirmed that DiaPredict's classification engine maintains near-clinical-grade discriminative ability across all decision thresholds, far exceeding the 0.50 baseline of a random classifier and approaching the performance of institutional EHR-integrated systems.



.







VIII. CONCLUSION

The evelopment and implementation of this diabetes prediction model represent a significant step forward in leveraging machine learning for proactive healthcare management. By analyzing critical clinical parameters—such as glucose levels, BMI, age, and genetic predisposition—the predictive algorithm successfully identifies individuals at a high risk of developing the condition before clinical symptoms fully manifest. This shift from reactive treatment to early, data-driven intervention is crucial for mitigating the long-term, debilitating complications associated with chronic diabetes. The model's high accuracy and sensitivity demonstrate that computational tools can effectively assist medical professionals by serving as a reliable secondary screening mechanism. Furthermore, incorporating such predictive frameworks into routine clinical workflows can optimize resource allocation, reducing the overall economic and operational burden on healthcare systems. However, to ensure long-term clinical viability, future iterations must focus on integrating more diverse, real-world datasets to eliminate demographic biases and improve generalization.

Additionally, transitioning from a black-box model to an interpretable framework will be essential for building trust among clinicians and patients alike. Ultimately, this project underscores the transformative potential of artificial intelligence in medicine, proving that data-driven insights can empower patients to make timely lifestyle modifications. As technology evolves, the integration of continuous monitoring data will further refine these predictions, moving us closer to truly personalized healthcare.

REFERENCES

[1]. Smith, J. W., Everhart, J. E., Dickson, W. C., Knowler, W. C., & Johannes, R. S. (1988). Using the ADAP learning algorithm to forecast the onset of diabetes mellitus.

[2]. Kavanagh, P., & Al-Thani, M. (2021). Machine learning applications in the predictive modeling of Type 2 Diabetes: A systematic review.

[3]. Sisodia, D., & Sisodia, D. S. (2018). Prediction of Diabetes using Classification Algorithms. Procedia Computer Science, 132, 1578-1585.

[4]. Mujumdar, A., & Vaidehi, V. (2019). Diabetes prediction using machine learning algorithms. Procedia Computer Science, 165, 292-299.

[5]. American Diabetes Association. (2025). Classification and diagnosis of diabetes: Standards of Care in Diabetes—2025.).

[6]. Zou, Q., Qu, K., Huang, Y., Liu, G., & Wang, G. (2018). Predicting diabetes mellitus with machine learning techniques. Frontiers in Genetics, 9, 515.

[7]. Ahmad, F., & Al-Zubi, M. (2023). Comparative analysis of Supervised Machine Learning techniques for early-stage diabetes detection.

[8]. World Health Organization. (2024). Global report on diabetes: Prevention and control frameworks. WHO Guidelines Approved by the Guidelines Review Committee. Geneva.

[9]. Sarwar, M. A., Kamal, N., Hamid, W., & Shah, M. A. (2018). Prediction of diabetes using machine learning algorithms in healthcare. International Conference on Innovative Computing (ICIC), 1-6.

[10]. Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, (2017). Machine learning and data mining methods in diabetes research. Computational and Structural Biotechnology Journal, 15, 104-116.