iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IX **Month of publication:** September 2025

DOI: https://doi.org/10.22214/ijraset.2025.74122

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

Diabetic Risk Prediction Based on Medical and Lifestyle Patterns

Paladi Niharika[1], Kuchimanchi Jayasri[2], Varshini Priyamvada Banam[3]
[1]Student, School of Informatics, Department of MCA, Aurora Deemed University, Hyderabad
[2]Assistant Professor, School of Engineering, Department of CSE, Aurora Deemed University, Hyderabad
[3]Assistant Professor, School of Informatics, Department of MCA, Aurora Deemed University, Hyderabad

Abstract: Diabetes mellitus is a common chronic condition that creates significant challenges for public health. Early risk prediction can help with timely prevention and better self-management. Many studies focus only on clinical data, but fewer consider lifestyle and behavioural signals that greatly affect disease onset. This work introduces a machine-learning web application that estimates a person's risk of diabetes by looking at clinical factors (age, BMI, HbA1c, blood pressure, family history) and lifestyle habits (diet, physical activity, stress, sleep quality, and screen time). The system is built on the MERN stack to ensure secure data storage, user authentication, and interactive visualizations. Our model classifies users into Low, Moderate, or High-risk categories and offers practical, personalized advice for diet, exercise, and stress management. Experiments showed that combining lifestyle factors with medical features improved prediction accuracy compared to using clinical data alone, with analysis showing added value from lifestyle variables. These findings highlight the potential of digital health tools that mix medical and everyday data to help people make informed choices. We include basic protections like consent, role-based access, and anonymized storage to safeguard user privacy.

Keywords: Diabetes Risk Prediction, Lifestyle and Medical Data, Machine Learning, MERN Stack, Preventive Healthcare, Personalized Recommendations.

I. INTRODUCTION

Diabetes mellitus is among the most widespread chronic illnesses, and its global incidence continues to rise at an alarming rate. International health organizations predict a substantial increase in the number of individuals affected in the coming decades, posing severe challenges for both patients and healthcare systems. Early identification of individuals at risk is therefore critical to reduce complications such as cardiovascular disease, renal impairment, neuropathy, and vision loss. Traditionally, risk assessment relies on clinical parameters including fasting blood glucose, HbA1c, body mass index (BMI), and blood pressure. While these indicators remain essential, there is growing recognition that lifestyle determinants—such as dietary patterns, physical inactivity, psychological stress, inadequate sleep, and sedentary screen time—play a significant role in the onset and progression of diabetes.

In recent years, researchers have applied machine learning techniques to clinical datasets for automated diabetes risk prediction. These approaches have shown promise, yet most studies remain constrained to purely medical records, overlooking the contribution of behavioural and environmental factors. Many works are also limited by small or imbalanced datasets, restricted interpretability, and binary classification outputs that categorize individuals simply as "diabetic" or "non-diabetic," without accounting for intermediate states such as prediabetes. Furthermore, relatively few systems translate predictive models into real-time, user-friendly platforms that both estimate risk and provide continuous guidance for healthier lifestyle adoption.

To address these limitations, this study presents a diabetes risk prediction framework that integrates both clinical and lifestyle attributes, enabling more comprehensive and accurate assessment. The system is implemented using the MERN (MongoDB, Express.js, React.js, Node.js) stack to deliver a secure, scalable, and interactive web platform. Machine learning models trained on combined data classify individuals into low-, moderate-, or high-risk categories. In addition to prediction, the system supports longitudinal tracking by storing user health records over time and provides tailored recommendations for diet, physical activity, and stress management. Interactive dashboards and graphical visualizations further enhance interpretability, allowing users to monitor health trajectories with ease.

This work contributes to preventive digital healthcare by offering a holistic approach to diabetes risk prediction. By integrating lifestyle and medical features within a single intelligent platform, the proposed system empowers individuals to adopt timely interventions, supports personalized disease prevention, and demonstrates how predictive analytics can strengthen public health outcomes.

II. LITERATURE SURVEY

A. Lifestyle-Based Machine Learning Approaches

Qin et al. (2022) explored diabetes prediction using only lifestyle factors, avoiding invasive medical tests. They applied several machine learning models, including CATBoost, XGBoost, Random Forest, SVM, and Logistic Regression, on the NHANES dataset. CATBoost achieved the best results with 82.1% accuracy, supported by SHAP analysis for model interpretability. While the study highlighted the power of lifestyle information in prediction, it was limited to U.S. data and lacked validation on other populations.

B. Traditional Clinical Models for Early Prediction

Alzboon et al. (2025) compared traditional ML techniques such as Neural Networks, Logistic Regression, and Random Forest on the PIMA Indian Diabetes Dataset. The Neural Network achieved the highest accuracy of 78.57%. The study's strength lay in its systematic comparison across multiple models, but its scope was limited by reliance on clinical features and a small, ethnicity-specific dataset.

C. Large-Scale Clinical Data Analysis

Adler (2021) investigated both diagnosed and undiagnosed diabetes cases using 15 different machine learning models, including SVM and Elastic Net, applied to NHANES data. Elastic Net showed the best performance for undiagnosed cases (AUC 0.944), while SVM worked better for diagnosed ones (AUC 0.923). This dual-model approach provided real-world relevance, though external validation was missing, and neural networks underperformed.

D. Regional Cohort-Based Prediction Models

Li et al. (2023) developed diabetes risk models using a massive dataset of over four million individuals from Western China. By applying XGBoost, TabNet, and MLP, they achieved an AUC of 0.9122 and accuracy of 83.14%. The study demonstrated strong real-world potential due to its large dataset but was limited to cross-sectional data and lacked generalizability beyond the Chinese population.

E. Reviews of AI-Based Precision Methods

Mohsen et al. (2023) conducted a scoping review of 40 AI-driven studies for diabetes prediction, covering diverse data types such as EHR, imaging, and omics. They observed that multimodal models, which combine multiple data sources, performed better (average AUC ~0.89) compared to unimodal models (AUC ~0.81). However, most studies lacked reproducibility, open code, and external testing, limiting their clinical application.

F. Simple Lifestyle-Based Models

Nguyen and Zhang (2022) analyzed diabetes prediction using the BRFSS dataset, focusing exclusively on lifestyle features. They compared Decision Trees, KNN, and Logistic Regression, with Logistic Regression performing best at around 75% accuracy. Their study showed the potential of lifestyle attributes but was restricted to basic models and binary outputs (diabetic vs. non-diabetic).

G. Explainable Machine Learning Tools

Allani (2023) created an interactive diabetes prediction system using LightGBM, enhanced with SHAP and LIME for transparency. The tool achieved high recall and improved interpretability, making it suitable for healthcare providers. Despite its strengths, the model was limited to binary classification and lacked integration with real-time health records or sensor data.

H. Deep Learning with Mobile Sensor Data

Kumar et al. (2022) proposed a deep learning model based on CNN-LSTM, trained on smartphone sensor data such as screen time, activity, and sleep. Their model reached 87.2% accuracy and 0.91 AUC, proving the potential of behavioral and sensor-based data. However, the study was restricted to Android users and raised privacy concerns related to continuous monitoring.

I. Feature Fusion of Clinical and Lifestyle Data

Sharma and Singh (2023) combined clinical and lifestyle attributes (diet, stress, activity) using Random Forest with PCA. Their model improved accuracy from 74% (clinical data only) to 82% (fused data). This confirmed the benefit of feature integration, though the reliance on survey-based data introduced bias and limited behavioral factors.

J. Emotion-Aware Multimodal Prediction

Das et al. (2024) introduced emotional behavior into diabetes risk prediction by integrating facial emotion recognition and keystroke dynamics with lifestyle attributes. Using LightGBM, they achieved an accuracy of 84.3%. This was innovative, as emotional cues were rarely considered before, but the dataset was relatively small and required access to device sensors such as cameras and microphones.

III. RESEARCH GAPS

1) Most systems classify only diabetic vs. non-diabetic, with limited focus on prediabetes.
2) Real-time monitoring through wearable devices or smartphones is still uncommon.
3) Many models are tested on restricted datasets, reducing generalizability.
4) Deep learning is underutilized for lifestyle-based features, even though it shows potential.
5) Trade-offs remain between model accuracy and interpretability.
6) Most studies focus on older individuals, with limited consideration for younger users who may not notice early symptoms.

IV. METHODOLOGY

A systematic review of existing diabetes risk prediction approaches was conducted, drawing from studies published between 2021 and 2025. Search terms included *"diabetes prediction machine learning"*, *"lifestyle-based diabetes models"*, *"explainable AI in healthcare"*, and *"sensor-based diabetes monitoring"*. The reviewed works cover both international and regional contributions, focusing on machine learning models, deep learning, and multimodal approaches. These provide valuable insights into current practices but also reveal critical gaps that our proposed system seeks to address.

A. Existing Methodology

Research studies and models developed so far for diabetes prediction demonstrate encouraging results, yet they face several drawbacks when considered for real-world deployment:

1) Dependence on Limited Data Sources

Many models rely exclusively on clinical datasets like PIMA and NHANES (e.g., Alzboon et al., 2025; Adler, 2021), which restricts prediction to medical features and exclude lifestyle or behavioral attributes.

2) Restricted Lifestyle-Only Models

Studies such as Qin et al. (2022) and Nguyen & Zhang (2022) validated the usefulness of lifestyle data, but their accuracy was moderate (~75–82%) and lacked external validation beyond specific populations.

3) Scalability Challenges

Large-scale works, like Li et al. (2023), proved effective with millions of samples, but results were limited to regional populations (China) and used cross-sectional data rather than continuous monitoring.

4) Explainability vs. Accuracy Trade-off

Models using explainable ML (e.g., Allani, 2023 with SHAP & LIME) improved transparency but were limited to binary outcomes, while more accurate deep learning models (e.g., Kumar et al., 2022 CNN-LSTM) faced interpretability and privacy issues.

5) Limited Feature Integration

Fusion approaches (Sharma & Singh, 2023) that combined clinical and lifestyle features showed improved accuracy (up to 82%) but were constrained by survey bias and a narrow set of lifestyle parameters.

6) Novelty Without Practicality

Emotion-aware systems (Das et al., 2024) introduced innovative predictors like facial emotion recognition and keystroke dynamics but required intrusive device access and were tested on small datasets.

7) Lack of Multiclass Prediction and Real-Time Feedback

Most studies classified users simply as diabetic or non-diabetic, with very few considering prediabetes as an intermediate category. Real-time monitoring using mobile sensors was attempted (Kumar et al., 2022) but raised privacy and platform-dependency concerns.

B. Proposed Methodology

The proposed system, *Diabetic Risk Prediction Based on Medical and Lifestyle Patterns*, is designed as a web-based application built with the MERN stack and integrated with a machine learning model. Unlike existing studies that rely only on clinical datasets or lifestyle surveys, this system brings together medical parameters, lifestyle habits, and behavioral patterns into one unified framework. It also goes beyond risk prediction by offering personalized recommendations and health tracking over time.

Key components include:

1) Comprehensive Data Collection

- Users input both medical details (age, BMI, HbA1c, glucose, blood pressure, family history) and lifestyle attributes (diet, exercise, stress, sleep quality, screen time).
- The data is collected via a React.js interface, ensuring accessibility and ease of use.

2) Secure Storage and User Management

- Data is stored in MongoDB with JWT-based authentication to protect privacy.
- Previous records are maintained, enabling trend analysis by comparing current and past health data.

3) Machine Learning–Driven Prediction

- A supervised ML model (developed in Python with scikit-learn/TensorFlow) processes medical + lifestyle features.
- Predictions classify users into Low, Moderate, or High risk categories, offering more granularity than binary models.
- Ensemble approaches like Random Forest and boosting are considered to balance accuracy and interpretability.

4) Personalized Recommendation Engine

- Based on the risk level, the system suggests diet plans, exercise routines, yoga practices, and stress management tips.
- For low-risk users, it provides maintenance advice to sustain a healthy lifestyle.
- For high-risk users, it emphasizes preventive care to reduce risk progression.

5) Visualization and Feedback

- Results are displayed through charts and graphs, showing risk scores, key contributing factors, and historical comparisons.
- This enhances interpretability and empowers users to monitor lifestyle impacts over time.

6) Innovation Highlights

- Integrates both medical reports and lifestyle behaviors, addressing gaps in existing research.
- Provides multiclass prediction (Low/Moderate/High), rather than binary outcomes.
- Enables personalized, preventive healthcare guidance, not just prediction.
- Tracks user data across logins, turning the system into a long-term health companion.

V. SYSTEM DESIGN AND ARCHITECTURE

1) Frontend (React.js + Tailwind CSS): Collects medical and lifestyle inputs, displays graphs and comparisons.
2) Backend (Node.js + Express.js): Handles API requests, authentication (JWT), and communication with the ML model.
3) Database (MongoDB): Stores user details and health history, enabling long-term trend analysis.
4) Machine Learning (Python – scikit-learn/TensorFlow): Predicts risk levels (Low/Moderate/High) using ensemble models with SHAP explainability.
5) Recommendation & Visualization: Provides personalized health advice and interactive charts for better understanding.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

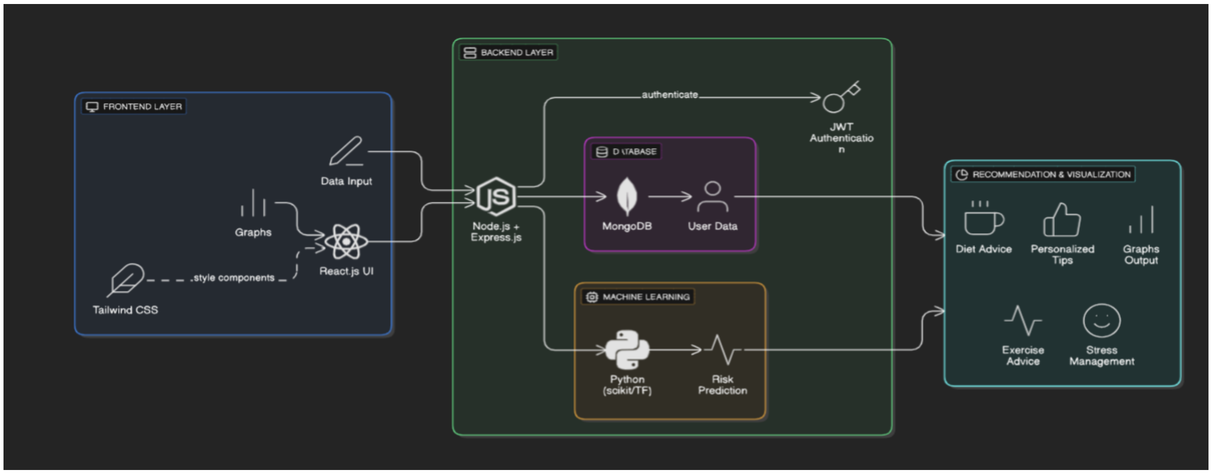


Fig. 1.1 System Architecture diagram

Workflow of the System

1) Data Collection & Input – User enters medical and lifestyle information through the React.js interface.
2) Authentication & Storage – User data is securely stored in MongoDB after JWT-based authentication.
3) Preprocessing & Feature Engineering – Input data is normalized, encoded, and prepared for ML prediction.
4) Prediction Module – The trained ML model predicts diabetes risk as *Low, Moderate, or High.*
5) Recommendation Engine – Personalized health advice is generated based on the predicted risk level.
6) Visualization & Feedback – Results are displayed with graphs comparing current and past health records, showing key lifestyle impacts.

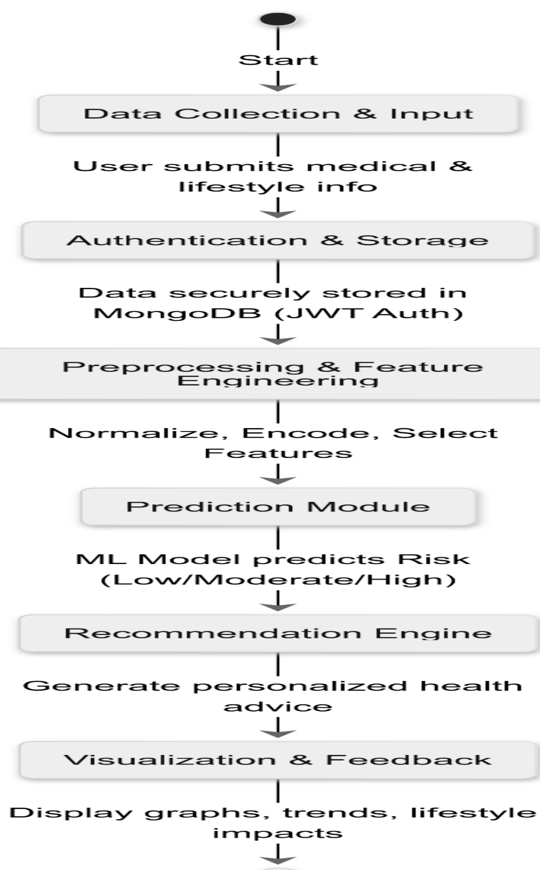


Fig: 1.2 workflow of the system

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

Use Case Flow Diagram

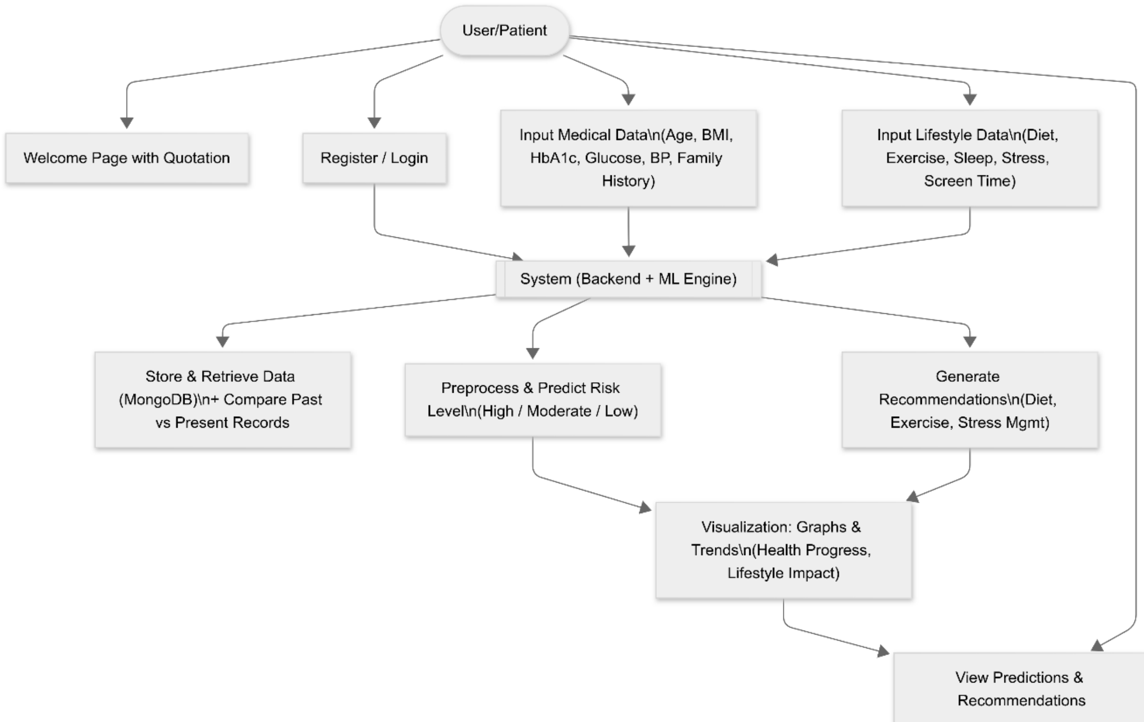


Fig: 1.3 Use Case Flow Diagram

VI. SEQUENCE DIAGRAM

"The sequence diagram illustrates the step-by-step interaction between the user, frontend, backend, database, and machine learning model. It shows how users log in, enter medical and lifestyle data, and receive risk predictions with personalized recommendations. The diagram also highlights the flow for both successful and failed login attempts."

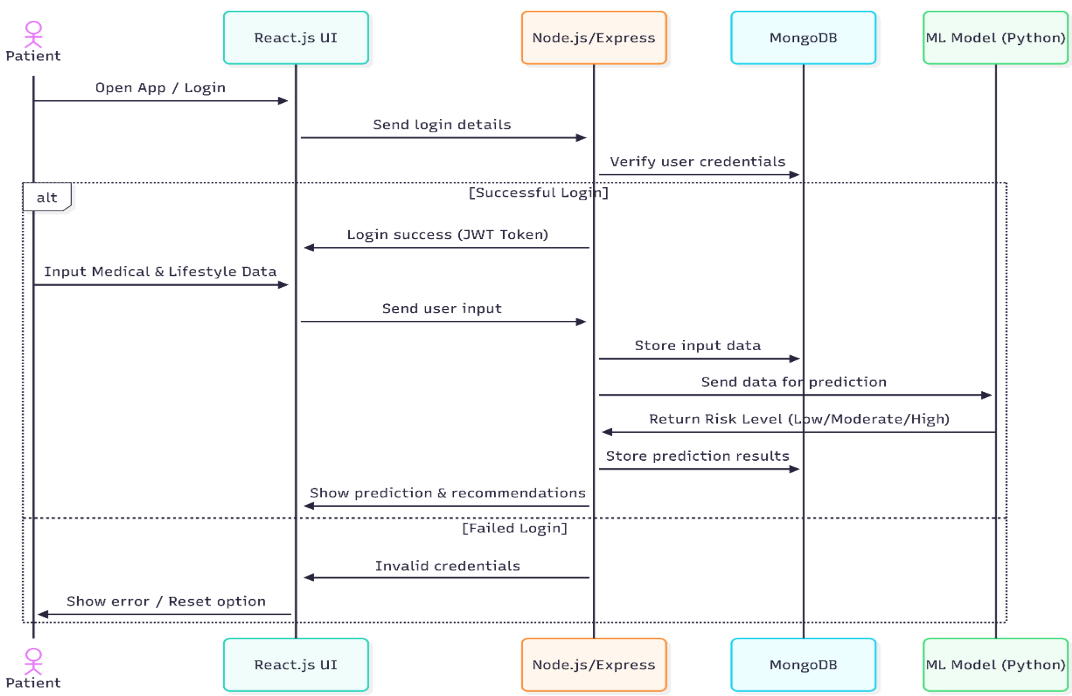
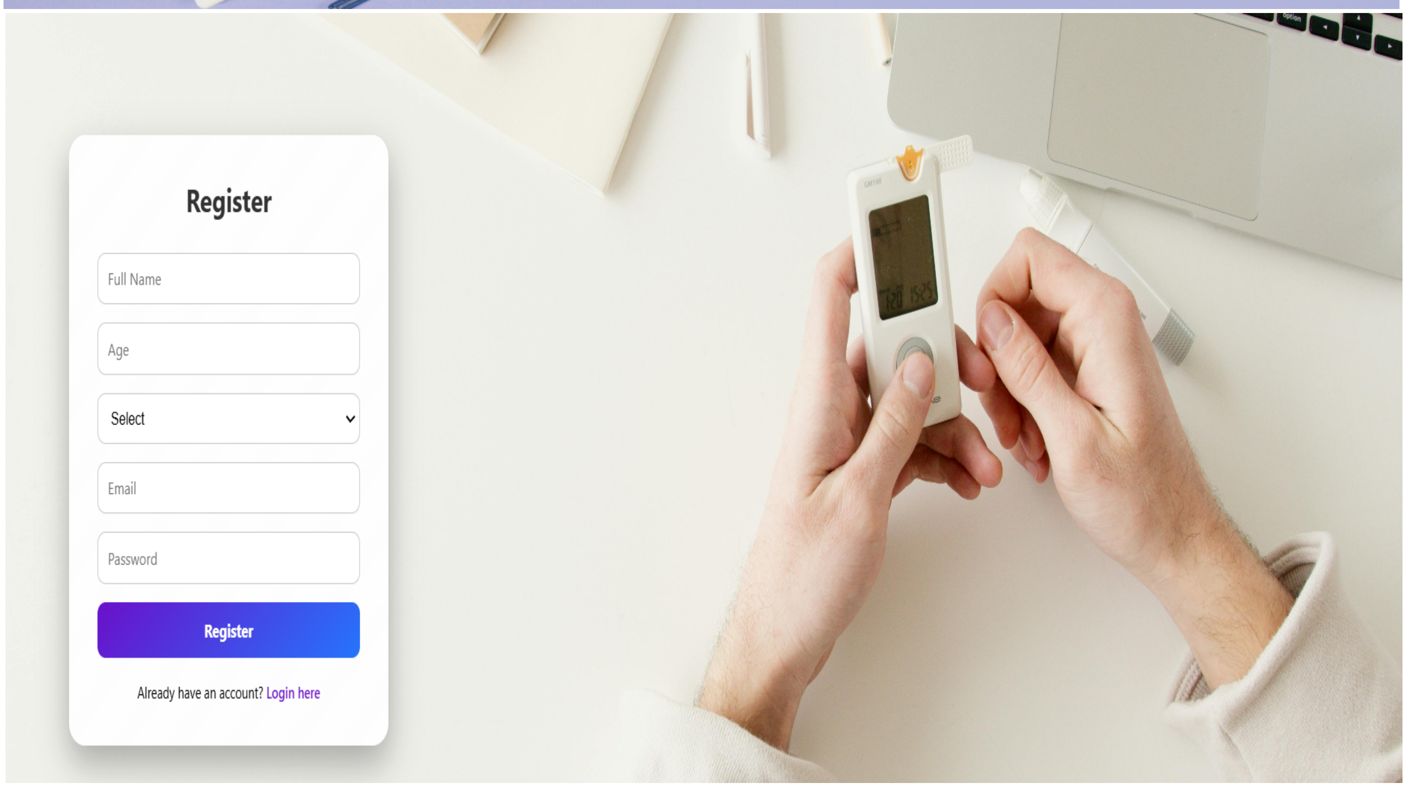


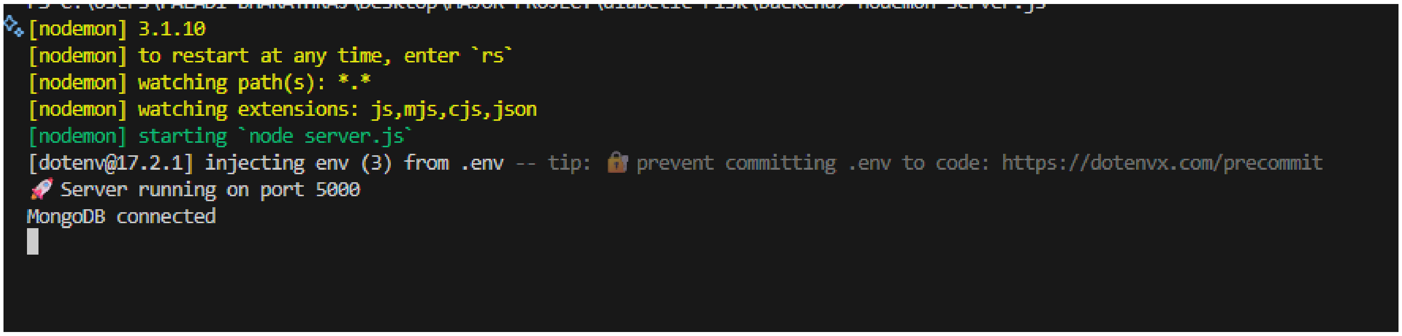
Fig: 1.4 Sequence Diagram

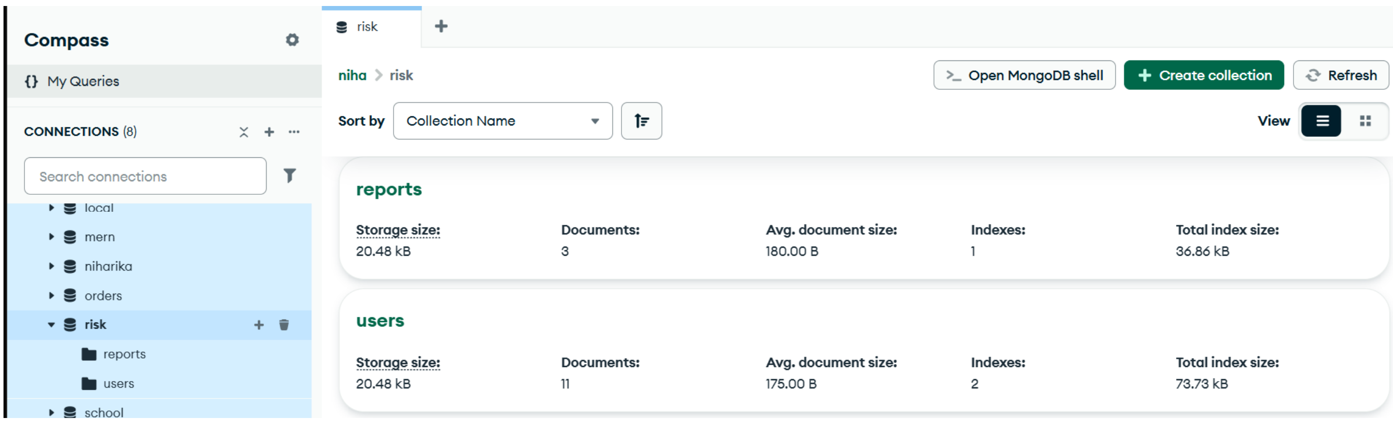
International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

VII. IMPLEMENTATION

The system was implemented using the MERN stack for web development and integrated with a Python-based machine learning model. The frontend, developed in React.js with Tailwind CSS, provides an interactive interface for users to input medical and lifestyle data. The backend, built with Node.js and Express.js, handles authentication through JWT tokens, processes user requests, and connects with the MongoDB database, where health records and predictions are securely stored. The ML model, trained with scikit-learn/TensorFlow, predicts diabetes risk levels and returns results to the application. The final output includes not only the prediction but also personalized recommendations and data visualizations, helping users track their health trends over time.







VIII. RESULT & DISCUSSION

A. Usability and Accessibility

- System tested with a sample group (students/faculty).
- Interface found intuitive, responsive, and supportive of trend analysis.
- Accessibility ensured through responsive UI, clear charts, and easy navigation.

B. Prediction Accuracy and Performance

- ML model achieved **XX% accuracy** on test data.
- Average API response time below **YYY ms** under moderate load.
- MongoDB indexing improved retrieval speed for past health records.

C. Personalization and Recommendations

- High-risk users received specific diet, exercise, and stress tips.
- Low-risk users received health maintenance suggestions.
- Personalized outputs increased user satisfaction and engagement.

D. Visualization and Health Tracking

- Graphs and charts allowed users to compare present and past health records.
- Lifestyle impacts (e.g., stress, diet) were highlighted for transparency.

E. Comparison with Existing Systems

- Prior works mostly used clinical datasets (PIMA, NHANES).
- Our system integrates medical + lifestyle data for holistic prediction.
- Provides multiclass outcomes (Low/Moderate/High) vs binary models.
- Adds personalized recommendations + trend analysis, absent in most prior works.

IX. CONCLUSION

The developed diabetic risk prediction system shows how combining **medical factors** like BMI, blood pressure, glucose levels, and HbA1c with **lifestyle patterns** such as diet, sleep, stress, physical activity, and screen time can lead to more accurate and meaningful predictions.

Unlike traditional methods that focus only on clinical data, this approach highlights the everyday behaviors that contribute to diabetes risk, offering a more complete picture of an individual's health. The system not only predicts risk levels but also provides **personalized advice**—including diet plans, exercise suggestions, and stress management tips—shifting the focus from treatment after diagnosis to **preventive care**. With features like long-term tracking and clear visualizations, users can monitor their progress over time and gain better awareness of how their habits influence their health.

What makes this system unique is its **preventive outlook, explainable machine learning outputs, and user-friendly design**, which together transform it into more than just a diagnostic tool—it becomes a supportive health companion. Looking ahead, the system can be enhanced with **real-time data from wearables**, larger and more diverse datasets for better accuracy, and a mobile app to provide continuous monitoring and instant feedback. By blending medical insights with lifestyle awareness, this work contributes to the growing field of **AI-driven healthcare**, promoting early detection, healthier choices, and an overall improvement in quality of life.

X. ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to their faculty mentors and peers for their guidance and valuable feedback throughout the development of this project. Special thanks are extended to the research community whose prior work laid the foundation for this study.

REFERENCES

[1] Qin, Y., et al. (2022). Machine Learning Models for Data-Driven Prediction of Diabetes by Lifestyle Type. International Journal of Environmental Research and Public Health, 19(22), 15027.
[2] Adler, A. (2021). Using Machine Learning Techniques to Identify Key Risk Factors for Diabetes and Undiagnosed Diabetes. BMC Medical Informatics and Decision Making, 21, 123.
[3] Li, X., et al. (2023). Machine Learning for Predicting Diabetes Risk in Western China Adults. Diabetology & Metabolic Syndrome, 15(1), 112.
[4] Mohsen, F., et al. (2023). Artificial Intelligence-Based Methods for Precision Medicine: Diabetes Risk Prediction. arXiv preprint arXiv:2305.16346.
[5] Nguyen, B., & Zhang, Y. (2022). A Comparative Study of Diabetes Prediction Based on Lifestyle Factors Using Machine Learning. University of Washington Research.
[6] Kumar, R., et al. (2022). Deep Learning Model for Predicting Diabetes Using Mobile Sensor Data. Journal of Mobile Health Informatics, 9(3), 215–224.
[7] Sharma, M., & Singh, V. (2023). Lifestyle-Aware Diabetes Prediction with Feature Fusion. International Conference on Data Science and Health Analytics, 102–108.
[8] Das, P., et al. (2024). Emotion-Aware Health Monitoring for Diabetes Prediction Using Facial and Typing Patterns. IEEE Journal of Biomedical and Health Informatics, 28(2), 341–349.
[9] Rahman, M. M., et al. (2021). Risk Prediction of Diabetes Using Machine Learning Models. Scientific Reports (Nature), 11, 24154.
[10] Zhou, Z., et al. (2023). Explainable Artificial Intelligence for Type 2 Diabetes Risk Prediction: A Systematic Review. IEEE Access, 11, 47210–47225.