



# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



**Impact Factor: 8.699**

**Volume 14, Issue 7, July 2025**

# **Diabetes Prediction System**

**Ms. K.P. Preetha Rhythami<sup>1</sup>, Mrs. P. Shenbagam<sup>2</sup>**

PG Scholar, Department of Master of Computer Applications, RVS College of Engineering, Dindigul,

Tamil Nadu, India<sup>1</sup>

Head & Assistant Professor, Department of Master of Computer Applications, RVS College of Engineering, Dindigul,

Tamil Nadu, India<sup>2</sup>

**ABSTRACT:** The **Diabetes Prediction System** is a machine learning–based web application developed to assess an individual’s risk of diabetes using health-related inputs. Built using **Python, Streamlit, and Scikit-learn**, the system uses 18 medical features such as glucose level, BMI, age, blood pressure, and insulin levels to predict whether a user is at risk of developing diabetes. A **random forest classifier** was trained on a well-preprocessed dataset, and the model was integrated into a **Streamlit interface** using joblib for real-time predictions. The platform includes a secure **user authentication system using SQLite**, interactive forms for input collection, and features such as **downloadable PDF reports** and **SHAP-based visual explanations**. It also provides users with additional insights via confusion matrices, class distribution, and feature importance analysis. The interface is simple, responsive, and easy to navigate, making it accessible for general users without technical backgrounds.

While the system does not replace clinical diagnosis, it serves as an early **decision-support tool** to promote awareness and encourage timely medical consultation. Future enhancements may include integration with real-time medical devices, cloud-based storage for user history, and deployment as a mobile app for wider accessibility. This project demonstrates the practical application of AI in healthcare and lays the groundwork for more robust, data-driven preventive health tools. This project not only highlights the predictive capabilities of machine learning in healthcare but also emphasizes the importance of accessibility and user empowerment. By offering an intuitive interface and real-time results, the system bridges the gap between complex medical analytics and everyday users. Its integration of explainable AI techniques ensures transparency in predictions, fostering trust and understanding among users, especially in a domain as sensitive as health diagnostics.

## **I. INTRODUCTION**

Diabetes is one of the most common and rapidly growing chronic diseases in the world today. Characterized by high blood glucose levels due to the body’s inability to produce or effectively use insulin, diabetes affects millions globally and is a leading cause of various health complications, including heart disease, kidney failure, and vision loss. In many cases, diabetes remains undiagnosed until severe symptoms arise, making early detection crucial for effective management and prevention.

With the rise of technology and healthcare analytics, machine learning (ML) has emerged as a powerful tool in predicting the onset of diseases like diabetes. ML can analyze large volumes of medical data, discover patterns, and make predictions more efficiently than traditional statistical methods. Integrating such ML capabilities into a user-friendly platform provides an opportunity to assist both patients and healthcare professionals in risk assessment, especially in resource-limited settings.

This project aims to develop a web-based application that can predict diabetes risk using machine learning algorithms. By leveraging Streamlit, a lightweight Python-based framework, and incorporating real-world health data, this system offers a practical solution that combines data science with medical awareness.

## **II. EXISTING SYSTEM**

In traditional healthcare practice, identifying diabetes risk often depends on clinical assessments such as fasting blood sugar tests, HbA1c levels, and physician interpretation. While these methods are medically sound, they are not without drawbacks:

- **Delayed Risk Detection:** Many individuals remain undiagnosed until symptoms are severe or irreversible damage has occurred.
- **Limited Accessibility:** Rural and remote populations may lack timely access to diagnostic centers and regular medical check-ups.
- **Manual Record-Keeping:** Traditional systems rely on paper records or basic digital entries, which lack automation and analytical support.
- **No Predictive Assistance:** Most current systems do not assist users with early predictions or risk forecasts unless a clinical consultation is done.
- **Low Awareness Tools:** Public awareness platforms are either generalized or do not personalize risk levels based on individual health metrics.

These challenges indicate the need for an intelligent, accessible, and user-friendly tool that provides early-stage insights into potential diabetes risk, helping individuals make informed health decisions without full reliance on clinical intervention.

### III. PROPOSED SYSTEM

To bridge the gap in early detection and empower users with accessible self-assessment, the proposed system leverages machine learning techniques within a web-based platform. It is designed to accept user health data through a form interface, analyze it using a trained model, and provide immediate prediction results along with interpretability insights.

#### Key Characteristics of the Proposed System:

- **Streamlit-Based User Interface:** A simple, interactive, browser-based form that collects health-related features.
- **ML-Driven Prediction:** A pre-trained Random Forest model processes the inputs and predicts the likelihood of diabetes.
- **Secure User Management:** A login and registration system backed by SQLite for personal usage and data isolation.
- **Visual Feedback:** Graphs such as feature importance and confusion matrix are shown to help users understand system accuracy.
- **Result Download:** The user can export their prediction as a structured PDF for offline reference.
- **Explainable AI (XAI):** SHAP (Shapley Additive Explanations) is used to show how each input affects the outcome, improving user confidence.

### IV. SYSTEM ARCHITECTURE

The diabetes prediction system employs a **web-based architecture** built using **Streamlit**, a powerful Python framework for creating interactive web applications. The architecture seamlessly integrates both the frontend and backend into a single platform, making it efficient and user-friendly. This system handles real-time diabetes prediction with interactive visual analytics, allowing users to enter their medical and lifestyle information, receive predictions, and explore data visualizations in an intuitive interface.

- **Web-based Application:** Users access the system through a browser, which makes it highly accessible across various devices without the need for complex installation procedures.
- **Frontend (Client-Side):**
  - **User Interaction:** The users interact with the system via a **Streamlit UI**, which presents a clear and responsive layout. Users input their health-related data through **Streamlit forms**, which is then processed by the backend.
  - **Visualization and Output:** Once the prediction is made, the results are displayed dynamically using **interactive widgets** in the browser. These visualizations are integral to understanding the model's predictions and the factors that influenced them.

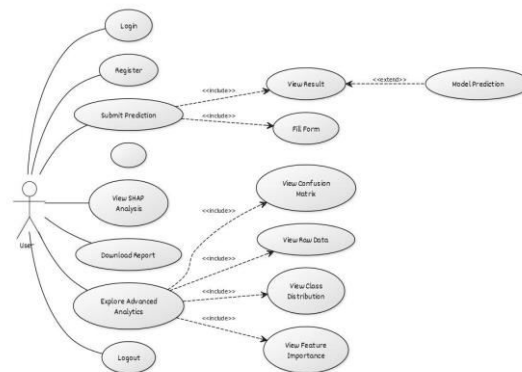
#### Backend (Server-Side):

- **Machine Learning Model:** The system loads a pre-trained **Random Forest model** (saved as a Joblib file) for prediction. The model is reused throughout the user session, optimizing both speed and resource consumption.
- **Data Management:** **SQLite** is used for user authentication and session management, securely storing credentials and allowing for login and registration functionalities.



○ **Prediction Pipeline:**

- **Input Collection:** Users provide 18 health and lifestyle features, which are then processed to ensure they are formatted correctly for model prediction.
- **Data Preprocessing:** The input data is cleaned and preprocessed to ensure it meets the model's requirements.
- **Prediction:** After processing, the **Random Forest model** predicts whether a user is at risk of diabetes (positive or negative).
- **Visualization:** Tools like **SHAP** (Shapley Additive exPlanations) are used to visually interpret the prediction, showing the influence of each feature on the outcome. Additional visual analytics include **feature importance** and **class distribution**.
- **PDF Report Generation:** After making a prediction, users can download a personalized **PDF report**, which includes the input data, prediction results, SHAP graph, and relevant timestamps.



## V. RESULTS

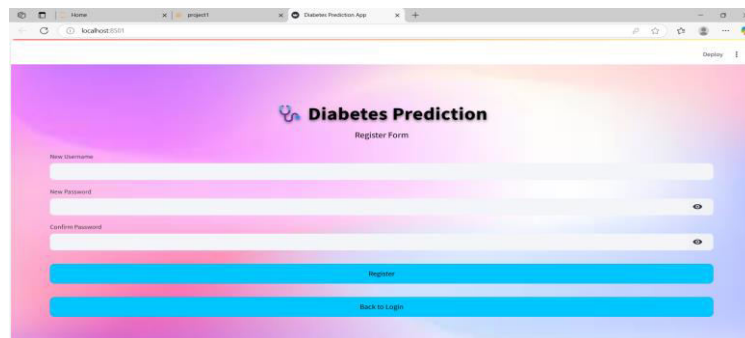


Fig 5.1 Login Screen

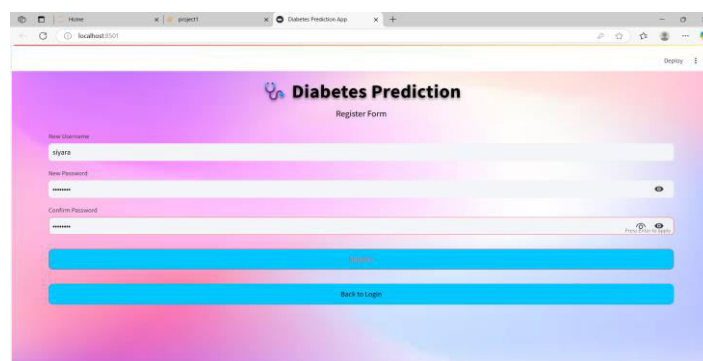
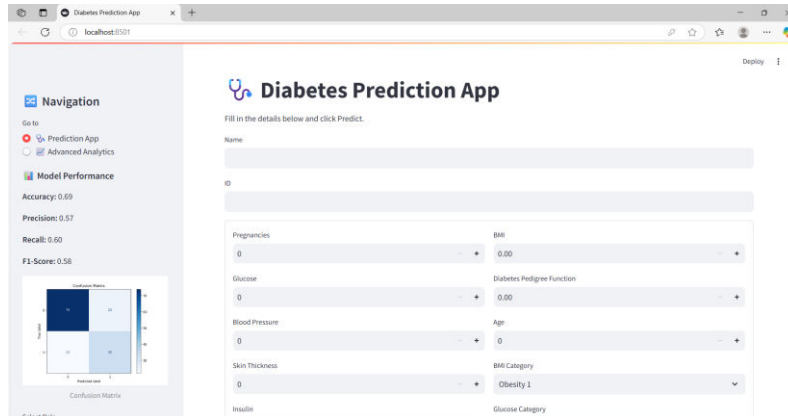
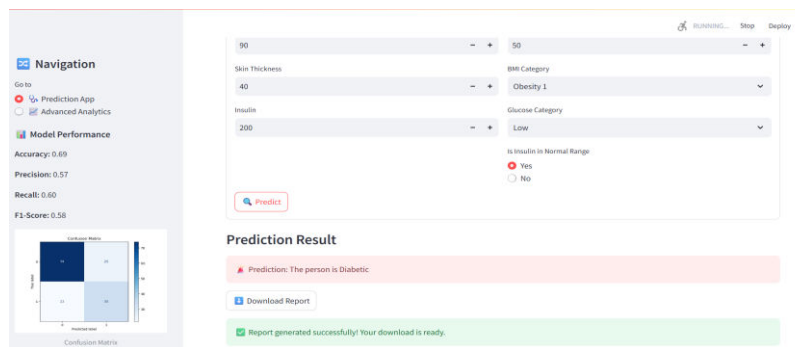


Fig 5.2 Register



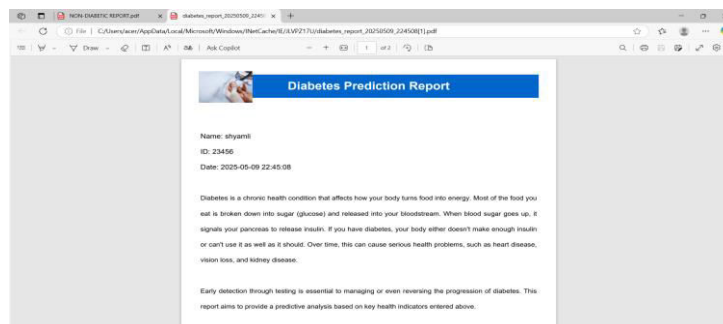
The screenshot shows the 'Diabetes Prediction App' interface. On the left, there is a navigation sidebar with options for 'Prediction App' and 'Advanced Analytics'. Below this, model performance metrics are listed: Accuracy: 0.69, Precision: 0.57, Recall: 0.60, and F1-Score: 0.58. A confusion matrix is also displayed. The main area is the 'Prediction Form' where users can input various health indicators. Fields include Name, ID, Pregnancies, BMI, Glucose, Diabetes Pedigree Function, Blood Pressure, Age, Skin Thickness, BMI Category, Insulin, and Glucose Category. A 'Predict' button is located at the bottom of the form.

**Fig 5.3 Prediction Form**



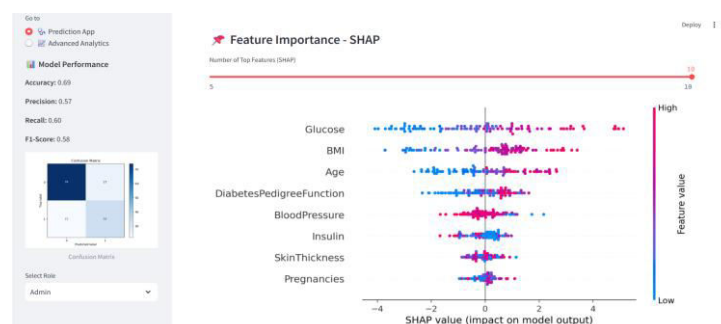
This screenshot shows the 'Prediction Result' section of the app. The 'Predict' button has been clicked, and the result is displayed: 'Prediction: The person is Diabetic'. Below this, there is a 'Download Report' button. A green message at the bottom states: 'Report generated successfully! Your download is ready.' The sidebar on the left remains the same as in Fig 5.3.

**Fig 5.4 Prediction Result**



The screenshot shows a PDF document titled 'Diabetes Prediction Report'. It includes patient information: Name: rshyam, ID: 23456, and Date: 2025-05-09 22:45:08. The report contains a detailed explanation of diabetes as a chronic health condition and the importance of early detection. It also mentions that the report aims to provide a predictive analysis based on the key health indicators entered.

**Fig 5.5 Report Generation**



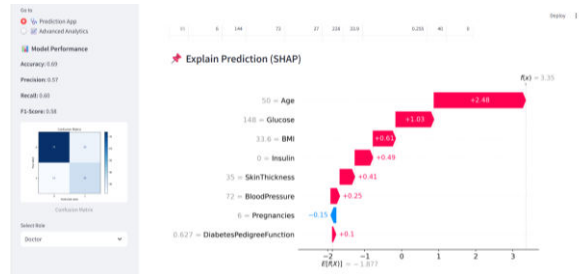


Fig 5.6 SHAP graph

## VI. CONCLUSION

The development of the **Diabetes Prediction System using Machine Learning and Streamlit** marks a significant step towards providing an accessible and informative health risk assessment tool. This project addresses the growing need for predictive healthcare solutions by leveraging machine learning algorithms to assess an individual's risk of diabetes based on personal health metrics.

Throughout the project, considerable effort has been put into selecting appropriate methodologies, integrating efficient algorithms, and designing an interactive user interface to ensure that users without technical expertise can easily navigate the system and understand their results.

## VII. FUTURE ENHANCEMENTS

### 1.Mobile App Integration:

- Develop a dedicated mobile version using frameworks like **React Native** or **Streamlit for Mobile** to enable prediction on the go.
- Integrate with health tracking apps for automatic data collection.

### 2.Advanced Predictive Models:

- Incorporate additional machine learning algorithms (like **XGBoost** or **SVM**) and dynamically choose the best model based on input data.
- Implement **severity prediction** instead of just binary classification.

### 3.Improved Model Update Mechanism:

- Introduce **real-time model retraining** with newly acquired data to maintain accuracy.
- Automate the process using **MLOps practices** for continuous integration and deployment.

### 4.Enhanced User Experience:

- Implement **multi-language support** for wider accessibility.
- Include **personalized health insights** based on user history and trend analysis.

### 5.Stronger Data Privacy and Security:

- Upgrade to a more secure database system like **PostgreSQL**.
- Implement **OAuth 2.0** for better user authentication and encrypted data storage.

### 6.Deployment on Cloud Platforms:

- Host the application on **AWS, Azure, or Google Cloud** to ensure scalability and high availability.
- Add offline functionality for areas with limited internet access.

### 7.Integration with Healthcare Systems:

- Connect with **Electronic Health Records (EHR)** to automatically fetch patient data.

#### REFERENCES

1. Berry, Jason, Jonathan Foote, and Tad Jones. Up from the Cradle of Jazz: New Orleans Music Since World War II. Athens: U of Georgia P, 1986.
2. Booth, Wayne C. "Kenneth Burke's Way of Knowing." *Critical Inquiry* 1 (1974): 1-22.
3. Breiman, Leo. "Random Forests." *Machine Learning* 45.1 (2001): 5-32.
4. FPDF: Free PDF Generator for Python. Retrieved from <https://pyfpdf.readthedocs.io>
5. Kaggle. Pima Indians Diabetes Database. Retrieved from <https://www.kaggle.com/datasets/mathchi/diabetes-dataset>
6. Matplotlib Developers. Matplotlib Documentation. Retrieved from <https://matplotlib.org/stable/index.html>
7. Mozilla Developer Network (MDN) Web Docs. HTML: HyperText Markup Language. Retrieved from <https://developer.mozilla.org/en-US/docs/Web/HTML>
8. Mozilla Developer Network (MDN) Web Docs. CSS: Cascading Style Sheets. Retrieved from <https://developer.mozilla.org/en-US/docs/Web/CSS>
9. Mozilla Developer Network (MDN) Web Docs. JavaScript. Retrieved from <https://developer.mozilla.org/en-US/docs/Web/JavaScript>
10. Pedregosa, Fabian, et al. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* 12 (2011): 2825-2830.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details