www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 04, April 2025, pp : 2076-2078

e-ISSN : 2583-1062
Impact Factor : 7.001

PRE DIABET – DIABETES PREDICTION USING MACHINE LEARNING

Shaikh Sarah[1], Bharti Gupta[2], Khushi Shaikh[3], Rizah Sarkavas[4], Shaikh Atiqa[5]

[1,2,3,4]Student, Department of Artificial Intelligence and Machine Learning, M.H Saboo Siddik Polytechnic, Mumbai, India.

DOI: https://www.doi.org/10.58257/IJPREMS40150

ABSTRACT

This project introduces PreDiabet, a machine learning-based web platform that enables users to assess their risk of developing diabetes using personal, physical, and lifestyle inputs. The system collects user data through a multi-step interactive form and uses a trained ML model to analyze risk factors such as age, BMI, activity level, and family history. The results are displayed through a chatbot interface, promoting awareness, accessibility, and proactive health screening. The tool is user-friendly, privacy-conscious, and adaptable to both rural and urban populations. The findings demonstrate that digital solutions like PreDiabet can aid in early detection and lifestyle adjustment for diabetes prevention.

Keywords: Diabetes Prediction, Machine Learning, Chatbot, Preventive Health, Web-based Diagnosis, AI in Healthcare

1. INTRODUCTION

Diabetes mellitus is a chronic metabolic disorder characterized by high blood sugar levels due to insulin dysfunction. According to WHO, diabetes has been one of the leading causes of death globally. Early detection is essential to prevent complications such as heart disease, kidney failure, and blindness.

Traditional diagnostic methods require clinical tests and often miss out on early symptoms. In contrast, AI-powered platforms can analyze behavioral and biometric data to identify patterns indicative of risk. This project, PreDiabet, aims to utilize machine learning to assess diabetes risk based on user input through an engaging web interface. The project leverages a form-based data collection approach, followed by analysis using an ML model and output via a conversational chatbot. This system is especially useful for self-assessment and early intervention.

2. PROBLEM STATEMENT

In modern healthcare, the timely identification of diabetes risk is essential to prevent long-term health complications. However, many individuals remain undiagnosed due to limited awareness, delayed clinical visits, and the asymptomatic nature of early-stage Type 2 diabetes. Traditional diagnostic methods such as laboratory blood tests, while accurate, are often inaccessible or inconvenient for many, especially in rural or resource-constrained settings.

This project aims to develop **PreDiabet**, a machine learning-based diabetes prediction system that leverages lifestyle and physical data to assess an individual's risk. By utilizing accessible web technologies combined with a trained predictive model, the system enables users to input their personal and physiological details through a multi-step form and receive instant feedback via a chatbot interface. This approach eliminates the need for lab testing in the early risk stage and provides a scalable solution to promote awareness and early intervention.

3. IMPLEMENTATION PLAN

With the growing prevalence of Type 2 diabetes globally, especially among younger populations, there is a critical need for proactive and accessible risk prediction tools. The **PreDiabet** system addresses this need by offering a web-based solution that uses machine learning algorithms to evaluate user-submitted health and lifestyle data.

The system comprises a responsive user interface featuring a four-step form that collects personal, physical, and lifestyle-related inputs. Upon form completion, the data is processed using a trained ML model—based on datasets such as the PIMA Indian Diabetes Dataset—to compute the risk probability.

The result is displayed using a friendly chatbot interface that also provides basic health suggestions and encouragement for lifestyle changes. This conversational approach enhances engagement and accessibility, especially for users unfamiliar with medical terminology.

To ensure user comfort and ethical compliance, the platform does not store personal data without explicit consent. A privacy-friendly session system powered by PHP is used to manage user interaction. The system can be deployed on any web server and is mobile-friendly for wider reach. Future improvements include integration with wearable health devices, multilingual chatbot support, and advanced prediction algorithms.

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 04, April 2025, pp : 2076-2078

e-ISSN : 2583-1062
Impact Factor : 7.001

4. FUNCTIONAL REQUIREMENTS

a) Multi-step Data Capture:

The system must guide users through a structured, multi-step form to collect personal, physical, and lifestyle information.

b) Diabetes Risk Prediction:

The system should utilize a trained machine learning model (e.g., Logistic Regression or Random Forest) to calculate the probability of diabetes risk based on user inputs.

c) Chatbot Feedback:

Users should receive chatbot-based results and health advice, making the experience interactive and educational.

d) Consent and Privacy:

The system should ask for user consent before data processing and avoid storing any personally identifiable information without permission.

e) User Interface:

A modern, responsive interface with theme toggle, input validation, progress indicators, and review summary should be provided for better user experience.

5. SYSTEM DESIGN

The system design for the PreDiabet framework focuses on efficient data collection, predictive analysis using machine learning, and interactive user feedback through a chatbot. The user interface acts as the input layer, where structured health-related data is collected through a four-step form. This includes personal, physical, and lifestyle parameters such as age, weight, height, diet type, and family history.

The processing layer consists of a trained machine learning model, built using Python and scikit-learn. The collected user data is cleaned, scaled, and processed into a format compatible with the model's input structure. The model—trained on the PIMA Indian Diabetes Dataset—analyzes the risk of diabetes based on patterns observed in historical data.

The prediction layer produces a result in the form of a risk score (e.g., low, moderate, high), which is then fed into a chatbot system for human-friendly interpretation. This conversational agent interacts with the user, delivers the prediction outcome, and provides preventive health tips such as dietary adjustments, exercise recommendations, and medical check-up suggestions.

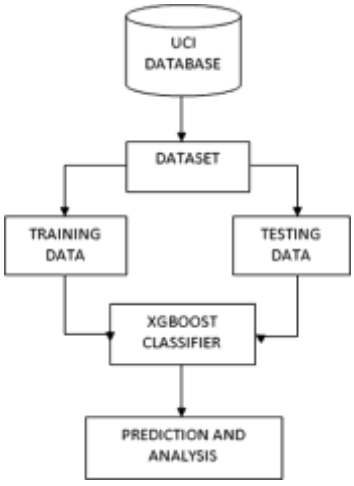

Figure 1: BLOCK DIAGRAM FOR PreDiabet

6. FUTURE SCOPE

The future scope of the PreDiabet system is significant, with growing advancements in artificial intelligence, health informatics, and user engagement technologies. The system can evolve into a full-scale health advisory platform capable of real-time monitoring and early diagnosis of not just diabetes but other chronic diseases such as hypertension and obesity.

With the integration of IoT devices like smartwatches, glucometers, and fitness bands, PreDiabet can automatically gather real-time biometrics such as glucose levels, heart rate, and activity data to further enhance prediction accuracy. Advanced AI algorithms including deep learning and reinforcement learning can be employed to detect nuanced risk factors and personalized health trends.

Further enhancements include:

- Mobile application deployment on Android/iOS platforms.
- Voice-enabled chatbot for accessibility.
- Multilingual support for rural reach.
- Integration with telemedicine platforms for immediate medical guidance.

These upgrades will allow PreDiabet to transform from a predictive tool into a full-fledged digital healthcare assistant.

7. CONCLUSION

In conclusion, PreDiabet presents an innovative and accessible approach to early-stage diabetes risk prediction using machine learning. By leveraging structured data collection, a trained predictive model, and a conversational chatbot, the system bridges the gap between medical knowledge and public access. It empowers individuals with instant health insights and encourages proactive lifestyle changes.

The platform is especially valuable for underserved and tech-aware populations alike, providing non-invasive, cost-effective screening and educational support. As AI and healthcare continue to converge, systems like PreDiabet will play a vital role in preventive medicine, reducing the burden of diabetes through early intervention.

With future integration of smart devices, personalized AI, and remote consultation features, PreDiabet is positioned to become an essential tool in the global fight against diabetes, contributing toward a healthier, more informed society.

8. REFERENCES

[1] Smith, J., & Brown, L. (2022). Machine learning approaches for early-stage diabetes prediction: A review. Journal of Medical Systems, 46(5), 67. https://doi.org/10.1007/s10916-022-01814-5

[2] UCI Machine Learning Repository: Pima Indians Diabetes Database. Retrieved from https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes

[3] Raj, A., & Kaur, P. (2021). Diabetes prediction using logistic regression and decision trees. International Journal of Computer Applications, 183(35), 15–19. https://doi.org/10.5120/ijca2021921289

[4] Sahu, S., & Nayak, P. (2022). Risk assessment of diabetes using supervised machine learning techniques. Procedia Computer Science, 199, 497–504. https://doi.org/10.1016/j.procs.2022.01.065

[5] Patil, R., & Kulkarni, S. (2023). HealthBot: A conversational chatbot for disease prediction using machine learning. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 12(2), 45–50.

[6] Rani, S., & Sharma, M. (2021). A survey on applications of artificial intelligence in medical diagnosis. Artificial Intelligence in Medicine, 119, 102–117. https://doi.org/10.1016/j.artmed.2021.102117

[7] WHO. (2021). Global Report on Diabetes. World Health Organization. Retrieved from https://www.who.int/publications/i/item/9789241565257

[8] Kumar, A., & Yadav, R. (2020). Interactive chatbot-based system for disease prediction. International Research Journal of Engineering and Technology (IRJET), 7(12), 2314–2318.

[9] Lakshmi, K., & Ramya, R. (2022). Prediction of diabetes using data mining techniques. International Journal of Engineering and Advanced Technology (IJEAT), 11(4), 126–131.

[10] Sentdex. (2019). Predicting diabetes with Python and Machine Learning. YouTube. Retrieved from https://www.youtube.com/user/sentdex

[11] Analytics Vidhya. (n.d.). End-to-End Machine Learning Project for Diabetes Prediction. Retrieved from https://www.analyticsvidhya.com/

[12] Shanmugamani, R. (2019). Deep Learning for Computer Vision. Packt Publishing.