Summary of the research paper as an outcome of PBL-II project

ON

"DiabBot: An AI-Powered Chatbot for Diabetes Diagnosis, Prevention and Management"

Submitted in partial fulfillment of the requirements for the degree of

BACHELOR OF TECHNOLOGY in COMPUTER SCIENCE & ENGINEERING

Submitted By

Soumin Mujumdar - 22070122190
Vidhi Binwal - 22070122249

UNDER THE GUIDANCE OF

Dr Dipti Theng

TABLE OF CONTENTS

Problem statement of the project

Diabetes mellitus (DM) is a chronic and progressively worsening metabolic disorder that affects millions of individuals worldwide, leading to serious health complications such as cardiovascular disease, kidney failure, neuropathy, and diabetic retinopathy if not detected and managed early. Traditional approaches to diagnosing and managing diabetes heavily rely on manual testing, clinical observations, and periodic check-ups by healthcare professionals, which can be time-consuming, prone to human error, and inaccessible to people in low-resource or rural settings. Moreover, with the global rise in diabetes prevalence, healthcare systems are increasingly strained, creating an urgent need for scalable, efficient, and accurate solutions to support early detection, continuous monitoring, and personalized disease management.

Recent advancements in machine learning (ML) and artificial intelligence (AI) offer promising tools to address these gaps by enabling automated data analysis, predictive modeling, and intelligent assistance for patients and healthcare providers alike. However, despite the wealth of research and technological progress, several persistent challenges remain, such as limited access to large and diverse datasets, variability in model performance across populations, and the lack of seamless integration between predictive algorithms and real-world clinical or self-care applications. There is also a pressing need to move beyond isolated algorithm development and toward holistic systems that incorporate cloud computing, wearable sensors, and interactive AI tools like chatbots, which can deliver real-time, personalized feedback and recommendations. Addressing these challenges is crucial to ensure that ML- and AI-driven diabetes solutions are not only scientifically sound but also practical, scalable, and accessible to all segments of the population, especially those most vulnerable.

Abstract of the Paper

Diabetes Mellitus (DM) affects over 537 million adults globally, representing one of the most pressing chronic health conditions of our time. Traditional approaches to diabetes diagnosis, prevention, and management face several limitations, including delayed detection, high medical costs, manual inefficiencies, lack of real-time monitoring, and low patient adherence to treatment plans. Addressing these challenges, we present DiabBot — an AI-powered chatbot designed to provide a personalized, accessible, and cost-effective solution for diabetes care. DiabBot integrates machine learning (ML), natural language processing (NLP), and Internet of Things (IoT) technologies to deliver real-time diabetes risk assessments, personalized health recommendations, continuous health metric monitoring, and streamlined communication with healthcare providers.

This paper outlines the complete development pipeline of DiabBot, detailing its design, functional requirements, backend and frontend integration, and the ML models (Random Forest, Logistic Regression, Neural Networks) that power its predictive capabilities. The system draws on the Pima Indian Diabetes Dataset for training and incorporates wearable device APIs for real-time health data collection. Through RASA NLU, DiabBot supports natural language interactions, allowing users to engage conversationally for risk assessments, lifestyle advice, and tailored exercise, diet, or medication recommendations. Importantly, the chatbot generates structured reports for healthcare providers, enhancing the coordination between patients and doctors.

The paper also highlights key performance metrics such as accuracy, precision, recall, and user satisfaction, reporting promising results during testing, including 92% risk assessment accuracy and 85% user satisfaction. Challenges like dataset limitations, hyperparameter tuning, data privacy, and cultural/language barriers are acknowledged,

with proposed future improvements such as integrating blockchain for data security, expanding multilingual NLP capabilities, and enhancing system scalability. Overall, DiabBot represents a significant step forward in AI-driven, patient-centered diabetes care, with the potential to extend its framework to other chronic diseases and transform the healthcare delivery landscape.

Summary of the Literature Review

Our literature review offers a thorough and comprehensive examination of over 25 scholarly articles, research papers, and technical reports that focus on the application of machine learning (ML) and artificial intelligence (AI) in the detection, diagnosis, and self-management of diabetes mellitus (DM). The sources reviewed come from a wide range of credible and authoritative platforms, including peer-reviewed medical journals, IEEE conference papers, healthcare informatics publications, and international research databases, ensuring a diverse and well-rounded evidence base. A significant portion of the review is dedicated to the datasets used across studies, with particular emphasis on the Pima Indian Diabetes Dataset (PIDD), which contains clinical variables such as glucose levels, body mass index (BMI), blood pressure, pregnancies, and insulin levels. Additionally, image-based datasets like Messidor, used to analyze retinal images for diabetic retinopathy detection, are discussed. The review also elaborates on pre-processing techniques employed by researchers, including normalization, outlier removal, adaptive thresholding, and feature extraction — all critical steps to improve the quality and reliability of the data fed into ML models.

The review then explores the different types of features used for analysis, highlighting both straightforward clinical features like patient demographics and medical histories, as

well as more advanced image-based attributes such as shape, texture, and color. These features are shown to significantly influence the performance of ML models. A wide array of machine learning algorithms, including K-Nearest Neighbor (KNN), Support Vector Machines (SVM), Convolutional Neural Networks (CNN), and deep learning models, are examined, with comparisons made between individual algorithm performances and the use of ensemble learning approaches that combine multiple algorithms for improved predictive accuracy. Importantly, the review also highlights the rising use of AI-powered intelligent assistants — such as chatbots, virtual health advisors, and voice-enabled systems — that provide real-time, personalized recommendations to patients, helping them manage their condition more effectively while simultaneously reducing the workload on healthcare professionals and systems.

Performance metrics used to evaluate these AI and ML models, such as accuracy, sensitivity, specificity, and area under the curve (AUC), are also addressed, emphasizing their importance in clinical settings where diagnostic precision is critical. Despite the significant advancements reported, the review identifies ongoing challenges, including the limited availability of large, high-quality datasets, the labor-intensive nature of manual hyperparameter tuning in deep learning models, and concerns about the generalizability of models across diverse patient populations. Looking forward, the review advocates for future research to focus on integrating IoT-connected wearable devices, cloud computing platforms, and advanced AI systems to create scalable, real-time, and accessible solutions for diabetes management — particularly aimed at improving care delivery in rural and underserved communities. This integration holds the potential to transform diabetes care from reactive, clinic-based treatment into proactive, continuous health monitoring and personalized intervention.

Brief description of the methodology

In this project, we focused on the design, development, and evaluation of **DiabBot**, an AI-powered chatbot created to assist in diabetes diagnosis, prevention, and management. We began by collecting data from the widely recognized Pima Indian Diabetes Dataset (PIDD), which includes key clinical variables such as age, glucose levels, body mass index (BMI), blood pressure, number of pregnancies, and medical history. To ensure the reliability and consistency of the input data, we applied a series of preprocessing steps, including normalization, outlier removal, and data cleaning, thereby preparing the dataset for accurate machine learning analysis.

We implemented and trained several machine learning models — specifically Random Forest, Logistic Regression, and Neural Networks — using the preprocessed dataset. These models were evaluated based on performance metrics such as accuracy, precision, recall, and F1 score to identify which approach offered the best predictive capability for diabetes risk assessment. We performed comparative analysis to select the most effective models for integration into our system.

For the chatbot interface, we developed a web-based application using React.js for the frontend, providing users with a clean, interactive, and user-friendly experience. The backend was built using Flask, allowing smooth communication between the chatbot interface and the machine learning models responsible for generating predictions and recommendations. The chatbot was designed to handle structured user inputs, collect relevant health details, and provide personalized feedback on diabetes risk, as well as tailored diet, exercise, and medication suggestions.

Finally, we conducted systematic testing to evaluate the system's overall performance, focusing on prediction accuracy and user satisfaction in various simulated scenarios.

Additionally, we ensured the chatbot could generate structured, readable reports for healthcare professionals, enhancing the coordination between patients and medical providers. We also reflected on key challenges such as data privacy, model generalizability, and system scalability, and we outlined directions for future improvements to strengthen the system further.

Summary of results

Our project successfully produced a functional AI-powered chatbot, **DiabBot**, designed for diabetes risk prediction, prevention advice, and personalized management recommendations. After training and testing multiple machine learning models, we observed that the Random Forest, Logistic Regression, and Neural Network models all achieved strong predictive performance, with the Neural Network showing the highest overall accuracy at 85%, followed by Random Forest at 82% and Logistic Regression at 78%.

During system testing, DiabBot demonstrated a diabetes risk assessment accuracy of 92%, reflecting its effectiveness in identifying individuals at risk based on input parameters such as age, BMI, glucose levels, and medical history. User satisfaction with the chatbot's recommendations and interface was also high, with an 85% satisfaction rating reported during usability testing. Additionally, the system was able to generate structured reports summarizing risk assessments and recommendations, supporting better communication between patients and healthcare providers.

Overall, the results confirm that DiabBot is a promising tool for delivering accurate, personalized diabetes-related insights in a user-friendly format. Despite these positive outcomes, we also identified areas for future improvement, including expanding the

system's ability to address diabetes-related complications and strengthening data security measures.