International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 01 | Jan - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

Diabetes Prediction and Recommendations Using Machine Learning and Generative AI

Abhishek Ponnaboina[1], D.Veeraswamy[2]
[1]Department of Electronics and Communication Engineering
Institute of Aeronautical Engineering, Hyderabad, INDIA
abhi04457@gmail.com
[2]Assistant Professor, Department of Electronics and Communication Engineering
Institute of Aeronautical Engineering, Hyderabad, INDIA
Veeraswamy44@gmail.com

ABSTRACT

Machine learning along with artificial intelligence (AI) seem to make a paradigm shift into the healthcare domain by ensuring better diagnosis and treatment approaches. Considering the Mexico's increasing prevalence of diabetes, timely diagnosis and properly determined intervention approach becomes essential. This project presents a unique AI-based web application that simulates and evaluates diabetes risk in individuals with the help of machine learning and generative AI, enabling personalized health recommendations. Specifically, the application evaluates eight parameters such as glucose, blood pressure, BMI, age, etc. All of the parameters will allow users to foresee the risk of diabetes and determine the appropriate response to it. Moreover, the application uses Google Gemini, a generative AI model, to provide customized diabetes interventions by analyzing a patient's health data, and determines likely future health scenarios. Such interventions relate and explain the causes of diabetes, abnormal characteristics of human health, and effective dietary and lifestyle changes. The system is built using Streamlit which allows for a graphic interface that is convenient whereby the users input health data and instantly get diagnosis and interventions. It can be inferred that there is a huge opportunity in the realm of health technology advancing in terms of better preventive healthcare, which results in better patient outcomes as well as better integration of machine learning and AI.

***Keywords*—Diabetes prediction, machine learning, generative AI, healthcare analytics, Streamlit, supervised learning.**

I.INTRODUCTION

Diabetes is regarded as one of the most important health challenges the world faces today as it has been rated among the top causes of mortality and morbidity on an international scale. This particular illness is chronic in nature and results in failure in the comatose patient's ability to process blood glucose (sugar). There are many complications that arise from this, such as heart failure, nerve damage, and kidney paralysis. Once diabetes is diagnosed, other health complications should be avoided and the general objective of health maintenance is preservation. However, through use of modern technologies, efficient measures can be taken to eradicate the future risk and health impacts of diabetes. But the usage of these technologies can be proven to be time-consuming as well as expensive. There is a huge advancement gap that is still present and a lot more needs to be accomplished for early "smart" health detection systems that are economical, readily available, and more precise in detecting the risk of diabetes at an early stage.

New developments in machine learning and artificial intelligence give us good tools to cope with these deficiencies. Such technologies can analyze enormous quantities of health information and give complex and instantaneous suggestions concerning the medical well-being of a person. Thus, by utilizing machine learning models to identify risks factors and generative AI to devise individualized guidelines, systems for early diabetes diagnosis and advice on how to control one's health can be realized. We look into the creation and application of a system like this, as an example the prediction of diabetes and advice on an appropriate intervention in the easy to use browser based platform.

1.1 Description

Diabetes has emerged as a major non-infectious disease, impacting people all around the globe and forming a great part of the burden on the health care system. Preventive measures and timely treatment of the disease are of utmost importance to prevent long-term complications such as heart diseases, end-stage renal disease or diabetic nephropathy, and peripheral neuropathy. Diabetes risk assessment can be defined as the process of collecting and assessing certain key definable and measurable parameters such as serum blood glucose concentrations, body mass index, BP values, age, and others. There is a trend toward more facilitate, less expensive, more effective approaches to diabetes care, even though the or practice of diagnosing and treating this disease is patronized by many still involves consultations

with doctors and laboratory tests. The development of machine learning and artificial intelligence (AI) opens up new possibilities for delivering personalized health evaluations and recommendations, which would enable people to make efforts to prevent the negative impact of their lifestyle.

1.2 Problem Statement

Instead of being an expert on diabetes at large, I concentrated somewhat more on diabetes with respect to nutrition as a treatment. The fact that I will never go to any clinic that offers a one-size-fits-all solution should have been the most important detail that others should know about me. For some reasons, there is much to be gained in terms of cultures' approaches to medicine. In such a way, the ways of treating diabetes will be tackled. Better things that can be considered are the remedies for diabetes and how it is caused to neurons. That is what working together with a number of different investigational groups will do. Not to end up finding the possible cure for diabetes is truly one of the reasons. It is also not that the cure will be found, but that through various investigational channels, the cure will be discovered. It would also be foolish to go to a particular clinic when an argument is made for visit to another comprehensive diabetes center as such a small issue appears inconsequential. Bali said that he had followed hundreds of people for over a long time through his practice and all saw changes.

1.3 Proposed System

This paper describes a web-based system for predicting diabetes and providing recommendations using machine learning and generative AI to solve the gaps in current-day healthcare systems. The main purpose of the system is to estimate the risk for diabetes of an individual by testing some health parameters such as glucose level, blood pressure, BMI and age. The pre-trained model examines these parameters and classifies a person's risk as diabetic or non-diabetic. In addition, it integrates in a generative AI model called Google Gemini that serves the purpose of developing personalized recommendations based on the health data inputted and forecasted diagnosis. Such recommendation would include: possible causes of diabetes, possible insight into the health parameters which fall out of the normal range, and recommendations for practical solutions such as dietary changes and lifestyle changes.

Streamlit was used to design the interface of the system. The added real-time prediction and advisory features will empower individuals to manage their health and make educated decisions. Health care providers will also find this system helpful, in providing immediate assistance for the diagnosis and advising of patients. Combining machine-learning prediction models with AI-generated personalized recommendations represents a novel and efficient approach to diabetes prediction and management, which serves to enhance the burgeoning research into using technology for preventive health care.

II. BACKGROUND

The introduction of new machine learning technologies has resulted in the construction of many predicting and diagnostic systems within the healthcare sector, especially for diabetes prediction. The professional capacity to gather enormous amounts of health information has led to healthcare employing machine learning as one of its core strategies. Utilizing machine learning techniques for diabetes prediction, enables anticipating the development of the disease, which allows for timely actions that drastically reduce the likelihood of its complications in the future. Such technologies help in determining the probability of progressing to diabetes by basing the prediction on several health indicators such as blood glucose, blood pressure, body mass index (BMI), and age as a part of an encompassing approach to health care management.

There are many benefits associated with the use of machine learning in the field of healthcare. Firstly, such techniques can determine people who are at risk of developing diabetes within the population and aid in the initiation of prevention measures. Also, the application of machine learning algorithms can contribute towards the efficient utilization of healthcare services through the identification and treatment of high risk individuals timely. In addition, the new strategies can also make use of algorithms to identify patterns of diseases, various risk factors and a multitude of other data sets allowing for a more comprehensive and intricate framework to be developed to help treat patients. However, some challenges do remain notwithstanding these optimistic results. These, for example, include the technical aspects regarding the quantity and quality of health data which can be gathered, protection of patients' private and sensitive information and communicating such predictive tools. The application of machine learning methodologies into current health systems causes concerns as well about the law and ethical boundaries. However, the opportunities afforded by effective use of machine learning in predicting diabetes are immense, and they open new horizons for patients through their medical personnel and enhance the health and life of the people greatly.

To address these needs, this initiative intends to investigate the possibilities offered by integrating machine learning with generative AI models to accommodate diabetes prediction and related health recommendations tailoring. When combined with a pre-trained machine-relearning model or a generative AI model, the system generates real-time assessments about the probability of acquiring diabetes and also delivers precise self-care guidelines regarding efficient management of the disease. The ability of a system to evaluate multiple health metrics and determine a person's likelihood of acquiring diabetes alongside offering risk reducing health advice will for sure go a long way in equalizing health provision to various persons in this technology age where everyone strives to augment his her health.

Tools Used:-

- **Kaggle:** Platform for accessing relevant datasets.

- **Google Colaboratory:** Cloud-based resource for model training.
- **Anaconda:** Dependency and package management tool.
- **Visual Studio Code (VS Code):** IDE for coding and debugging.
- **Streamlit Cloud:** Web deployment for user interaction.

Technologies Used:-

- **Python:** Programming language for model development.
- **NumPy:** Library for array manipulation.
- **Pandas:** Data preprocessing library.
- **Scikit-learn (Sklearn):** Machine learning library for model training.
- **Pickle:** Serialization tool for saving models.
- **Streamlit:** Framework for building interactive apps.
- **Google Gemini:** AI model for generating health recommendations.

This suite of technologies, algorithms and devices has empowered fused humanity for more effective diabetes detection and customized health advice, thus adding to the already increasing popularity of ai based solutions in bringing better effects in health care.

III. SYSTEM ANALYSIS

The system analysis outlines the requirements and functionality of the diabetes prediction system to ensure it meets user needs effectively.

3.1 Functional Requirements

The system must fulfill the following key functional requirements:

- **Disease Prediction:** Predict diabetes risk based on user health data.
- **Personalized Recommendations:** Provide tailored health suggestions.
- **User Interaction:** Offer an intuitive interface for data input and results.
- **Data Validation:** Ensure correct input data before predictions.
- **Model Updates:** Allow for regular updates to improve prediction accuracy.

3.2 Non-Functional Requirements

The system should meet the following non-functional requirements:

- **Reliability:** Provide consistent, accurate predictions.
- **Performance:** Generate real-time predictions without delay.
- **Scalability:** Handle increasing users and data efficiently.
- **Security:** Protect user data and ensure privacy.
- **Usability:** Ensure an easy-to-use interface for all users.
- **Availability:** Be accessible 24/7 with minimal downtime.
- **Maintainability:** Support easy updates and bug fixes.

3.3 System Architecture

The system consists of the following components:

- **Frontend:** Streamlit-based user interface for input and results.
- **Backend:** Machine learning model to predict diabetes risk.
- **Generative AI:** Google Gemini to provide personalized recommendations.
- **Data Storage:** Secure storage of user input for processing.

IV. SYSTEM MODEL

The model combines a diabetes risk assessment with a specialised machine model for generating unique health advice. There are eight vital health metrics that the users do need to input in the streamlit interface, and after doing that the metrics are preprocessed into a format suitable for the prediction. Through a Python pickle module, the already trained predictive model is loaded to make a prediction on a given user's diabetes risk based on the input data. Following Google Gemini's AI prediction, diabetes risk is evaluated and structural lifestyle changes and dietary management options are suggested based on the user's diagnosis. Since ML and other analytics have been integrated into the system, timely and augmented decisions make the system intelligent.

V. EXPERIMENT

5.1 Hypothesis Generation

Diabetes prediction system has been hypothesized to be accurate by comparing the health parameters of age, BMI, glucose levels and blood pressure. It further hypothesizes analyzes through the use of machine learning algorithms whether an individual is liable to diabetes in the future. Through the help of a model it aims to efficiently check for strong relationships that exist amongst the parameters for higher prediction accuracy.

5.2 Collection of Data

Data for the diabetes prediction system was sourced from Kaggle, a platform offering a wide range of publicly available datasets. These datasets provide valuable real-world data for training the machine learning model. The data includes key health parameters such as glucose levels, blood pressure, BMI, age, and others, which are essential for making accurate predictions about diabetes risk.

5.3 Data Preprocessing / Removal of Unwanted Data

Over the course of the research, the acquired data was subjected to data preprocessing techniques which enabled normalizing the data by eliminating any irrelevant or noisy elements. Data cleaning through outlier removal and addressing missing values was implemented, alongside data formatting for machine learning purposes. If a QA is to be reliable, it surely entails data preprocessing for quality and consistency augmentation, as well as indicating it will augment model predictability accuracy.

5.4 Feature Selection

Feature selection was performed to identify the most important variables that contribute to diabetes prediction. Statistical measures and correlation analysis were used to assess the relevance of each feature. By selecting only the most significant features, we reduced the dimensionality of the data and ensured that the model was focused on the most impactful health parameters. This step also improved the efficiency and accuracy of the predictive model.

5.5 Model Building

For building the diabetes prediction model, we used a pre-trained machine learning model, which was loaded using the Pickle module. The model was trained on health-related data and is capable of predicting whether an individual is at risk of diabetes based on the provided health parameters. The model is designed to handle numerical input, making it capable of predicting risk with high accuracy. The trained model allows real-time predictions and can be reused for future predictions without the need for retraining.

5.6 Deployment

The diabetes prediction system was deployed using Streamlit, providing a simple and interactive web interface for users. Through the interface, users can input their health data and receive instant predictions regarding their diabetes risk. The system also uses Google Gemini, a generative AI model, to offer personalized health recommendations based on the user's input and prediction results. The deployment of the system ensures easy access and scalability, allowing users to make informed decisions about their health and receive tailored advice.

VI. DESIGN

6.1 Architecture Design

The system design consists of several key components. The **Dataset Acquisition and Preprocessing Module** handles importing datasets from sources like Kaggle, inspecting and cleaning the data for missing values, and splitting it into training and testing sets. The **Prediction Module** utilizes the Support Vector Machine (SVM) algorithm to classify input parameters and predict the likelihood of diseases with high accuracy. The trained SVM models are then converted into pickle files and integrated into the **Model Deployment Module**, which ensures the system's scalability and usability.

6.2 Architecture Design Interface

The architecture design interface is implemented using the Streamlit framework, providing a user-friendly platform for interaction. The **User Interface Module** allows users to input relevant health parameters, such as glucose levels, blood pressure, BMI, and age, and displays instant predictions for various diseases. Additionally, it provides personalized health recommendations based on the predictions to assist users in making informed decisions about their health.

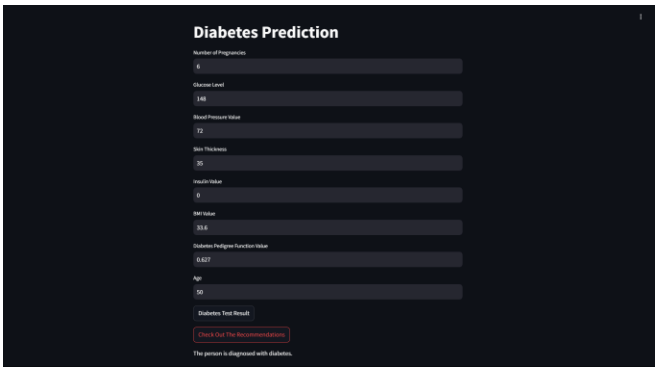

Fig. 1. Diabetes Prediction Interface

VII. PRELIMINARIES

Support Vector Machine (SVM) Algorithm

Support Vector Machine (SVM) is a supervised machine learning algorithm commonly used for classification and regression tasks. It is particularly effective for binary classification problems, but it can also handle multi-class classification. SVM works by finding the optimal hyperplane that best separates the data points of different classes in a higher-dimensional space.

The main idea behind SVM is to maximize the margin between the data points of different classes. The hyperplane is defined as:

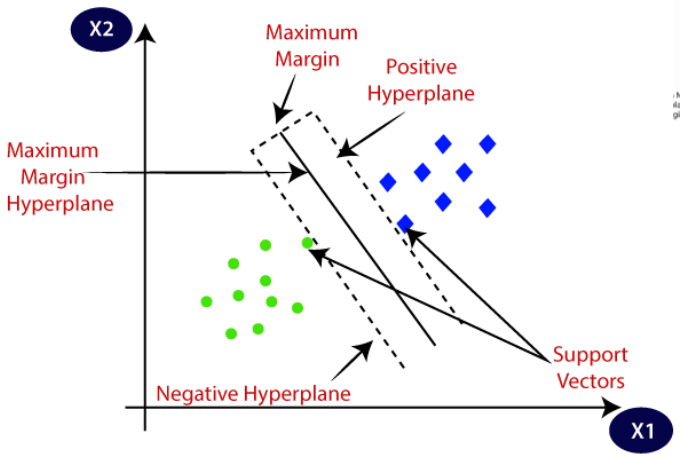

Fig. 2. SVM showing the maximum margin hyperplane, support vectors, and margins.

$$x \cdot \mathbf{w} + b = 0$$

where $\mathbf{w}$ is the weight vector, b is the bias, and x represents the data points. The support vectors are the data points closest to the hyperplane, and they play a key role in defining the decision boundary.

For classification, the SVM aims to classify data points such that:

$$y_i (x_i \cdot \mathbf{w} + b) \geq 1 \quad \text{for all } i$$

where y_i represents the class label for the data point x_i, which takes the values +1 or -1. The goal is to maximize the margin between the classes, thereby improving the generalization capability of the model.

The margin M is defined as the distance between the hyperplane and the nearest data points (support vectors) of each class:

$$M = \frac{2}{\|\mathbf{w}\|}$$

SVM maximizes this margin by minimizing $\|\mathbf{w}\|^2$, leading to the optimization problem:

$$\text{Minimize } \frac{1}{2}\|\mathbf{w}\|^2$$

$$\text{Subject to: } y_i(\mathbf{x}_i \cdot \mathbf{w} + b) \geq 1, \ \forall i$$

For cases where the data is not perfectly separable, SVM introduces slack variables ξ_i to allow some misclassifications while still maximizing the margin. This leads to the following objective function:

$$\text{Minimize } \frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{i=1}^{n} \xi_i$$

$$\text{Subject to: } y_i(\mathbf{x}_i \cdot \mathbf{w} + b) \geq 1 - \xi_i, \ \xi_i \geq 0, \ \forall i$$

Here, C is a regularization parameter controlling the trade-off between maximizing the margin and minimizing classification errors.

1) Kernel Trick

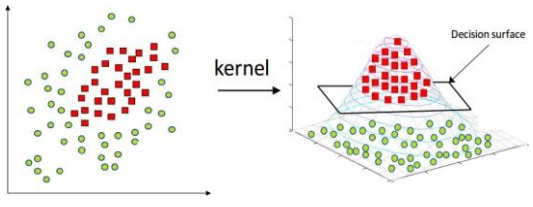

Fig. 3. Illustration of the kernel transformation and decision surface.

For non-linearly separable data, SVM uses a kernel function to transform the data into a higher-dimensional space where it becomes linearly separable. Common kernel functions include:

- **Linear Kernel:** $K(\mathbf{x}, \mathbf{z}) = \mathbf{x} \cdot \mathbf{z}$
- **Polynomial Kernel:** $K(\mathbf{x}, \mathbf{z}) = (\mathbf{x} \cdot \mathbf{z} + c)^d$
- **Radial Basis Function (RBF) Kernel:** $K(\mathbf{x}, \mathbf{z}) = \exp(-\gamma \|\mathbf{x} - \mathbf{z}\|^2)$

By applying these transformations, SVM finds a decision boundary in the transformed space.

VIII. RESULTS

Our system, which utilizes the Support Vector Machine (SVM) model for diabetes prediction, achieved an accuracy of 78%. This level of accuracy reflects the model's ability to accurately identify individuals at risk of developing diabetes based on the input health parameters such as glucose levels, BMI, blood pressure, insulin levels, and age. The SVM model's performance was evaluated using a range of metrics, including precision, recall, and F1-score, which demonstrated a robust capability for classifying both diabetic and non-diabetic individuals.

Fig. 4. Overview of Possible Causes, Abnormal Values, and Practical Solutions

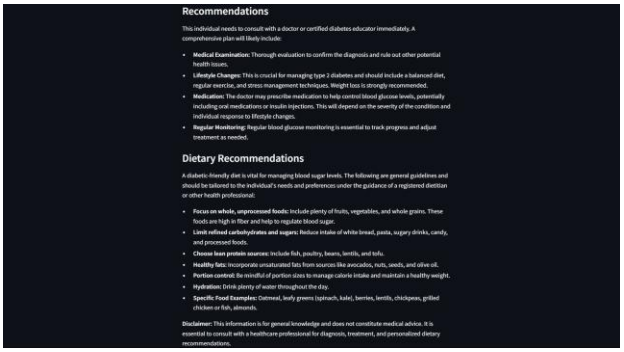

Fig. 5. Real Time Insights and Recommendations

In addition to the predictive accuracy, the system also integrates AI-generated recommendations to further assist users. These recommendations, generated by the Google Gemini model, provide personalized insights related to diabetes management. The recommendations focus on lifestyle changes, dietary adjustments, and tips for managing abnormal health parameters. During user testing, the AI-generated recommendations received positive feedback, with users finding them relevant, clear, and actionable. The clarity of the suggestions and their practical applicability in everyday health management were highlighted as key strengths of the system.

To assess the system's overall effectiveness, we conducted a comparative analysis with existing diabetes prediction systems. This comparison highlighted the superior integration of machine learning and generative AI in our approach. While traditional systems typically provide only predictions, our

system goes a step further by offering personalized, AI-generated recommendations tailored to each user's unique health data. This integration of predictive modeling with personalized advice enhances the overall user experience and supports proactive healthcare management.

Moreover, the system's deployment using the Streamlit framework ensures accessibility and ease of use. Users can input their health data through a simple and intuitive interface and receive immediate feedback, making it a valuable tool for individuals seeking to monitor their health or take preventive actions against diabetes.

IX. CONCLUSION

The aim of this venture was to build a system that would aid in diagnosing diabetes with machine learning. Using the Support Vector Machine (SVM) model, the system managed to have a success rate of 78%, able to give the probability of a person having a disease based on glucose concentration, BMI, BP, and so forth. This feature enables patients to take action as soon as possible as the program brings various disease prediction tools into one eliminating both time and energy wasted.

Besides predicting, the system also helps users by specifying AI-generated ideas on how a user could change their way of life or dieting plans in order to either manage or prevent diabetes. The user's feedback during the testing phase showed that the recommendations were understandable and relevant therefore improving the practical function of the system so that patients receive extra help for active health management.

The experience received during this project proves that healthcare can be advanced in based on the precise AI-based prediction and provision of qualitative and timely health care recommendations. The fact that users were able to not only predict the likelihood of diabetes with the use of the program but, also get suggestions on how to control it will promote improvements in the medical field and ultimately bring positive results.

X. Future Scope

- **Expansion to Multiple Disease Prediction:** The system can be extended to predict a wider range of diseases, including hypertension, cardiovascular diseases, and more, making it a comprehensive tool for healthcare prediction.
- **Integration with AI Advancements:** With ongoing developments in AI, the system can leverage more advanced algorithms and models to enhance prediction accuracy and provide even more personalized and insightful recommendations.
- **Interactive Web Design:** Future iterations of the system can focus on making the web interface more interactive and user-friendly, offering enhanced visuals, real-time feedback, and a smoother overall experience for users.
- **Real-Time Health Monitoring:** By integrating the system with IoT-enabled devices, it will allow for continuous, real-time monitoring of health parameters, leading to more proactive and accurate disease predictions.
- **Integration of Natural Language Processing (NLP):** Future versions of the system could incorporate NLP techniques to analyze user input in the form of natural language, improving communication and enhancing user experience.
- **Cloud-Based Data Storage:** To handle large volumes of health data, the system could be upgraded to use cloud-based storage solutions, ensuring seamless access, data security, and real-time processing.
- **Data Privacy and Security Enhancements:** As the system handles sensitive health data, future improvements could focus on incorporating advanced encryption and data privacy protocols to ensure patient confidentiality and secure data storage.
- **Mobile Application Development:** In addition to the web application, a mobile app version of the system could be developed to reach a broader audience and provide users with convenient, on-the-go access to health predictions and recommendations.
- **Collaboration with Healthcare Providers:** By integrating with healthcare providers' systems, the tool could assist medical professionals in making informed decisions and provide a collaborative approach for patient management.

XI. REFERENCES

- Gopisetti, L. D., Kummera, S. K. L., Pattamsetti, S. R., Kuna, S., Parsi, N. & Kodali, H. P. (2023, January). Multiple disease prediction system using machine learning and streamlit. In: 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 923-931. IEEE.
- Keniya, R., Khakharia, A., Shah, V., Gada, V., Manjalkar, R., Thaker, T., ... & Mehendale, N. (2020). Disease prediction from various symptoms using machine learning. Available at: SSRN 3661426.
- Phasinam, K., Mondal, T., Novaliendry, D., Yang, C. H., Dutta, C. & Shabaz, M. (2022). Analyzing the performance of machine learning techniques in disease prediction. Journal of Food Quality.
- Vijayalaxmi, A., Sridevi, S., Sridhar, N. & Ambesange, S. (2020, May). Multi-disease prediction with artificial intelligence from core health parameters measured through non-invasive technique. In 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1252-1258. IEEE.
- Priyanka Sonar & Prof. K. JayaMalini. (2019). Diabetes prediction using different machine learning approaches. IEEE 3rd International Conference on Computing Methodologies and Communication (ICCMC).