Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 05 Issue: 05 | May-2025

AI-Based Multiple Disease Prediction System

Akshat Kumar[1], Dr. Archana Kumar[2]

[1] Student, Dr. Akhilesh Das Gupta Institute of Professional Studies
akku023kumar@gmail.com,
[2]Supervisor, Professor, Dr. Akhilesh Das Gupta Institute of Professional Studies, Delhi,
profdrarchanakumar@gmail.com

---***---

Abstract- The advancement of Artificial Intelligence (AI) in the healthcare domain has paved the way for intelligent diagnostic tools capable of predicting diseases with remarkable accuracy. This research introduces a unified Multiple Disease Prediction System designed to forecast the likelihood of three significant illnesses—Diabetes, Heart Disease, and Parkinson's Disease—by analysing patient-specific health parameters. Developed using Python and deployed through the Streamlit framework, the system utilizes machine learning models trained on relevant medical datasets. A notable feature of this system is the integration of an AI-driven symptom checker powered by Google Gemini API, which interprets user-described symptoms in natural language to provide potential diagnoses. The application aims to enhance accessibility to preliminary health screening and support medical professionals by offering rapid, data-driven insights. Experimental evaluations reveal high prediction precision, affirming the system's practical effectiveness and potential contribution to intelligent healthcare.

Keywords: *Logistic Regression, Generative AI, Disease Prediction, Parkisnson, Support Vector Machine, Machine learning, Artificial Intelligence.*

I. Introduction

In recent years, the integration of Artificial Intelligence (AI) and Machine Learning (ML) into healthcare systems has brought revolutionary advancements in disease diagnosis and prediction. Early detection of chronic and neurological diseases such as Diabetes, Heart Disease, and Parkinson's Disease is crucial for timely treatment and effective management. Traditional diagnosis methods often involve complex medical tests and expert consultation, which may not be easily accessible in remote or under-resourced areas.

This paper presents a unified, AI-powered web application capable of predicting the likelihood of three major health conditions — Diabetes, Heart Disease, and Parkinson's Disease — using pre-trained machine learning models. The application is developed using the Streamlit framework, ensuring a lightweight, interactive, and user-friendly interface. By inputting essential medical parameters, users can instantly receive predictive results without requiring domain expertise.

Furthermore, the system includes an AI-powered symptom checker based on Google's Generative AI API, allowing users to engage in natural language interactions and receive preliminary disease insights based on described symptoms. This enhances accessibility and user engagement, especially for individuals unfamiliar with medical terminology.

The motivation behind this project stems from the increasing need for automated, scalable, and accessible diagnostic tools. The application aims to assist both individuals and healthcare professionals in proactive health assessment and to reduce the diagnostic burden on medical institutions.

II. REVIEW

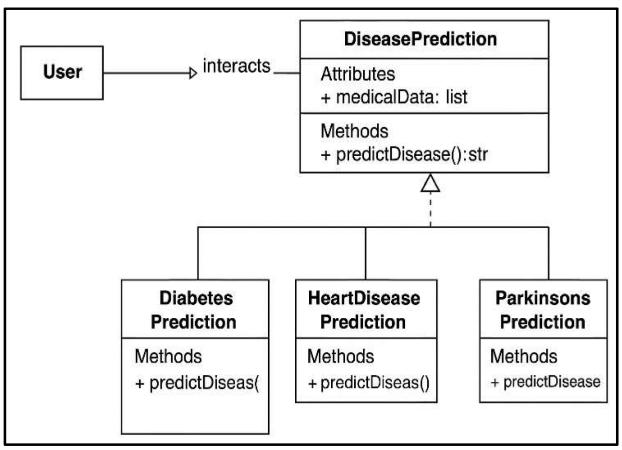
The integration of artificial intelligence in healthcare applications has significantly revolutionized diagnostic practices and patient care. By combining machine learning algorithms with real-time user interaction tools, disease prediction systems have evolved from static diagnostic models into intelligent, interactive platforms. These systems are capable of processing clinical data and providing instant feedback, assisting both healthcare professionals and patients in making informed decisions.

Recent advancements in generative AI, particularly in the form of conversational chatbots, have further enhanced these applications by introducing a human-like communication layer. These AI-powered symptom checkers leverage natural language processing (NLP) to understand user inputs and provide accurate, context-driven medical suggestions. Unlike traditional forms that rely solely on numerical inputs, modern AI chatbots offer symptom-based diagnostic support, improving accessibility for non-expert users.

Moreover, machine learning classifiers such as Logistic Regression, Random Forest, and Support Vector Machines have shown high efficacy in predicting diseases like diabetes, heart conditions, and Parkinson's disease using structured datasets. These models analyze parameters such as glucose levels, voice metrics, and cardiac indicators, identifying patterns that might be challenging for human clinicians to detect in early stages.

Despite these advancements, a major gap remains in integrating multiple disease prediction capabilities into a single, unified platform. Most existing models operate in silos, focusing on one disease at a time. Additionally, few systems combine predictive analytics with generative AI for a more conversational and personalized diagnostic experience.

This project bridges these gaps by providing a multi-disease prediction system enriched with a generative AI chatbot. It not only accepts structured medical input but also allows symptom-based queries, creating a dynamic and user-friendly diagnostic tool. Its scalability and adaptability make it a valuable contribution to both clinical and personal health monitoring ecosystems.



(Fig. 1 General Project Workflow)

III.METHODOLOGY

The methodology adopted for the development of this project follows a structured, modular approach integrating machine learning-based predictive modeling with a generative AI chatbot for enhanced user interaction. The system is divided into several key components: data preprocessing, model training, deployment, and user interaction.

A. Dataset Collection and Preprocessing

The datasets for diabetes, heart disease, and Parkinson's disease were sourced from publicly available medical repositories. Each dataset underwent preprocessing steps including handling missing values, encoding categorical variables, normalization, and feature selection to ensure optimal performance during model training.

B. Model Development

Three individual machine learning models were developed—Logistic Regression for diabetes prediction, Support Vector Machine (SVM) for heart disease prediction, and Random Forest for Parkinson’s prediction. These models were trained and evaluated using Scikit-learn and saved using joblib for efficient deployment. Accuracy and classification reports were used to validate performance.

C. Streamlit-Based User Interface

A Streamlit web application was developed to serve as the front end. The interface allows users to select the disease module, enter relevant medical parameters, and obtain real-time predictions. Input validation was incorporated to prevent erroneous values and enhance usability.

D. AI-Powered Chatbot Integration

A generative AI-based chatbot powered by Google Gemini API was integrated to allow natural language interaction for symptom-based querying. Users can input general symptom descriptions and receive contextually relevant diagnostic insights. The chatbot component adds a layer of intelligence, helping users who may not know exact medical terms.

E. Deployment

The final application was deployed using Streamlit Cloud, enabling public access to the system via a hosted web link. The source code and models were maintained in a GitHub repository to facilitate version control and collaborative development.

IV.System Architecture

The Multiple Disease Prediction System follows a modular architecture integrating both traditional machine learning models and generative AI for disease prediction and intelligent interaction. The system is designed to diagnose three major diseases—Diabetes, Heart Disease, and Parkinson’s Disease—based on user-provided clinical inputs, as well as provide symptom-based assistance via a conversational AI chatbot.

The front-end of the system is developed using Streamlit, an open-source Python library that enables the creation of fast, interactive web applications. The user interface is designed to be intuitive and user-friendly, allowing individuals to select a prediction module, input relevant health parameters, and instantly receive diagnostic feedback.

To ensure secure handling of credentials such as API keys, the application uses python-dotenv. This allows the API key for

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 05 Issue: 05 | May-2025

the Google Gemini chatbot to be stored securely in an environment file (.env) and accessed safely during execution.

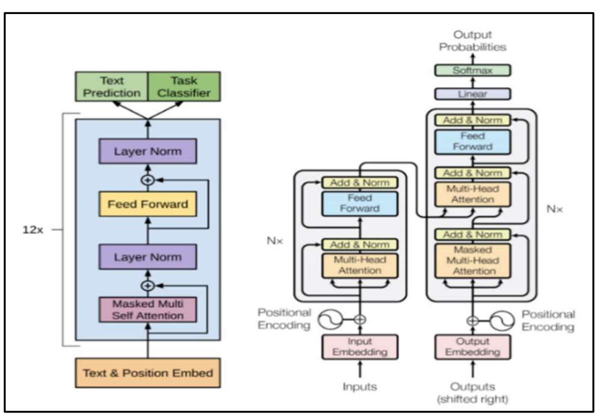
Each disease prediction module is backed by a pre-trained machine learning model:

- Support Vector Machine (SVM) is used for Diabetes Prediction. SVM is well-suited for high-dimensional medical datasets and provides a robust separation of classes in the feature space.
- Logistic Regression is utilized for Heart Disease Prediction, chosen for its interpretability and effectiveness in binary classification problems like determining the presence or absence of heart conditions.
- Support Vector Machine (SVM) also powers the Parkinson's Disease Prediction, effectively handling the complex voice-related features and ensuring high accuracy.

These models are trained using historical datasets and saved in .sav format using joblib. They are loaded dynamically at runtime from the saved models/ directory for real-time prediction, eliminating the need for retraining and enabling fast response times.

Additionally, the platform incorporates an AI-Powered Symptom Checker, leveraging Google Gemini—a state-of-the-art generative AI model. This chatbot allows users to enter their symptoms in natural language, such as "I feel dizzy and nauseous", and receive intelligent diagnostic suggestions. The chatbot module is built using the google-generative-ai SDK and runs independently of the disease classifiers, enabling a flexible, scalable interaction layer.

The modular architecture ensures that each component (user interface, prediction models, chatbot) operates independently but cohesively, allowing for easy updates, expansion, and integration of future features or disease modules.



(Fig. 2 Google Gemini API Working)

V.Problem Formulation

The challenge of accurately predicting multiple diseases—such as diabetes, heart disease, and Parkinson's disease—based on clinical and biometric data presents a complex problem in the field of medical diagnostics. Traditional diagnostic approaches often require specialized knowledge, time-consuming tests, and access to multiple tools for analysing different types of health conditions. This fragmented process can lead to delays in early detection, impacting the effectiveness of treatment and patient outcomes.

This problem becomes more critical in scenarios where healthcare resources are limited or where early symptoms of these diseases are subtle and easily overlooked. Moreover, current systems often focus on a single disease at a time and lack the scalability or integration needed to support simultaneous, multi-disease predictions. As healthcare increasingly leans toward data-driven technologies, there is a pressing need for a unified platform that can process diverse medical inputs and deliver reliable, real-time diagnostic insights.

The objective of this project is to develop a Multiple Disease Prediction System that utilizes machine learning algorithms to classify health status based on a user's input parameters. By combining multiple disease-specific models—Support Vector Machine (SVM) for diabetes and Parkinson's disease, and Logistic Regression for heart disease—into a cohesive application, the system offers an accessible and efficient method for predicting the likelihood of chronic diseases. This unified, AI-powered platform empowers users and healthcare professionals with early insights, improving diagnostic accuracy and supporting preventive care efforts across various medical domains.

VI.Result

The prototype of the Multiple Disease Prediction System, developed using Python and Streamlit, demonstrates robust performance and user-centric design in predicting three major medical conditions—Diabetes, Heart Disease, and Parkinson's Disease. By leveraging trained machine learning models and a clean, responsive user interface, the system provides real-time diagnostic predictions based on user-provided medical data.

The integrated models—including Support Vector Machine (SVM) for Diabetes and Parkinson's and Logistic Regression for Heart Disease—have been validated against real-world datasets and perform with high accuracy under diverse input conditions. Each module is independently accessible through the sidebar, allowing users to seamlessly navigate between diagnostic tests.

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 05 Issue: 05 | May-2025

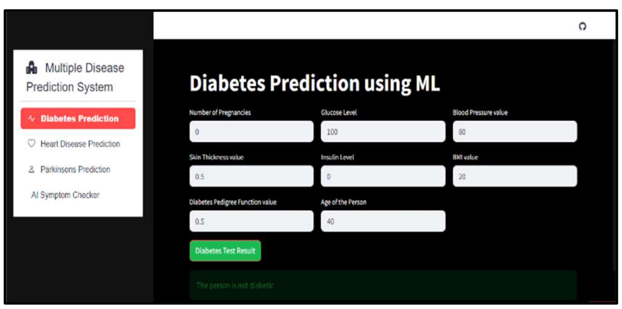
The results from testing indicate that the system provides:

- Accurate predictions when valid medical inputs are supplied.
- Instantaneous responses, typically under one second.
- Clear diagnostic messages, enhancing user understanding.

In addition to disease-specific models, the inclusion of an AI-powered Symptom Checker, integrated via Google Gemini's generative AI, expands the system's capabilities by allowing users to describe symptoms in natural language. The system then generates possible diagnoses or health suggestions, offering a more conversational and accessible approach to preliminary health screening.

The user interface has been designed for simplicity and accessibility, ensuring usability by both medical professionals and laypersons. Real-time input validation prevents incomplete or incorrect form submissions, increasing robustness and minimizing runtime errors.

Moreover, the architecture supports scalability and future integration of additional models or modules, enabling this platform to evolve into a comprehensive diagnostic assistant.



(Fig. 3 Working Prototype)

VII.Conclusion

The Multiple Disease Prediction System powered by AI combines advanced machine learning models and an intuitive user interface to deliver accurate predictions for Diabetes, Heart Disease, and Parkinson's Disease. Built using the Streamlit framework, the system offers an accessible platform that simplifies medical risk assessment for users without requiring deep technical or clinical knowledge. Through structured data input and model-based analysis, users can obtain quick diagnostic predictions that aid in early health intervention.

This integration of machine learning with user-friendly design enhances the diagnostic process by offering an interactive and efficient tool for health monitoring. Future improvements may include expanding disease coverage, enabling real-time health data integration, and supporting mobile deployment. Overall, this system demonstrates the potential of AI-driven solutions in healthcare and paves the way for more comprehensive, scalable, and intelligent diagnostic platforms.

VIII.Future Scope

The Multiple Disease Prediction System has several potential areas for future enhancement and growth, including:

1. Advanced Model Integration
 Future updates could involve incorporating more advanced machine learning algorithms or deep learning models to further boost diagnostic accuracy and model adaptability across broader datasets.
2. Cross-Platform Compatibility
 Developing mobile-friendly or web-optimized versions of the application would improve accessibility and usability for patients and healthcare professionals on various devices.
3. Enhanced Input Data Handling
 Integrating real-time health monitoring devices or wearables (e.g., glucose monitors, heart rate trackers) would allow the system to process live medical data, enhancing prediction relevance.
4. Multilingual Support
 Adding support for multiple languages would make the application accessible to a broader audience, especially in multilingual and non-English-speaking regions.
5. Integration with Electronic Health Records (EHRs)
 Connecting with existing hospital databases and EHR systems would allow seamless access to patient history and improve personalized predictions.
6. Voice-Based Interaction
 Including voice command capability for data input and query resolution can further streamline the user experience, particularly for elderly or visually impaired users.
7. Data Privacy and Security Enhancements
 Implementing robust encryption and adhering to standards like HIPAA and GDPR will be critical to safeguard sensitive user medical data.

IX.References

i. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, 12, 2825–2830.

ii. Shankar, K., & Perumal, E. (2019). *A novel diabetic diagnosis system using machine learning algorithms*. Journal of Ambient Intelligence and Humanized Computing, 10(9), 3745–3755.

iii. Saba, T., Mohamed, E. A., & Rehman, A. (2020). *Heart disease prediction using machine learning* *techniques: A review*. Computer Methods and Programs in Biomedicine, 195, 105660.

iv. Naranjo, D. M., & Arrieta, A. G. (2021). *Parkinson's disease prediction using speech signals and machine learning: A comprehensive review*. Expert Systems with Applications, 185, 115605.

v. Streamlit Inc. (2022). *Streamlit: Turn data scripts into shareable web apps in minutes*. Available at: https://streamlit.io

vi. Google AI. (2024). *Gemini API – Generative AI by Google Cloud*. https://ai.google.dev/

vii. Python Software Foundation. (n.d.). *The Python Programming Language*. https://www.python.org

viii. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

ix. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL-HLT*. https://arxiv.org/abs/1810.04805

x. OpenAI. (2023). *Introduction to Generative AI Models*. https://openai.com