Multiple disease prediction using machine learning

Prof. S.D. Jadhav[1], Aarti H. Mulik[2], Arpita R. Sartale[3], Rajlaxmi S. Patil[4], Sanika L. Mohite[5]
Lecturer, Computer Science Engineering, Tatyasaheb Kore Institute of Engineering and Technology Polytechnic, Warananagar, Maharashtra, India
Students, Computer Science Engineering, Tatyasaheb Kore Institute of Engineering and Technology Polytechnic, Warananagar, Maharashtra, India

***Abstract*—The Multi-Disease Prediction System (MDPS) aims to utilize the power of machine learning (ML) algorithms, particularly Logistic Regression and Support Vector Machines (SVM), to predict the likelihood of developing multiple diseases. The system is designed with a Streamlit interface, making it user-friendly and accessible for individuals to predict the onset of conditions such as diabetes, heart disease, and Parkinson's disease. The goal is to create an efficient and accurate tool that not only assists in early detection but also promotes personalized healthcare by identifying the risk factors related to these diseases.This approach considers key health indicators like blood pressure, pulse rate, cholesterol levels, and heart rate to predict disease risks. Through a model that achieves reliable accuracy and precision, the system identifies underlying patterns and provides predictions based on these health parameters. The research emphasizes the potential impact of machine learning on public health, particularly in enhancing disease prediction and management.While many algorithms exist for predicting individual diseases, the unique focus of this paper is the development of a system capable of predicting multiple diseases simultaneously. The system, trained using sample data, allows for more comprehensive health assessments, contributing to proactive healthcare decisions and better disease prevention strategies.**

***Index Terms*—Streamlit, Machine Learning, Diabetes, Heart Disease, Parkinson's Disease, SVM, Logistic Regression.**

I. INTRODUCTION

This project is aimed at developing a system that utilizes machine learning to predict the likelihood of three prevalent diseases: heart disease, diabetes, and Parkinson's disease. The system will process patient data, including details such as age, symptoms, and medical history, to generate predictions. A range of machine learning models—such as Naïve Bayes, K-Nearest Neighbors (KNN), Decision Trees, Random Forest, and Support Vector Machines (SVM)—will be employed to make these predictions. The outcomes will be displayed through a user-friendly web application built with Streamlit. This approach aims to assist both healthcare providers and patients in gaining early insights into potential health conditions.

II. PROBLEM DEFINITION

A. Literature Review

The primary focus of this project is the simultaneous prediction of multiple diseases, which contrasts with traditional methods where each disease is predicted separately. In the healthcare domain, existing approaches typically involve creating distinct models for conditions like heart disease, diabetes, or Parkinson's. In contrast, the Multi-Disease Prediction System uses machine learning algorithms to assess the likelihood of a patient suffering from several diseases based on their medical history and reported symptoms. The ultimate goal of this system is to provide accurate predictions for conditions such as heart disease, kidney disease, and diabetes. By enabling early identification, it facilitates timely treatment and contributes to better healthcare outcomes and patient management.

III. SCOPE

The Multi-Disease Prediction System utilizes machine learning algorithms to predict several diseases, including heart disease, diabetes, and Parkinson's disease, based on patient data. This system is designed to promote early detection and timely treatment, offering a user-friendly interface that can be used by healthcare providers. By providing accurate predictions and personalized recommendations, the system aims to enhance healthcare outcomes. It is also

designed with scalability in mind, enabling the inclusion of additional diseases and real-time data integration in the future.

Key Features:

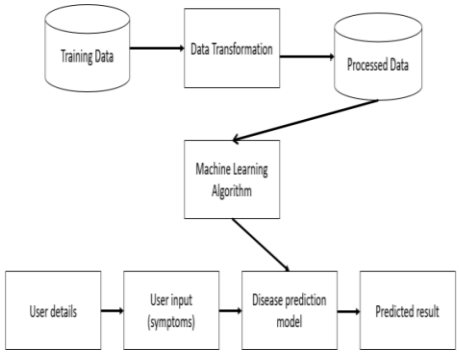
- Multi-Disease Prediction: Simultaneously predict multiple diseases (e.g., heart disease, diabetes, Parkinson's) based on patient information.
- Early Detection: Facilitate the early identification of at-risk individuals for prompt intervention.
- Improved Healthcare Outcomes: Support enhanced patient care through reliable predictions and tailored healthcare suggestions.
- User-Friendly Interface: Provide a simple and intuitive platform for healthcare providers or patients to input data and receive results.
- Integration of Multiple Models: Use various machine learning techniques to predict different diseases in parallel, overcoming the limitations of single-disease models.
- Future Scalability: Enable future extensions to include additional diseases and integrate real-time health data.

IV. METHODOLOGY

The process of predicting multiple diseases at the same time involves several key steps, ensuring that the system is both comprehensive and efficient. The methodology for this project is outlined as follows:

1. Data Collection: The first step involves gathering data from multiple sources, including electronic health records (EHRs), public health databases, and medical literature, specifically for diabetes, heart disease, and Parkinson's disease.
2. Data Preprocessing: This phase includes data cleaning and transformation, which are necessary to convert the raw data into a format suitable for machine learning algorithms. These tasks are essential to ensure that the data is accurate and ready for model training.
3. Model Selection: During this phase, the most appropriate machine learning models for each disease are selected. These models are then trained using the preprocessed data, and evaluated based on various performance metrics like accuracy and precision to improve their prediction capabilities.
4. Data Splitting: The preprocessed data is divided into training and testing datasets. The models are trained using the training set, and their performance is assessed using the testing set to determine their effectiveness.
5. Deployment and Integration: After training, the models are deployed as part of a cloud-based system, integrated into an interactive web application. This application allows users to input relevant health parameters, select diseases for prediction, and receive results, making it more accessible and user-friendly.

V. BLOCK DIAGRAM



Dig.: Multiple disease prediction using machine learning

VI. SYSTEM ANALYSIS

Functional Requirements

- The system should offer an intuitive interface that allows users (patients) to easily input important information, including details about their symptoms, medical history, demographics, and other relevant factors.
- Once the user provides their input, the system will process the data using integrated prediction models and display the predicted diseases. The results should be shown in a clear and understandable format for the user.

Non-Functional Requirements

- The system must provide a range of values or confidence intervals related to predicted disease outcomes. This helps users understand the level of uncertainty in the predictions, which can guide

decisions on whether to seek further medical consultation or interventions.

- The system should be dependable and consistent in its performance, ensuring that predictions and user interactions are accurate and reliable.

VII. CONCLUSION

This research aimed to develop a machine learning-based system for predicting multiple diseases, including heart disease**,** diabetes, and Parkinson's disease**.** By leveraging the SVM algorithm, we achieved the following prediction accuracies**:** 78% for diabetes**,** 89% for Parkinson's**,** and *85% for heart disease***.** The system facilitates early disease detection, allowing both healthcare providers and patients to make better-informed decisions. With machine learning models in place, the system helps optimize patient care, healthcare resources, and supports early disease intervention. In conclusion, this study highlights the potential of machine learning to improve healthcare outcomes and create more efficient healthcare systems.

REFERENCE

[1] Katarya, R., & Srinivas, P. (2020). A survey on predicting heart disease at early stages using machine learning. In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 302–305).

[2] Poudel, R. P., Lamichhane, S., Kumar, A., et al. (2018). Risk prediction of type 2 diabetes mellitus using data mining techniques. Journal of Diabetes Research, 2018, Article 1686023.

[3] Ismaeel, S., Miri, A., & Chourishi, D. (2015). Application of the Extreme Learning Machine (ELM) technique for heart disease diagnosis. In 2015 IEEE Canada International Humanitarian Technology Conference (IHTC) (pp. 1–3).

[4] Gavhane, A., Kokkula, G., Pandya, I., & Devadkar, K. (2018). Heart disease prediction through machine learning techniques. In 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 1275–1278).

[5] Deo, R. C. (2015). Machine learning in medicine. Circulation, 132(20), 1920–1930. Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. Wadsworth and Brooks.

[6] Parashar, A., Gupta, A., & Gupta, A. (2014). Machine learning techniques for predicting diabetes. International Journal of Emerging Technologies and Advanced Engineering, 4(3), 672–675.

[7] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. Wadsworth and Brooks.

[8] Paniagua, J. A., Molina-Antonio, J. D., Lopez-Martinez, F., et al. (2019). Heart disease prediction using random forests. Journal of Medical Systems, 43(10), 329.