MULTIPLE DISEASE PREDICTION SYSTEM

Project Reference No.: *48S_BE_2475*

College : Shri Madhwa Vadiraja Institute of Technology and Management, Udupi
Branch : Artificial Intelligence And Data Science Engineering
Guide(s) : DR. Ganesha Prasad
Student(s) : Ms. Disha P U
Ms. HITHASHREE
Ms. Neha Dinesh Shettigar
Ms. Nidhi Prabhu K

Keywords:

Multiple disease prediction, Diabetes prediction, heart disease prediction, Parkinson's disease, Streamlit, Naïve Bayes, Support vector machine.

Introduction:

The Multiple Disease Prediction System is a web-based application developed using the Streamlit framework, aimed at assisting users in the early detection of common yet critical health conditions. This system focuses on predicting the likelihood of three major diseases: Diabetes, Heart Disease, and Parkinson's Disease. Healthcare professionals often rely on various diagnostic tests and clinical assessments to identify these diseases. However, such procedures can be time-consuming and sometimes inaccessible to many. To address this our system leverages machine learning algorithms to provide quick, reliable, and user-friendly predictions based on medical input parameters. Key features of the system include a simple and interactive Streamlit UI for ease of use, Separate modules for each disease prediction and Immediate results based on user-provided health data. This project aims to serve as a supportive tool for individuals and healthcare providers, promoting early diagnosis and timely medical intervention.

Objectives:

1. To develop robust machine learning models in order to predict the likelihood of diabetes, Parkinson's disease, and heart disease.

2. To create an intuitive and accessible web application using Streamlit that allows users to input their medical data and receive disease predictions easily.
3. To evaluate and compare the performance of different machine learning models in predicting multiple diseases.
4. To implement data visualization tools to display the relationships and distributions of key features related to each disease.
5. To provide fast and reliable predictions that can help in early medical intervention.
6. To continuously improve prediction accuracy through model optimization and feedback.

Methodology:

<u>Collecting and preparing data:</u> Data cleaning is the process of preprocessing data to deal with outliers and missing values while maintaining consistency between datasets. The datasets related to diabetes, Parkinson's disease, and heart conditions are collected which may be sourced from public health datasets, research papers, or medical repositories.

<u>Feature Selection and Engineering:</u> In this step we select pertinent features from each dataset that have the potential to predict for a given illness. Feature engineering is used to produce new features that could increase prediction accuracy. We have used statistical measures and correlation analysis techniques to assess the importance and relevance of each feature.

<u>Model Choice:</u> Here we select the best machine learning models for predicting each disease. In our system, Support Vector Machine (SVM) algorithm is utilized for Diabetes disease prediction and Parkinson's disease prediction, while Naïve Bayes algorithm is employed for heart disease prediction, as these algorithms have demonstrated the best accuracy for their respective diseases.

<u>Evaluation of the Model:</u> Model evaluation involves several key steps to assess a predictive model's performance and reliability. Initially, the dataset is split into training and testing sets (80:20), to ensure the model is evaluated on unseen data. The model is trained on the training set, learning the patterns within the data, and then evaluated on the testing set to see how well it performs on new data.

Streamlit integration: For creating an interface that is easy to use, we make use of Streamlit (Python framework). Streamlit enables users to enter personal health information, such as age, gender, and other medical background information to predict the disease.

Result and Conclusion:

Implemented Support Vector Machine (SVM) models for diabetes and Parkinson's disease prediction, and a Naive Bayes model for heart disease prediction. Achieved high accuracy, precision, and recall metrics, ensuring reliable and accurate predictions for each disease.

Developed a Streamlit-based interface with straightforward input forms for each disease. Users can enter their medical parameters and obtain immediate prediction results, enhancing the application's usability and accessibility.

Conducted a thorough evaluation of the models using metrics such as accuracy, precision, recall, and F1-score.

Integrated interactive visualizations such as histograms, scatter plots, and correlation matrices. These visualizations will help users and clinicians gain deeper insights into the data, improving the interpretability and decision-making process based on the predictions.

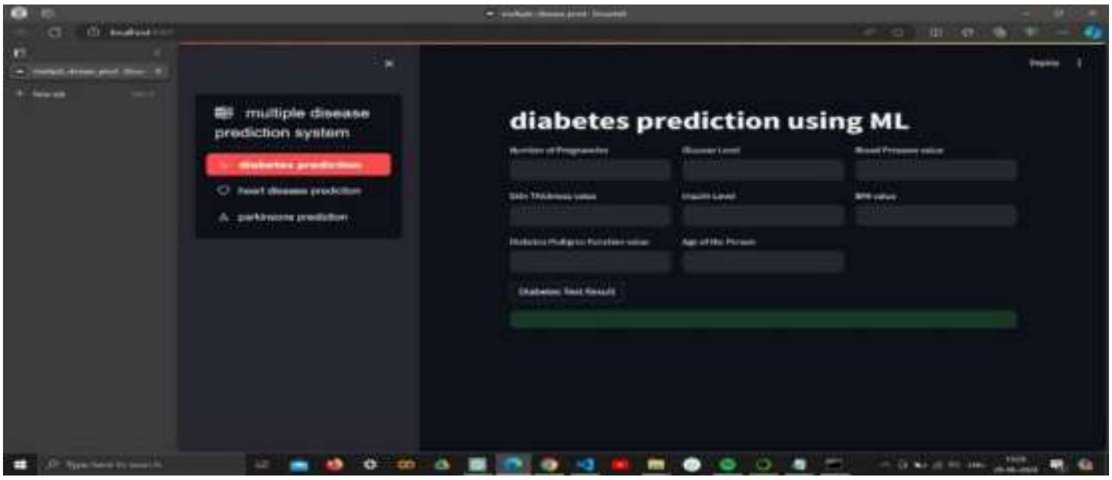

Figure 1: snapshot of the web application

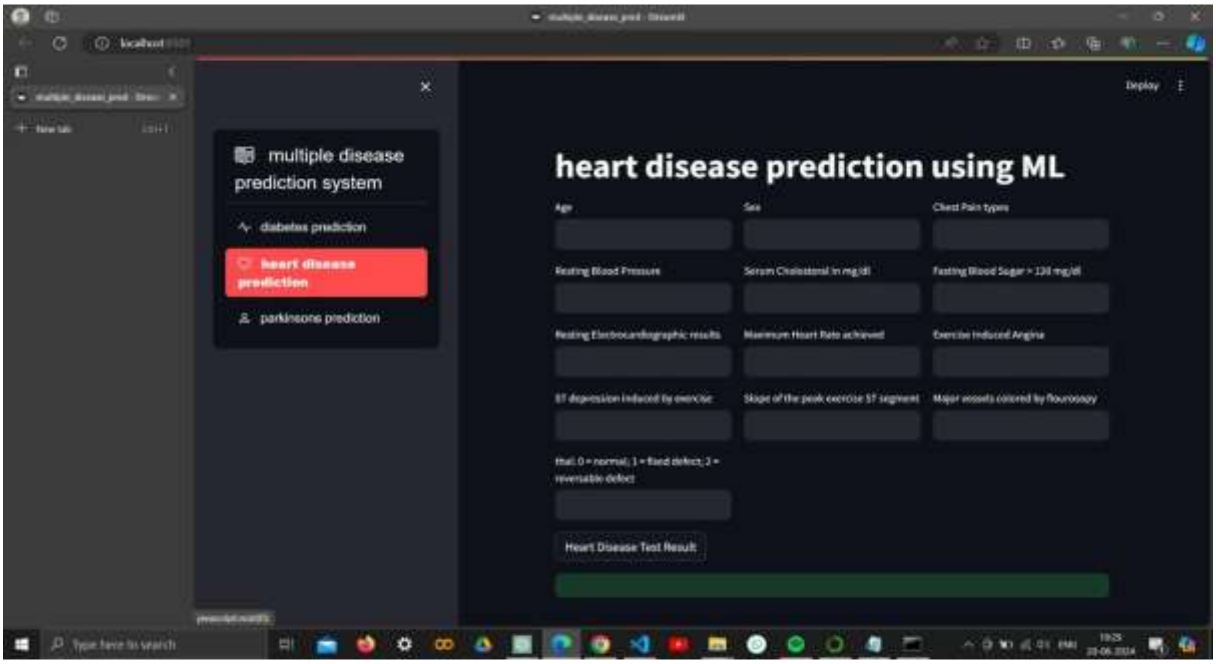

Figure 2: snapshot of the web application

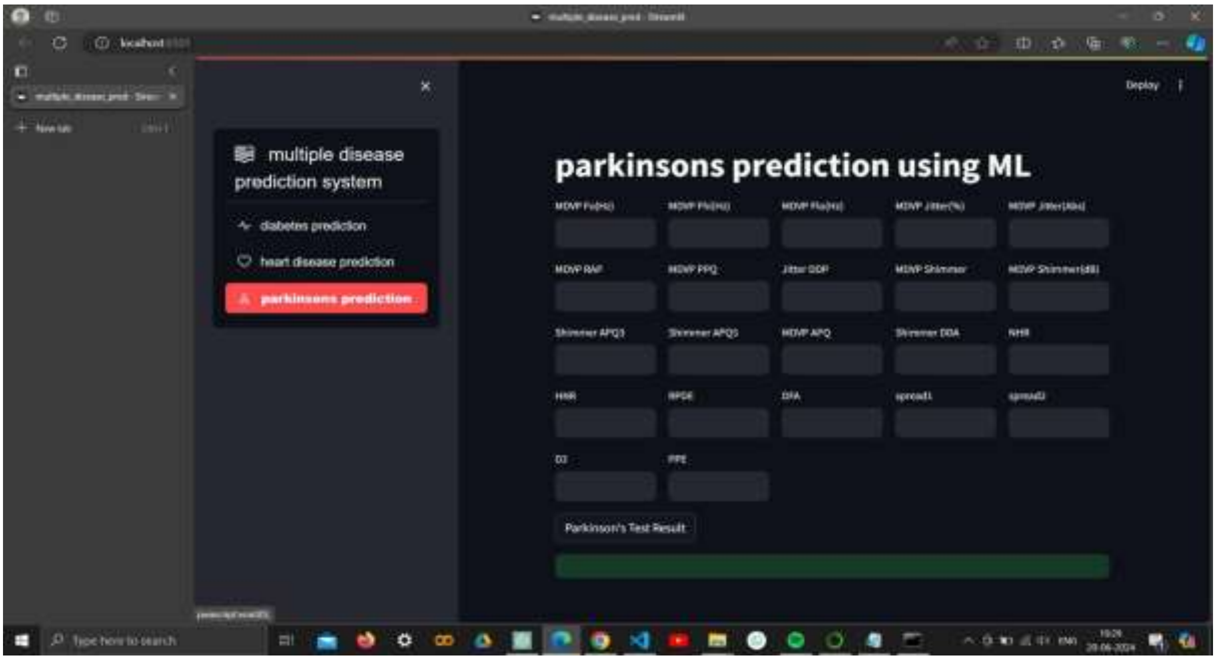

Figure 3: snapshot of the web application

Project Outcome & Industry Relevance:

Project Outcome: Successfully developed a web-based application using the Streamlit framework for predicting Diabetes, Heart Disease, and Parkinson's Disease. Implemented and integrated machine learning models with good prediction accuracy based on real-world healthcare datasets. Designed a simple, interactive, and user-friendly interface using Streamlit framework that allows users to input health data and receive instant results.

Industry Relevance: The application of ML in healthcare is of significant relevance due to the increasing amount of medical data and the need for efficient data analysis to improve patient outcomes. Accurate and timely disease prediction can lead to early interventions, reducing mortality rates and healthcare costs. Both Naive Bayes and SVM have shown promise in various studies, making them suitable candidates for further exploration and application in the medical field.

Future Scope:

The future scope of this project includes:

1. Addition of More Diseases: In the future, we can expand the system by incorporating additional chronic diseases into the existing web application. This would enable users to predict a broader range of diseases and further enhance the system's usefulness in healthcare.
2. Accuracy Improvement: As part of ongoing research and development, we can strive to improve the accuracy of disease predictions. By refining the machine learning algorithms, optimizing feature selection, and incorporating more comprehensive datasets, we can reduce false predictions and increase the overall accuracy of the system. This would ultimately contribute to lowering the mortality rate by enabling timely interventions and treatments.
3. Mobile Application Development: Developing a mobile application version of the multiple disease prediction system would enhance accessibility and convenience for users. It would allow individuals to access the system on their smartphones, providing real-time disease predictions and empowering them to take proactive measures for their health anytime, anywhere.