DISEASE PREDICTION SYSTEM USING MACHINE LEARNING.

Subashini N UG Student, Benita Jenisar M UG Student

St. Mother Theresa Engineering College, Thoothukudi

ABSTRACT

Chronic diseases are depicted as conditions that require proceeding with medical consideration, limit everyday exercises, or both. Chronic diseases, for example, heart disease, diabetes and Parkinson are the main sources of death and disability on the planet. It is a critical challenge to detect these diseases by regular clinical data analysis. Early prediction of these chronic diseases would help in saving multiple lives. The ability to predict multiple diseases simultaneously can significantly improve early diagnosis and treatment, leading to better patient outcomes and reduced healthcare costs. We present an overview of various machine learning models and data sources commonly used for disease prediction. Additionally, we discuss the importance of feature selection, model evaluation, and the integration of multiple data modalities for enhanced disease prediction. Our project highlight the potential of machine learning in multi-disease prediction and its potential impact on public health. This training model takes a sample data and train itself for predicting disease. Chronic Heart Disease will be predicted using Logistic Regression, Diabetes can be predicted using Support Vector Machine (SVM), Parkinson Disease can be predicted using Support Vector Machine (SVM).An integrated model of the listed algorithms is proposed to predict the listed chronic diseases.

Keywords: Heart disease, Diabetes, Parkinson, SVM, Logistic Regression, Early prediction, Machine Learning

I. INTRODUCTION

In recent years, the field of machine learning has witnessed remarkable advancements and applications in various domains, including healthcare. The ability to predict multiple diseases simultaneously using machine learning models has the potential to revolutionize medical diagnostics and improve patient outcomes. This research paper explores the utilization of the Support Vector Machines (SVM) model to predict the presence of three prevalent diseases: heart disease, diabetes, and Parkinson's disease. Cardiovascular diseases, diabetes, and Parkinson's disease are significant public

health concerns that impose a considerable burden on individuals and healthcare systems worldwide. Early detection and accurate diagnosis of these diseases play a vital role in improving patient prognosis, optimizing treatment plans, and reducing healthcare costs. Machine learning, with its ability to analyze vast amounts of data and identify complex patterns, offers promising avenues for multi-disease prediction.

II. EXISTING SYSTEM

Health has always been an important factor in our lives. With time, people are getting more cautious about health. Clinical assessments rely on the experience of the healthcare professional, leading to potential variability in diagnosis, especially for early-stage cases. Traditional methods often miss early detection, delaying intervention and impacting disease management. Predicting the disease has traditionally relied on a combination of medical history, physical examinations, and various tests. These methods primarily consider major risk factors, neglecting individual variations and potential genetic susceptibility. Manual It can show what to use and what not to use from the given datasets and the final results. It predicts probable diseases by mining data diagnosis may take more time to predict the disease and start the treatment to the patient.

III. PROPOSED SYSTEM

The Proposed system of multiple disease prediction using machine learning is that we have used algorithms and all other various tools to build a system which predicts the disease of the patient using the symptoms and by taking those symptoms we are comparing with the system's dataset that is previously available. By taking those datasets and comparing with the patient's disease we will predict the accurate percentage disease of the patient. The classification of those data is done with the help of machine learning algorithms such as Logistic regression and Support Vector Machine (SVM). Then the data goes in the recommendation model, there it shows the risk analysis that is involved in the system and it also provides the probability estimation of the system such that it shows the various probability like how the system behaves when there are n number of predictions are done and it also does the recommendations for the patients from their final result and also from their symptoms like sets such as Diabetes, Parkinson's disease and heart disease.

IV. LOGISTIC REGRESSION

Logistic regression is a statistical method that can be used to predict the probability of a binary outcome (e.g., presence or absence of heart disease) based on a set of independent variables (e.g., age, blood pressure, cholesterol levels). It is a popular choice for heart disease prediction because it is relatively easy to interpret and can provide valuable insights into the relationships between risk factors and heart disease.

- **Data collection:** Collect data on a set of individuals, including information on their heart disease status (e.g., diagnosed with heart disease or not) and a set of potential risk factors (e.g., age, sex, blood pressure, cholesterol levels, smoking status, etc.).
- **Data preprocessing:** Clean and prepare the data for analysis. This may involve handling missing values, outliers, and categorical variables.
- **Model training:** Split the data into a training set and a testing set. Train the logistic regression model on the training set, using the independent variables to predict the heart disease status.
- **Model evaluation:** Evaluate the performance of the model on the testing set. This typically involves calculating metrics such as accuracy, precision, recall, and AUC (area under the ROC curve).
- **Interpretation:** Interpret the coefficients of the logistic regression model to understand how each independent variable is associated with the risk of heart disease.

V. SUPPORT VECTOR MACHINE

Support Vector Machines (SVMs) are powerful machine learning algorithms that can be used for classification tasks, such as predicting whether an individual has diabetes or Parkinson's disease based on certain features. Here, I'll provide a general overview of how you might use SVMs for these two medical prediction tasks.

Diabetes Prediction:

- **Data Preparation:**
  - Obtain a dataset with relevant features such as glucose levels, BMI, age, blood pressure, etc.
  - Split your dataset into training and testing sets.
- **Feature Scaling:**
  - Normalize or standardize the features to ensure that they are on a similar scale.
- **Model Training:**
  - Use the training data to train an SVM classifier.

- Choose an appropriate kernel function (e.g., radial basis function, linear) and tune hyper parameters using techniques like cross-validation.

- **Model Evaluation:**
  - Evaluate the trained model on the testing set using metrics like accuracy, precision, recall, and F1-score.
- **Hyper parameter Tuning:**
  - Experiment with different kernel functions and SVM parameters to find the best combination for your specific dataset.

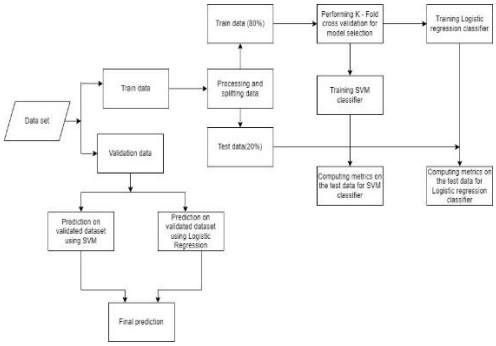
Parkinson's disease prediction:

- **Data Preparation:**
  - Obtain a dataset with relevant features such as speech characteristics, tremor measurements, age, etc.
  - Split your dataset into training and testing sets.
- **Feature Scaling:**
  - Normalize or standardize the features to ensure that they are on a similar scale.
- **Model Training:**
  - Use the training data to train an SVM classifier.

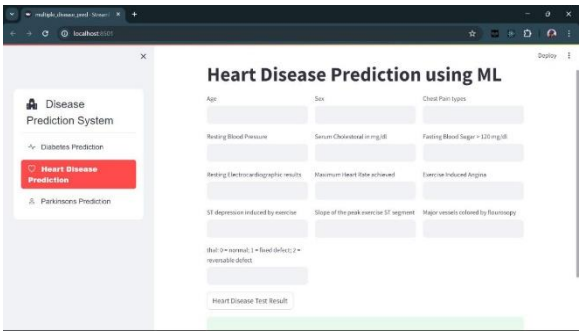
- Choose an appropriate kernel function (e.g., radial basis function, linear) and tune hyper parameters using techniques like cross-validation.

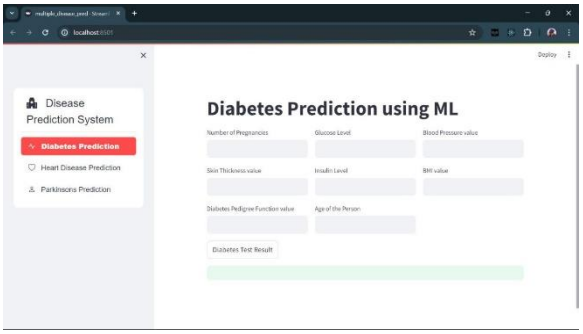
- **Model Evaluation:**
  - Evaluate the trained model on the testing set using metrics like accuracy, precision, recall, and F1-score.
- **Hyper parameter Tuning:**
  - Experiment with different kernel functions and SVM parameters to find the best combination for your specific dataset.

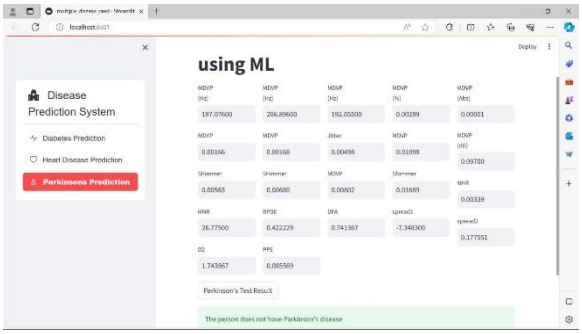
VI. SYSTEM ARCHITECTURE

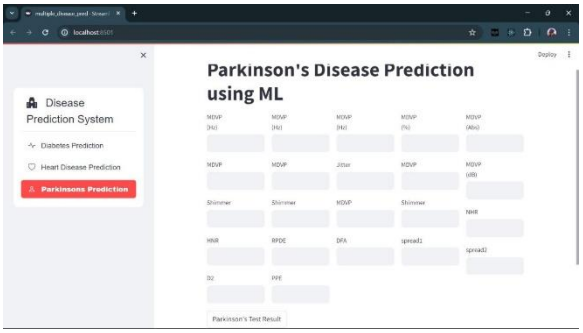


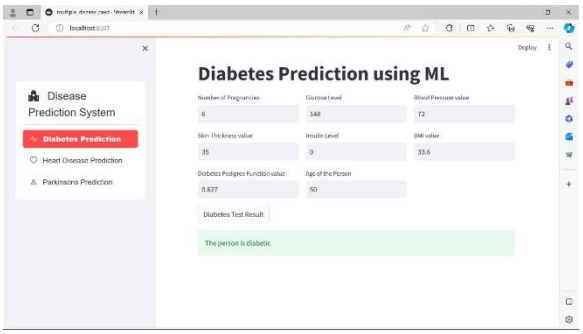
VII. RESULTS

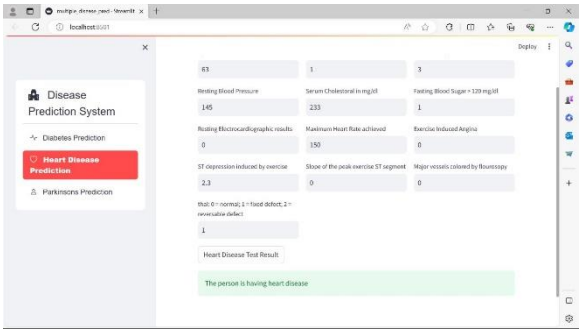












VIII. CONCLUSION

Nowadays Heart, Diabetes diseases are the life-threatening diseases affecting millions of people around the world every year. Hence, early prediction of diseases can benefit patients and healthcare professionals by providing the information they need to minimize death and reduce costs. We can conclude that machine learning can play an important role in our healthcare system. Traditionally, diagnosis of the disease was performed by standard procedures and doctor's intuitions which had limitations and led to costly expenses, but with machine learning models, diagnosis can be done on large datasets cost-effectively. By using our system disease prediction takes less time and patients get treatment at earliest stage.

FUNDING INFORMATION:

This research received no specific grants from any agency in the public, commercial, or not-for-profit sector

COMPETING INTEREST DECLARATION:

Subashini N[1] is the author of this research and student of St. Mother Theresa Engineering College. Benita jenisar M[2] is the co-author of this research and student of St. Mother Theresa Engineering College.

IX. REFERENCES:

- Rinkal Keniya et al., *Disease prediction from various symptoms using machine learning.*, 2020.
- Rudra A. Godse et al., "Multiple Disease Prediction Using Different Machine Learning Algorithms Comparatively", *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 8, no. 12, pp. 50-52, 2019.
- Rayan Alanazi, "Identification and prediction of chronic diseases using ma chine learning approach", *Journal of Healthcare Engineering*, vol. 2022, 2022.
- K. Arumugam et al., "Multiple disease prediction using Machine learning algorithms", *Materials Today: Proceedings*, 2021.
- Indukuri Mohit et al., "An Approach to detect multiple diseases using machine learning algorithm", *Journal of Physics: Conference Series*, vol. 2089, no. 1, 2021.
- Aishwarya Mujumdar and V. Vaidehi, "Diabetes prediction using machine learning algorithms", *Procedia Computer Science*, vol. 165, pp. 292-299, 2019.