Group 49: Parkinson’s Disease Detection

- Vats Pratap Singh

Contents -

What is Parkinson’s Disease?

Parkinson Disease is a brain neurological disorder. It leads to shaking of the body, hands and provides stiffness to the body. No proper cure or treatment is available yet at the advanced stage. Treatment is possible only when done at the early or onset of the disease. These will not only reduce the cost of the disease but will also possibly save a life.

Objective

Parkinson's disease is a progressive neurodegenerative disorder that affects millions of individuals worldwide. Early diagnosis of Parkinson's disease is crucial for timely intervention and effective management of symptoms. However, the early stages of Parkinson's disease can be challenging to diagnose accurately, often leading to delayed treatment and progression of the condition.

The problem addressed in this project is the development of a machine learning-based solution using SVM algorithm to predict Parkinson's disease based on patient data. By utilizing features such as voice recordings, clinical scores, and biomedical data, the aim is to create a predictive model that can distinguish between individuals with Parkinson's disease and those without. This model will help in:

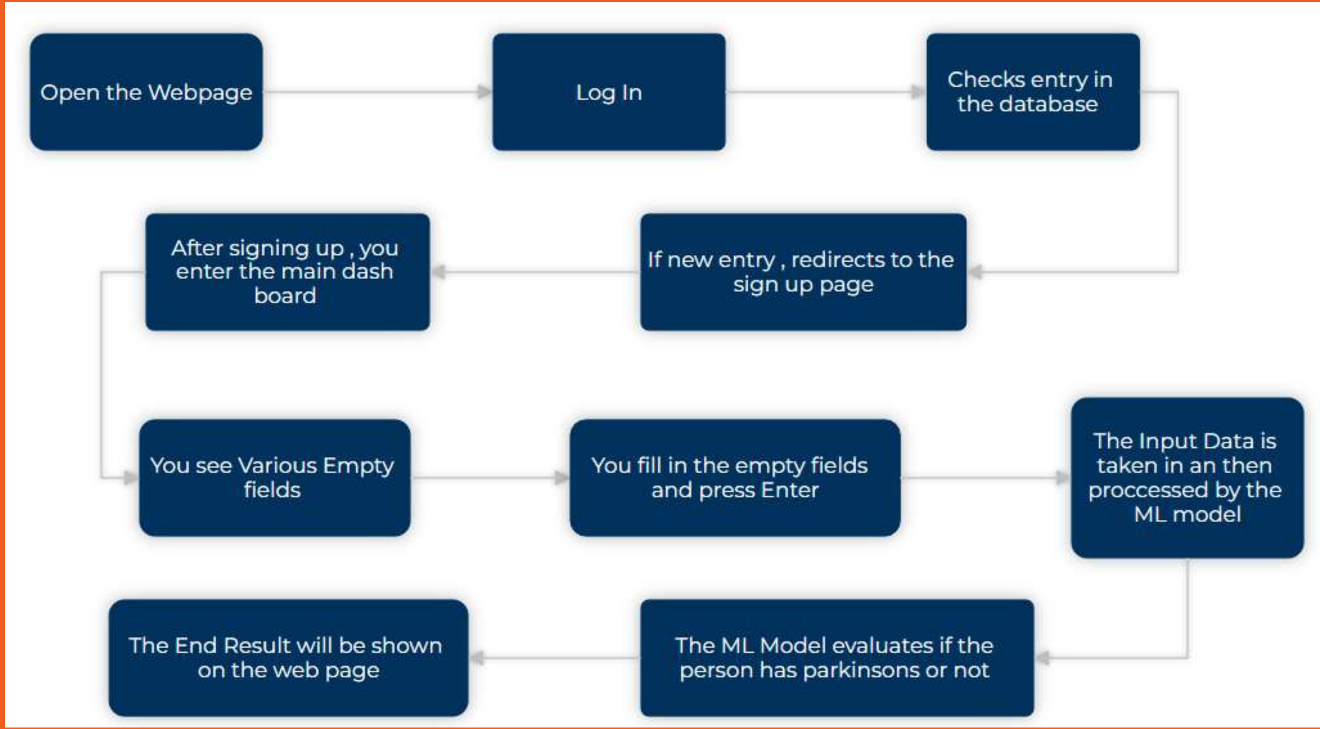
- **Early Detection:** Identifying individuals at risk of developing Parkinson's disease at an early stage.
- **Precision Medicine:** Tailoring treatment plans based on individual risk factors and disease progression.
- **Improving Patient Outcomes:** Enabling timely interventions and therapies to improve the quality of life for individuals with Parkinson's disease.

Problem Statement

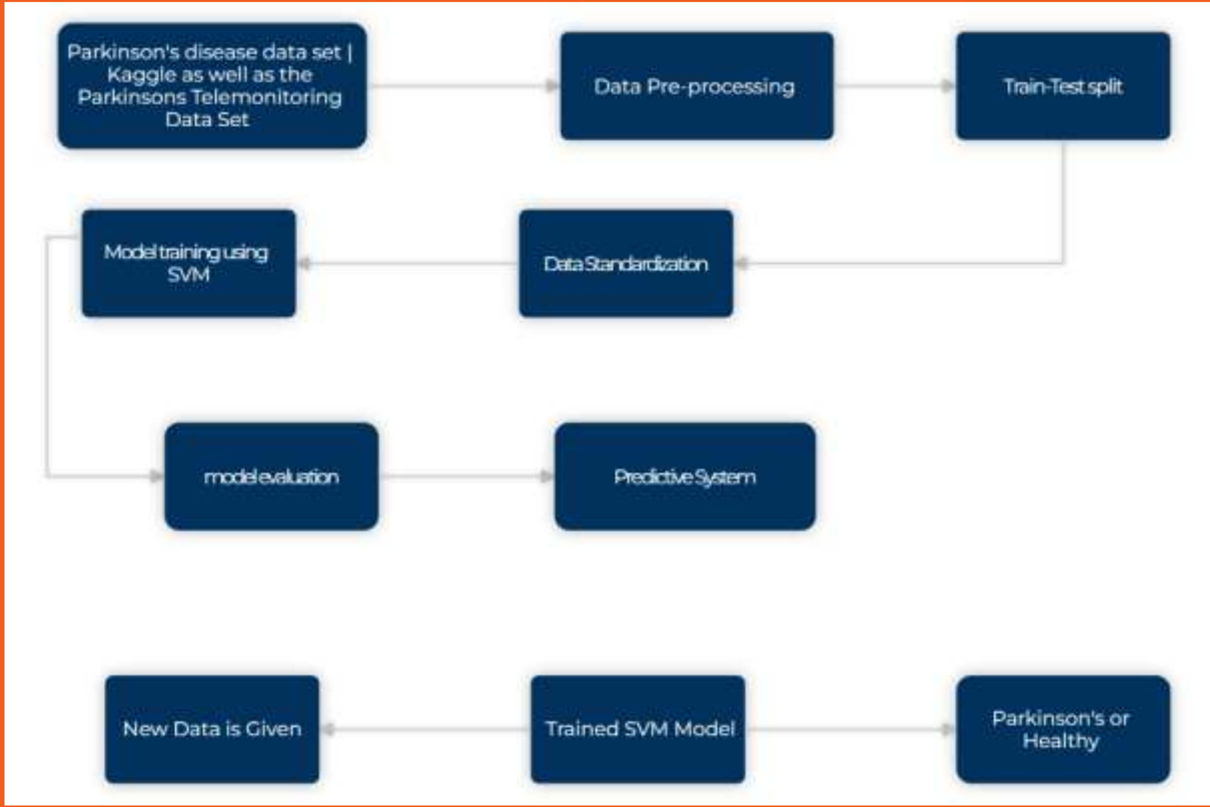
The aim of this Parkinson's disease prediction project is to develop a reliable and accurate machine learning model that can assist in the early diagnosis and prediction of Parkinson's disease based on specific features extracted from patient data. The primary goal is to create a tool that can aid healthcare professionals in identifying individuals who are at risk of developing Parkinson's disease, potentially enabling early intervention and treatment strategies.

WorkFlow

For the End Result :



For The Machine Learning Model :



Proposed Work & Methodology

1. Collection of Parkinson's Data

- The first step is to gather a dataset that contains information related to Parkinson's disease. This dataset may include features such as patient age, gender, medical history, symptoms, and various clinical measurements.

2. Data Preprocessing

- This step involves cleaning the dataset to handle missing values, outliers, and inconsistencies. It also includes standardizing or normalizing the data to ensure all features are on a similar scale.
- Features may be selected or engineered to improve the model's performance.

3. Feature Extraction Techniques:

- MinMax Scaling (MinMaxScaler):
  - MinMax scaling is a technique used to scale numeric features to a specific range, typically between 0 and 1. It ensures that all features have the same scale, preventing some features from dominating due to their larger magnitude.
  - The MinMaxScaler from sklearn.preprocessing library is applied to scale the feature variables.
- Standardization (StandardScaler):
  - Standardization is a technique used to transform the data such that it has a mean of 0 and a standard deviation of 1. It centers the data around zero and scales it based on the standard deviation.
  - The StandardScaler from sklearn.preprocessing library is used to standardize the feature variables.

4. Train-Test Split

- The dataset is then divided into two subsets: a training set and a testing set. The training set is used to train the SVM model, while the testing set is used to evaluate its performance.

5. Trained Algorithm

- Then after using Support Vector Machine (SVM) Algorithm in which when entering new data we can easily classify more accurately if a person has parkinsons or is healthy

Novelty

The provided project focuses on predicting Parkinson's disease based on various voice features. While the prediction of Parkinson's disease isn't novel in itself, the project's novelty lies in several aspects:

- **Voice-Based Diagnosis:** Parkinson's disease prediction using voice features is an emerging field. Analyzing voice characteristics to detect neurological disorders offers a non-invasive and cost-effective diagnostic approach.
- **Feature Engineering:** The project involves extracting relevant features from voice recordings. These features may include jitter, shimmer, and other acoustic parameters, which are indicative of Parkinson's disease. Selecting and engineering these features for optimal performance is a crucial aspect of the project.

- **Model Evaluation and Interpretation:** The project evaluates models not only based on traditional accuracy metrics but also on precision, recall, F1-score, and other performance indicators. Additionally, visualizations such as confusion matrices and precision-recall curves aid in model interpretation and selection.
- **Scalability and Accessibility:** By leveraging tools like Colab (Google Colaboratory), the project promotes accessibility and scalability. It allows researchers and practitioners to replicate, extend, and collaborate on the project with ease.
- **Potential for Early Diagnosis:** Successful implementation of the predictive model can contribute to early diagnosis and intervention for Parkinson's disease. Early detection enables timely medical interventions, potentially improving patient outcomes and quality of life.

Overall, the project's novelty lies in its interdisciplinary approach, combining elements of biomedical research, signal processing, machine learning, and healthcare technology to address a critical healthcare challenge.

Real Time Usage

- The Parkinson's disease prediction system developed in the project offers real-time usage by providing:
- Early Detection and Monitoring: Enables clinicians to detect Parkinson's disease early and monitor its progression by analyzing voice samples during routine check-ups or telehealth sessions.
- Remote Monitoring: Facilitates remote monitoring of Parkinson's disease progression, allowing patients to record voice samples for analysis, especially useful for those with mobility issues or in remote areas.
- Home-Based Screening: Allows individuals to self-assess their risk of Parkinson's disease by providing voice samples through user-friendly interfaces, promoting proactive healthcare-seeking behavior.
- Research and Clinical Trials: Supports research studies and clinical trials by assessing therapy effectiveness, monitoring disease progression, and identifying potential biomarkers.
- Population Health Management: Assists health organizations in allocating resources more efficiently by identifying regions or demographic groups with higher prevalence rates of Parkinson's disease, supporting early diagnosis and intervention.

Support Vector Machine (SVM) Algorithm

It is a supervised machine algorithm. Image classification and handwriting recognition are where the support vector machine comes in handy. It sorts the data in one out of two categories and displays the output with the margin between the two as far as possible.

Datasets

- Parkinson’s Data Set | Kaggle
  - This dataset is composed of a range of biomedical voice measurements from 196 people, 148 with Parkinson's disease (PD). Each column in the table is a particular voice measure, and each row corresponds to one of 196 voice recordings from these individuals ("name" column). The main aim of the data is to discriminate healthy people from those with PD, according to the "status" column which is set to 0 for healthy and 1 for PD.

- Parkinsons Telemonitoring Data Set
  - This dataset is composed of a range of biomedical voice measurements from 42 people with early-stage Parkinson's disease recruited to a six-month trial of a telemonitoring device for remote symptom progression monitoring. The recordings were automatically captured in the patient's homes. Each row corresponds to one of 5,875 voice recording from these individuals.

- **Matrix column entries (attributes):**
  - **Jitter(Abs), Jitter:RAP, Jitter:PPQ5, Jitter:DDP** - Several measures of variation in fundamental frequency
  - **Shimmer,Shimmer(dB),Shimmer:APQ3,Shimmer:APQ5,Shimmer:APQ1,Shimmer:DDA** - Several measures of variation in amplitude
  - **NHR, HNR** - Two measures of the ratio of noise to tonal components in the voice
  - **status** - The health status of the subject (one) - Parkinson's, (zero) - healthy
  - **RPDE** - nonlinear dynamical complexity measure
  - **DFA** - Signal fractal scaling exponent
  - **PPE** - nonlinear measures of fundamental frequency variation

The Predictive System

- **SVM Model Initialization:**
  - The SVM model is initialized with a linear kernel and probability estimates enabled.
- **Training:**
  - The SVM model is trained using the fit function on the training data (X_train, y_train).
- **Standardization:**
  - The StandardScaler is fitted to the training data to standardize the input features.
  - The input data (input_data) is transformed (scaled) using the trained scaler.

- **Prediction:**
    - The standardized input data (std_data) is used to make predictions using the SVM model.
    - The predicted value (prediction) is then printed.
- **Output Interpretation:**
    - If the predicted value is 0, it means the model predicts the person does not have Parkinson's disease.
    - If the predicted value is 1, it means the model predicts the person has Parkinson's disease.

This project essentially demonstrates how to use a trained SVM model to predict whether a person has Parkinson's disease based on the input features provided. It involves initializing the model, training it, standardizing input data, making predictions, and interpreting the prediction results.

Result

1. Accuracy

- Accuracy is one of the most straightforward metrics. It measures the overall correctness of the model in terms of correctly classified instances out of all instances.
- An accuracy of 1.0 (100%) means all predictions were correct.
- Higher accuracy generally indicates better model performance, especially when classes are balanced.

2. Precision

- Precision focuses on the number of correctly predicted positive cases out of all cases predicted as positive. It's about the "purity" of the positive predictions.
- High precision means the model is good at avoiding false positives.
- It's useful when the cost of false positives is high.

3. Recall (Sensitivity or True Positive Rate)

- Recall measures the ability of the model to correctly identify positive instances from all actual positives. It's about the "completeness" of the positive predictions.
- High recall indicates that the model finds most of the positive cases.
- Useful when the cost of false negatives is high.

4. F1-Score

- The F1-score is the harmonic mean of precision and recall. It provides a balance between precision and recall, especially when classes are imbalanced. It's a single metric that combines both precision and recall into one number.
- The F1-score reaches its best value at 1 and worst at 0.
- It's a useful metric when you want to seek a balance between precision and recall.

Now for comparison we also used the K-Nearest Neighbors (KNN) algorithm

K-Nearest Neighbors (KNN)-

The K-Nearest Neighbors (KNN) algorithm is a type of supervised machine learning algorithm used for both classification and regression tasks. It is considered one of the simplest machine learning algorithms and falls under the category of instance-based learning or lazy learning.

❖ Parkinson's Data Set | Kaggle | USING SVM

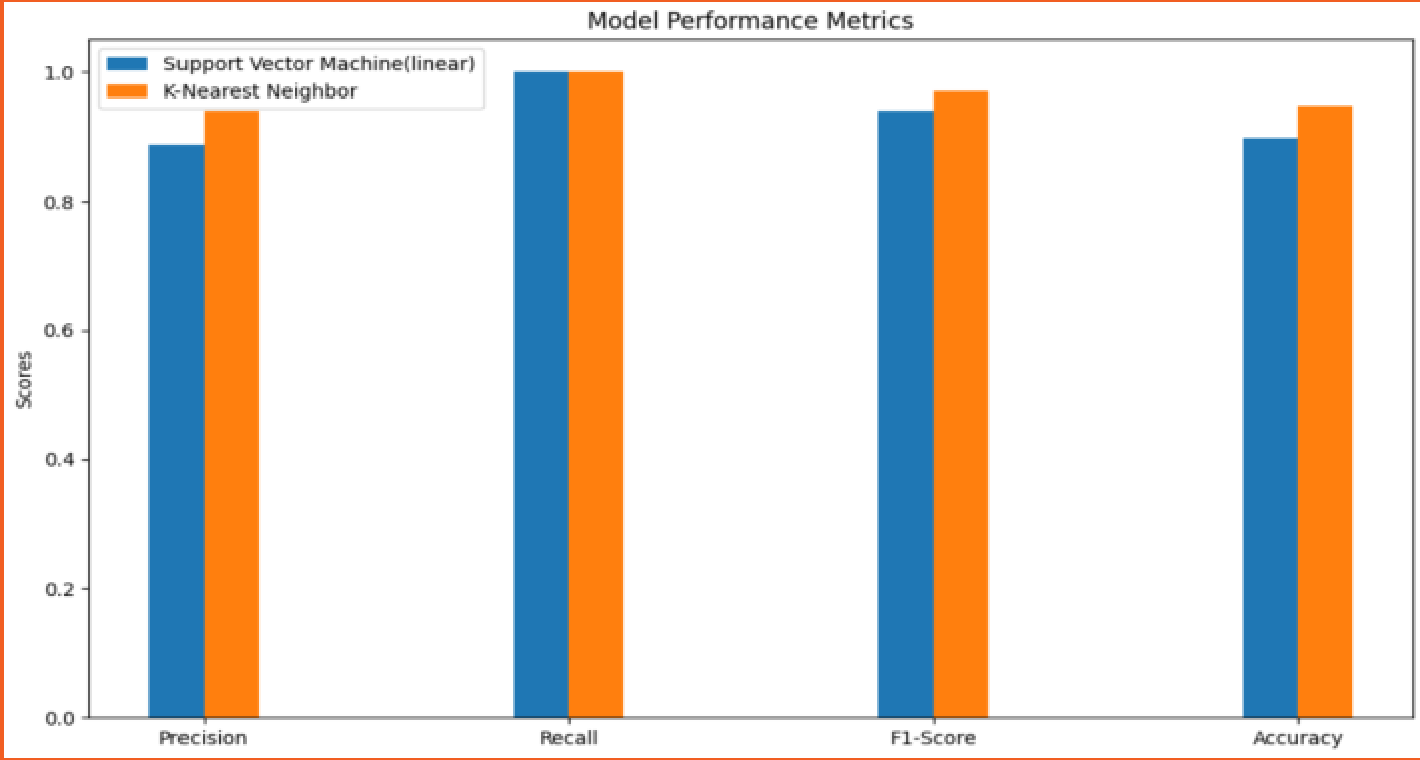
- ➢ Training Accuracy Score : 83.33
- ➢ Cross Validation Score : 76.79
- ➢ Testing Accuracy Score : 87.18
- ➢ Precision Score is : 86.49
- ➢ Recall Score is : 100.0
- ➢ F1-Score Score is : 92.75

- **Parkinson's Data Set | Kaggle | USING KNN**
  - Training Accuracy Score : 95.51
  - Cross Validation Score : 89.04
  - Testing Accuracy Score : 87.18
  - Precision Score is : 90.91
  - Recall Score is : 93.75
  - F1-Score Score is : 92.31

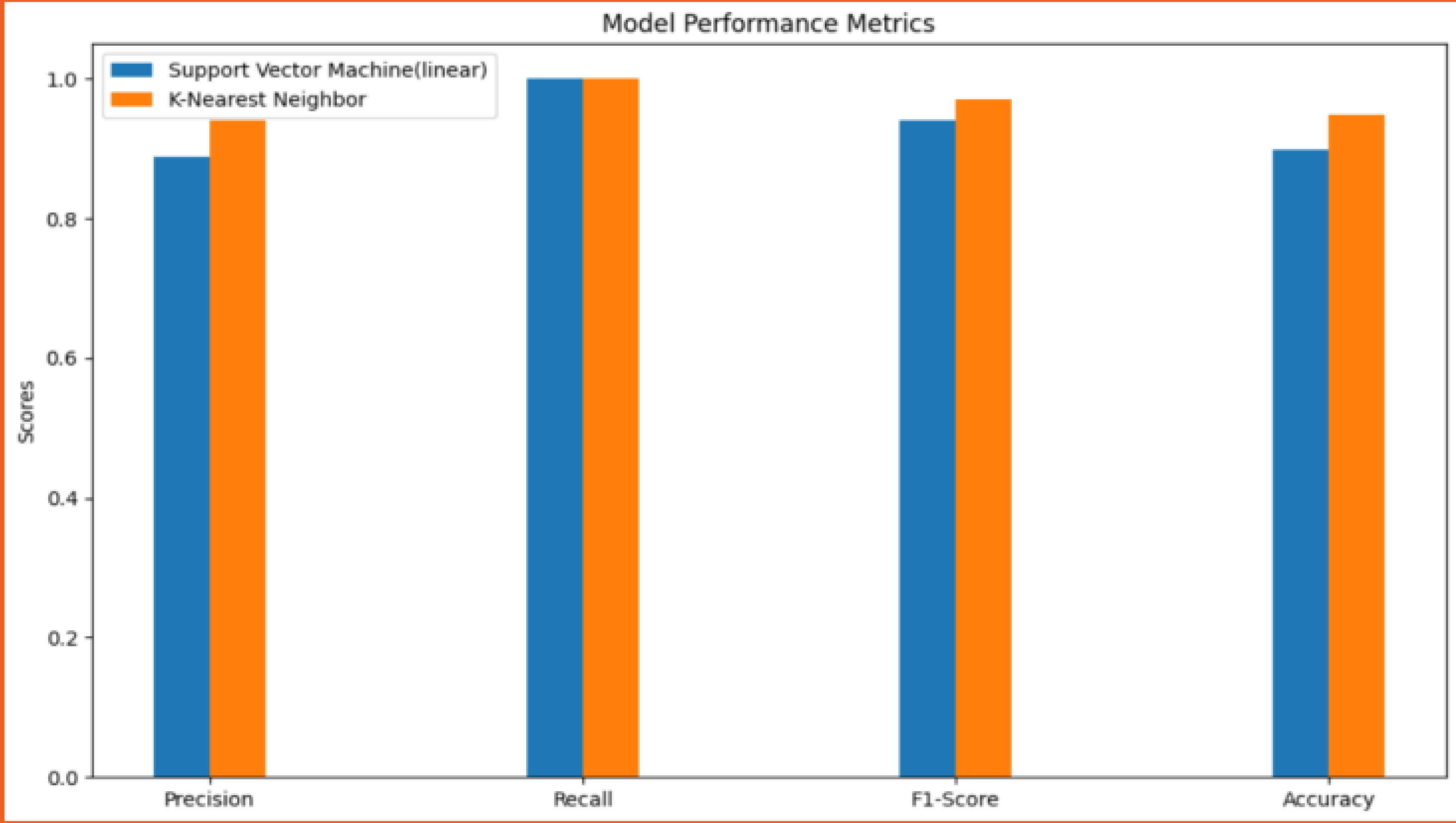
- **Parkinsons Telemonitoring Data Set | USING SVM**
  - Training Accuracy Score : 71.09
  - Cross Validation Score : 70.98
  - Testing Accuracy Score : 70.72
  - Precision Score is : 97.62
  - Recall Score is : 10.68
  - F1-Score Score is : 19.25

- **Parkinsons Telemonitoring Data Set | USING KNN**
  - Training Accuracy Score : 89.26
  - Cross Validation Score : 83.0
  - Testing Accuracy Score : 81.87
  - Precision Score is : 77.85
  - Recall Score is : 62.24
  - F1-Score Score is : 69.18

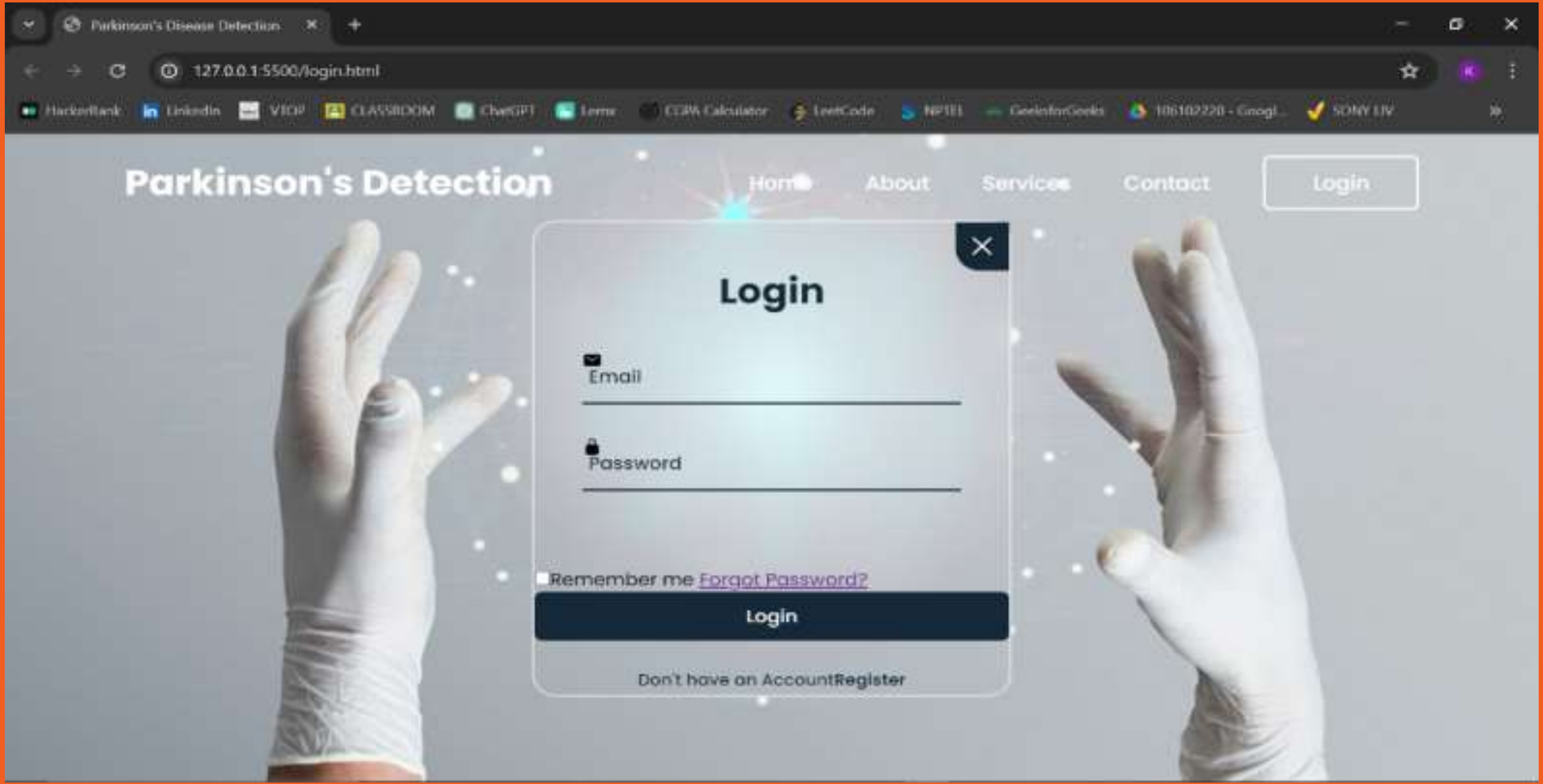
❖ Parkinson’s Data Set | Kaggle



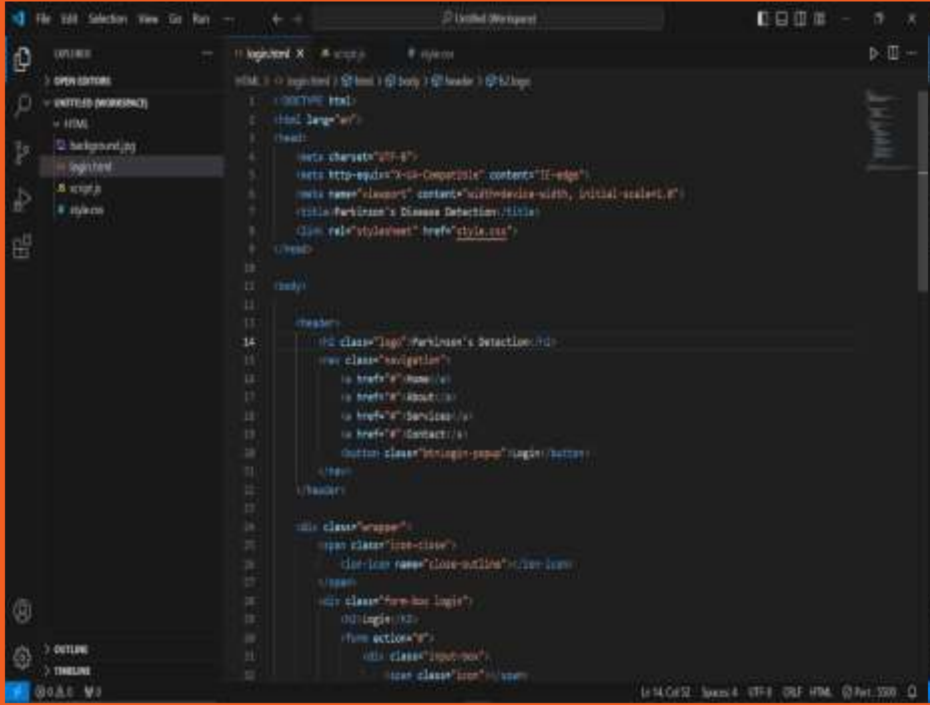
❖ Parkinsons Telemonitoring Data Set

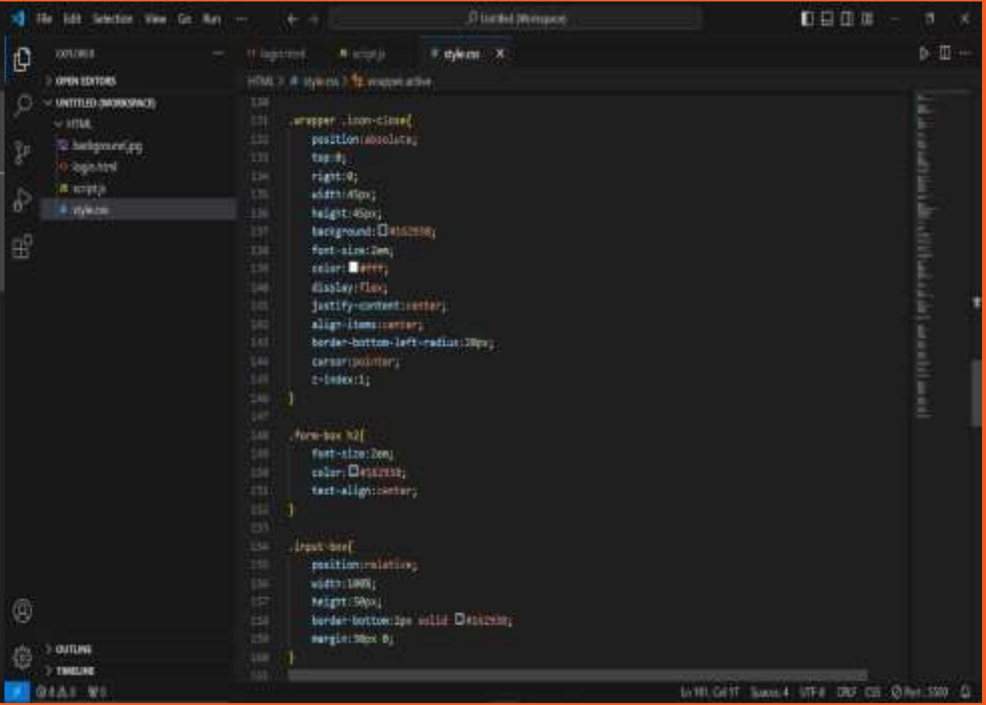


Website


Parkinson's Disease Detection
127.0.0.1:5500/login.html
Parkinson's Detection
Home
About
Services
Contact
Login
Login
Email
Password
Remember me Forgot Password?
Login
Don't have an AccountRegister

Code for website





THANK YOU