PARKINSON'S DISEASE PREDICTION USING MACHINE LEARNING

C.Surekha ,A.Sathwika Goud, Ellumvilayil Shajimon Shaliya, Ch.Padmavathi,P. Manoj Yadav

Computer Science and Engineering (Internet of Things)

Hyderabad Institute of Technology And Management

Medchal,India

ABSTRACT

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by motor and non- motor symptoms, posing significant challenges for early detection and accurate diagnosis. Machine learning (ML) techniques have emerged as powerful tools for predicting PD onset, enabling timely interventions and personalized treatment strategies. The use of machine learning enables objective assessment, personalized medicine, and efficient analysis of large-scale biomedical data. This research contributes to advancing Parkinson's disease diagnosis, management, and research, ultimately improving patient outcomes and healthcare efficiency.

Keywords: Parkinson's disease, Machine learning, biomedical data.

1. INTRODUCTION

Parkinson's disease is a chronic and progressive neurodegenerative disorder that primarily affects the motor system of the human body. First described by Dr. James Parkinson in 1817, this disease is characterized by a range of motor symptoms, including tremors, muscle rigidity, bradykinesia (slowness of movement), and postural instability. These symptoms arise due to the gradual loss of dopamine-producing neurons in a region of the brain called the substantia nigra. As dopamine levels decrease, the brain's ability to control and coordinate muscle movements deteriorates, leading to the hallmark motor disturbances associated with the disease.

Despite extensive research, Parkinson's disease remains incurable. The exact cause of the disease is still not fully understood, though it is believed to result from a combination of genetic and environmental factors. Current treatments focus primarily on managing symptoms and improving the quality of life for patients. Medications, physical therapy, and surgical interventions can help alleviate symptoms, but they do not halt the progression of the disease.

Early detection of Parkinson's disease is crucial as it allows for timely intervention, which can significantly improve patient outcomes. However, diagnosing the disease in its early stages is challenging due to the subtlety and variability of initial symptoms. Traditional diagnostic methods rely heavily on clinical evaluations and subjective assessments, which can lead to delays in diagnosis and treatment.

In recent years, the advent of machine learning (ML) and data analytics has opened new avenues for the early detection and prediction of Parkinson's disease. Machine learning algorithms can analyze vast amounts of data, identify patterns, and make predictions with high accuracy. By leveraging various data sources, such as speech recordings, handwriting samples, and movement data, these algorithms can detect subtle changes that may indicate the onset of Parkinson's disease long before noticeable symptoms appear.

The primary objective of this project is to develop a machine learning model capable of accurately predicting the onset of Parkinson's disease based on patient data. By integrating and analyzing diverse data types, the model aims to identify early indicators of the disease and provide a reliable tool for early diagnosis. This project will explore different machine learning techniques, evaluate their performance, and determine the most effective approach for predicting Parkinson's disease.

Through this project, we aim to contribute to the growing field of medical data science and offer a potential solution for the early detection of Parkinson's disease. By improving early diagnosis, we hope to enhance patient outcomes, enable more effective treatment planning, and ultimately improve the quality of life for those affected by this debilitating condition.

2. MOTIVATION

Parkinson's disease (PD) is a chronic and progressive movement disorder, meaning that symptoms continue and worsen over time. Understanding the motivation behind studying Parkinson's disease can help in grasping the broader implications and potential impact of research and advancementsinthis field.

Motivation for Studying Parkinson's Disease

1. **Improving Patient Outcomes:**

Early Diagnosis: Early detection can significantly improve the quality of life for patients by allowing for timely interventions and treatments.

Better Treatments: Research can lead to the development of new therapies that are more effective in managing symptoms and slowing disease progression.

2. **Understanding Disease Mechanisms:**

Pathophysiology: Studying the underlying mechanisms of PD can help in identifying potential targets for new treatments.

Genetic Factors: Understanding genetic predispositions can lead to personalized medicineapproaches.

3. **Technological Advancements:**

Medical Imaging: Advancements in imaging techniques can improve the accuracy of PD diagnosis and monitoring.

Machine Learning: Utilizing machine learning algorithms to predict PD can lead to innovative diagnostic tools.

4. **Social and Economic Impact:**

Reducing Healthcare Costs: Early and accurate diagnosis can reduce the long-term healthcare costs associated with managing PD.

Improving Quality of Life: Enhancing patient care and developing better treatment options can significantly improve the quality of life for PD patients and their families.

5. **Contributing to Scientific Knowledge:**

Interdisciplinary Research: Studying PD involves various fields such as neurology, genetics, bioinformatics, and more, contributing to a broader scientific understanding.
Collaborative Efforts: Research in PD often involves collaborations between institutions, fostering a global effort in tackling the disease.

6. **Personal Motivations:**

Family and Personal Connections: Many researchers are motivated by personal connections to individuals affected by PD.
Passion for Neurology: A deep interest in understanding and treating neurological disorders can drive researchers to focus on PD.

By addressing these motivations, research in Parkinson's disease can lead to significant advancements in how the disease is diagnosed, treated, and managed, ultimately improving the livesof those affected.

3. LITERATURE SURVEY

A literature survey typically involves summarizing and synthesizing existing research on a particular topic. Here's a structured approach to cover the key aspects of Parkinson's disease:

1. **Introduction to Parkinson's Disease**

- Definition and Symptoms: Parkinson's disease (PD) is a neurodegenerative disorder characterized by motor symptoms like tremors, rigidity, bradykinesia, and postural instability, as well as non-motor symptoms such as cognitive impairment, depression, and autonomic dysfunction.
- Epidemiology: Discuss the prevalence and incidence rates globally. For instance, PD affects approximately 1% of the population over 60 years old.

2. **Pathophysiology**

- Neuropathological Features: The degeneration of dopamine-producing neurons in the substantia nigra part of the brain.
- Molecular Mechanisms: Involvement of α-synuclein aggregation, mitochondrial dysfunction, oxidative stress, and neuroinflammation.
- Genetic Factors: Genes such as SNCA, LRRK2, PARK2, and GBA associated with PD.

3. **Diagnosis**

- Clinical Diagnosis: Based on medical history, symptoms, and neurological examinations. Imaging Techniques: MRI, PET, and DaTscan to support diagnosis.
- Biomarkers: Research on biomarkers in blood, CSF, and imaging for early detection and progression monitoring.

4. **Treatment Strategies**

Pharmacological Treatments:

- Dopaminergic Treatments: Levodopa, dopamine agonists, MAO-B inhibitors.

- Non-dopaminergic Treatments: Anticholinergics, NMDA receptor antagonists.

Non-Pharmacological Treatments:

- Surgical Interventions: Deep brain stimulation (DBS).
- Rehabilitation: Physical therapy, occupational therapy, speech therapy.
- Emerging Treatments: Gene therapy, stem cell therapy, and neuroprotective strategies.

5. **Management of Non-Motor Symptoms**

Cognitive impairment, mood disorders, sleep disturbances, autonomic dysfunction. Integrated care approaches to manage these symptoms.

6. **Research and Future Directions**

- Advances in Understanding PD: Recent findings in the molecular and genetic basis of PD.
- Innovative Therapies: New drug developments, personalized medicine, and advancements in gene and cell therapies.
- Clinical Trials: Summary of ongoing and recently completed clinical trials.

7. **Challenges and Opportunities**

- Current Challenges: Heterogeneity in disease presentation, lack of biomarkers, and side effects of long-term treatment.
- Future Opportunities: Improved diagnostic tools, development of disease-modifying therapies, and better understanding of disease mechanisms.

4. EXISTING MODEL

Existing models for predicting and diagnosing Parkinson's disease (PD) primarily utilize various machine learning algorithms and deep learning techniques. These models are designed to analyze patterns in data to identify potential indicators of PD. Here are some commonly used models and approaches:

Machine Learning Models

✓ **Support Vector Machines (SVM):**

Application: SVMs are used to classify PD patients based on features extracted from speech, gait, or other biomedical data.
Advantages: High accuracy with a good separation margin, effective in high-dimensional spaces.

✓ **Random Forests (RF):**

Application: RFs are used for both classification and regression tasks in PD prediction by analyzing features from medical records or wearable sensors.
Advantages: Robust to overfitting, can handle large datasets and missing values effectively.

✓ **K-Nearest Neighbors (KNN):**

Application: KNN is used for classifying PD patients based on symptom severity or other clinicalfeatures.

Advantages: Simple to implement, effective for small to medium-sized datasets.

- ✓ **Logistic Regression (LR):**

Application: LR models are used for binary classification tasks to predict the presence or absence of PD based on clinical data.
Advantages: Easy to interpret, effective for binary classification problems.

5. PROPOSAL METHOD

Parkinson's disease (PD) is a neurodegenerative disorder characterized by motor symptoms such as tremors, bradykinesia, and dyskinesia. Early diagnosis is crucial for effective management and treatment. This proposal outlines a method to predict Parkinson's disease using machine learning algorithms based on patient symptoms and imaging data.

Objectives :

Develop a machine learning model to predict Parkinson's disease based on symptoms and imaging data.
Achieve high accuracy, sensitivity, and specificity in PD prediction.

Provide a user-friendly interface for clinicians to input patient data and receive predictions.

Data Collection :

- Data Sources

Clinical Data: Collect data on patient symptoms, including bradykinesia, dyskinesia, tremors, andother relevant clinical features.
Imaging Data: Obtain MRI, DAT scans, or other relevant imaging data that show brain abnormalitiesassociated with PD.
Public Datasets: Utilize publicly available datasets such as the Parkinson's Progression MarkersInitiative (PPMI) and other similar repositories.

- Data Preprocessing

Cleaning: Remove any missing or inconsistent data.

Normalization: Normalize clinical and imaging data to ensure consistent scales. Feature Extraction: Extract relevant features from both clinical and imaging data.

Methodology

- Feature Selection

Statistical Analysis: Perform statistical tests to identify the most significant features correlatedwithPD.
Dimensionality Reduction: Use techniques like Principal Component Analysis (PCA) to reduce the feature space while retaining important information.

- Model Selection

Machine Learning Models: Evaluate multiple models such as Support Vector Machines (SVM),Random Forests (RF), and Logistic Regression (LR).
Deep Learning Models: Implement Convolutional Neural Networks (CNN) for imaging data analysisand Recurrent Neural Networks (RNN) for time-series symptom data.

- Training and Validation

Data Split: Split the data into training, validation, and test sets (e.g., 70% training, 15% validation,15% testing).
Model Training: Train the selected models on the training data.

Hyperparameter Tuning: Optimize model hyperparameters using techniques like Grid Search orRandom Search.
Cross-Validation: Use k-fold cross-validation to ensure the model generalizes well to unseen data.

- Model Evaluation

Metrics: Evaluate the models using accuracy, sensitivity, specificity, precision, and F1- score. Confusion Matrix: Analyze the confusion matrix to understand the model's performance inpredicting true positives, false positives, true negatives, and false negatives.

- Model Deployment

Flask API: Develop a Python Flask API to serve the model predictions.

User Interface: Create a user-friendly web interface for clinicians to input patient data and receivepredictions.
Integration: Ensure the system integrates seamlessly with existing healthcare databases and systems.

Expected Outcomes

A robust machine learning model capable of accurately predicting Parkinson's disease based on clinical and imaging data.
A web-based application that allows clinicians to input patient data and receive predictions with highaccuracy.
Improved early diagnosis of Parkinson's disease, leading to better patient outcomes and more effective treatment plans.

Timeline

- Phase 1 (1-2 months): Data collection and preprocessing.
- Phase 2 (2-3 months): Model development and training.
- Phase 3 (1-2 months): Model evaluation and refinement.
- Phase 4 (1 month): Deployment and user interface development.
- Phase 5 (1 month): Testing, integration, and final adjustments.

6. SOFTWARE REQUIREMENTS

SOFTWARE:

1. Pycharm
2. Jupyter Notebook (anaconda 3)
3. IDLE (Python 3.11.1 64-bit)

7. EXPLANATION OF SOFTWARE

SOFTWARE APPLICATIONS

1. PyCharm

PyCharm is an Integrated Development Environment (IDE) developed by JetBrains specifically for Python. It provides a comprehensive set of tools and features that facilitate the development, debugging, and maintenance of Python applications.

Key Features:

- Code Editor: Provides intelligent code completion, syntax highlighting, and codenavigation.
- Debugging: Powerful debugger with breakpoints, watches, and stepping features.
- Refactoring: Tools for safe and efficient code refactoring.
- Integration: Integrates seamlessly with version control systems (e.g., Git), databases, and webdevelopment frameworks (e.g., Flask, Django).
- Project Management: Helps in organizing project files and dependencies.
- Testing: Supports various testing frameworks like pytest, unittest, and nose.

Usage in the Project:

- Code Development: Writing and maintaining the Python code for the machine learning modelsand the Flask API.
- Debugging: Identifying and fixing bugs in the code.
- Integration: Managing project dependencies and integrating different componentsof theproject.

2. Jupyter Notebook (Anaconda 3)

Jupyter Notebook is an open-source web application that allows you to create and share documentsthat contain live code, equations, visualizations, and narrative text. Anaconda is a distribution of Python and R for scientific computing and data science, which includes Jupyter Notebook and other useful packages.

Key Features:

- Interactive Coding: Allows for real-time code execution and visualization.
- Documentation: Combines code with rich text elements such as paragraphs, equations, andplots.
- Data Visualization: Easy integration with libraries like Matplotlib, Seaborn, and Plotlyfor datavisualization.
- Environment Management: Anaconda simplifies package management and deployment.

Usage in the Project:

- Exploratory Data Analysis (EDA): Analyzing and visualizing the dataset to understand theunderlying patterns and features.
- Model Development: Prototyping machine learning models and experimenting withdifferentalgorithms.
- Visualization: Creating plots and charts to illustrate data insights and model performance.
- Documentation: Documenting the development process, findings, and results.

3. **IDLE (Python 3.11.1 64-bit)**

IDLE (Integrated Development and Learning Environment) is the default IDE that comes withPython. It is a simple and lightweight environment for writing and running Python code.

Key Features:

- Simple Interface: Provides a straightforward interface for writing and executingPythonscripts.
- Interactive Shell: Comes with an interactive Python shell for testing code snippets anddebugging.
- Basic Debugging: Includes a basic debugger with step-through code execution.
- Lightweight: Easy to install and use, making it suitable for small scripts and learningpurposes.

Usage in the Project:

- Quick Testing: Running small code snippets and testing individual functions.
- Debugging: Performing basic debugging tasks.
- Learning and Prototyping: Useful for beginners to learn Python and for quickprototyping ofideas.

Important Note:

Each ofthese components plays a crucial role in the development process:

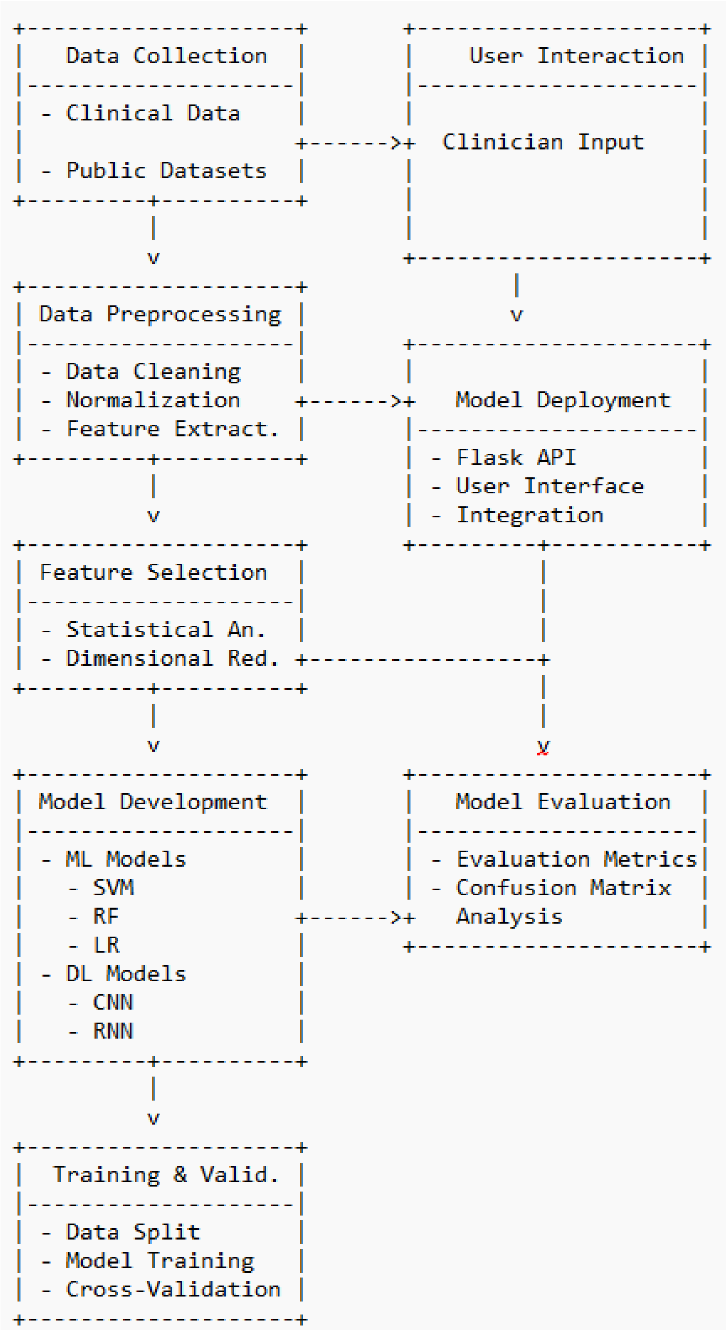
- PyCharm: Ideal for large-scale project development, offering advanced features for coding,debugging, and integration.
- Jupyter Notebook: Excellent for data analysis, visualization, and iterative

development ofmachine learning models.

- IDLE: Useful for quick tests and learning Python, providing a simple environment to executeand debug small scripts.

By leveraging these tools, you can efficiently develop, test, and deploy your machine learningmodel for predicting Parkinson's disease.

BLOCK DIAGRAM

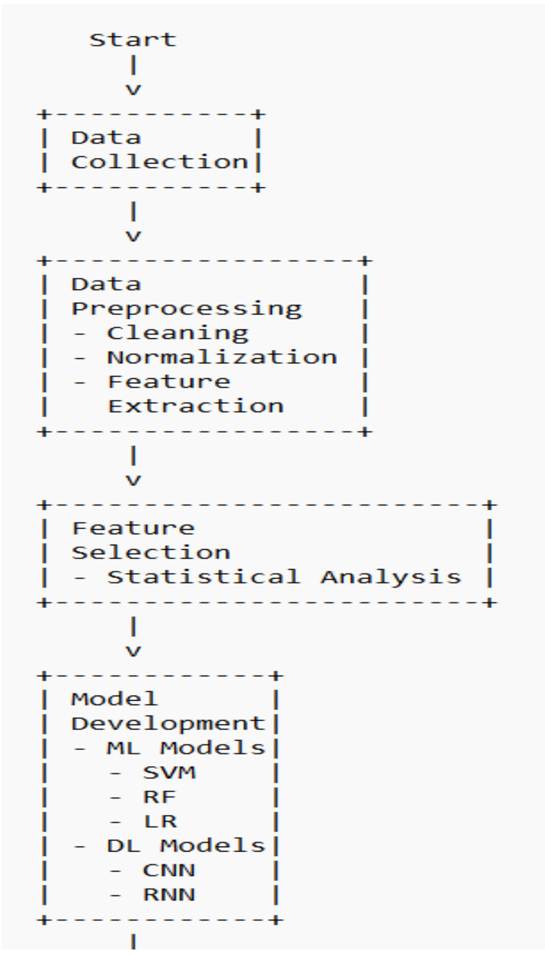


Explanation of the Block Diagram

1. Data Collection: Clinical data and public datasets are collected.
2. Data Preprocessing: The collected data is cleaned, normalized, and features are extracted.
3. Feature Selection: Significant features are selected using statistical analysis and dimensionalityreduction techniques.
4. Model Development: Machine learning and deep learning models are developed usingtheselected features.
5. Training and Validation: The models are trained and validated using training and validationdatasets, and cross-validation.
6. Model Evaluation: The performance of the models is evaluated using various metrics andconfusion matrix analysis.
7. Model Deployment: The best-performing model is deployed using a Flask API, and a userinterface is created for clinicians to input data and receive predictions.
8. User Interaction: Clinicians input patient data, and the system provides predictions about thelikelihood of Parkinson's disease.

This block diagram and its components provide a comprehensive overview of the method for predicting Parkinson's disease using machine learning.

FLOW CHART



```
+--------------------+
| Training &         |
| Validation         |
| - Data Split       |
| - Model Training   |
| - Cross-Validation |
+--------------------+
     |
     v
+------------+
| Model      |
| Evaluation |
| - Metrics  |
| - Confusion|
|   Matrix   |
+------------+
     |
     v
+--------------+
| Model        |
| Deployment   |
| - Flask API  |
| - User       |
|   Interface  |
| - Integration|
+--------------+
     |
     v
+-------------+
| User        |
| Interaction |
| - Clinician |
|   Input     |
| - Prediction|
|   Output    |
+-------------+
     |
     v
   End
```

8. ADVANTAGES

1. Early Diagnosis

- Timely Intervention: Early detection of Parkinson's disease (PD) can lead to timely medical intervention, potentially slowing disease progression and improving the quality of life for patients.
- Proactive Management: Early diagnosis allows for proactive management of symptoms and better planning of long-term care strategies.

2. Improved Accuracy

- Advanced Algorithms: Machine learning models can analyze complex patterns in data that might be missed by traditional diagnostic methods, leading to more accurate predictions.
- Reduction in Misdiagnosis: By using a data-driven approach, the likelihood of misdiagnosis can be reduced, ensuring that patients receive appropriate treatment sooner.

3. Personalized Treatment Plans

- Tailored Therapies: Predictive models can help in creating personalized treatment plans based on individual risk factors and disease progression, leading to more effective treatments.
- Data-Driven Decisions: Clinicians can use model predictions to make informed decisions about treatment options and adjustments.

4. Enhanced Research and Understanding

- Insights into Disease Mechanisms: Analyzing large datasets can provide insights into the underlying mechanisms of Parkinson's disease, contributing to research and development of newtherapies.

- Identification of Risk Factors: Models can help identify potential risk factors and early symptoms associated with Parkinson’s, contributing to better understanding and prevention strategies.

5. Cost-Effectiveness

- Reduced Healthcare Costs: Early and accurate diagnosis can reduce the need for expensive diagnostic tests and treatments, leading to lower healthcare costs.
- Efficient Resource Allocation: Predictive models can help healthcare providers allocate resources more efficiently, focusing on patients who are at higher risk.

6. Patient Empowerment

- Increased Awareness: Providing patients with early predictions can increase their awareness and understanding of their health, empowering them to take proactive measures.

- Better Planning: Early knowledge of the disease can help patients and their families plan for the future, including lifestyle adjustments and long-term care arrangements.

7. Scalability and Accessibility

- Widespread Application: Once developed, the model can be deployed in various healthcare settings, making advanced diagnostic tools accessible to a larger population.
- Remote Diagnosis: The model can be integrated into telemedicine platforms, allowing for remote diagnosis and monitoring, especially beneficial in areas with limited access tospecialized healthcare.

8. Objective Assessment

- Consistency: Machine learning models provide a consistent and objective assessment, reducing variability and subjectivity in diagnosis.
- Data Integration: The ability to integrate various data sources (clinical, imaging, genetic) into a single predictive model enhances the overall assessment accuracy.

9. Continuous Improvement

- Learning from Data: Machine learning models can continuously improve as they are exposed to more data, leading to progressively better performance over time.
- Adaptability: Models can be updated and refined based on new research findings and clinical data, ensuring they remain relevant and accurate.

9. DATA SET INFORMATION

This dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). Each column in the table is a particular voice measure, and each row corresponds one of 195 voice recording from these individuals ("name" column). The main aim of the data is to discriminate healthy people from those with PD, according to "status".

Attribute Information:

This dataset is called the "Parkinson's Disease Classification" dataset and consists of measurements from voice recordings of 195 patients with Parkinson's disease. The dataset has 24 features, which are described below:

- Name: The name of the patient as a string.
- MDVP:Fo(Hz): The average fundamental frequency measured in Hz.
- MDVP:Fhi(Hz): The highest fundamental frequency measured in Hz.

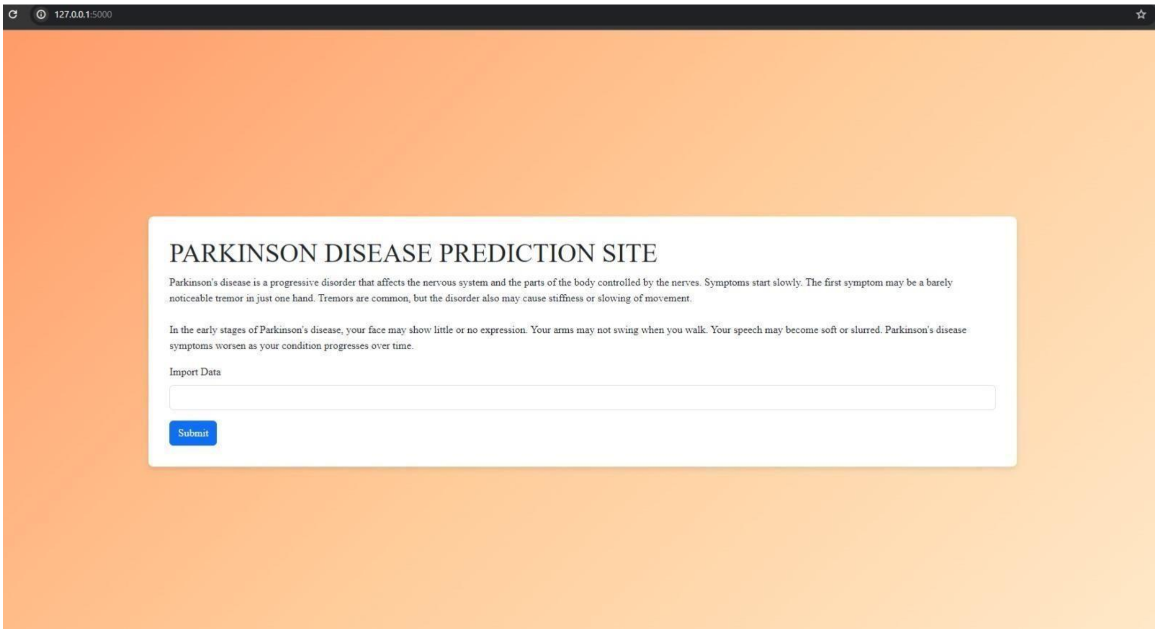
- MDVP:Flo(Hz): The lowest fundamental frequency measured in Hz.
- MDVP:Jitter(%): The percentage of the absolute jitter, which is defined as the variation in the fundamental frequency, relative to the fundamental frequency.
- MDVP:Jitter(Abs): The absolute jitter, measured in Hz.
- MDVP:RAP: The relative amplitude perturbation, which is defined as the average absolute difference between consecutive signal periods divided by the average signal period.
- MDVP:PPQ: The pitch period perturbation quotient, which is similar to the RAP but only takes into account the pitch periods.
- Jitter:DDP: The average absolute difference between consecutive signal periods divided by three Times the average signal period.
- MDVP:Shimmer: The amplitude variation of the signal in dB.
- MDVP:Shimmer(dB): The amplitude variation of the signal in dB.
- Shimmer: APQ3: The amplitude perturbation quotient measured in dB, which only takes into accountthe first three harmonics.
- Shimmer: APQ5: The amplitude perturbation quotient measured in dB, which only takes into account the first five harmonics.
- MDVP:APQ: The amplitude perturbation quotient measured in dB, which takes intoaccount allharmonics.
- Shimmer:DDA: The average absolute difference between consecutive signal periods divided by the signal length, measured in dB.
- NHR: The ratio ofthe noise to the harmonic components in the voice.
- HNR: The ratio ofthe energy in the harmonic components to the energy in the noise. I
- status: A binary variable that indicates whether the patient has Parkinson's disease (1) or
- not (0). RPDE: The relative average perturbation entropy, which is a measure of the unpredictability of the signal.
- DFA: The detrended fluctuation analysis, which measures the scaling behavior of thesignal.
- Spread1: A nonlinear measure of the speech signal.
- Spread2: Another nonlinear measure of the speech signal.
- D2: The correlation dimension, which measures the fractal dimension of the signal.
- PPE: The pitch period entropy, which measures the disorder and complexity of the signal.

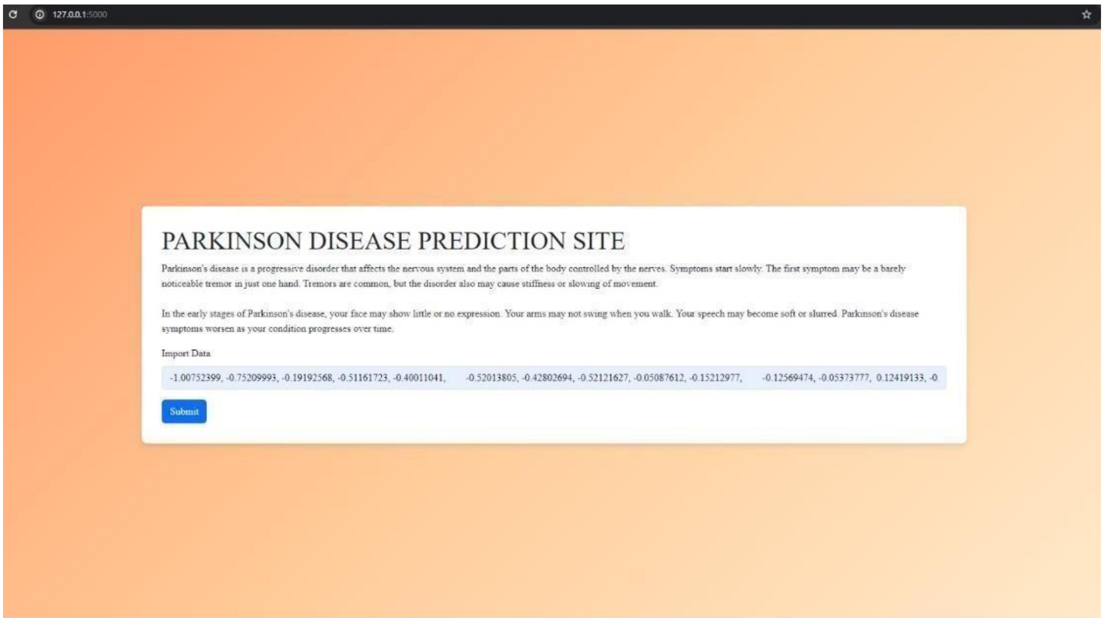
10. APPENDIX

```
import numpy as np import
pandas as pd
parkinson = pd.read_csv('parkinsons.data')
parkinson.head()
parkinson.shape parkinson.isnull().sum()
parkinson.duplicated().sum()
parkinson.info() parkinson.describe()
parkinson['status'].value_counts()X.shape
Y.shape
from sklearn.model_selection import train_test_split from
sklearn.metrics import accuracy_score
from sklearn.linear_model import LogisticRegression from
sklearn.neighbors import KNeighborsClassifier from sklearn.svm
import SVC
from sklearn.naive_bayes import GaussianNB
fromsklearn.tree import DecisionTreeClassifier
fromsklearn.ensemble import RandomForestClassifiermodels = {
"lg": LogisticRegression(), "knc":
KNeighborsClassifier(),"svc": SVC(),
"gnb": GaussianNB(),
'dtc': DecisionTreeClassifier(), 'rfc':
RandomForestClassifier(),
}
svc = SVC() svc.fit(X_train,y_train)
y_predict = svc.predict(X_test)
accuracy_score(y_test,y_predict)
np_data = np.asanyarray(input_data)
prediction = svc.predict(np_data.reshape(1,-1))
```

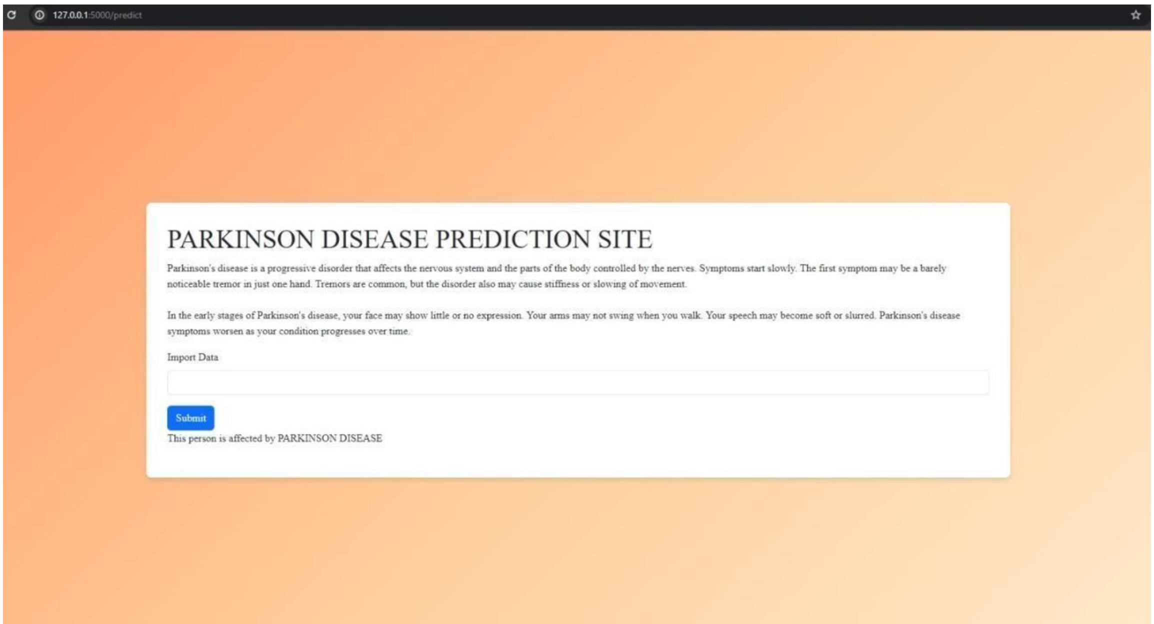
International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

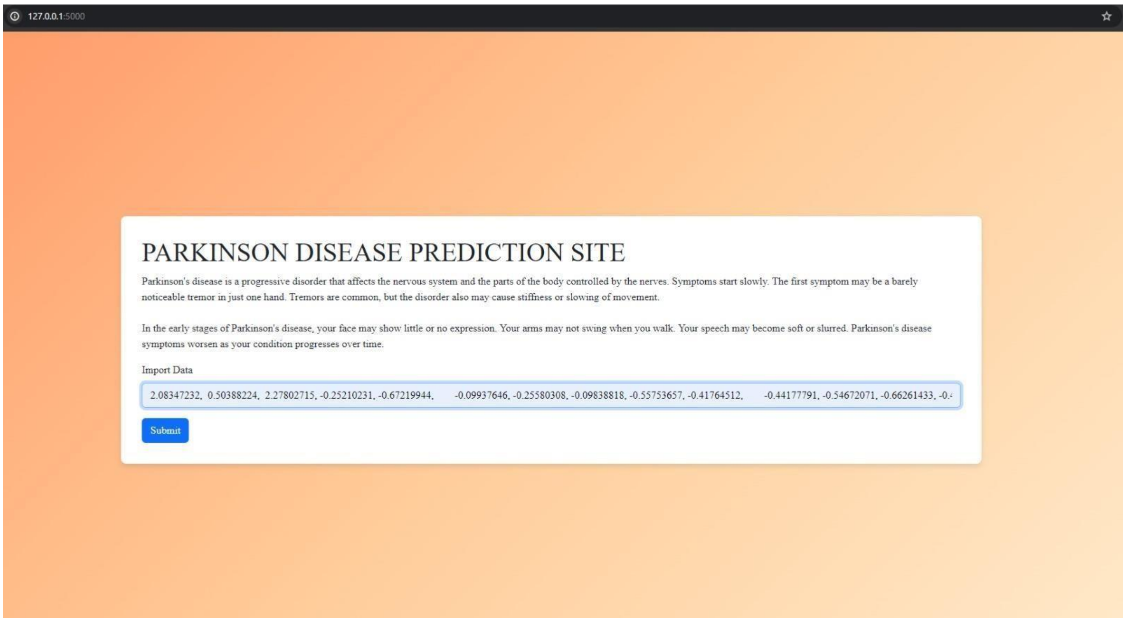
11. OUTPUT

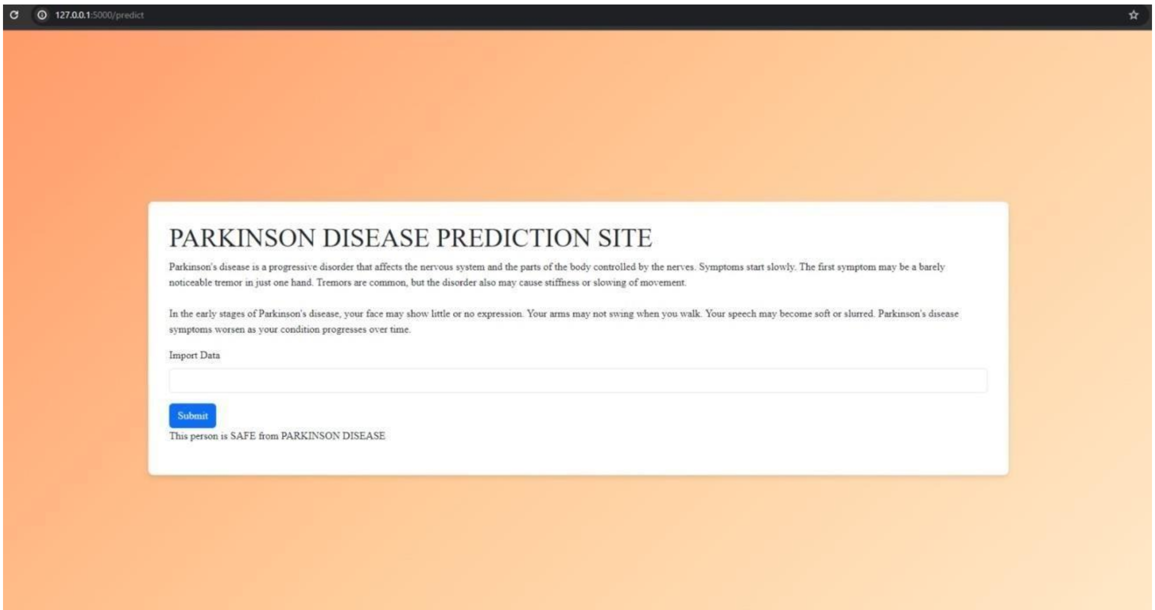




International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930







12. REFERENCES

1. Kalia LV, Lang AE. Parkinson's disease. Lancet. 2015 Aug 29;386(9996):896–912.https://doi.org/10.1016/S0140-6736(14)61393-3.

2. Tanner CM, Goldman SM. Epidemiology of Parkinson's disease. Neurol Clin. 1996May;14(2):317–35. https://doi.org/10.1016/S0733-8619(05)70259-0.

3. Nussbaum RL, Ellis CE. Alzheimer's disease and Parkinson's disease. N Engl J Med. 2003Apr 3;348(14):1356–64. https://doi.org/10.1056/NEJM2003ra020003.

4. Saiki M, Baker A, Williams-Gray CH, Foltynie T, Goodman RS, Taylor CJ, et al. Association of the human leucocyte antigen region with susceptibility to Parkinson's disease. J Neurol Neurosurg Psychiatr.2010 Aug 1;81(8):890–1.

5. Gao X, Chen H, Schwarzschild MA, Ascherio A. Use of ibuprofen and risk of Parkinson disease. Neurology. 2011;76(10):863–9.

6. Williams-Gray CH, Wijeyekoon R, Yarnall AJ, Lawson RA, Breen DP, Evans JR, et al. Serum immunemarkers and disease progression in an incident Parkinson's disease cohort (ICICLE-PD). Mov Disord. 2016 Jul 1;31(7):995–1003.