CAPSTONE PROJECT 2
PARKINSON DISEASE DETECTION
BY SAI TEJA PGA36

1. Executive Summary:

Parkinson's disease (PD) is a progressive neurological disorder that affects movement and motor control. Early detection is crucial for managing the disease and improving patient outcomes. Traditional diagnostic methods rely heavily on clinical assessments, which can be subjective and may lead to delays in diagnosis. Machine learning (ML) offers the potential to improve the accuracy and speed of Parkinson's disease detection by analyzing various biomarkers and patient data.

2. Problem Statement:

Background: Parkinson's disease (PD) is a chronic and progressive neurological disorder that primarily affects movement, often leading to tremors, rigidity, and bradykinesia (slowness of movement). Early and accurate diagnosis is critical for managing the disease and improving patient outcomes. However, traditional diagnostic methods are often subjective, time-consuming, and reliant on clinical expertise, which can lead to delays in diagnosis and variability in outcomes.

Objective: To develop and validate a machine learning-based approach that can accurately and efficiently detect Parkinson's disease using patient data, such as motor symptoms, voice recordings, and movement analysis. The goal is to create a model that enhances the diagnostic process, enables early detection, and can be easily integrated into clinical practice to improve patient outcomes.

Scope: This project aims to leverage machine learning techniques to develop a reliable and accurate model for detecting Parkinson's disease (PD) based on patient data. The scope of the project includes data collection, model development, validation, and potential integration into clinical settings.

3. Data Sources:

Primary Data: Original data collected directly from patients, such as medical records, neurological assessments, wearable sensor data, and imaging scans..

Secondary Data: Publicly available datasets like the UCI Parkinson's Disease Dataset, Parkinson's Progression Markers Initiative (PPMI), and mPower study data, which are reused for analysis.

4. Methodology:

Data Preprocessing: Handle missing data, remove outliers, and ensure consistency in data formats across sources.

Model Development: Experiment with various machine learning models, including Support Vector Machines (SVM) Effective for high-dimensional data like voice features and Random Forest Suitable for handling diverse feature sets and reducing overfitting and Neural Networks Ideal for detecting complex patterns in multi-modal data (e.g., voice, movement) and Logistic Regression Useful as a baseline model for binary classification.

Model Training: Train selected models using the training dataset, optimizing for early and accurate detection of Parkinson's disease.

Model Evaluation: Evaluate models using metrics such as accuracy, precision, recall, F1 score, and ROC-AUC to assess their detection capabilities.

5. Expected Outcomes:

- A machine learning model that accurately detects Parkinson's disease with high performance metrics (e.g., accuracy, precision, recall).
- A user-friendly prototype application for real-time Parkinson's detection based on patient data.
- A comprehensive report detailing model performance, insights, and recommendations for clinical use.

6. Tools and Technologies:

- Jupyter notebook:For data Preprocessing exploratory data analysis
- Machine Learning Frameworks : TensorFlow,XGBOOST,Scikit-learn
- Programming Language : Python
- Data visualization:Matplotlib,Seaborn for model performance analsis

7. Risks and Challenges:

- **Data Availability and Quality:** Limited or poor-quality data can lead to inaccurate model predictions and hinder model effectiveness.

- **Model Generalizability:** Ensuring the model works well across different patient populations is crucial to avoid overfitting and maintain clinical relevance.

- **Data Privacy and Compliance:** Handling sensitive patient data requires adherence to privacy regulations like GDPR or HIPAA to avoid legal and ethical issues.

- **Bias and Fairness:** Addressing biases in the data is essential to ensure fair and unbiased predictions, particularly across diverse demographic groups.

8. Conclusion:

This project aims to develop a machine learning model for early detection of Parkinson's disease, which can enhance diagnosis and patient care. By using data like clinical records, voice, and movement patterns, the model can make accurate predictions. While the potential benefits are significant, challenges like data privacy, model fairness, and generalizability must be addressed. Future efforts should focus on refining the model and integrating it into real-world healthcare settings.