Project Report:

Detection of Parkinson's Disease

Introduction:

Parkinson's disease is a degenerative condition of the brain that is associated with motor symptoms such as slow movement, tremor, rigidity, and imbalance. It also leads to other complications including cognitive impairment, mental health disorders, sleep disorders, and pain and sensory disturbances. Early detection and prediction of Parkinson's disease can significantly improve patient outcomes and provide them with better care and support. In this project, we aim to develop a machine learning model to predict the presence of Parkinson's disease based on certain parameters extracted from speech signals.

Data:

We used a dataset containing various features extracted from speech signals of individuals with and without Parkinson's disease. The dataset consists of 756 entries and 755 columns. Each entry represents a sample, and the columns represent different features. Some of the important features include gender, PPE (Pitch Period Entropy), DFA (Detrended Fluctuation Analysis), RPDE (Recurrence Period Density Entropy), numPulses, tqwt\_kurtosisValue (Tuned Q-Factor Wavelet Transform), locPctJitter, etc.

Data Preprocessing:

Before building the machine learning model, we performed some data preprocessing steps. First, we dropped the 'id' column as it was not relevant to our analysis. Then, we separated the features (X) and the target variable (y) from the dataset. X contains all the columns except the 'class' column, which represents the presence or absence of Parkinson's disease.

Exploratory Data Analysis:

We conducted exploratory data analysis to gain insights into the dataset. We computed descriptive statistics of the dataset, which provided information about the mean, standard deviation, minimum, maximum, and quartiles of each feature. We also examined the first 100 rows of the dataset to get a glimpse of the feature values.

Data Cleaning:

To ensure data quality, we checked for missing values in the dataset. Fortunately, there were no missing values in any of the columns. Additionally, we removed any duplicate entries from the dataset.

Correlation Analysis:

To understand the relationships between different features, we computed the correlation matrix of the dataset. We visualized the correlation matrix using a heatmap, which helped us identify any strong correlations between features. This analysis was important to determine if any features were highly correlated and could be potentially dropped to avoid multicollinearity.

Model Development:

For the prediction of Parkinson's disease, we applied machine learning algorithms. We split the dataset into training and testing sets and trained various models on the training data. The models considered for this project included decision trees, random forests, support vector machines, logistic regression, and neural networks. We evaluated the performance of each model using appropriate evaluation metrics such as accuracy, precision, recall, and F1 score.

Results:

The results obtained from the machine learning models indicated that the random forest algorithm achieved the highest accuracy and F1 score for predicting Parkinson's disease. This model showed promising results in detecting the disease based on the provided features extracted from speech signals.

Conclusion: In this project, we developed a machine learning model to detect Parkinson's disease based on speech signal features. The model showed good performance in predicting the presence of the disease. Early detection of Parkinson's disease can significantly improve patient outcomes by enabling timely interventions and personalized care. Further research and validation on larger datasets can enhance the accuracy and reliability of the predictive models for Parkinson's disease detection.