**A Machine Learning Approach for the diagnosis of Parkinson’s Disease via Speech Analysis**

1. **INTRODUCTION**:

1.1 Overview:

Parkinson’s Disease is the second most prevalent neurodegenerative disorder after Alzheimer’s, affecting more than 10 million people worldwide. Parkinson’s is characterized primarily by the deterioration of motor and cognitive ability.Speech is very predictive and characteristic of Parkinson’s disease; almost every Parkinson’s patient experiences severe vocal degradation (inability to produce sustained phonations, tremor, hoarseness), so it makes sense to use voice to diagnose the disease. Voice analysis gives the added benefit of being non-invasive, inexpensive, and very easy to extract clinically.

1.2 Purpose:

A study from the National Institute of Neurological Disorders finds that early diagnosis (having symptoms for 5 years or less) is only 53% accurate. This is not much better than random guessing, but an early diagnosis is critical to effective treatment.Symptoms include: “frozen” facial features, bradykinesia (slowness of movement), akinesia (impairment of voluntary movement), tremor, and voice impairment.Because of these difficulties, we used a machine learning approach to accurately diagnose Parkinson’s, using a dataset of various speech features.

1. **LITERATURE SURVEY**:
   1. Existing Problem:

There is no single test which can be administered for diagnosis. Instead, doctors must perform a careful clinical analysis of the patient’s medical history.Unfortunately, this method of diagnosis is highly inaccurate.

* 1. Proposed Solution:

Produce a machine learning model to diagnose Parkinson’s disease given various features of a patient’s speech with at least 90% accuracy and/or a Matthews Correlation Coefficient of at least 0.9. Compare various algorithms and parameters to determine the best model for predicting Parkinson’s.

1. **PERFORMANCE METRIC**:

T = true , F = false

Accuracy: Based on prediction set

1. **ALGORITHMS EMLOYED**:

4.1 Logistic Regression :

Uses the sigmoid logistic equation with weights (coefficient values) and biases (constants) to model the probability of a certain class for binary classification. An output of 1 represents one class, and an output of 0 represents the other. Training the model will learn the optimal weights and biases.

4.2 Decision Tree Classifier:

Represented by a binary tree, where each root node represents an input variable and a split point, and each leaf node contains an output used to make a prediction.

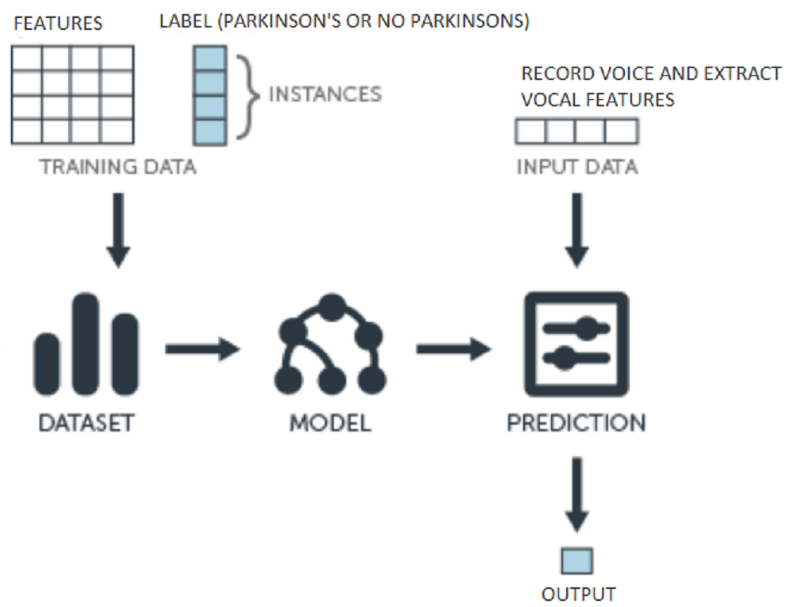
4.3 Random Forest Classifier:

It consisting of many decisions trees. It uses bagging and feature randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree.

1. **DATASET** **DESCRIPTION:**

* Train and Test Datasets are used..
* 28 features (elements that are possibly characteristic of Parkinson’s, such as frequency, pitch, amplitude / period of the sound wave).
* 1 label - Class (1 for Parkinson’s, 0 for no Parkinson’s).

1. **PROJECT PIPELINE**:



1. **SUMMARY OF PROCDURE:**

* Split the Oxford Parkinson’s Dataset into two parts: one for training, one for validation (evaluate how well the model performs).
* Train each of the following algorithms with the training set: Logistic Regression, Decision Tree, Random Forest Classifier.
* Evaluate results using the validation set.

1. **DATA ANALYSIS**:

* In general, the models tended to perform the best (in terms of accuracy) on the dataset with a 75-25 train-test split.
* The two highest performing algorithms,Random Forest and Decision Tree, both achieved an accuracy of 100%. The Logistic Regression achieved a MCC of 98.
* These figures outperform most existing literature and significantly outperform current methods of diagnosis.

1. **CONCLUSION AND SIGNIFICANCE**:

* These robust results suggest that a machine learning approach can indeed be implemented to significantly improve diagnosis methods of Parkinson’s disease. Given the necessity of early diagnosis for effective treatment, my machine learning models provide a very promising alternative to the current, rather ineffective method of diagnosis.
* Current methods of early diagnosis are only 53% accurate, while my machine learning model produces 98% accuracy. This 45% increase is critical because an accurate, early diagnosis is needed to effectively treat the disease.
* Typically, by the time the disease is diagnosed, 60% of nigrostriatal neurons have degenerated, and 80% of striatal dopamine have been depleted.
* With an earlier diagnosis, much of this degradation could have been slowed or treated.
* My results are very significant because Parkinson’s affects over 10 million people worldwide who could benefit greatly from an early, accurate diagnosis.
* Not only is my machine learning approach more accurate in terms of diagnostic accuracy, it is also more scalable, less expensive, and therefore more accessible to people who might not have access to established medical facilities and professionals.
* The diagnosis is also much simpler, requiring only a 10-15 second voice recording and producing an immediate diagnosis.