Name: Romil Shah
NUID: 00167359

Parkinson's Disease detection based on Speech and Gait signal analyses using Machine Learning algorithms

1. Idea:

Detection or prediction of Parkinson's disease has always been based on medical history of a person. It has recently changed to use of expensive devices. But as medical conditions suggest, the voice and walk pattern is affected.

I began starting from collection of the walking data using 'MYO Armband' which was in terms of 'pitch-yaw-roll' to find out the x-y-z locations whilst walking. On further double differentiation, I was able to get the acceleration in each direction. In place of that I am planning to use a dataset that has the same features. This dataset has the same x-y-z acceleration data in order to check if a person freezes whilst walking or not and is that related to Parkinson's or not. I am also using the speech data to find out a person is affected by Parkinson's or not.

In this project, I am trying to analyze the walk and talk pattern of a person in order to classify if a person is a Parkinson's patient or not. I am planning to explore the most suitable machine learning technique that may be used for the proper detection.

2. Background:

Parkinson's disease is a central nervous system disorder that affects the motion of the human body that include shaking, rigidity, slow movement and difficulty in walking [1]. As the disease advances, thinking and behavioral problems start taking place that affects a person's psychology and ultimately leads to depression. Parkinson's disease affects about 200,000 people on an average annually. It is an incurable disease but can be contained with treatment [2]. Parkinson's disease can be diagnosed on the basis of the symptoms that include muscle spasms, stiffness, movement coordination problems, dizziness, impaired speech, voice box spasms etc. Ways to diagnose it is based on the patient's medical history as well as neurological examination [3]. There is no lab test that will clearly identify it and brain tests are used to rule out other diseases. Traditional diagnosis of Parkinson's Disease involves a neurological history of the patient and observation of the motor skills in various situations. Since there is no definitive laboratory test to diagnose Parkinson's Disease, diagnosis is often difficult, particularly in the early stages when motor effects are not yet severe. Monitoring progression of the disease over time requires repeated clinic visits by the patient. There is no cure, but pharmacological treatment to manage the condition includes dopaminergic drugs. But all the identification is mainly dependent on the medical history. This is why I decided to work on the speech pattern and gait pattern recognition on basis of which Parkinson patients and healthy patients can be classified as the motor and speech are the parts that have the most effects.

There are several researches based on the diagnosis of Parkinson's disease based on medical instruments or based on history of the patient. The most recent is the one using neuroimaging techniques. There have been a few notable advances in the use of machine learning for the

detection and classification of patients and healthy beings. One of the notable work is formation of new machine learning algorithms specially for prediction of Parkinson's [4].

One such research that has caught my attention is where they are using sensors on the patient's body and analyzing the pattern of the patient whilst walking [5]. Here they are analyzing the parts when the gait 'freezes' i.e. when the patient has problems whilst walking and suddenly stops or has a balance problem. This is when I planned to use the MYO Armband and try analyzing the gait pattern myself. They have also devised an algorithm called the FOG detection or Freezing of Gait detection.

I had devised a technique using the 'MYO Armband' and hacked it to be used for walk pattern recognition using real time data training and detection of an irregularity in the walking pattern on basis of which Parkinson can be classified [6]. I would like to take this further and train the dataset properly to be used in classification of healthy and diseases patients.

Initially I had an idea of using the speech signals but after going through a research paper specially based on the usage of data to analyze patients, I decided to use the gait signal analysis too. There are various tools available for investigating the time-series data of the patients obtained at clinical research centers. For the linear and nonlinear analysis of signals, using linear and nonlinear dynamic parameters, several software packages such as CDA (Chaos Data Analyzer Programs), NLyzer (Nonlinear Analysis in Real Time) TISEAN (Nonlinear Time Series Analysis), WFDB Software, MATLAB Software and Physio Toolkit Software are available [7]. I am planning to use Machine Learning techniques to classify the walk pattern and speech pattern data in order to obtain a clear classification between the type of patients using the inbuilt functions and toolboxes or packages available in R, MATLAB and Python. Using ANN, Fuzzy Logic, SVM, K-Means, AdaBoost etc. on the EMG and EEG Data of the patients we can classify them and do an early diagnose.

Using the methods, I expect to discover an easier and faster way of diagnosing Parkinson's disease. This way an initial way of disease identification can be done based on simple EEG and EMG data. This can save a lot of time, efforts and money. A patient's health can be analyzed by saving long-term costs. The expected results will help the research community work more on learning the various patterns that can be observed in the different data collected from sensors for different types of patients. This can help the community grow more complex algorithms to identify the difference in more efficient way. Also a web-based or mobile-application based real time data can be fetched, analyzed and prediction can be made after using the training dataset to train the models.

Whilst working on the dataset there might be several complications that may arise due to various factors. For example, multiple datasets for a given disorder often exist, collected from different sources and using slightly different features [8]. Combining them in some effective way into a large, cohesive dataset would result in a more robust and well-trained learner.

I am planning to use the available datasets on Parkinson's disease based on walking patterns [9] and that of the EMG [10], EEG and speech signals [11]. I would test the different algorithms on the available dataset and test them on the live data. I am also planning to make a web application that would help in taking in real time data from around the world to be analyzed and as more data rolls in, the better it would be as the accuracy of the classification and prediction would increase.

3. References:

[1] https://en.wikipedia.org/wiki/Parkinson%27s_disease

[2] https://www.gstatic.com/healthricherkp/pdf/parkinson_s_disease.pdf

[3] "Parkinson's disease: clinical features and diagnosis", http://www.ncbi.nlm.nih.gov/pubmed/18344392

[4] "New machine-learning algorithms for prediction of Parkinson's disease", https://www.researchgate.net/publication/234131546_New_machine-learning_algorithms_for_prediction_of_Parkinson's_disease

[5] "Wearable assistant for Parkinson's disease patients with the freezing of gait symptom", http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5325884

[6] http://devpost.com/software/myowalk

[7] "Data Processing for Parkinson's Disease: Tremor, Speech and Gait Signal Analysis", http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6150330

[8] "A Machine Learning Approach to Diagnosis of Parkinson's Disease", http://scholarship.claremont.edu/cgi/viewcontent.cgi?article=1784&context=cmc_theses

[9] https://physionet.org/physiobank/database/gaitndd/

[10] https://physionet.org/physiobank/database/tremordb/

[11] https://archive.ics.uci.edu/ml/datasets/Parkinsons

[12] https://archive.ics.uci.edu/ml/datasets/Daphnet+Freezing+of+Gait

4. Data sources:

I am planning to use the following two datasets:

1. Gait dataset:
    a. Gait Dynamics in Neuro-Degenerative Disease Data Base:
        https://physionet.org/physiobank/database/gaitndd/
    b. Daphnet Freezing of Gait Data Set:
        https://archive.ics.uci.edu/ml/datasets/Daphnet+Freezing+of+Gait
2. Voice dataset:
    a. Parkinsons Data Set:
        http://archive.ics.uci.edu/ml/datasets/Parkinsons
3. Additional: Brain simulation dataset:
    a. Effect of Deep Brain Stimulation on Parkinsonian Tremor:
        https://physionet.org/physiobank/database/tremordb/