Detection of Early Stages of Parkinson's Disease by Analyzing fMRI Data and Machine Learning Approaches

by

Ahmed Hasin Neehal
16101142
Md. Nura Azam
16101169
Md. Sazzadul Islam
16101161
Md. Ishrak Hossain
16101166

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
December 2019

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Ahmed Hasin Neehal 16101142	Md. Nura Azam 16101169
Md. Sazzadul Islam 16101161	Md. Ishrak Hossain 16101166

Approval

The thesis/project titled "Detection of Early Stages of Parkinson's Disease by Analyzing fMRI Data and Machine Learning Approaches" submitted by

1. Ahmed Hasin Neehal (16101142)
2. Md. Nura Azam (16101169)
3. Md. Sazzadul Islam (16101161)
4. Md. Ishrak Hossain (16101166)

Of Fall, 2019 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on December 26, 2019.

Examining Committee:

Supervisor:
(Member)

Mohammad Zavid Parvez, PhD
Assistant Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Mahbubul Alam Majumdar, PhD
Professor and Chairperson
Department of Computer Science and Engineering
Brac University

Abstract

Parkinson's Disease is a progressive nervous system brain disorder which affects motor neuron loss control and movement coordination. Parkinson's symptoms are shown gradually and get worse over time. Its signs and symptoms can be different for everyone. There may be minor early signs and they may go unnoticed. Therefore, early detection of Parkinson's disease might significantly improve life style by giving proper treatment. Moreover, doctors may suggest regulating certain regions of your brain and improve the symptoms. In recent years, the use of Functional Imaging in neurodegenerative diseases has increased, with applications in basic pathophysiology research, support in determination, or evaluation of new medications. In our research we used fMRI data of eight early PD patients. Resting-state fMRI images were collected for analyzing the data and feature extraction. Time series data were generated for each subject based on voxel intensity. In addition, STFT was used to measure the time frequency function. Furthermore, SVM classifier was used for the classification and prediction of the early stage of PD. Using our proposed method, we have achieved 100% sensitivity, specificity, and accuracy considering seven subjects, however, one subject was exceptional whereas we have achieved 99.76% accuracy, 100% specificity and 99.53% sensitivity. Finally, this process is a well-structured model for predicting the early stages of PD. It may help to the doctors for diagnosis of the disease at its early stages and the patients should receive better treatment.

Keywords: Parkinson's Disease; Functional Imaging; fMRI; Voxel intensity; Machine Learning; SVM Classifier; STFT

Dedication

We would like to dedicate our work for all the patients of Parkinson's Disease for whom we have worked so far. We also want to dedicate our work to our parents, without whom we could never come so far in life.

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis have been completed without any major interruption.
Secondly, to our supervisor Dr Mohammad Zavid Parvez for his kind support and advice in our work. He helped us whenever we needed help.
And finally to our parents without their throughout support it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

β Beta

ω Omega

ε Epsilon

$BOLD$ Blood Oxygen Level Dependent

DFT Discrete Fourier Transform

$DTFT$ Discrete-Time Fourier Transform

$fMRI$ functional Magnetic Resonance Imaging

PD Parkinson's Disease

$PPMI$ Parkinson's Progression Markers Initiative

PSD Power Spectral Density

$R-fMRI$ Resting-state functional Magnetic Resonance Imaging

$STFT$ Short-Time Fourier Transform

SVM Support Vector Machine

Chapter 1

Introduction

Parkinson's disease is a progressive neurological disorder characterized by movement disorders [39]. A hormone in the brain called dopamine enables the body's smooth and synchronized muscle movements. Dopamine is formed in the "substantia nigra" region of the brain. The substantia nigra cells begin to die in Parkinson's and the dopamine levels are reduced. The signs of Parkinson's beginning to appear when the dopamine level drops to 60 to 80 percent.
After Alzheimer's disease, Parkinson's disease is the second most common neurodegenerative disorder associated with age. Seven to 10 million people worldwide are estimated to have the disorder of Parkinson's [10]. The prevalence of the disease ranges from 41 per 100,000 in the fourth decade of life to over 1,900 per 100,000 among those 80 years of age.
Generally, the prevalence of the disorder, or the rate of newly diagnosed cases, increases with age, although it can stabilize in people over 80. Before age 50, an additional 4% of people with Parkinson's are diagnosed. Men are 1.5 times more likely than women to have Parkinson's [23].
At any point, Parkinson's may affect anyone. The cause of Parkinson's is a long-standing subject for study around the world, and there are many hypotheses that include causes such as climate, oxidative stress, genes, etc [42].
In this chapter, we will discuss the motivation for our work, our contributions towards this research and the thesis orientation which will cover the contents of each chapter in this paper.

1.1 Motivation

Parkinson's disease, being the second most common disease in the world, is ever on the rise everyday. Many researches have led to this fact [36]. So, it is obvious to detect early stage of PD for faster prevention of the disease. Nowadays many computational tools are used to assist doctors in making decision about the patients' medication. To detect early stages of PD, it is important to use computational tools in present days.
Our motive is to enhance the detection process of early stages of PD. Previously, we have found that there are several research works on this subject but they had different approach for the whole process. So, we decided to make a contribution to this cause with a view to establishing a unique approach of our own. Most often Parkinson's Disease remain undetected in human body and symptoms are diagnosed

at the later stages of the disease. At that stage, it is quite incurable and the patients are the worst sufferers of rigidity, bradykinesia and postural instability. So, we have come up with the method of detecting the disease at its early stages so that we get a perfect accuracy of the results.
We have chosen fMRI data for our research as each of the fMRI images represent a specific range of voxel information which gives us a scope of analyzing time series of data points. As a result, we have decided to add a new dimension to this research work which we have not found in any other research works so far. Previously, we have found that researchers use image processing approaches while we are initiating a model to use signal processing concepts in this paper. We have discussed our contributions in this subject in the next section (1.2).

1.2 Major Contribution

Most researches in the field of detecting Parkinson's disease include fMRI image processing techniques which finds the Region of Interest by locating the active region of the brain when any activity is initiated. These researches include Machine Learning approaches on cognitive states of human brain by brain pattern extraction, Graph Theoretical Metrics on R-fMRI, Cognitive impairment, Joint Regression model.
We have initiated a unique approach in this research which include analyzing resting state fMRI image data. Each voxel represents a time series data which is used for calculating time frequency domain. In this process, STFT is used to get the localized frequency information of the signal. We get a frequency vector and a time vector from the output of the STFT function, which is later used to classify the data using SVM classifier. SVM classifier gave us the best accuracy in the prediction of the early stages. As we have taken the data from early stage PD patients, we have been able to do the feature extraction for our desired outcome of detecting early stages of PD.

1.3 Thesis Orientation

The subsequent sections of the thesis have been organized as follows. Chapter 2 is the literature review which contains related research works and existing approaches relevant to our proposed model. Chapter 3 includes all the background information related to our work and how we are using the resources to get the desired output. In chapter 4, we have described the proposed model of our works along with relevant graphs and figures. It includes our overall working methodologies, Dataset information about patients and preprocessing criteria about dataset will also be included. Algorithms related to our prediction will also be described in this chapter. The predicted results and relevant discussions are showed in chapter 5. Lastly, the summary of the report and conclusion as well as some future workplan is done in chapter 6. In future work part, we will talk about our future ambition on this work.

Chapter 2

Literature Review

For our research work, we had to study a bunch of earlier research paper related to Parkinson's disease and fMRI data which helped us to find our way to accomplish our research work. According to a paper written by R Prashanth, Sumantra Dutta roy, Prava K. Mandal and Shantanu Ghosh Parkinson disease is a neurological disease [31]. The disease occurs due to the loss of dopaminergic neurons which leads to substantial amount of dopamine decrease in the striatum. The visible presence of PD are rigidity, bradykinesia and postural instability. They also stated that there are only few visible symptoms which are mainly at the later stage of the disease, so they worked on the early detection of that disease so that it can be diagnosed at early stage. The used the smell identification test data of university of pennsylvania to analyse their research work. They tried to discuss about a method which helps to detect parkinson disease at early stage in which hey used 193 PD data and 156 normal data. In these data they used logistic regression in which they observed enhanced classification. In this research they basically tried to classify the non motor and imaging marker in the preclinical stage and they did almost a near classification using SVM.

We went through another paper which is written by Amirali Kazeminejad, Soroosh Golbabaei, Hamid Soltanian-Zadeh which depicts about graph theoretical analysis for automatic diagnosis of PD [34]. In this work, they used fMRI data which is in the resting state and contains data of 18 healthy and 19 PD affected patients. After data processing 90 regions of interest were identified and the average time series of each region was identified. They used R-fMRI data to examine functional change in PD affected brain and normal healthy brain state. In state of focusing on motor neuron activities they focused on large scale alteration in brain of a PD affected patient.

We looked into another paper which is written by Rana Fayyaz Ahmad, Aamir Saeed Malik, Nidal Kamel, Faruque Reza they wanted to find out different state of cognitive brain state change [27]. In this study they tried to explain fMRI based approach to classify the cognitive state of PD affected brain. Further they classified the cognitive states with fMRI data between the two cognitive states with fMRI data between two cognitive load assistants with 70% accuracy and for that they used SVM classifier.

A third benefit indicates the recognition of events that can only be shown by the subject. An example of such an occurrence is the accidental transition between the perception of ambiguous visual objects as in the face-vase illusion or between 2D and

3D perception of 2D stereo grams; . A fourth benefit is that the methods associated with events require certain inventive designs that can not be prevented easily.
For our research, we went through another paper written by K.Niazmand, A. Kalaras, H. Dai, T.C. Lueth who used acceleration sensors to detect body tremor and the data was processed using spectral analysis [16].They mentioned about the different investigation methods for their analysis and one of these are STFT. STFT use only a small section of the signal at a time. The time signals they are getting from the accelerometer transformed into time signals. For each data point they are calculating STFT for each data point with an interval of 2 seconds counting backwards from that point.
A paper written by Ze Wang, Anna R. Childress, Jiongjiong Wang, and John A. Detrea combines a machine learning algorithm, the support vector machine (SVM), and the random effect model [7]. A multivariate and brain response model-free approach is important to explore the multivariate complexity of fMRI data and understand the differences between the subject brain response. They used SVM to obtain a total spatial discrimination map (SDM) of the brain, describing the difference between the opposing experimental conditions.
Sangkyun Lee, Sebastian Halder, Andrea Ku bler, Niels Birbaumer and Ranganatha Sitaram discussed functional mapping of fMRI data using Support Vector Machine(SVM) classifier [15]. For fMRI research using SVM, an elegant way to interpret SVM training effects could be to overlay the weight vector on brain pictures. Pre processing was performed with SPM5 and MATLAB was used for classification. They stated in their research work that effect mapping could be a different option in the multi derivative fMRI data analysis by discriminability and distribution of data. In their paper they evaluated different multi derivative functional mapping.
We went through another paper which described about Methods For 367 patients with Parkinson's disease and phenotypically normal imaging data and 165 controls without neurological disease, they built a model for disease classification using data from the Parkinson's Progression Marker Initiative (PPMI) research. Olfactory function, genetic risk, Parkinson's family history, age and gender were selected by incremental logistic regression. We also establishes a reliable, non-invasive approach for distinguishing from controls against patients with Parkinson's disease. The studies they analyzed differ in nature, recruiting and execution that might be useful in other studies.
One of the paper discussed about Using multi-modal baseline neuroimaging data, a joint regression and classification scheme for PD diagnosis [9]. Specifically, in a unified multi-task feature selection model, a new feature selection method was developed through relational learning. Their experimental results show that multi-modal data can effectively improve performance in the recognition of class labels compared to single modal data. They proposed that their method will greatly enhance accuracy in the estimation of clinical scores and also outperform state-of - the art methods. They proposed that for further medical research and treatment, the specified brain regions can be identified.
Another paper written by Hristo ZHIVOMIROV integrates the development of two Short-Time Fourier Transform (STFT) and Inverse Short-Time Fourier Transform (ISTFT) software routines together with know-how on their practical implementation.
The suggested algorithms and the corresponding novel MATLAB functions form a

combined analysis-synthesis pair and assist in the study, processing, re-synthesis and visualization of non-stationary signals in real-world time-frequency.The Examples are given to confirm the accuracy of the algorithms and routines, including analysis, spectral modification and re-synthesis of non-stationary signals, along with example for the out-performance of the proposed analysis function compared to the respecting built in MATLAB function.

Another paper talked about cognitive impairment in Parkinson's disease is a common non-motor symptom. The pathophysiology of the underlying deception remains unclear.

All the research work previously done related to detection or classification has been done by image processing approach. In our research approach we used signal processing approach for our study, which is a new dimension in this field.

Chapter 3

Background Study

After Alzheimer's, the second most common neurodegenerative disease that affects a patient's movement is Parkinson's Disease (PD) [5], [43]. With a barely noticeable tremor in side, the symptoms slowly begin. The disorder creates stiffness and slowing movements of the patient. His/her face may show little expression in the early stages of PD. At the time of walking, the arms may not swing and voice may become slurred. Although this disorder cannot be completely cured, the patient's condition may be improved by proper medication. The signs and symptoms may vary from person to person and in the early stages may remain mild, which may remain unnoticeable.
PD is characterized primarily by four motor symptoms (tremor, stiffness, bradykinesia and postural instability) and four non-motor symptoms (depression, sleep, odor and cognition) [17]. The patient's life can be measurable for these symptoms. Early detection of PD can therefore be an important factor in controlling the progression of disease and in alleviating mental and physical distress.
A variety of objective measures for the differential diagnosis of PD have been implemented over the past two decades, including a selection of olfactory, electrophysiological and neuropsychological assessments [4], [26], [37]. Neuroimaging, however, is the most advanced field of offering an objective evaluation. Functional magnetic resonance imaging (fMRI) is the most effective tool to study PD due to its high temporal resolution. Usually, Resting State fMRI is used to monitor neurological networks. BOLD signals with low frequency oscillations are perceived as a marker of simultaneous neuron activity.

3.1 Human Brain

The human brain is the largest brain of all vertebrates compared to the weight of approximately 3.3 lbs of the body.The average male is 1,274 cubic centimeters in brain volume while the female brain is 1,131 cubic centimeters in volume.It contains approximately 86 billion nerve cells (neurons) called 'gray matter' [8]. It also contains billions of nerve fibers (axons and dendrites) called 'white matter'.
According to the Mayfield Clinic, the main part of the human brain is the cerebrum separated into two hemispheres. The brainstem lies below, and the cerebellum sits behind it (see Figure 3.1) [25]. The cerebral cortex is the outermost layer of the cerebrum, composed of four lobes: the frontal, parietal, temporal and occipital lobes. The human brain grows, like all vertebrate brains, from three parts known as the

forebrain, midbrain, and hindbrain. Each of these includes ventricles called fluid-filled cavities. The forebrain grows into the cerebrum and underlying structures; the midbrain is part of the brainstem; and the hindbrain causes brainstem and cerebellum regions to grow [11]. The cerebral cortex is considered the center of complex thinking and is significantly expanded in human brains. Vision perception takes place in the back of the skull in the occipital lobe. The temporal lobe processes sound and expression, including the hippocampus and amygdala, respectively, which play roles in memory and emotion. For spatial orientation and movement, the parietal lobe combines information from different senses.

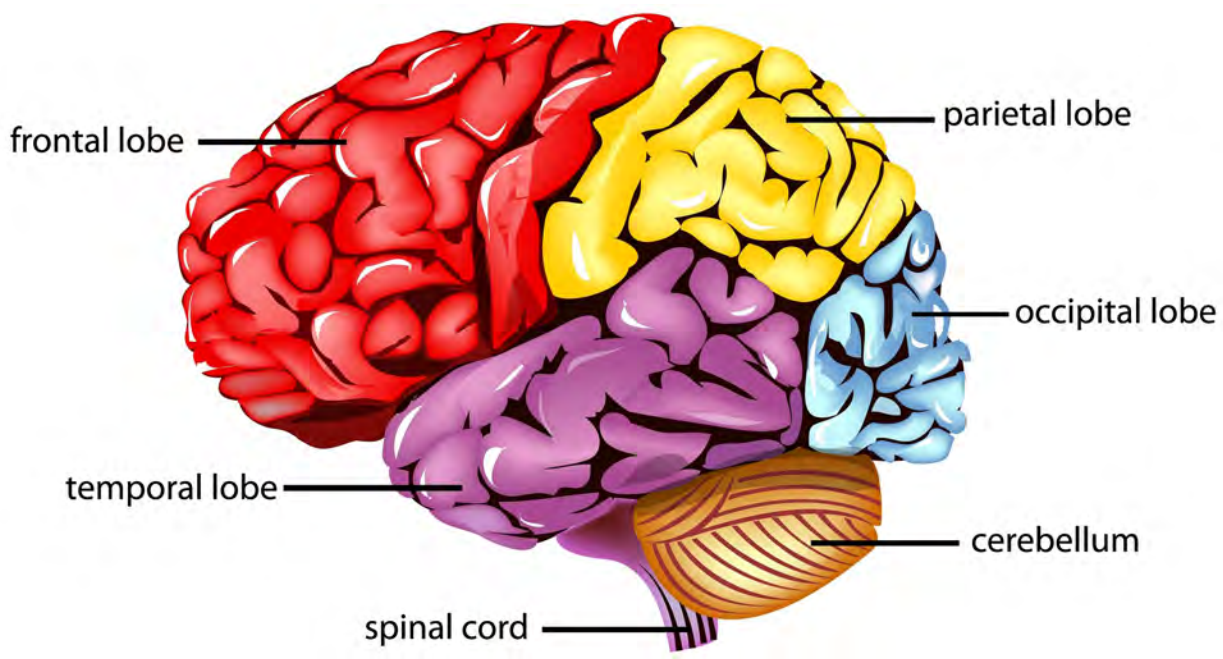

Figure 3.1: Structure of a Human Brain

In the following section we are going to discuss about the Parkinson's Disease.

3.2 Parkinson's Disease

Parkinson's Disease (PD) is the second most common neurodegenerative disorder following Alzheimer's disease that affects a patient's movement. With a barely noticeable tremor in side, the symptoms slowly begin. The condition produces the patient's stiffness and slow motion. In the early stages of PD, his/her face can display little emotion. At the time of walking, the arms may not swing, and speech may become slurred. Although it is not possible to completely cure this disease, proper treatment can improve the patient's condition. The signs and symptoms may vary from person to person and in the early stages can remain mild, which may remain unnoticeable. PD is characterized primarily by four motor symptoms (tremor, stiffness, bradykinesia, and postural instability) [17] and four non-motor symptoms (depression, sleep, odor, and cognition). The major symptoms are following:

1. **Bradykinesia:** This symptom is associated with slow motion, making simple tasks difficult and time consuming. Patients find difficulties to get out of chair and their steps become shorter when they walk.

2. **Loss of automatic movements:** Patients' ability to perform involuntary gestures such as blinking, waving arms, laughing etc. has decreased.

3. **Tremor:** A tremor is an involuntary wave of quivering [32]. It occurs at rest. Parkinson's disease tremor typically begins in one hand, foot, or leg and may eventually affect both sides of the body. Tremor is distracting, however. It also attracts attention, so that during conferences people tend to hold their bad hand in a pocket or sit on it. Despite movement vanishing, it turns out that tremor does not seriously interfere despite daily living activities.

4. **Speech Changes:** Patients hesitate before starts talking. Sometimes they speak softly, quickly and slur.

5. **Writing Changes:** Patients writings become smaller and they find it hard to write.

6. **Rigid Muscles:** In any part of your body, muscle stiffness can occur. The rigid muscles can be painful, restricting the range of motion of the patient.

In the following section we are going to write about fMRI:

3.3 Functional Magnetic Resonance Imaging

There are many brain imaging techniques for detecting and analyzing brain activity. Some of them are invasive and some are non-invasive. Invasive means there may occur some damages in the brain when the images are taken and non-invasive refers to the opposite. Here is the big advantage of fMRI which is a non-invasive technique that is used to measure neural activity by observing changes in the blood flow in the brain while performing a task. Additionally, fMRI can measure the magnetic properties of oxygenated and deoxygenated blood [19], [28]. Moreover, fMRI can be used efficiently for mapping brain activity. fMRI is used very widely for studying brain connectivity patterns during rest and activation while performing various tasks for Parkinson's disease affected patients.
fMRI can not directly measure neural activity [2]. Electrical signals are transmitted from one nerve to another nerve when an action is performed. During an activity, the concentration of blood oxygen level changes in a specific region in the brain. The fMRI system reads the ratio of oxygenated and deoxygenated blood as an indication of the level dependent response of the blood [3].
Based on the BOLD response, neural activity can be visualized while performing a specific task [18]. Finally, abnormal changes in the brain due to Parkinson's disease can be detected using fMRI techniques.
In the following section, we are going to discuss about machine learning approaches:

3.4 Machine Learning

Machine Learning is a subset of AI by which a machine or program learns from experience and make predictions based on the experience. It mainly focuses on making data driven decisions rather than being explicitly programmed for performing certain task [38]. These algorithms work in such a way that they learn from their mistake and improve over time when they are tested to new data. Moreover, data are separated into training set and test set by which algorithms are trained by training data and predicted data are compared to test data for checking accuracy [14], [41]. Machine learning algorithms are trained until a satisfactory accuracy results are obtained.
Types of Machine Learning:
Machine Learning algorithms can be categorized in three different types [41],

1. Supervised Learning
2. Unsupervised Learning
3. Reinforcement Learning

3.4.1 Supervised Learning

Supervised algorithms of machine learning are designed for example learning. The term "supervised" learning comes from the fact that this form of algorithm training is like having an instructor supervise the entire process. The training data must consist of inputs coupled with the right outputs when training a supervised learning algorithm. The algorithm will scan in the data for patterns that correspond with the desired outputs during training. A supervised learning algorithm can take in new unknown inputs after training and will decide which mark will identify the new inputs on the basis of prior training data. A supervised learning model's goal is to predict the right label for new input data presented (See Figure 3.2).

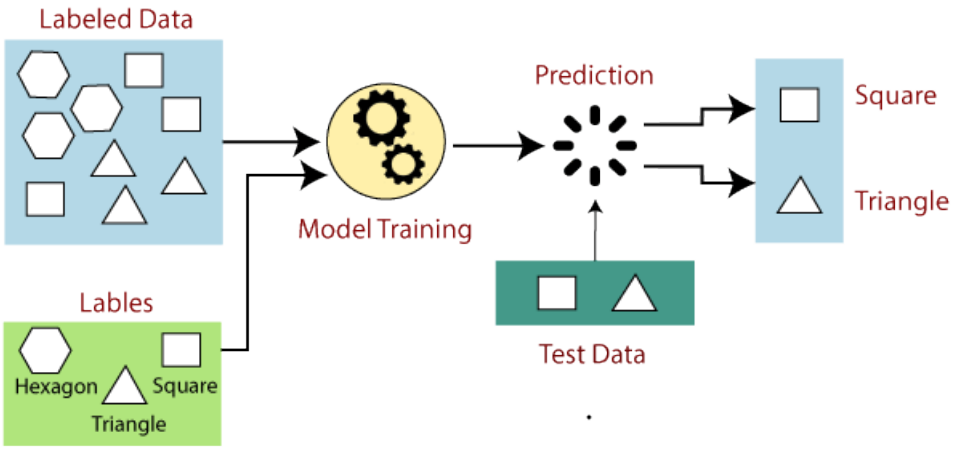

Figure 3.2: Supervised Learning

In supervised learning, there are input variables (x) and an output variable (Y) using which an algorithm is formed to learn the mapping function from input to output Y= f(x). The target is to calculate the mapping function in such a way so

that the best result is derived and whenever new input data (x) is gathered, the output variables (Y) can be predicted for that amount of data.
Supervised learning consists of two types - Classification and Regression. A regression problem is when the output variable is a real or constant value. Example- Salary, Weight etc. It is possible to use many different models, the simplest of which is linear regression. It tries to fit the best hyper-plane data that passes through the points.
A classification problem is when a category is the output variable. A model of classification tries to draw an inference from observed values. A classification model may attempt to predict the value of one or more outcomes given one or more inputs.

3.4.2 Unsupervised Learning

Unsupervised learning is a type of machine training that uses information that is unlabelled and not classified and that allows the algorithm to function without guidance on that information. Unsupervised learning algorithms allow to perform tasks that are more complex than supervised learning. Unsupervised learning, however, may be more volatile compared to other natural methods of learning (See Figure 3.3).

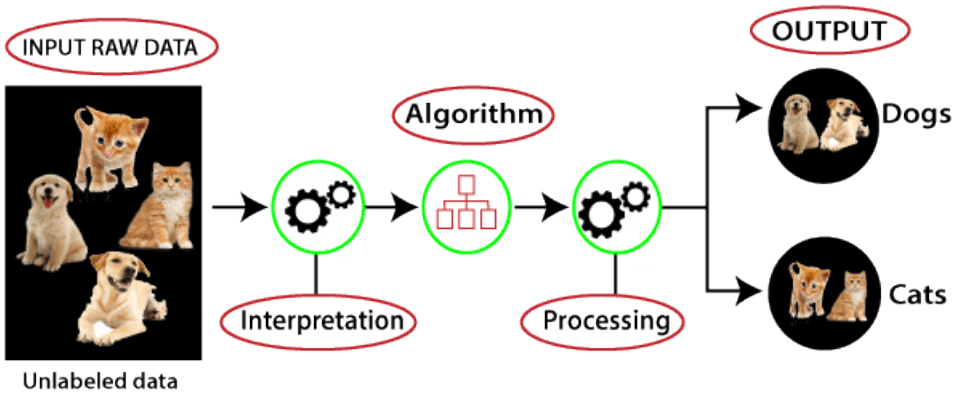

Figure 3.3: Unsupervised Learning

Unsupervised Learning is of two types- Clustering and Association problems. Clustering is specifically concerned with finding a structure or pattern in an uncategorized data collection. Clustering algorithms will analyze the data and, if they exist in the data, will find natural clusters (groups). You can also adjust how many clusters should be found by your algorithms.
Association rules allow you to set up associations within large databases between data objects. This unsupervised technique involves the discovery in large databases of important relationships between variables.

3.4.3 Reinforcement Learning

Reinforcement learning is teaching the models of Machine Learning to make a series of decisions. In an unpredictable, potentially complex environment, the agent learns to reach a goal. It is about how automated agents in an environment can take action.

It is part of the deep learning method that lets you optimize any accumulated reward component (See Figure 3.4).

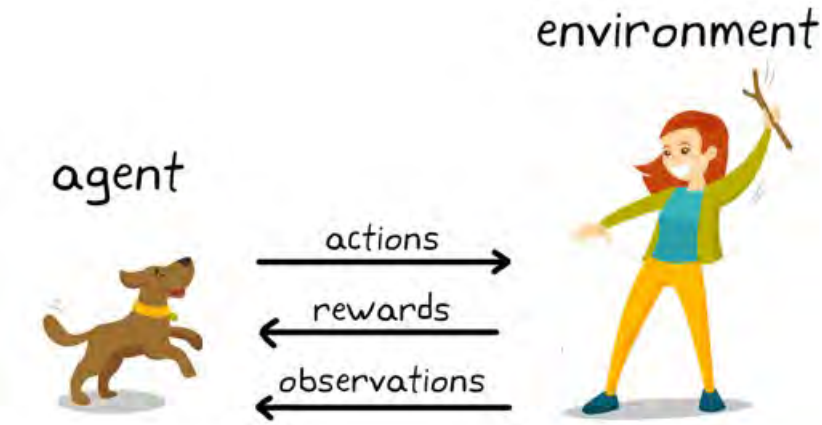

Figure 3.4: Reinforcement Learning

Chapter 4

Proposed Model

In this paper, several steps were performed for accomplishing better accuracy. In the following sections, we are going to describe each steps sequentially (see details in Figure 4.1). Firstly, we got our fMRI dataset from Parkinson's Progression Markers Initiative (PPMI) database (http://www.ppmi-info.org/data). Then we categorized and selected some specific subjects related to our research. Secondly, we extracted time series using voxel-time series methodology from the selected subjects and removed null values from the time series. Thirdly, we extracted time frequency feature as energy spectral density using Short Term Fourier Transform (STFT). In addition, we used SVM classifier to classify our subjects. Finally, we were able to predict the early stage with an outstanding accuracy.

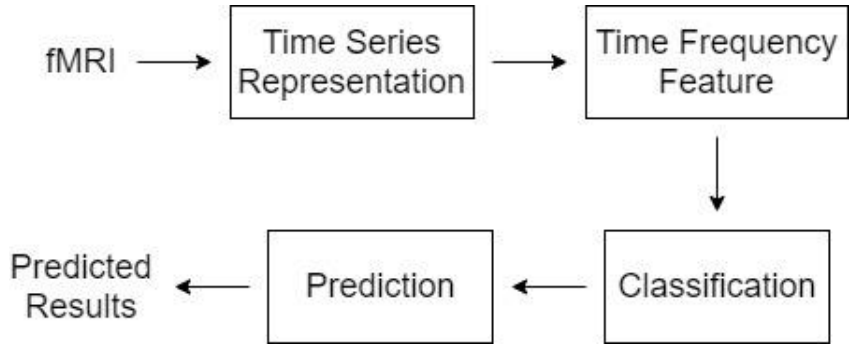

Figure 4.1: Proposed Model

4.1 fMRI Dataset

Technology in medical science has advanced tremendously that it's now became very easy to dissect human body into thin images by computer scanning. Moreover, three-dimensional models of tissues and organs can be created so that it may help doctors to find abnormalities and diagnose disease properly. Functional Magnetic Resonance Imaging (fMRI), a newly invented scanning technology which advanced the medical science one step farther. The biggest advantage of fMRI technology is that it is not limited to diagnose disease only, rather it might help doctors to analyze neural activities inside the brain very efficiently. fMRI can be used as lie detector as it can detect what we're thinking and feeling. Furthermore, fMRI is a non-invasive technique where strong magnetic field and radio waves are used by which detailed image of the brain can be visualized easily. However, fMRI doesn't create images

of organs and tissues like Magnetic Resonance Imaging (MRI), rather it detects neural activity directly by looking at the blood flow throughout the brain. High technology computers are used to capture these changes in the blood flow which may help doctors to relate between certain tasks and brain activities. The changes in the blood flow are visualized in images termed as slice in fMRI. Moreover, the brain is sliced into almost thousand three dimensional slices (see Figure 4.2) which together makes a four dimensional image within some seconds. In addition, every three dimensional slices are made of small equal cubes which are named as voxels in fMRI. As a result, doctors can detect neural activity by analyzing changes in the consequent voxels.

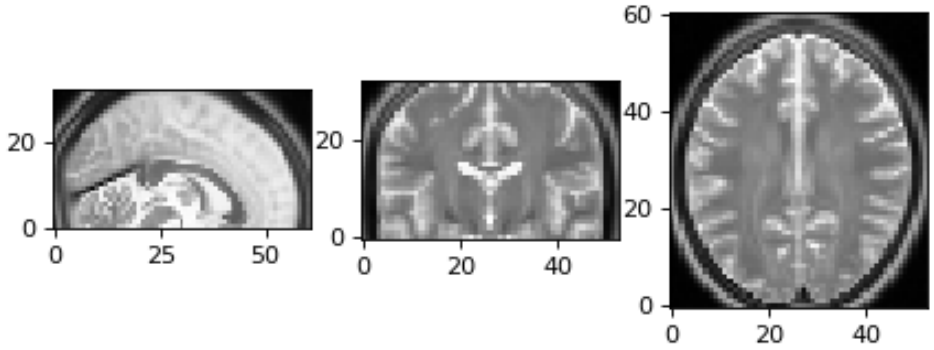

Figure 4.2: fMRI Image of a brain volume

We used resting state fMRI data which is found to be most accurate for research in the recent times. The data used in our research were from the Parkinson's Progression Markers Initiative (PPMI) database (http://www.ppmi-info.org/data). The PPMI is a landmark, large-scale, comprehensive, observational, international, multicenter study that recruits de novo (early-untreated) PD patients and age-matched healthy normal subjects to identify PD progression biomarkers. As fMRI contains a large amount of slices and our study is based on early stages, 8 patients were taken for our analysis of which 7 were male and 1 was female and everyone was in prodromal (early stage) stage (see details in Table 4.1). Data were taken for 2 years as maximum differences between them. For example: subject 'X' has data of 3 years like at the age 63, 64 and 65. We took data of the age 63 and 65 for our research so that we can detect and predict at the first year when the Parkinson's disease starts to appear. Finally, the aim of our dataset is to predict at the early stages of the Parkinson's disease so that doctor can take proper steps for betterment of patients.

Table 4.1: Initial Data Table

Serial No.	Sex	Subject ID	Age	Initial Data Shape (3- D)
1	Female	60036	71	(68, 66, 40, 210)
			73	(68, 66, 40, 210)
2	Male	60006	82	(68, 66, 40, 210)
			84	(68, 66, 40, 210)
3	Male	60035	66	(68, 66, 40, 210)
			68	(68, 66, 40, 210)
4	Male	60043	68	(68, 66, 40, 210)
			70	(68, 66, 40, 202)
5	Male	60044	73	(68, 66, 40, 210)
			75	(68, 66, 40, 210)
6	Male	60073	65	(68, 66, 40, 210)
			67	(68, 66, 40, 210)
7	Male	60074	74	(68, 66, 40, 210)
			76	(68, 66, 40, 210)
8	Male	60075	66	(68, 66, 40, 210)
			68	(68, 66, 40, 210)

4.2 Time Series Representation

Time series represents a series of data taken in consecutive time differences. By analyzing time series data, characteristics and meaningful statistics about the data can be obtained. Moreover, time series data is very useful to predict values by observing previous continuous values. fMRI scans brain as a function of time which can treat fMRI data as a time series data. In fMRI, the images consist of many three dimensional cubes called as voxels. Different voxels have different intensity which means every voxel represents a specific time which combinedly led to a time series data (See Figure 4.3) [29].

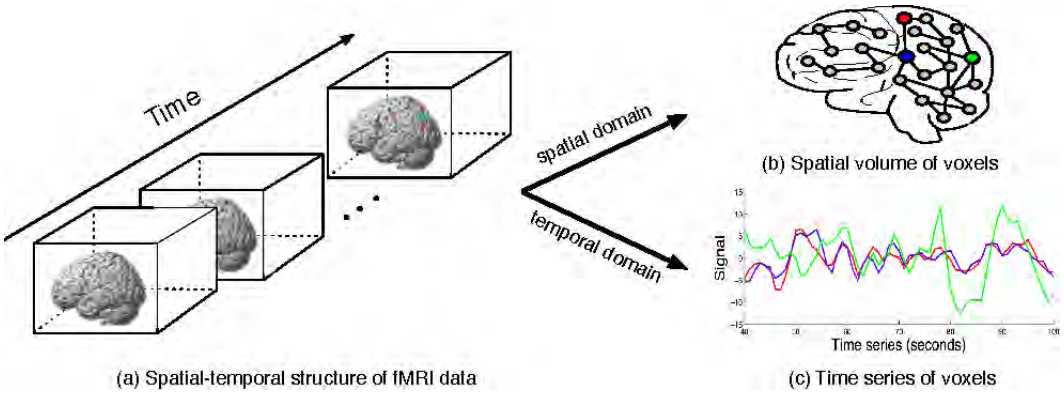

Figure 4.3: Voxel to Time Series Representation

There are many null values included in the time series data as some of the voxels remain outside of the brain. Those null values are removed for better understanding the data. Only the non zero values are included for training and testing machine learning models. Additionally, Time series analysis is a very efficient tool in recent

researches. However, we tried a new dimension for analysis of time series data which is signal processing approach because time series analysis led very efficient results in signal processing technology.
Table 4.2 shows the shapes of the time series data for each of the subjects.

Table 4.2: Time Series Data Table

Serial No.	Sex	Subject ID	Age	Shape of Time Series Data
1	Female	60036	71	(210, 179520)
			73	(210, 179520)
2	Male	60006	82	(210, 179520)
			84	(210, 179520)
3	Male	60035	66	(210, 179520)
			68	(210, 179520)
4	Male	60043	68	(210, 179520)
			70	(202, 179520)
5	Male	60044	73	(210, 179520)
			75	(210, 179520)
6	Male	60073	65	(210, 179520)
			67	(210, 179520)
7	Male	60074	74	(210, 179520)
			76	(210, 179520)
8	Male	60075	66	(210, 179520)
			68	(210, 179520)

4.3 Time Frequency Feature Calculation using Short time Fourier Transform

By analyzing the time series data, time-frequency representation has been done using STFT. From STFT, we have calculated PSD using Fourier Transform methods. The process will be described in the sub sections below.

4.3.1 STFT

STFT is a sequence of Fourier transforms of a windowed signal. It provides the information of time-localized frequency. It is only applicable on variable frequency components of a signal [1], [40]. The Standard Fourier Transform collects information of the frequency throughout the whole signal time interval.
The general mathematical function of STFT is-

$$X_m(\omega) = \sum_{n=-\infty}^{\infty} x(n)w(n - mR)e^{-j\omega n} = \mathrm{DTFT}_\omega(x \cdot \textsc{Shift}_{mR}(w))$$

where $x(n)$ denotes the input signal at time n, $w(n)$ means the window function of length M, X_m denotes DTFT of windowed data centered about time mR and R

means the hop size between successive DTFTs. The hop size is calculated by the difference between the window length M and the overlap length L.
STFT generates a matrix with its complex STFT coefficients having frequency across the rows and time across the columns. The method also generates a frequency vector and a time vector which is later used in finding the power spectral density.
The STFT spectral option furnishes Fourier spectral information for non-stationary data, i.e., for data whose spectral information is affected by time [21]. The STFT is often used to assess whether or not a signal is stationary.
In our research, we used STFT for our feature extraction where it was used for determining the PSD of the signal. In the next part, we will discuss about PSD.

4.3.2 PSD

PSD is the power content ratio per signal frequency. Usually, PSD is used to describe random broadband signals. The spectral resolution used to digitize the signal normalizes the magnitude of the PSD [33]. The shape of the extracted features are shown in table 4.3.
According to Welch's periodogram method, the time signal is divided into continuous blocks and average of squared- magnitude DFTs of the signal blocks is taken and PSD is estimated through this calculation [22]. Let $x_m(n) = x(n + mN)$, $n = 0, 1, \ldots, N-1$, denote the mth block of the signal $x \in C^{MN}$, where M denotes the number of blocks. Then the PSD estimation by Welch is calculated by-

$$\hat{R}_x(\omega_k) = \frac{1}{M} \sum_{m=0}^{M-1} |DFT_k(x_m)|^2 = \left\{ |X_m(\omega_k)|^2 \right\}_m$$

where "$\{\cdot\}_m$" means time averaging across blocks or frames of data indexed by m.

Table 4.3: Extracted Feature Table

Serial No.	Sex	Subject ID	Age	Shape of Extracted Feature (2-D)
1	Female	60036	73	(420,2)
2	Male	60006	84	(420,2)
3	Male	60035	68	(420,2)
4	Male	60043	70	(412,2)
5	Male	60044	75	(420,2)
6	Male	60073	67	(420,2)
7	Male	60074	76	(420,2)
8	Male	60075	68	(420,2)

Here PSD generated the feature vector which is used for the classification through machine learning. In the next section, we will be focusing on the process of classification using the machine learning technique SVM.

4.4 Classification Using SVM Classifier

After figuring out the power spectral density using STFT we have got the extracted feature of each subject of patients. With this extracted feature, we need to use a

classifier to classify the early PD subjects and the normal subjects. The aim of using a classifier is to identify the subject states where it is safe or early. To do this, we used cross-validation machine learning approaches. The main challenge was to locate the mapping with the unknown test sets that generalized the training sets. Using cross-validation, we partitioned our training sets and test sets. We used 5-fold cross-validation for our experiment where we randomly divided each dataset into 5 splits from which 4 were used for training purposes and the fifth was used for testing. We used SVM-based classifier for classification, since SVM is a good time series domain classifier. SVM is a possible problem-solving tool in linear and nonlinear classification, function estimation, and learning methods based on the kernel [13]. By using the SVM classifier we can get the maximum margin hyperplane and minimum operational error. It also maximizes the efficiency of classification [6]. SVM's biggest drawback is its higher computational burden of restricted programming for optimization. [35]. But this drawback can be overcome by LS-SVM based classifier. LS-SVM is the extended version of SVM and is closely linked to regularization networks and Gaussian systems, as well as providing primal-dual definitions. So we will use LS-SVM in our experiment.
For non-linear problem, the kernel function introduced in the decision function can be defined as

$$f(x) = sign[\sum_{i-1}^{N} a_i y_i K(x, x_i) + b]$$

where $K(x, xi)$ is a kernel function, α_i is the Lagrange multiplier [12], b is the bias term and y_i the training output pairs. RBF kernel is used in our experiments and this function can be defined as

$$k(x, x_i) = exp(- \left\| x - x_i \right\|^2 / 2\sigma^2)$$

where σ controls the width of the RBF function [30].

4.5 Prediction

The aim of our research was to predict the early stages of Parkinson's disease by using machine learning models. In accordance to that, supervised learning methods were used to construct the classifier for our research. In short, to generate a desired output from a collection of input data, a supervised machine-learning algorithm was trained. Support Vector Machine (SVM) which is a good classifier [24] in signal processing domain, was used for classification. In addition, MATLAB was used for developing our algorithm. Throughout our research, eight subjects were used for training and testing of which seven were male and the other was female. Data were taken for two years as maximum difference between the years for each subject. For example: subject 60006 had data for consecutive three years like age 82, 83 and 84. Data of age 82 and 84 were taken for training and testing our classifier. Moreover, data of second year were used for training the classifier and first year's data were used for checking the accuracy of the predictor. Additionally, five fold cross validation, a technique for checking the accuracy of the predictor was used for our research. Finally, the predicted output was very accurate compared to the test data set.

Chapter 5

Results and Analysis

5.1 Results

In our paper, the estimation of Parkinson's Disease early stages is proposed on the basis of eight fMRI data from the patient. The fMRI images are converted into time series representation for each voxels in the image. By using STFT technique, features are extracted from the time series representation. Based on the extracted features, an SVM classifier is used to classify the prodromal level. A regulatory-based decision-making approach is implemented to improve accuracy.
In our experiment, we use 8 patient's data of 2 different ages. So, we have a total 16 sets of data. Among those 16 sets of data we trained some and test with the rest of them. We measure average value after training and testing all features to determine sensitivity, specificity and accuracy. The sensitivity, specificity and accuracy are calculated by the following equations:

$$Sensitivity = (TP|(TP + FN)) * 100$$
$$Specificity = (TN|(TN + FP)) * 100$$
$$Accuracy = ((TP + TN)|(TP + TN + FP + FN)) * 100$$

where TP and TN represent, respectively, the total number of true positive events observed and true negative events. The FP and FN are both false positives and false negatives.

Table 5.1: Predicted Results

Serial No.	Sex	Subject ID	Sensitivity(%)	Specificity(%)	Accuracy(%)
1	Female	60036	100	100	100
2	Male	60006	100	100	100
3	Male	60035	100	100	100
4	Male	60043	100	100	100
5	Male	60044	100	100	100
6	Male	60073	99.53	100	99.76
7	Male	60074	100	100	100
8	Male	60075	100	100	100

The proposed method provides high sensitivity, specificity and accuracy. Table 1 shows that the sensitivity, specificity and accuracy are 100% in most of the cases. Only in one subject id (60073), the sensitivity is 99.53% and accuracy is 99.76% but the specificity is still 100%. To compare with [20], the accuracy, sensitivity and specificity is 86.96%, 78.95%, 92.59% respectively which is not higher than the proposed method. The proposed method provides 100% accuracy according to the tests, while the method in [31] provides 96.40% accuracy. The experiment therefore shows that the approach proposed achieves the highest accuracy in predicting the outcome.

5.2 Analysis

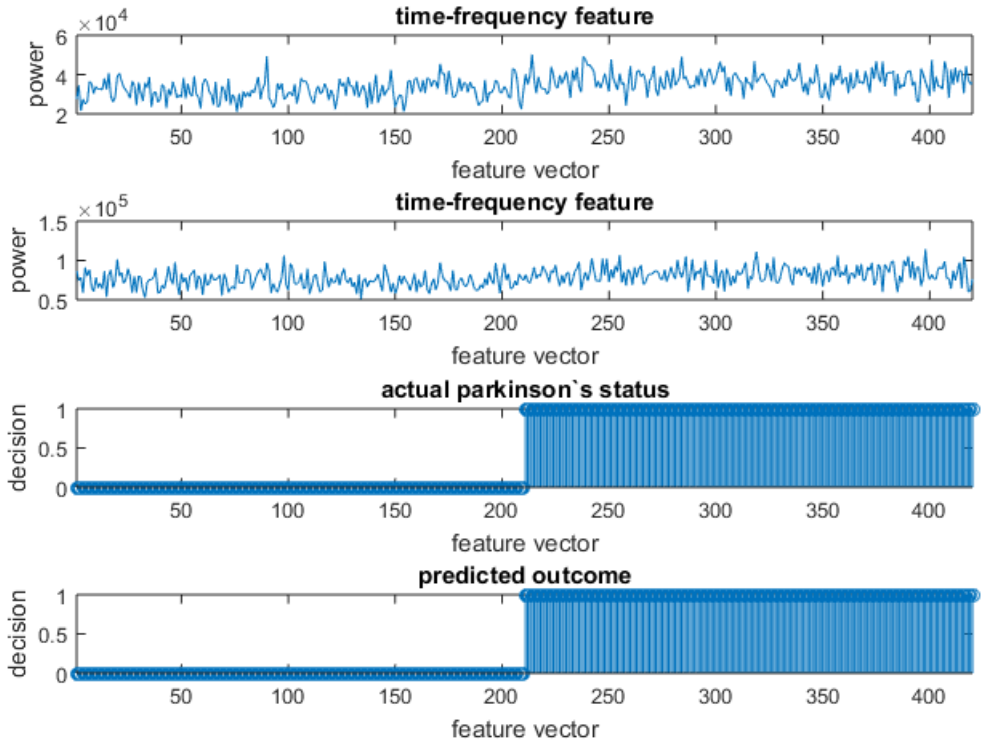

Figure 5.1: Time Frequency feature and predicted outcome for Subject ID: 60006

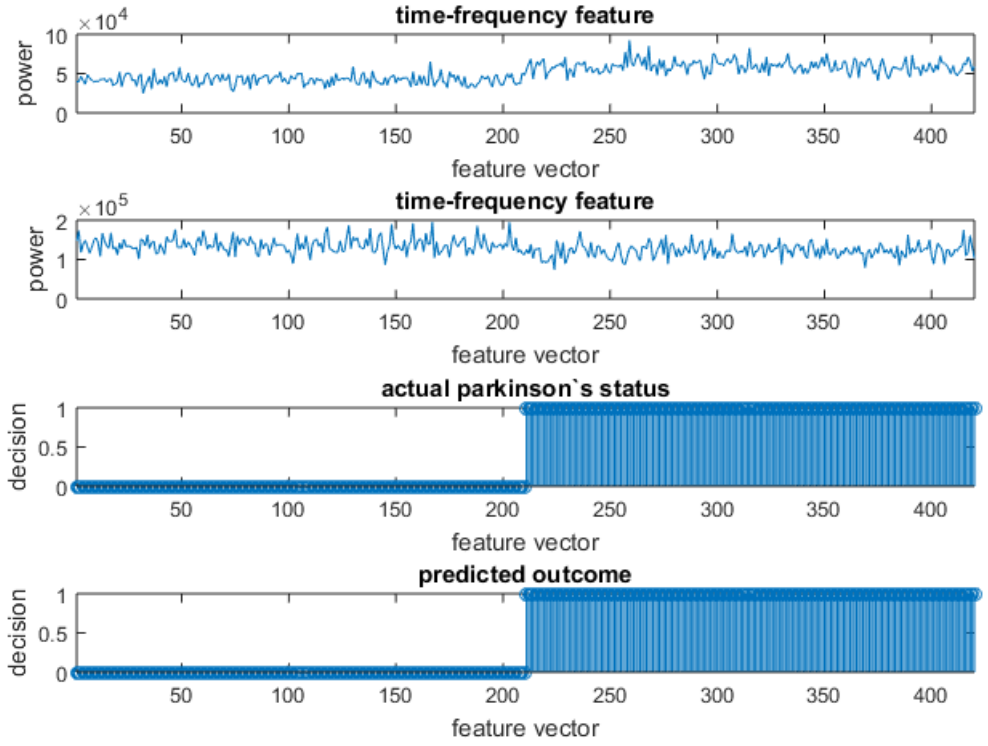

Figure 5.2: Time Frequency feature and predicted outcome for Subject ID: 60044

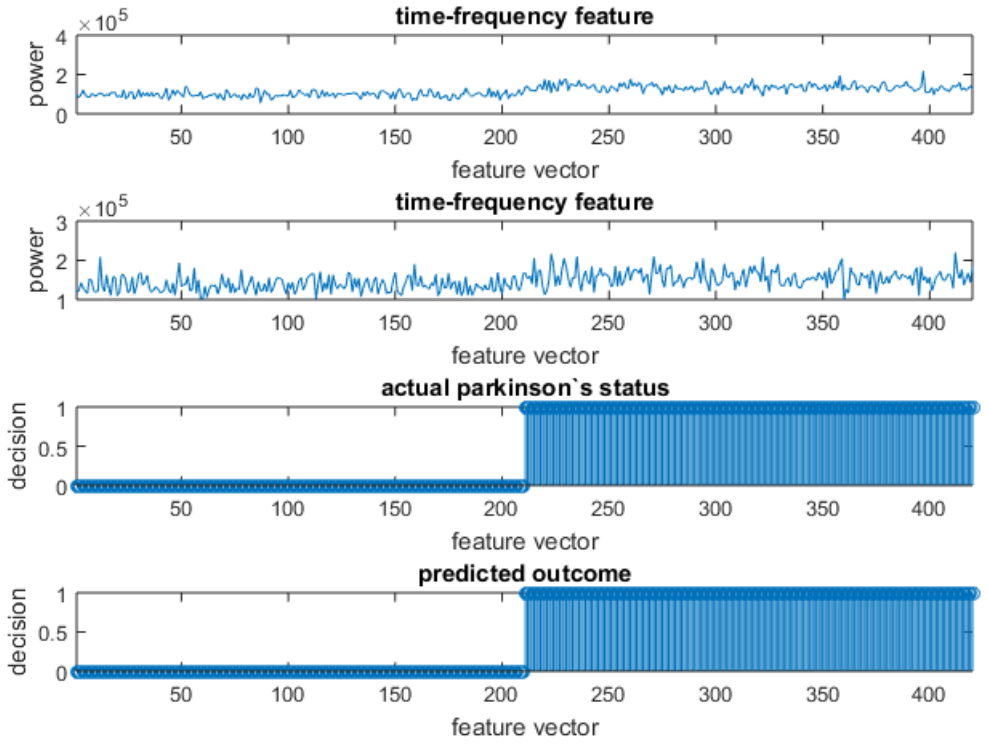

Figure 5.3: Time Frequency feature and predicted outcome for Subject ID: 60073

In this research, time frequency feature is extracted from the voxel intensity where zero values are removed for every subjects. After that a feature vector power spectral density is obtained using STFT. In this analysis, data of two ages from 8 subjects are taken for training and testing. In Figure 5.1-5.3, firstly time frequency feature is shown for both the subjects where the first 210 values are for early age and later 210 values are for the second years data. First years data are labeled as zero vector and second years data are labeled as one. Moreover, the predicted results are accurate compared to the actual data of the early stages of PD. However, one subject was exceptional whereas we have achieved 99.76% accuracy, 100% specificity and 99.53% sensitivity.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

Parkinson's disease is a neurocognitive brain disorder that can result into rigidity, brandy kinesia, postural instability. As the condition of a PD affected patient get worse day by day the early detection process of Parkinson's disease is significant. For our research, we used fMRI data of eight Prodromal PD patients. Moreover, we collected the fMRI image of those PD patients. Time series data was calculated relying on the voxel intensity. Time frequency feature was calculated using STFT. As SVM is a widely used classifier, it was used for the classification and prediction of early PD. We used 5 fold cross validation procedures to detect the accuracy of our proposed machine learning based system. We have used signal processing approach for our research work which has not been done in earlier works. Maximum work has been based on image processing approach, where the accuracy in results were less than what we have found here. We have showed that we can certainly do our feature extraction of fMRI data based on signal processing approach and it can provide us with better prediction results for predicting the early stages of PD.
We hope that our research work will have a great effectiveness in diagnosis of patients of Parkinson's Disease. The early detection will surely enable the doctors to take preventive measures for the patients. This is why, we have worked for detecting the early stages of the disease.

6.2 Future Work

The main goal of our research is to detect the early stages of PD and we have achieved result with a good accuracy. Since, eight subjects were taken into consideration for conducting the research, our future work would be to consider more subjects as possible. In addition, we have a plan to use more signal processing concepts as we found this domain as a new dimension in our research.

References

[1] D. Griffin and Jae Lim, "Signal estimation from modified short-time fourier transform", in *ICASSP '83. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 8, Apr. 1983, pp. 804–807. DOI: 10.1109/ICASSP.1983.1172092.

[2] D. Cox and R. Savoy, "Functional magnetic resonance imaging (fmri) "brain reading": Detecting and classifying distributed patterns of fmri activity in human visual cortex", *NeuroImage*, vol. 19, pp. 261–70, Jul. 2003. DOI: 10.1016/S1053-8119(03)00049-1.

[3] N. Logothetis and B. Wandell, "Interpreting the bold signal", *Annual review of physiology*, vol. 66, pp. 735–69, Feb. 2004. DOI: 10.1146/annurev.physiol.66.082602.092845.

[4] D. Calne, "A definition of parkinson's disease", *Parkinsonism related disorders*, vol. 11 Suppl 1, S39–40, Jul. 2005. DOI: 10.1016/j.parkreldis.2005.01.008.

[5] S. M. Hague, S. Klaffke, and O. Bandmann, "Neurodegenerative disorders: Parkinson's disease and huntington's disease", *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 76, no. 8, pp. 1058–1063, 2005, ISSN: 0022-3050. DOI: 10.1136/jnnp.2004.060186. eprint: https://jnnp.bmj.com/content/76/8/1058.full.pdf. [Online]. Available: https://jnnp.bmj.com/content/76/8/1058.

[6] H. Wang and D. Hu, "Comparison of svm and ls-svm for regression", vol. 1, Nov. 2005, pp. 279–283, ISBN: 0-7803-9422-4. DOI: 10.1109/ICNNB.2005.1614615.

[7] Z. Wang, A. R. Childress, J. Wang, and J. A. Detre, "Support vector machine learning-based fmri data group analysis", *NeuroImage*, vol. 36, no. 4, pp. 1139–1151, 2007, ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2007.03.072. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811907002145.

[8] G. Beattie, "What we know about how the human brain works.", in. Jan. 2008.

[9] B. He and Z. Liu, "Multimodal functional neuroimaging: Integrating functional mri and eeg/meg", *IEEE Reviews in Biomedical Engineering*, vol. 1, pp. 23–40, 2008, ISSN: 1941-1189. DOI: 10.1109/RBME.2008.2008233.

[10] J. Jankovic, "Parkinson's disease: Clinical features and diagnosis", *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 79, no. 4, pp. 368–376, 2008, ISSN: 0022-3050. DOI: 10.1136/jnnp.2007.131045. eprint: https://jnnp.bmj.com/content/79/4/368.full.pdf. [Online]. Available: https://jnnp.bmj.com/content/79/4/368.

[11] R. Cloninger, "Evolution of human brain functions: The functional structure of human consciousness", *The Australian and New Zealand journal of psychiatry*, vol. 43, pp. 994–1006, Nov. 2009. DOI: 10.3109/00048670903270506.

[12] M. Paul, M. Frater, and J. Arnold, "An efficient mode selection prior to the actual encoding for h.264/avc encoder", *IEEE Transactions on Multimedia*, vol. 11, pp. 581–588, Jun. 2009. DOI: 10.1109/TMM.2009.2017610.

[13] S. Abe, *Support Vector Machines for Pattern Classification*. Jan. 2010. DOI: 10.1007/978-1-84996-098-4.

[14] T. Ayodele, "Introduction to machine learning", in. Feb. 2010, ISBN: 978-953-307-034-6. DOI: 10.5772/9394.

[15] S. Lee, S. Halder, A. Kübler, N. Birbaumer, and R. Sitaram, "Effective functional mapping of fmri data with support-vector machines", *Human Brain Mapping*, vol. 31, no. 10, pp. 1502–1511, 2010. DOI: 10.1002/hbm.20955.

[16] K. Niazmand, K. Tonn, A. Kalaras, S. Kammermeier, K. Boetzel, J. H. Mehrkens, and T. C. Lueth, "A measurement device for motion analysis of patients with parkinson's disease using sensor based smart clothes", in *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, May 2011, pp. 9–16. DOI: 10.4108/icst.pervasivehealth.2011.246014.

[17] H. Shafique, "Causes of parkinson's disease: Literature review", *Journal of Parkinsonism and Restless Legs Syndrome*, p. 5, Dec. 2011. DOI: 10.2147/JPRLS.S37041.

[18] T. Arichi, G. Fagiolo, M. Varela, A. Melendez-Calderon, A. Allievi, N. Merchant, N. Tusor, S. Counsell, E. Burdet, C. Beckmann, and D. Edwards, "Development of bold signal hemodynamic responses in the human brain", *NeuroImage*, vol. 63, pp. 663–73, Jul. 2012. DOI: 10.1016/j.neuroimage.2012.06.054.

[19] A. Dimoka, "How to conduct a functional magnetic resonance (fmri) study in social science research", *MIS Quarterly*, vol. 36, pp. 811–840, Sep. 2012. DOI: 10.2307/41703482.

[20] D. Long, J. Wang, M. Xuan, Q. Gu, X. Xu, D. Kong, and M. Zhang, *Automatic classification of early parkinson's disease with multi-modal mr imaging*, 2012.

[21] R. D. Peters, *Tutorial on power spectral density calculations for mechanical oscillators (with an exhaustive discussion of units)*, Jan. 2012. [Online]. Available: http://physics.mercer.edu/hpage/psd-tutorial/psd.html.

[22] D. R. Mehra, "Power spectrum estimation using welch method for various window techniques", *International Journal of Scientific Research Engineering Technology*, vol. 2, pp. 389–392, Sep. 2013.

[23] R. Katzenschlager, "Parkinson's disease: Recent advances", *Journal of neurology*, vol. 261, Apr. 2014. DOI: 10.1007/s00415-014-7308-9.

[24] M. Z. Parvez and M. Paul, "Epileptic seizure detection by analyzing eeg signals using different transformation techniques", *Neurocomputing*, vol. 145, pp. 190–200, 2014.

[25] S. Budday, P. Steinmann, and E. Kuhl, "Physical biology of human brain development", *Frontiers in cellular neuroscience*, vol. 9, p. 257, Jul. 2015. DOI: 10.3389/fncel.2015.00257.

[26] S. Rewar, "A systematic review on parkinson's disease (pd)", *Indian Journal of Research in Pharmacy and Biotechnology*, vol. 3, pp. 2321–5674, Apr. 2015.

[27] R. F. Ahmad, A. S. Malik, H. U. Amin, N. Kamel, and F. Reza, "Classification of cognitive and resting states of the brain using eeg features", in *2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, May 2016, pp. 1–5. DOI: 10.1109/MeMeA.2016.7533741.

[28] M. Filippi, *fMRI Techniques and Protocols*. Jan. 2016, ISBN: 978-1-4939-5609-8. DOI: 10.1007/978-1-4939-5611-1.

[29] G. Ma, L. He, C.-T. Lu, P. S. Yu, L. Shen, and A. B. Ragin, "Spatio-temporal tensor analysis for whole-brain fmri classification", in *SDM*, 2016.

[30] M. Z. Parvez and M. Paul, "Epileptic seizure prediction by exploiting spatiotemporal relationship of eeg signals using phase correlation", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 1, pp. 158–168, Jan. 2016, ISSN: 1558-0210. DOI: 10.1109/TNSRE.2015.2458982.

[31] R. Prashanth, S. Dutta Roy, P. Mandal, and S. Ghosh, "High-accuracy detection of early parkinson's disease through multimodal features and machine learning", *International Journal of Medical Informatics*, vol. 90, Mar. 2016. DOI: 10.1016/j.ijmedinf.2016.03.001.

[32] S. Chandra, M. Srivastav, and N. Chauhan, "Recent advances in parkinson disease", *Pharmaceutical and Biological Evaluations*, vol. 4, p. 141, Jun. 2017. DOI: 10.26510/2394-0859.pbe.2017.22.

[33] A. Goavec, M. Zarudniev, R. Vauché, F. Hameau, J. Gaubert, and E. Mercier, "An efficient method of power spectral density estimation for on-chip ir-uwb transmitter self-calibration", *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 64, no. 3, pp. 686–695, Mar. 2017, ISSN: 1558-0806. DOI: 10.1109/TCSI.2016.2617887.

[34] A. Kazeminejad, S. Golbabaei, and H. Soltanian-Zadeh, "Graph theoretical metrics and machine learning for diagnosis of parkinson's disease using rs-fmri", in *2017 Artificial Intelligence and Signal Processing Conference (AISP)*, Oct. 2017, pp. 134–139. DOI: 10.1109/AISP.2017.8324124.

[35] F. Sun, X. Li, H. Liao, and X. Zhang, "A bayesian least-squares support vector machine method for predicting the remaining useful life of a microwave component", *Advances in Mechanical Engineering*, vol. 9, no. 1, p. 1 687 814 016 685 963, 2017. DOI: 10.1177/1687814016685963. eprint: https://doi.org/10.1177/1687814016685963. [Online]. Available: https://doi.org/10.1177/1687814016685963.

[36] Y. Brazier, *Parkinson's disease: Early signs, causes, and risk factors*, Oct. 2018. [Online]. Available: https://www.medicalnewstoday.com/articles/323396.php#causes.

[37] S. Esmail, "The diagnosis and management of parkinson's disease", Dec. 2018.

[38] V. Jatana, *Machine Learning For Beginners*. Oct. 2018.

[39] M. Lawton, Y. Ben-Shlomo, M. T. May, F. Baig, T. R. Barber, J. C. Klein, D. M. A. Swallow, N. Malek, K. A. Grosset, N. Bajaj, and et al., *Developing and validating parkinson's disease subtypes and their motor and cognitive progression*, Dec. 2018.

[40] L. Li, H. Cai, H. Han, Q. Jiang, and H. Ji, *Adaptive short-time fourier transform and synchrosqueezing transform for non-stationary signal separation*, 2018. arXiv: 1812.11292 [eess.SP].

[41] X. Teng and Y. Gong, "Research on application of machine learning in data mining", *IOP Conference Series: Materials Science and Engineering*, vol. 392, p. 062 202, Aug. 2018. DOI: 10.1088/1757-899X/392/6/062202.

[42] I. Zahoor, A. Shafi, and E. Haq, "Pharmacological treatment of parkinson's disease", in. Dec. 2018, pp. 129–144, ISBN: 9780994438164. DOI: 10.15586/codonpublications.parkinsonsdisease.2018.ch7.

[43] J. Garza-Ulloa, "Update on parkinson's disease", *Bio/Technolgy*, vol. 2, May 2019. DOI: 10.34297/AJBSR.2019.02.000614.