Intelligent Parkinson's Disease Classification and Progress Monitoring using Non-invasive Techniques

by

Sara Soltaninejad

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Computing Science

University of Alberta

Abstract

Parkinson's Disease (PD) is the second major neuro-degenerative disorder caused by dopaminergic loss in the brain region known as the Basal Ganglia (BG) especially Substantia Nigra (SN). The major symptoms of this disease are motor abnormalities. In addition, it also contains non-motor symptoms, such as cognitive deficits, which may show at early stages of PD. Physical exam, demographic characteristics, and neuroimaging analysis have been commonly used to diagnose PD. However, such assessments by clinicians are subjective, time-consuming, expensive, and prone to error. Furthermore, the underlying causes are often not easily identified, especially in an early stage of the disease. Thus, establishing fast and accurate diagnosis can be difficult, even by experienced specialists. In my research, I have studied smart, non-invasive, and practical approaches to help clinicians assess PD, intending to diagnose and monitor the disease progression.

The first section of this research is gait analysis for PD. We propose a non-invasive approach using movement data captured by Kinect to monitor the motor deficits of PD patients. The motion data of standard exercises, normally performed in rehabilitation sessions by physiotherapists, are collected including Step Length (SL), tremor, and Time Up & Go (TUG). The standard medical Unified Parkinson's Disease Rating Scale (UPDRS) is used by a physical therapist to determine the level of severity as the ground truth. The general framework after obtaining the motion data includes two steps; feature extraction from the kinematic motion data, and classification using Random Forest (RF)

(for SL and tremor data) and K-means (for TUG data). Freezing of Gait (FOG) is one of the most incapacitating motor symptoms for PD, especially in the later stages of the disease. FOG is a short absence or reduction in the ability to walk for PD patients that may result in a fall, decrease in patients' quality of life, and even death. Existing FOG assessments by doctors are based on a patient's diaries and experts' manual video analysis, which give subjective, inaccurate, and unreliable results. In the present research, an automatic FOG assessment system (Kin-FOG) is designed for PD patients to provide objective information to neurologists about the FOG condition and the symptoms' characteristics. The proposed FOG assessment system utilizes Kinect for capturing data. The analysis of foot joint trajectory of the motion captured by Kinect is used to find the FOG episodes. The standing mode is similar to a FOG episode. Thus, Kin-FOG uses the gradient displacement of the angle between the foot and the ground for discriminating between FOG and standing modes.

Clinical characteristics of the patients have essential information for the specialist in PD diagnosis. However, assessment of this information by doctors can also be subjective, vulnerable to human errors and inefficient. Therefore, the second section of my research targets automatic, early and non-invasive assessment of PD using clinical properties with machine learning. Feature selection is conducted by Mean Decrease Impurity (MDI) method. In the classification step, RF is used as a classifier model for two goals; PD Diagnosis, and PD progress monitoring with Hoehn & Yahr (H&Y) scale.

Neuroimaging has been successfully used for diagnosing the neurological disease. Magnetic Resonance Imaging (MRI) is one of the most popular methods due to its non-invasive nature and high resolution images. The most common MRI sequences are T1-weighted and T2-weighted scans. In the third section of this research, data analysis is conducted for PD diagnosis using MR images with machine learning. The proposed method follows four steps; preprocessing,

feature extraction, feature selection, and classification. Preprocessing pipeline is performed using different medical libraries; Freesurfer, SPM/CAT12, and FSL. Thereafter, subcortical and region-based features are extracted from the preprocessed MR images. Feature selection or dimensionality reduction is performed by Principal Component Analysis (PCA). In the classification, two important machine learning algorithms are used, Support Vector Machine (SVM) and RF. Furthermore, we assess four deep-learning based models that classify patients based on the biomarkers in MRI data. In the last part of the neuroimaging analysis, SN region in MRI T2 and T1 are evaluated using Local Binary Pattern (LBP) and Histogram Oriented Gradient (HOG) features. For each subject, few slices around the center of a MRI DICOM volume (midbrain area) are selected as Slice of Interest (SOI). In each slice, the area around SN becomes the Region of Interest (ROI) for further analysis. RF and SVM, with or without PCA, are used for classification. The ability of the proposed system is compared for these two popular MR imaging modules based on the different number of scans.

To obtain more experimental data, our research group has started collaboration with Arizona State University (ASU), which provided us with the motion data for PD and HC subjects. The dataset consisted of motion properties for a group of PD patients that may or may not have FOG. The objective is to use machine learning methods for classifying the data into two groups; FOG and Non-FOG.

Preface

The contents of this thesis has been published or are under review in the journals and conferences. The contents of Chapter 3 presents the work in Parkinson's Disease (PD) body movement monitoring using a smart sensor based non-invasive technique and has been published in IEEE Healthcom 2018. Chapter 4 details our method of automatic simulated Freezing of Gait (FOG) assessment system for PD and has been published in Sensors journal. Chapter 5 is learning demographic and clinical features for automatic PD monitoring and classification which has been published in IEEE EMBS 2019.

Chapters 6,7,8 and 9 provide our contributions using neuroimaging analysis using MRI T1 and T2. Chapter 6 is presenting the results of PD identification using T1 images that has been published in ICSM 2018 conference. Chapter 7 is the result of our study for PD classification using region-based analysis for T1 MR images. This work is going to be submitted to the ICME conference. Chapter 8 is on study of different deep learning models for PD classification and has been published in ICSM 2019. In chapter 9, the mid-brain scan assessment is done for PD patients using MR T1 and T2 images. This work is going to be submit to the Medical & Biological Engineering & Computing journal. The proposed method for T2 MRI data has been published in IEEE BIBE 2019 conference. I choose to use first person plural throughout this thesis to honor the contributions of my advisors and collaborators on my various work.

Publications during PhD study

- **Sarah Soltaninejad**, Irene Cheng, Anup Basu, Parkinson's Disease Mid-brain Assessment using MR T2 Images, BIBE 2019.
- **Sarah Soltaninejad**, Irene Cheng, Anup Basu, Kin-FOG: Automatic Simulated Freezing of Gait (FOG) Assessment System for Parkinson's Disease, Sensors 2019, 19(10), 2416.
- **Sarah Soltaninejad**, Irene Cheng, Anup Basu, Automatic Classification and Monitoring of Denovo Parkinson's Disease by Learning Demographic and Clinical Features, Engineering in Medicine and Biology Conference (EMBC), 2019: 3968-3971.
- Chris West, **Sarah Soltaninejad**, Irene Cheng, Assessing the Capability of Deep-Learning Models in Parkinson's Disease Diagnosis, International Conference on Smart Multimedia (ICSM), 2019.
- Harsh Sharma, **Sarah Soltaninejad**, Irene Cheng, Automated classification of Parkinson's Disease using Diffusion Tensor Imaging Data, Reverse EXPO, 2019.
- **Sara Soltaninejad**, Andres Rosales-Castellanos, Fang Ba, Mario Alberto Ibarra-Manzano, Irene Cheng, Body movement monitoring for parkinson's disease patients using a smart sensor based non-invasive technique, IEEE International Conference on E-health Networking, Application & Services (IEEE-Healthcom), 2018.
- **Sarah Soltaninejad**, Irene Cheng, Anup Basu, Towards the identification of parkinson's disease using only T1 MR Images, International Conference on Smart Multimedia (ICSM), 2018.

Acknowledgements

Undertaking this PhD has been a truly life-changing experience for me and it would not have been possible to do without the support and guidance that I received from many people.

Firstly, I would like to thank my advisors Dr. Anup Basu and Dr. Irene Cheng, whose expertise was invaluable in the formulating of the research. You are truly a great inspiration for me. Accept my heartfelt gratitude for your time, support, and patience.

I gratefully acknowledge the funding received towards my PhD from the Alberta Innovates Graduate Student (AIGS) scholarship. I appreciate the collaborative efforts of my colleagues Gabriel Lugo Bustillo, and Andres Rosales from the University of Guanajuato, Mexico, for collecting the data. I would like to thank Chris West at the University of British Columbia, and Harsh Sharma at the University of Alberta. I want to thank you for your excellent cooperation and for all of the opportunities I was given to conduct my research and further my dissertation. I am also thankful to my friends at the University of Alberta especially Multimedia Research Center (MRC): Xinyao, Nasim, Subhayan, Esmat, Housam, Pouneh, Chenqiu, Harini, Arjun, Telly for all the fun discussions on life and research over coffee.

I would like to thank my thesis committee: Dr. Osmar Zaiane, Dr. Pierre Boulanger, and Dr. Bin Zheng, for their valuable guidance. You provided me with the tools that I needed to choose the right direction and successfully complete my dissertation. My sincere thanks to Dr. Fang Ba, for her valuable guidance, timely suggestions and support throughout my research project work.

I acknowledge the people who mean a lot to me, my parents, Rostam and Khavar, for showing faith in me and giving me liberty to choose what I desired.

You are always there for me. It's my fortune to gratefully acknowledge the support of my family, Forogh, Maryam, Arezoo, Hashem, and Sadegh. I owe thanks to a very special person, my husband, Amir for his continued and unfailing love, support and understanding during my pursuit of Ph.D degree that made the completion of thesis possible. You were always around at times I thought that it is impossible to continue, you helped me to keep things in perspective. In addition, I would like to thank my in law family, Zaman, Salimeh, and Aria for their wise counsel and sympathetic ear.

Lastly, I would like to thank every single person, who directly and indirectly supported me with my Ph.D. journey.

Contents

List of Tables

List of Figures

Acronyms

ADL Activities of Daily Living

AD Alzheimer's Disease

BG Basal Ganglia

CAD Computer Aided Diagnosis

CAT Computational Anatomy Toolbox

CNN Convolutional Neural Networks

COM Center of Mass

CP Cerebral Palsy

CSF CerebroSpinal Fluid

DBM Deformation-Based Morphometry

DL Deep Learning

DOP Direction of Progress

DSS Decision Support System

DTI Diffusion Tensor Imaging

DWT Discrete Wavelet Transform

EEG Electroencephalography

EMG Electromyography

ET Essential Tremor

FA Fractional Anisotropy

FFT Fast Fourier Transform

FN False Negative

FOG Freezing Of Gait

FP False Positive

FSL FMRIB Software Library

GB Gradient Boosting

GM Gray Matter

GV Gait Velocity

HAY Hoehn & Yahr

HC Healthy Control

HOG Histogram of Oriented Gradients

HR Heart Rate

ICA Independent Component Analysis

IMU Inertial Measurement Units

JFSS Joint Feature-Sample Selection

KNN K-Nearest Neighbor

LBP Local Binary Pattern

LPP Locality Preserving Projection

LR Logistic Regression

MBFA Moore-Bachlin FoG Algorithm

MDI Mean Decrease Impurity

MDS Movement Disorder Society

MD Movement Disorder

MNI Montreal Neurological Institute

MRI Magnetic Resonance Imaging

MRMR Minimum Redundancy and Maximum Relevancy

MSA Multiple System Atrophy

NB Naive Bayes

NMS-MRI Neuromelanin Sensitive Magnetic Resonance Imaging

NMS Non-Motor Symptoms

PCA Principal Component Analysis

PD Parkinson's Disease

PET Positron Emission Tomography

PPMI Parkinson's Progression Markers Initiative

PSD Power Spectral Density

PSP Progressive Supranuclear Palsy

QSM Quantitative Susceptibility Mapping

RBF Radial Basis Function

RBM Region-Based Morphometry

RFE Recursive Feature Elimination

RFR Radio Frequency

RF Random Forest

RIPPER Repeated Incremental Pruning to Produce Error Reduction

RN Red Nucleus

ROC Receiver Operating Characteristic

ROI Region Of Interest

RT Repetition Time

SBM Surface-Based Morphometry

SBR Striatal Binding Ratio

SCN Structural Covariance Networks

SDK Software Development Kit

SL Stride Length

SN Substantia Nigra

SOI Slice Of Interest

SPECT Single Photon Emission Computed Tomography

SPM Statistical Parametric Mapping

STD Standard Deviation

SVM Support Vector Machine

SWEDD Scan Without Evidence of Dopaminergic Deficit

SW Simple Walking

TE Time to Echo

TN True Negative

TP True Positive

TTC Time To Complete the Task

TUG Timed Up & Go

U-AMS Unconstrained Activity Monitoring System

UPDRS Unified Parkinson's Disease Rating Scale

VBM Voxel Based Morphometry

WM White Matter

WOST Without Standing

WST With Standing

WWT Walking With Turning

Chapter 1

Introduction

Movement disorders are a group of nervous system (neurological) abnormalities that cause uncontrollable body movements, lack of coordination, muscle tightness, or increased difficulty in walking or hand use which may be voluntary or involuntary. Movement disorders are neurological and require an accurate diagnosis, along with options for treatment, for patients to remain in control of their bodies and lives. There are different movement disorder diseases among which the most prevalent ones include: Essential Tremor (ET), Parkinson's Disease (PD), Multiple System Atrophy (MSA), Progressive Supranuclear Palsy (PSP), and Dementia with Lewy bodies (diffuse Lewy Body Disease).

PD, the second most common movement disorder, is a progressive neurodegenerative disease that mostly affects seniors [106]. The incidence of PD, or the rate of newly diagnosed cases, generally increases with age, although it can stabilize in people who are older than 80. An estimated 4 percent of people with PD are diagnosed before the age of 50. Men are 1.5 times more likely to have the disease than women [73]. PD is caused by the loss of dopaminergic neurons in the part of the brain known as the Substantia Nigra (SN). The substantia nigra (SN) is located in Basal Ganglia (BG) structure in the midbrain that plays an important role in reward and movement. Thus, the most important symptoms, which people with PD manifest are motor symptoms including tremor, bradykinesia, postural impairments, and rigidity [92]. In many patients, subsequent non-motor problems may arise such as cognitive, mood, sleep, olfactory, with dementia commonly occurring in the advanced stages of the

disease [73].

PD can be difficult to diagnose accurately, particularly in the early stages of the disease when symptoms resemble other medical conditions, and misdiagnoses occur occasionally [138]. There are currently no blood or laboratory tests that have been proven to help in diagnosing PD. The diagnosis is based on the medical demographic characteristics, a physical assessment that is conducted by interviewing and observing the patient, and neurological imaging as complementary information.

There are three general research problems which will be addressed in this thesis.

- Specialists use physical examination in their clinics to diagnose PD or evaluate the patient's situation if they already have PD. This process relies on expertise, time consuming, prone to human error, and might be subjective. Therefore, our goal is to design a computer aided system using machine learning methods to make this evaluation process more efficient, easier, faster, and more accurate.

- Another approach to evaluate PD is to use clinical and demographic information. However, doctors need to spend a lot of time to assess this large amount of data, which makes this process difficult, time consuming, and can be inaccurate. Thus, the second objective of this research is to make this assessment easier, accurate, and more automated using machine learning techniques.

- Neuroimaging is another source of information for clinicians to assess PD patients. However, the lack of visible sign of PD in Magnetic Resonance Imaging (MRI) makes it not useful as a solo source for PD assessment. Therefore, we are going to design a Computer Aided Diagnosis (CAD) system using MRI as a first step to discover relevant biomarkers in the brain to diagnose PD.

The present research has three sections. The first section is gait analysis for automatic PD assessment using non-invasive kinematic motion data. The gait

analysis section has two parts. In the first part, machine learning techniques are used for PD diagnosis and progress monitoring using the collected motion data with Kinect in different experiments done by PD patients. In the second part of gait analysis section, an automatic Freezing Of Gait (FOG) assessment system is designed to provide objective information to neurologists about the FOG condition and the symptom's characteristics of their PD patients. In the second section of this thesis, the demographic and clinical characteristics are used to facilitate diagnosis and disease level determination. Machine learning methods are used for establishing these purposes. In the last part, the wide range of analysis from traditional computer vision techniques to deep learning methods has been used to diagnose PD with MRI, especially at an early stage of the disease. Two popular types of MRI, T1 and T2, are used for PD classification. In this part, a CAD system is proposed with just a few midbrain scans. The SN is selected as Region Of Interest (ROI) for further analysis using machine learning and computer vision techniques. The ability of this system is evaluated with different number of slices for T1 and T2. The output of the system is compared for these two imaging modalities.

There has been increasing interest in recent years on using smart sensor technology, e.g., Kinect and Leap Motion, to capture and analyze human body movements, to benefit not only games but also health care and rehabilitation applications. The motivation of the first part of the motion analysis section is proposing a non-invasive approach using movement data captured from Kinect to diagnose and monitor motor deficits for PD patients. In the first part of gait analysis section of our research, we captured and evaluated simple exercises, normally performed in rehabilitation sessions by the physical therapist: Stride Length (SL), Tremor, and Timed Up & Go (TUG). The standard medical Unified Parkinson's Disease Rating Scale (UPDRS) scale is used by a physical therapist to determine the level of severity as the ground truth. Depending on the type of experiment, different gait features are extracted. Random Forest (RF) is used for classification for Tremor and SL, and K-means is used for clustering the TUG data.

Although there are a lot of motor symptoms for PD, they can vary greatly

from individual to individual—both in terms of intensity and how they progress. However, some of them are common in most PD patients. FOG is one of the most debilitating motor symptoms in patients with PD as it may lead to falls and loss of independence. There are many characteristics related to FOG that seem to differ from the cardinal features of PD and is still largely unexplored. The current clinical FOG assessments are based on patient self-report diaries and manual video analysis by neuroscientists. Both these practices are time-consuming, imprecise, and subjective. Therefore, in the second part of the gait analysis section, the goal is designing an automatic FOG detection system, which provides the essential information of PD patients to the doctors and care providers. This system could have remote application for the patients and with more capabilities such as disease progress visualization and other useful medical information.

Since the specialist must assess much demographic information and clinical properties for the patients, the diagnosis process is very time consuming, subjective and difficult. Thus, having an automated CAD system for evaluating the demographic and clinical features in PD diagnosis and progress monitoring is helpful for the medical community. On the other hand, there are some widely used rating scales for PD diagnosis and progress monitoring, such as UPDRS, Hoehn & Yahr (HAY) scale, Schwab and England, Activities of Daily Living (ADL) scale, PD-Non-Motor Symptoms (NMS) questionnaire and NMS survey. Since these scales may not give reliable results, a CAD technology needs to be designed for accurate detection and progress monitoring of PD [39].

Motor symptoms have been used for proposing high accuracy automatic and semi-automatic PD diagnosis and disease stage determination. However, most of the motor signs of PD will appear at a later stage of the disease where most of the dopaminergic cells are lost. Moreover, any neuroprotective therapy initiated at such a late stage may have fewer substantial effects on disease progression. Thus, it is crucial to find out some valid and objective biomarkers to distinguish early PD patients from the healthy population [90]. In recent years, neuroimaging has been increasingly used as an objective method for

the early diagnosis of PD and other neurological diseases [90]. There are some published research for diagnosing, progress monitoring and assessment of PD using different neuroimaging data such as MRI [130], Positron Emission Tomography (PET) [89], Single Photon Emission Computed Tomography (SPECT) [36], and Diffusion Tensor Imaging (DTI)[17]. MRI can provide evidence of structural changes that occur as a result of dopamine neurons loss in the SN, as well as the loss of non-dopaminergic neurons in other brain regions. Therefore, MRI methods that are sensitive to detect these tissue changes may prove useful as a biomarker for PD [158]. Because of its high-resolution contrast, ready availability, non-invasive nature and no need for any pharmaceutical injections, many have used structural MRI methods in their pursuit of developing a biomarker for PD [39]. However, there are still plenty of unexplored approaches in this domain. The analysis in this study has been done for two popular types of MR Images: T1 and T2. In this research, different medical packages were used for preprocessing MR T1 data including Freesurfre, FMRIB Software Library (FSL), and Statistical Parametric Mapping (SPM)/Computational Anatomy Toolbox (CAT)12. The preprocessed data are used for different analysis in order to classify patients with and without PD.

The brain has three main parts: Gray Matter (GM), White Matter (WM) and CerebroSpinal Fluid (CSF). In our research, the region-based analysis is established for PD diagnosis using using GM and WM in the MRI T1 images. These regions are extracted by preprocessing MRI T1 data with SPM/CAT12. In addition, sub-cortical features are analyzed with machine learning techniques to distinguish between PD and Healthy Control (HC) T1 samples. The sub-cortical features are obtained by Freesurfer library used for MRI preprocessing pipeline. After feature extraction, Support Vector Machine (SVM) and RF methods are used as classifiers for neuroimaging analysis with or without Principal Component Analysis (PCA) as feature selection for dimensionality reduction. Furthermore, a group of 2D and 3D Deep Learning (DL) models are evaluated for PD classification using T1 MR data. These models are applied on preprocessed MR T1 samples generated by the FSL package. As deep learning

methods require a lot of data for training that may not be practical to acquire, a data analysis has been done by computer vision techniques and machine learning methods for both T1 and T2 data. These analysis are done for SN region by selecting just a few midbrain scans of the brain. The main goal behind this part of our research was comparing the ability of the proposed system for T1 and T2 MR image based on the different number of MR scans instead of the whole 3D volume.

1.1 Contributions

The contributions of the present thesis can be categorized into three parts:

- Gait Analysis for PD:
  - Designing a body movement monitoring system for PD patients using a smart sensor based non-invasive technique. The proposed system facilitates the diagnosis and progress monitoring of the disease using a Kinect, and can be used by not only the patients but also the specialists and care providers.
  - Designing an automatic simulated FOG assessment system for PD (Kin-FOG). Kin-FOG helps the doctors to obtain the required information about FOG for their PD patients in a faster and easier manner .
- Clinical Data Analysis for PD
  - Designing an automatic classification and monitoring of Denovo PD by learning demographic and clinical features. The proposed system uses machine learning strategies for an early diagnosis and progress monitoring using patient's clinical information.
- Neuroimaging Data Analysis for PD:
  - Toward the identification of PD using T1 MR images. The proposed method uses sub-cortical features including volume, area, mean

curvature, thickness, and their fusion with different machine learning techniques for PD diagnosis just with T1 MR Images.

- PD Classification using GM/WM volume of T1 MR Images. We propose a classification model that can classify subjects with and without PD based on structural T1 MRI GM/WM volume features and their fusion with machine learning methods.
- Assessing the Capability of Deep-Learning Models in PD Diagnosis. We propose and assess four deep-learning based models that classify patients based on biomarkers found in structural T1 MRI.
- PD midbrain assessment using MR T1 and T2 images. We contribute a system for PD diagnosis using Local Binary Pattern (LBP) and Histogram of Oriented Gradients (HOG) features and their fusion with machine learning techniques using only a few midbrain scans from MR T1 and T2 images. The capability of the system is compared for the two types of images and also for number of scans.

1.2 Challenges

In this research, machine learning methods are used for PD diagnosis and progress monitoring using different types of healthcare data. There are several obstacles for integration of machine learning in healthcare today. The main challenge is the ability to make a dataset which have the necessary size and quality of samples for training the state-of-the-art machine learning models. The data collection process is hard, time consuming and sometimes dangerous for the patients. Furthermore, cleaning and preparation of the medical data is one of the essential phase before machine learning analyses due to the inconsistency in format and quality of data [57].

Another major challenge to effective healthcare data analytic is skewed data class distribution, which is referred to as the imbalanced classification problem [175]. An imbalanced classification problem occurs when there is a big gap between the number of samples in different classes in a dataset. For example, in our research, the imbalanced classification problem exists since

the number of samples for HC group has significantly fewer observations than the PD class. The former is usually called a minority class, and the latter, a majority class. In this study, we develop our proposed methods for fixing the imbalance problem before going to the data analysis part.

Once the dataset is built, there are still several other challenges that must be overcome before proceeding with the data analysis. In motion analysis, walking styles of PD patients differ across subjects (including diverse motor anomalies) [99]. Conventional gait assessment for patients is mainly conducted by a clinician via visual inspection and observation which produce results that are subjective and rely on the observer's experience [127]. Thus, it is crucial to get the knowledge of the rehabilitation process by quantitative gait analysis, and understand to what extent does a patient recover from the disease [127]. On the other hand, automatic prediction of FOG is essential for generating the cueing only when a FOG event occurs for PD patients. FOG investigation is challenging because of its unpredictable and unreliable nature [159]. The current treatments and cueing techniques are just temporary solutions for limiting a FOG event time. Moreover, there are many unknown aspects of the pathophysiology of FOG and its relation to PD. In the neuroimaging part of this research, MRI data is used for PD assessment. Lack of clear visual sign for PD in MR images make this process extremely challenging. In addition, there are a few key challenges such as the identification of features and how to efficiently select the most important ones and combine them into our model.

1.3 Organization of the thesis

The rest of the thesis document is organized as follows: In Chapter 2, we cover the background research and related works in the areas of gait analysis, demographic and clinical analysis, and neuroimaging research. In Chapter 3 and 4, we detail our study for PD diagnosis and progress monitoring with gait analysis, and the proposed Kin-FOG system for PD. Chapter 5 covers our findings on automatic analysis of demographic and clinical information using machine learning methods for PD diagnosis and PD severity determination. In

Chapter 6, 7,8 and 9 we list our approaches in neuroimaging analysis for PD assessment. In Chapter 6 we present the results of a study using sub-cortical features from MRI T1 images for PD classification. Chapter 7 presents our study for PD diagnosis using GM and WM volume. In Chapter 8, we report our results of different deep learning models with preprocessed MR T1 images for PD diagnosis. Chapter 9, presents our research for comparing the ability of T1 and T2 MR images with just a few midbrain scans for PD classification. Finally, in Chapter 10 we state our conclusions and indicate directions for future work.

Chapter 2

Background and Related Work

There have been considerable research in PD assessment for diagnosis and progress monitoring by different types of data from human motion to neuroimaging. The publications are categorized based on the data type is used. Three types of data has been assessed, including kinematic human motion, clinical, and demographic data and neuroimaging data.

2.0.1 Assessment of Parkinson's Disease using Gait Analysis

Current assessment of PD patients' physical performance and their disease progress is based on visual assessment of a trained physical therapist or occupational therapist which can be subjective and inaccurate due to human error. In recent years, sensors and computer technology have gained popularity for motion capturing (Mocap) which results in better clinical and behavioral assessment, and more efficient therapeutic decision [105]. Motor analysis can be performed with the help of wearable sensors [119], [124]. However, wearable devices are vulnerable to an uncontrolled environment, affected by noises and undesirable movements. We adopt an alternative approach in favor of using non-invasive sensors such as Microsoft Kinect. Due to its affordable cost, users can purchase their own and capture movements at their own convenience. There is research taking advantage of the versatility of Kinect to provide remote assistance to elderly individuals [21], [52], which provides a mean outside the clinic or rehabilitation center to assess progress of patients. Remote assessment can

reduce visits to clinics and transportation costs, leading to shorter wait-time for those who have urgent needs.

It is non-invasive because only the joint trajectories, not the faces, are used for analysis. Many researchers believe that Kinect offers a simple to use and non-invasive assessment method for PD [21], [136], suitable for the elderly population. Kinect has been tested as a reliable system for healthcare applications. Mobini et al. [101] demonstrated its feasibility for home-based rehabilitation of stroke patients. Hondori et al. [105] showed Kinect to be far more accurate than RGB vision based systems, and less costly and more accurate compared to optoelectronic systems. When it comes to wearable and non-wearable systems, Muro-de-la-Herran et al. [107] indicated that while wearable devices offer a convenient solution for a minimum number of specific individualized tasks, non-wearable systems are better suited for more complex in-depth analysis, which requires simultaneous correlation analysis of a higher number of parameters.

The improvement of algorithms together with advancing sensor technology has made progress in computer assisted techniques towards assessment of motion data. For example, extensive studies have demonstrated that Machine Learning techniques provide effective tools for diagnosis [58], [117], [120], [123]. However, machine learning techniques, especially deep learning, require a very large training dataset, which is often unavailable from clinics and rehabilitation centers. Our challenge is how to efficiently apply machine learning techniques for movement analysis given the data constraint.

In the past decade, smart sensors, especially wearable ones have increasingly become a tool for assessment of motor symptoms such as FOG in PD and other movement disorders. This is because of improvements in computational power of small devices [30]. The proposed methods for FOG detection using these sensors are categorized into two groups depending on the type of the signal and the analysis method used. The first group uses electrical signals and the second group is based on gait information [169]. The first group assesses FOG by monitoring the physiological changes of signals, such as Electroencephalography (EEG) [66] and Electromyography (EMG) [11]. Despite a large number of

studies that investigated the use of wearable sensors to detect gait disturbances, such as FOG and falls, there is little agreement regarding the most effective system design, e.g., types of sensors, number of sensors, location of the sensors on the body, and signal processing algorithms [43]. In [66], it is shown that during FOG episodes, the total amount of EMG activity is reduced in the lower limb muscles. The detection and monitoring of FOG with gait information including kinematic and kinetics is more convenient and easier [169]. The proposed methods for FOG detection is based on shallow machine learning algorithms which are applied to the signals acquired from the sensors. The researchers aim to extract some features which can distinguish FOG episodes from normal gait. As the first attempt in this field, Han et al. [64], explored FOG episodes and movement abnormalities in PD patients and other movement disorders by using an Unconstrained Activity Monitoring System (U-AMS). The proposed method used wavelets to discriminate between FOG and normal walking. Fast Fourier Transform (FFT) and amplitude analysis were used as the features to classify between FOG and non-FOG regions. The dataset consisted of 5 PD and 2 normal subjects. The authors reported that the frequency response range for the 2 patients with accelerometers at the ankle was between 6 and 8 Hz. Moore et al. [102], presented a device for ambulatory monitoring of FOG using the frequency characteristics of vertical leg movement. They defined the freeze index (FI) as the power of the considered body segment acceleration signal in the "freeze" band (3–8 Hz) divided by the power of the signal in the "locomotor" band (0.5–3 Hz). A FOG event was detected when FI exceeded a certain threshold [156]. The experimental results were subject-dependent, and showed 78% correct detection of FOG (true positive rate) and 20% false positive rate.

The proposed methods for FOG detection can be real-time or offline. The first category is suitable for real-time applications. In 2009, Bachlin et al. [19] presented the Moore-Bachlin FoG Algorithm (MBFA) method, which is a real-time FOG detection method, to fix the latency limitation of the method by Moore et al. [102]. They introduced a new term, Power Index (PI), defined as the addition of the Walking Band (WB) and the Freezing

Band (FB) to indicate the amount of movement. In their paper, there are two thresholds which are Freezing Threshold (FTH) and Power Threshold (PTH). FOG episodes are determined if $FI > FTH$ and $PI > PTH$. This FOG detection method has low computational cost and good performance. Furthermore, once the FOG episodes are detected the person gets auditory signals until resuming walking. The limitation of this work is the low number of patients. They reported 73.1% and 81.6% for sensitivity and specificity, respectively. During this time, the Daphnet dataset was created by the same authors for the evaluation of FOG detection methods. However, the conditions defining the data collection protocol, such as the limited set of activities and the clinical settings, may overestimate the results of the approaches tested on it compared to performances obtained using real data. In 2010, Maidan et al. [121] tested the hypothesis that Heart Rate (HR) increases during FOG and just before FOG is going to happen. To evaluate these hypotheses, HR and HR variability were considered to be the features for the subjects who carried out tasks that frequently provoke FOG. The dataset in their paper has 15 healthy older adults, 10 patients with PD who experienced FOG, and 10 patients who did not. Moreover, their results suggest that action observation has a positive additional effect on recovery of walking ability in PD patients with FOG. Delval et al. [44], proposed a FOG detection system based on gait analysis. The data for this paper was gathered by walking a group of subjects (10 PD with FOG, 10 PD without FOG and 10 Control) on a motorized treadmill while avoiding unexpectedly appearing obstacles. Treadmill walking was videotaped, and FOG episodes were identified by two independent experts. Gait was also analyzed using detailed kinematics properties. Knee joint signals were processed using time frequency analysis with combinations of sliding FFT and wavelet transform. This approach using the time frequency features detected even very brief FOG with acceptable sensitivity (75–83%) and specificity ($\geq$ 95%).

Mazilu et al. [100] presented a novel FOG monitoring system based on the work of Bachlin et al. [19]. They used a smartphone and a wrist acceleration sensor for capturing the motion data. This paper was the first to use machine learning for the detection of FOG. The features for doing this classification were

mean, standard deviation, entropy, energy, FI, and power of the acceleration signal. The ML algorithms used were RF, Naive Bayes (NB) and K-Nearest Neighbor (KNN). The best result obtained was 66.25 and 95.83 for sensitivity and specificity with RF. In the next year, they presented another automatic FOG detection system using wearable sensors. The feature learning and unsupervised methods were compared by time domain and statistical features from the motion data. The latter one has better performance by up to 8.1% in terms of F1-measure. The authors performed multi-class analysis since the pre-FOG is considered to be the new class (FOG vs. pre-FOG vs. normal locomotion). They also used auditory cueing at the end of their analysis for warning the patient about FOG episodes. Another real-time FOG detection method is presented by Zhao et al. [174]. An accelerometer integrated pant (MiMed-Pant) is used for capturing the gait information which sends the information to a wireless computer. The FOG detection algorithm is based on the frequency and time analysis and Power Spectral Density (PSD) features. The proposed method is executed every 0.5 s and the results are shown within the same time period. This is one of the positive aspects of this method. Evaluation is performed over 8 PD subjects in which the real FOG is determined for them by a specialist using recorded video. The limitation of this work is that the patient needs to wear the cumbersome MiMed-Pant for using the technique.

EEG signals are used in the paper presented by Ardi [65] for FOG detection. The EEG signals are obtained using a 4 channel wireless EEG system developed for 26 PD patients at the Parkinson's Disease Research Clinic at the Brain and Mind Research Institute, in University of Sydney. Preprocessing is done for the motion data by applying a high-pass and a low-pass filter for removing noise and unnecessary information. In this research, a Discrete Wavelet Transform (DWT) based on dyadic scales and positions is used for feature extraction. In the classification phase, a three-layer Back Propagation Neural Network (BP-NN) is used, with 56% of the data used for training, 25% for validation, and 19% for testing. Experimental results show 0.75% accuracy for the proposed FOG detection system. In 2013, Moore et al. [103], proposed assessing different

numbers of sensors in different locations for gait analysis with the purpose of FOG detection. The MBFA algorithm was used as the FOG detection algorithm with different sensor configurations. The data is captured by 7 sensors for 25 PD patients. Their analysis shows that by using all the sensor maximum accuracy is achieved for FOG detection which has high sensitivity (86.2%) and specificity (82.4%) when considered to be a binary test for the presence or absence of FOG.

In 2015, Zack et al. [171], presented a method for FOG detection using a triaxial linear waist-mounted accelerometer. Experiments for their paper include walking rapidly with short steps and rapid full turns in both directions, conducted on 23 patients. Two independent experts identified FOG episodes using offline video analysis (gold standard). In the proposed method FI is used as the discriminator, but a general threshold instead of a numerical one is applied to find the FOG episodes. Receiver Operating Characteristic (ROC) curves were drawn to determine a global FI threshold to distinguish between FOG and non-FOG episodes within the different tasks and for all tasks together. In addition to the global FI threshold, they calculated the sensitivity and specificity of the FI threshold for each subject. Combining all tasks together, sensitivity of 75% and specificity of 76% were obtained.

Tay et al. [155] presented a real-time PD monitoring and biofeedback system for FOG using low-cost wearable sensors (3 for each sensor on the neck, right ankle, and left ankle). Their proposed method detected FOG based on kinematic gait analysis using the gait cycle time and frequency features. The dataset consisted of 8 PD subjects, 5 of whom have FOG episodes. Moreover, this system benefits PD patients from the periodic cueing to pace their steps after a FOG occurrence; hence, improving their gait performance. The system includes local storage capability which is useful for FOG detection when patients are outside their home or a clinic. However, no quantitative results were given.

In 2017, Rodríguez et al. [137] presented a novel approach for FOG detection using machine learning techniques based on daily activities of PD patients in a real environment. Their paper has the advantage that the data is captured in a real environment. They extracted 55 FOG-related features from 21 PD

patients using just one waist-worn triaxial accelerometer. SVM with leave-one-out cross validation was used for classification. Evaluation was done in two cases, including general (user-independent) and specific (user-dependent). Experimental results show significant improvement in accuracy of the personalized model compared to the generic model; with enhancement in the geometric mean by 7.2%. Following this work, Sama et al. [142] reduced the number of extracted features to 28 for the same dataset. They also evaluated the extracted features by using 8 types of classifiers with greedy subset selection process, 10-fold cross validation, and different window sizes for signal analysis. The results show that the proposed method detects FOGs at a patients' home with 91.7% sensitivity and 87.4% specificity, enhancing the results of former methods by 5% to 11% and providing a more balanced rate of true positives and true negatives.

Most of the papers presented use wearable sensors for the collection of motion data because of their application in real life environments. However, using wearable sensors in an out-of-lab environment requires users to put the sensors in the correct positions. Such flexibility can cause variations in data capture which impacts the gait assessment quality. This is one of the major challenges for self-administration of wearable sensors by users in an out-of-lab environment without any supervision [42]. In [42], a framework is proposed for quantifying the variations resulting from using wearable sensors for data capture in a free-living environment. Even though there are many factors that affect data capture by wearable sensors, they consider the four more important sources of variations including mounting location, mounting leg, sensors, and speed. They statistically model this problem by using just one healthy non-elderly subject with a full-factorial design of 48 factors combinations. In [42] the influence of these four factors and their interaction on gait features derived by wearable sensors are characterized. As a result of their analysis, mounting location and gait speed were demonstrated as the most dominant sources of variation. There are many studies that use non-wearable sensors, such as Kinect as an assessment tool for gait analysis for PD patients. In [151], Kinect was used for body movement monitoring for PD patients with machine learning

techniques. The proposed method can monitor gait abnormalities for PD patients with high accuracy not only for diagnosis but also for disease severity detection. To the best of our knowledge, there is only one paper [10] that used Microsoft Kinect for FOG detection. In their paper, a novel system is designed based on gait analysis of human motion to detect FOG and falling for PD patients. The novelty of their method is in monitoring and improving the mobility using laser-based visual cues, which is called Kinect4FOG. An algorithm was developed to monitor the behavior of a subjects' gait cycle and the number of footsteps within a given time interval to estimate the occurrence of a FOG. A foot-off event is considered to have occurred when the knee angle of one foot has decreased to less than a specific threshold. The authors evaluated their method using over 15 PD subjects with FOGs who performed motion experiments using Microsoft Kinect V2. There was positive feedback on the proposed system in domestic usability, but it had limitations in outdoor use. This paper reported that 86.6% were satisfied with the FOG detection system whereas 13.3% neither agree nor disagreed. The reported results were based on people's feedback which are subjective. However, no objective results are reported.

Feature learning methods have also been used for automatic FOG determination. Among these, the most popular one these days is DL. DL models learn feature extractions that can easily handle multimodal data, missing information, and high-dimensional feature spaces. Just a few papers are proposed for FOG detection using DL including [30], [31]. Julia et al. [31] proposed a DL model which uses a 6-layer Convolutional Neural Networks (CNN) network for FOG detection. The input for this network is the stack of compressed form of the current signal window at time t and time $t-1$ which consists of 9 gyroscopes and acceleration features. The current method is better than all the state-of-the-art methods for FOG detection, achieving performances of about 90% for geometric mean criteria. However, DL-based methods need more computational resources than others in real-time applications for FOG detection. Also, DL needs a dataset that includes a large and representative set of examples to successfully train the network and obtain meaningful results.

Qiu et al. [127] presented a low-cost, intelligent, and light-weight platform for providing quantitative gait assessment information of stroke patients. They used multi-sensor fusion using accelerometer, gyroscope and magnetometer for taking advantage of the individual sensors and avoiding their weaknesses. They reported different gait parameters for stroke patients such as step length, number of steps, and other gait cycle parameters. Moreover, they provided qualitative results which have useful information for the specialists about the situation of their patients. Another gait assessment system was proposed for equestrian sport evaluation by Wange et al. [163]. This paper, similar to Qiu [127], used fusion of sensors for the rider on 7 parts of their body but no sensor was placed on the horse. They also used Northern Digital Camera for capturing videos of the riders for evaluating their system. Pelvis joint was selected as the reference for gait analysis to distinguish between professional and novice riders. Until now, most of the proposed methods for FOG detection tries to warn patients before FOG happens. However, neurologists for assessing PD patients need to have an automatic, accurate, and fast system for getting the FOG information of the patients. In addition, most of the proposed FOG detection methods must tape a video while they are capturing the data and then do manual video analysis for generating the ground truth labels to evaluate their algorithm.

2.0.2 Assessment of Parkinson's Disease using Demographic and Clinical Characteristics

There are different criteria and scales for rating PD severity, treatment evaluation and quality of life [157]. These clinical properties contain motor and non-motor biomarkers for the patients. The most applicable scores for evaluating the situation of a PD patient are *HAY* and the *UPDRS*. *HAY* is a commonly used by neuroscientists and researchers due to its simplicity in administration and reliable data representation. This scale was originally described in 1967 including Stages 1 through 5. It has since been modified with the addition of Stages 1.5 and 2.5 to account for the intermediate course of Parkinson disease [70]. In essence, *HAY* stages are correlated with motor decline, deterioration

in quality of life, and neuroimaging studies of dopaminergic loss. The *UPDRS* was first proposed in 1980*s* [128] and is divided into 4 parts: Part I, *Mentation, Behavior, and Mood* (4 items); Part II, *Activities of Daily Living* (13 items); Part III, *Motor Examination* (27 items); and Part IV, *Complications of Therapy* (11 items). In 2001 [133], the Movement Disorder Society (MDS) proposed MDS-UPDRS for addressing the weaknesses of the first *UPDRS*, including insensitivity to early motor impairment and inability to capture symptomatic impacts on daily living. The revised version of MDS-UPDRS was published in 2008 [60], which retains the 4-subscale structures and 5-point severity scales [71]. Although, much efforts have been expended on PD diagnosis using neuroimaging techniques [150], human motion analysis [151], and speech assessment [172], there is insufficient study focusing on the analysis of demographic with clinical features for PD assessment. In [126], SVM and Logistic Regression (LR) were used to build a classification model. This model is based on the four Parkinson's Progression Markers Initiative (PPMI) clinical features including *Striatal Binding Ratio (SBR)* for the left and right putamen, and *SBR* for the left and right caudate, all of which are computed from the SPECT imaging process. In [126], SVM classifier with Radial Basis Function (RBF) kernel is used for differentiating the subjects between early PD and HC. They have also estimated the risk of PD using LR which demonstrates high degree of consistency with statistical significance. However, the dataset used in [126] is imbalance (the number of PD samples is much more than the number of HC samples) and the paper fails to explain how the technique can be extended to the general population with this insufficient HC samples.

Kostas et al. [157] proposed a Decision Support System (DSS) for estimating the HAY score from the UPDRS-III using machine learning techniques including feature wrapper selection method and the Repeated Incremental Pruning to Produce Error Reduction (RIPPER) algorithm [40]. The results of the DSS, range from $84.1 - 87.0\%$. In their paper, they claimed that for the first time, a DSS based on data mining has been developed to facilitate HAY stage estimation for PD patients in everyday practice. But, it is only based on the motor evaluation scores of UPDRS-III.

2.0.3 Assessment of Parkinson's Disease using Neuroimagings

There have been many studies on the assessment of PD using a variety of neuroimaging methods. In this part, the papers for PD assessment using MRI T1 and T2 are reviewed.

Focke et al. [53] proposed a method for PD classification by MR Images using GM and WM individually with the SVM classifier. Voxel Based Morphometry (VBM) has been used for preprocessing and feature extraction. The results reported poor performance for classification based on GM and WM with 39.53% and 41.86% accuracy respectively. Babu et al. [18] proposed a CAD system for diagnosing PD. Their method includes three general steps: feature extraction, feature selection, and classification. In the first part, VBM is used for GM to construct feature data. For the feature selection, Recursive Feature Elimination (RFE) is used to select the most discriminative features. In the last step, projection based learning and meta-cognitive radial basis function was used for classification, which results in 87.21% accuracy. The potential biomarker for PD is identified as the superior temporal gyrus. The limitation in this work is that VBM is univariate and RFE is computationally expensive. Salvatore et al. [141] proposed a method that used PCA for feature extraction. The PCA was applied to normalized skull stripped MRI data and SVM was used as the classifier, resulting in 85.8% accuracy. Rana et al. [131] extracted features over the three main tissues of the brain consisting of WM, GM and CSF. Then, they used t-test for feature selection and in the next step, SVM for classification. This resulted in 86.67% accuracy for GM and WM and 83.33% accuracy for CSF. In their other work [129], a graph-theory based spectral feature selection method was applied to select a set of discriminative features from the whole brain volume. A decision model was built using SVM as a classifier with leave-one-out cross-validation, giving 86.67% accuracy. The method proposed in [130] was not focused on just individual tissues (GM,WM and CSF); rather, it considered the relationship between these areas because the morphometric changes in one tissue might affect other tissues.

3D LBP was used as a feature extraction tool that could produce structural and statistical information. After that, Minimum Redundancy and Maximum Relevancy (MRMR) relevance with t-test are used as a feature selection method to get the most discriminative and non-redundant features. At the end, SVM is used for classification giving 89.67% accuracy. In [122], the low level features (GM, cortical volume, etc.) and the high level features (ROI connectivity) are combined to perform a multi-level ROI feature extraction. Then, filter and wrapper feature selection method is followed up with multi kernel SVM to achieve 85.78% accuracy for differentiation of PD and HC data. Adeli et al. [1] proposed a method for early diagnosis of PD based on the Joint Feature-Sample Selection (JFSS) procedure, which not only selects the best subset of the most discriminative features, but it also chooses the best sample to build a classification model. They utilized the robust regression method and further develop a robust classification model for designing the CAD for PD diagnosis. They have used MRI and SPECT images for evaluation on both synthetic and publicly available PD datasets which provided very accurate classification results. Haijun et al. [87], presented a multi-class classification (PD, Scan Without Evidence of Dopaminergic Deficit (SWEDD) and HC) framework by using multi-modal data including MRI and DTI for PD assessment. They proposed a novel sparse feature selection technique which is the combination of Fisher's LDA and Locality Preserving Projections (LPP) that uses the local and global information of the data. In preprocessing, SPM were used for bias correction, denoising, normalization of MRI and DTI Images. After that, MRI images are segmented to WM, GM and CSF parts producing 116 tissue volumes as the feature matrix and DTI gives 16 Fractional Anisotropy (FA) intensity values for each subject. In the feature selection part, a novel method was utilized followed by SVM classification. Some demographic information have been used beside the selected features to improve the classification accuracy. In the experimental results, they used PPMI dataset showing remarkable effect of their novel feature selection in distinguishing the data to PD, HC, and SWEDD which in the best case is 78.37 ± 8.11. They have also presented the regions in the brain most affected by PD and also their correlation together. Ozkan

et al. [115] proposed an efficient covariates and contrast on classification of PD using structural MRI. The general framework has three steps which are preprocessing, feature selection and classification. The preprocessing is done by using VBM for normalization, modulation, smoothing and parcellating the MRI to GM and WM. In the next step, PCA is used for dimension reduction of the combination of GM and WM. Then, the statistical model is built using t-contrast and f-contrast. After that, the total intracranial volume (eTIV), age, sex and the combination of them are used as features in the classification step for GM, WM and GM+WM data using SVM with 10-fold cross validation. The PPMI dataset has been used for evaluation of the proposed method with 40 PD and 40 HC subjects. The classification results using f-contrast show a superior performance for GM, WM, and the combination of them compared to t-contrast. In [150], PD classification using MR T1 images is conducted. In their paper [150], Freesurfer used for MRI preprocessing phase. After that T1 volume voxels are used as features for the classification. Three different classifiers including RF, SVM and LR are compared based on their ability to distinguish between PD and HC for MR T1 Images. The PPMI dataset is used for evaluation of the proposed method. The proposed CAD system proved promising results for assisting in diagnosing PD. Lee et al. [86] proposed a machine learning based method to identify disease-related spatial covariance patterns of grey matter volume as an aid in the classification of PD using T1-weighted structural MRI scans (70 PD, 70 Control). They conducted a Structural Covariance Networks (SCN)-based classifier frame work with Independent Component Analysis (ICA) as feature selection. They have done Leave-One-Out-Cross-Validation for classification. In the validation, the classifier had moderate generalization ability, with a mean sensitivity of 81%, specificity of 69% and overall accuracy of 75%. Furthermore,certain individual SCNs were also associated with disease severity.

In the case of deep-learning-based classification of PD, it is important to consider several distinct elements of the problem. The first and arguably the most important of these elements is the overall accuracy of the method. Regardless of the elegance or apparent efficacy of a model, one which yields

poor results is ultimately not useful in a practical setting. Along with accuracy comes the simplicity of the method, as well as the availability of the data. These two points are of particular importance in this task, as the domain of PD classification is largely medical rather than computational. Data is often not abundant, and any tool built to assist clinicians must be readily applicable. It is important to remember that a clinician is not a data scientist. With this in mind, we will assess new work in the field of deep-learning-based PD classification.

Choi et al. propose a model to classify PD patients based on CNN and SPECT imaging [36]. SPECT imaging is highly specialized, and requires an injection of a radioactive isotope to monitor its uptake in different areas of the body. This specialization is reflected in the fact that over a one year period of England's NHS imaging operations, there were approximately 100 times more MRI scans performed than SPECT [45]. As well, because of the injection of a tracer, the technique may be considered somewhat invasive. This suggests that a model based on SPECT imaging may lack applicability in common medical use. Despite this, the model achieves a 96.0% accuracy with 100% sensitivity. This is coupled with a relatively large dataset (in medical terms) of 624 preprocessed samples, which suggests a relatively robust model. It is important to consider that there is a significant class imbalance, with 431 PD to 193 HC samples. Class imbalance in the training set is detrimental to the performance of CNN s because they tend to over-classify the majority class [27]. The authors do no upsampling or downsampling to correct this class imbalance, but they perform data-augmentation in the form of left/right flipping to increase the total training sample size. In summation, Choi et al. put forward a promising but somewhat niche deep-learning based model to classify PD.

Long et al. [123] propose a machine-learning based approach to classification of PD from resting-state functional MRI (rsf-MRI). What makes rsf-MRI different from sMRI is that it detects subtle changes in blood-flow between areas of the brain, effectively allowing one to observe areas of activity due to higher metabolism. Meanwhile, sMRI only detects general anatomical features

in the brain, while ignoring activity. The researchers segment into separate GM, WM and CSF, and then extract 116 features for classification. The actual classification process is done with a hyperbolic tangent kernel, and the researchers achieve 87% classification accuracy. These results are quite good, although it is important to consider that the experiment was done with a very small number of samples (17 PD and 27 HC). This is not as much of a problem for rsf-MRI, because the data is rich and one can extract many features from very few scans. They discover several ROI, some of which are supported from previous clinical findings. and it would be of significant interest to explore further research in rsf-MRI-based PD detection.

Iron deposits would create local inhomogeneities in the magnetic field which in turn, would result in loss of the T2 signal. Iron distribution is clearly mapped as signal hypointensity (darkness) on T2-weighted image due to local-field heterogeneities produced by ferritin [15], [48]. The increase in tissue iron is significant in magnitude, it is observable in postmortem parkinsonian brain [15], [20] and can be evaluated with imaging techniques in living subjects. Polonen [125] proposed a shape analysis system using T2 MRI images. The SN region and the area around that are selected as ROI in the slice. Then for each ROI, the segmentation is done using three different methods including thresholding, region growing and k-means. After that, they used different shape features such as area, width, height, and etc to compare the SN area between the PD and HC subjects. The used p-value as a criteria for the shape analysis for different types of segmentation. Their analysis is done for the left and right hemisphere of the brain. The size of the group was small and most of the patients and controls were female. This is a problem because the prevalence of PD is significantly higher in men and the progression of the disease appears to be slower in women. However, they have mentioned that $3D$ dimensional shape analysis and morphometry have provided more comprehensive results from several different brain structures including SN. Recent developments in MRI techniques have offered new research opportunities to visualize substantia nigra pathology in PD. Neuromelanin-based techniques such as neuromelanin-sensitive imaging and visualization of nigrosome-1 have emerged as important

Research Works	Year	Dataset (HC/PD)	Classifier	Data Type	Utilize Region	Best Performance (Accuracy)
Fock et al. [53]	2011	22,21	SVM	MRI T1	WM,GM	41.86
Salvator et al. [141]	2014	28,28	SVM	MRI T1	WM	85.8
Babu et al. [18]	2014	112,127	PBL-McRBFN	MRI T1	GM	87.21
Rana et al. [131]	2015	30,30	SVM	MRI T1	GM,WM,CSF	86.67
Rana et al. [129]	2016	30,30	SVM	MRI T1	GM,WM,CSF	88.89
Adeli et al. [1]	2016	169,374 (PPMI)	JFSS	MRI T1	GM,WM,CSF	81.9
Peng et al. [122]	2017	103,69 (PPMI)	SVM	MRI T1	Whole Brain	85.87
Rana et al. [130]	2017	30,30	SVM	MRI T1	GM,WM,CSF	95
Ozkan et al. [115]	2018	40,40 (PPMI)	SVM	MRI T1	WM,GM	75
Shinde et al. [147]	2019	39,40	CNN-DL	NMS-MRI	SN	80
Cheng et al. [35]	2019	77,87	SVM	QSM	SN	88

Table 2.1: Comparison of the proposed studies for PD classification using MRI.

candidate neuroimaging biomarkers in PD which have particularly been of interest in the past two decades [35], [95], [154]. Cheng et al. [35] explored the ability of radiomic features of nigrosome-1 in SN, based on Quantitative Susceptibility Mapping (QSM) to differentiate IPD patients from HC. First, ROIs of the SN were manually drawn, and subsequently, volumes of interest were segmented. Then, 105 radiomic features of bilateral selected volumes were extracted. The proposed feature selection method combined analysis of variance, random forest, and recursive feature elimination, which results in 40 features. The selected features were further utilized to classify IPD patients from HC using the SVM classifier with 10 rounds of 3-fold cross-validation. The classification results from SVM were: accuracy: 0.88, sensitivity: 0.89, and specificity: 0.87. Deep learning is a branch of machine learning which is successful in many applications in computer vision problems. In the case of deep-learning-based classification of PD, Shinde et al. [147] proposed a method for SN assessment in order to PD classification. Convolutional neural nets with discriminative localization (CNN-DL) is used to create prognostic and diagnostic biomarkers of PD from Neuromelanin Sensitive Magnetic Resonance Imaging (NMS-MRI). The reported classification accuracy is 80%. Moreover, it has the capability to locate the most discriminative regions on the NMS-MRI which is mentioned as the left SN. Brief review and comparison of methods employed by recent studies that have used machine learning and statistical learning techniques to classify PD from MRI modalities are presented in Table 2.1.

Based on the literature review papers, the SN analysis using different imaging sequence are done based on the 3D volume of the image. Although, using 3D analysis on MR images considered all the required information in a different dimension, but makes the computational complexity much more higher compared to 2D analysis. Whereas, 3D analysis does not always result in higher accuracy at the end. On the other hand, not all the slices in the 3D volume of the imaging data have the useful knowledge for PD assessment. Therefore, having high performance classification system with fewer number of slices which are related to PD is our general goal in this research.

Chapter 3

Body Movement Monitoring for Parkinson's Disease Patients Using A Smart Sensor Based Non-Invasive Technique

3.1 Introduction

PD is a progressive neuro-degenerative disorder caused mainly by lack of dopamine in the brain. Dopamine is a neurotransmitter involved in movement, motivation, memory, and other functions. Dopamine loss in PD brain is a cause of motor deficiency in PD patients. These motor deficiency can be categorized to tremor, slowness of movement (bradykinesia), rigidity and gait problems which are cardinal features of this disease [76]. These conditions constitute a major source of concern in elderly population as it impacts the quality of life on not only the patients, but also their families and care providers. To date, the gold standard of diagnosis for PD is still based on clinical features and the UK brain bank criteria [59]. Response to dopaminergic medication, such as levodopa [76] is helpful in diagnosing the condition. It is important to mention that currently there is no definitive test for diagnosis of PD, and thus obtaining absolute conclusion can be challenging, especially in the early course of the disease[76].

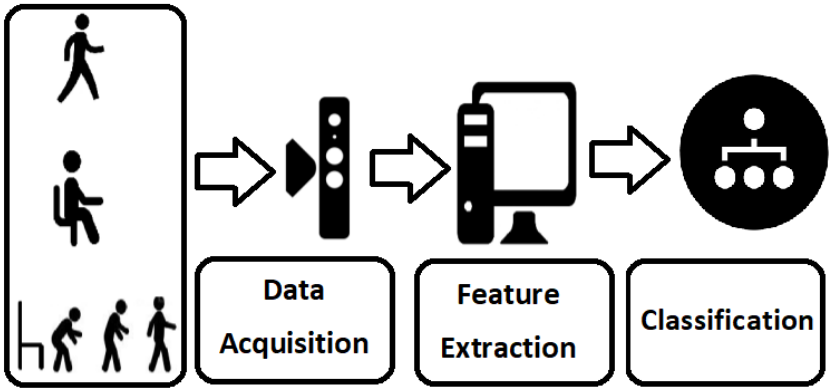


Figure 3.1: The proposed body movement monitoring system for PD based on kinematic data analysis.

3.2 Motivation

There are some medical evaluation with a thorough physical examination for diagnosing and prognosis of PD. However, the evaluation results might be subjective based on the different doctor's opinions. In addition, the physical assessment is a time consuming process not only for specialists but also for the patients. Therefore, in this part of my thesis, a CAD system is proposed for diagnosing and monitoring PD using kinematic motion data obtained from the PD standard exercises supervised by physical therapists. The proposed system uses machine learning techniques to make the assessment process easier and faster for the specialists and the PD patients.

3.3 Methodology

In this section, we first describe the data acquisition process and then explain the proposed method on PD analysis using kinematic motion data. The proposed PD monitoring system is shown in Figure 3.1.

3.3.1 Data Acquisition

In our implementation, Microsoft Kinect 2.0 is used to capture the motion data generated from the body joints. Microsoft Kinect 2.0 is the latest version released by Microsoft when we performed our experiments. It delivers significant

improvements over its predecessor in term of accuracy and resolution. Kinect 2.0 employs a RGB camera with a resolution of 1920×1080 at 30 fps and a depth sensor with an output resolution of 512×424 based on time of flight to perform depth calculation. The depth range is 0.5 to 4.5 meters. Kinect 2.0 is able to detect up to 6 people, with each represented by 25 joints [153].

3.3.2 Exercises Protocol

A set of experiments was conducted in the Neuroscience Unit at the University of Alberta Hospital. The goal was to capture individual motion data using Kinect V2, while the subject was performing a sequence of gait tasks. These standard tasks: Tremor, SL and TUG, are used in rehabilitation sessions as assessment methods to determine the severity of motor symptoms in PD patients.

- **Tremor**: Tremor is the most common and easily recognized symptom of PD [76]. Tremors are unilateral at onset and progress to contralateral side over time. Tremor can be categorized into rest, postural and kinetic. Rest tremor occurs when the body part is at complete rest against gravity. Postural tremor takes place while maintaining a position against gravity. Kinetic tremor can occur during normal movement. In this exercise, the subjects were asked to sit and keep their hand straight in-front of themselves.

- **SL**: PD patients tend to demonstrate a reduced and variable SL, as well as overall low velocity [88]. This characteristics is considered of great use for analysis because it is a very common symptom of PD. In this exercise, the patients were asked to walk perpendicularly to the optical axis of the Kinect sensor from one point to another point.

- **TUG**: In the TUG exercise, the patients were asked to rise up from a chair, walk 3 meters, turn, walk back and sit down [109]. The TUG test is a common measure of gait and balance in PD patients, as it is highly correlated with functional mobility, gait speed and other lower extremity functions. [109][69].

3.3.3 Gait Analysis

In order to analyze Kinect-based motion data, the subject's body skeleton is represented by a graph which contains a set of nodes (joints) J and vectors (bones connecting dominant joints) V. The graph model G is defined in Eq.3.1

$$G = \{J, V\} \quad where \quad J = \{j_1, ..., j_n\}, V = \{v_1, ..., v_e\} \tag{3.1}$$

where n is the number of joints in the body model and e is the number of connected edges.

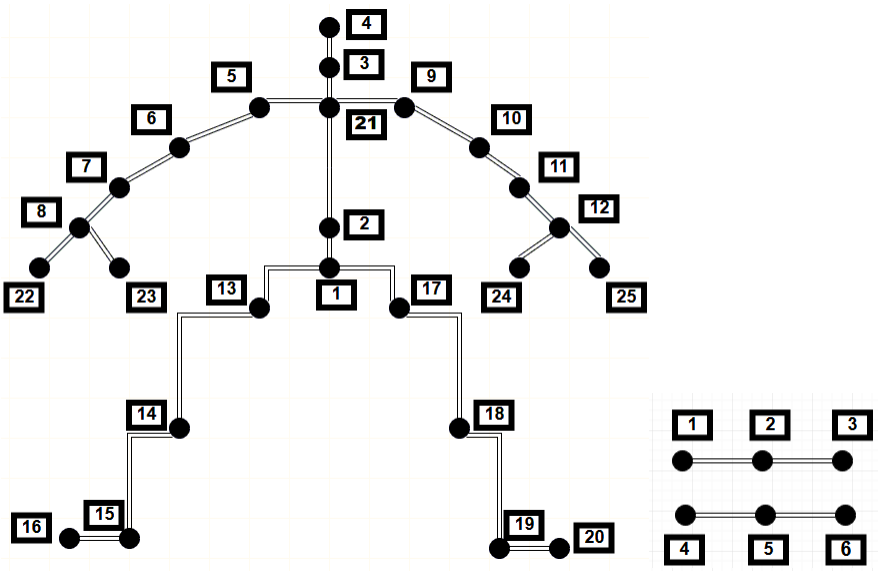
In the TUG and SL evaluations, the whole body skeleton is captured but in the Tremor evaluation, only hand movements are monitored. A graph model composed of 25 skeleton joints is used for analysis. Figure 3.2(a) shows the graph model for Tremor with three joints recorded per hand: 1,4-Elbow. 2,5-Wrist. 3,6-Hand Tip. Figure 3.2(b) shows the graph model for the SL and TUG experiments, involving 25 joints: 1- Spine Base, 2- Spine Mid, 3- Neck, 4- Head, 5- Left Shoulder, 6- Left Elbow, 7- Left Writs, 8- Left Hand, 9- Right Shoulder, 10- Right Elbow, 11- Right Writs, 12- Right Hand, 13- Left Hip,14- Left Knee, 15- Left Ankle, 16- Left Foot, 17- Right Hip,18- Right Knee, 19- Right Ankle, 20- Right Foot, 21- Spine Shoulder, 22- Hand Tip Left, 23- Thumb Left, 24- Hand Tip Right, 25- Thumb Right.

Our goal is to monitor the level of severity for PD patients using the kinematic motion data obtained from the tremor and TUG exercises. The SL motion data is used for classification of the subjects to PD and HC. Gait data analysis is carried out for each exercise by following two steps: Feature Extraction and Classification, as explained below.

Feature Extraction from Tremor Data

The tremor motion magnitude relative to the frame rate is an effective cue for classifying the subjects based on their severity in PD.

We compute the rates of joint displacement in three spatial axes. This is achieved by obtaining a displacement vector d based on frame-to-frame Euclidean distance (Eq.3.2). The displacement for each subject is defined in



(a) (b)

Figure 3.2: Kinematic motion data graph models. (a) SL and TUG graph model. (b) Tremor graph model for the two hands.

Eq.3.3.

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \tag{3.2}$$

$$Disp_i = \begin{pmatrix} d_{1,1} & \dots & d_{1,n-1} \\ \vdots & \dots & \vdots \\ d_{m,1} & \dots & d_{m,n-1} \end{pmatrix} \tag{3.3}$$

$Disp_i$ is a displacement matrix for the i^{th} subject, $d_{1,1}$ shows the euclidean distance between the position (x, y, z) of frame 2 and frame 1 for the first joint. m is the number of joints and n is the number of frames. In our analysis, the $L1$ norm distance is also used to compute hand displacement between two time frames. Eq.3.4 shows how $L1$ distance is obtained from two vectors with n time instances (frames). Thus, the same displacement matrix can be obtained for the $L1$ method.

$$d_{L1}(v_1, v_2) = ||v_1 - v_2|| = \sum_{i=1}^{n} |v_1(i) - v_2(i)| \tag{3.4}$$

At the end, the euclidean displacement mean and standard deviation (STD), together with the $L1$ displacement mean and STD, are used as features for tremor motion data assessment.

Feature Extraction for SL

PD patients often have shorter step length compared to healthy subjects (HC). This experiment helps to distinguish PD subjects from healthy individuals. Perumal et al. [124] measure the SL by analyzing the stance phase and swing phase. The stance phase starts at the point of heel strike, and ends at toe-off, while the swing phase takes place between the toe-off instance and the next heel strike. However, Perumal's method requires wearable force sensor to determine the points of heel strike and toe-off. Our technique is more convenient because only the acquired Kinect data is used for analysis and no wearable device is needed. Given a subject's walking distance, there is a sequence of Step Length Vectors (SLV) defined by Eq.3.5.

$$SLV = [SL_1, \ldots, SL_w] \quad \rightarrow STL_i = \sum_{k=1}^{w} SLV_k/w \tag{3.5}$$

An average is computed for each patient based on the stride sequence (SLV) as shown in Eq.3.5. The output is STL_i, where w is the total number of steps completed by a subject. The same process is performed by healthy subjects. PD subjects can be identified by comparing their average SL with those of healthy subjects.

The number of frames, which reflects the time taken by each subject (Time To Complete the Task (TTC)) is also recorded as an important feature from the stride-length motion data.

Feature Extraction for TUG

In this exercise, the subject is initially sitting down and is asked to stand up, walk, return and sit down on the chair. Movement disorder neurologists find that, in general, there is a correlation between the time required to complete the TUG action cycle and the stage of Parkinson's disease [109],[104]. In this paper, the time between the start (standing up) and end positions (sitting down) is used as a TUG feature for analysis. In this evaluation, we focus on three specific joints, i.e., Spine-Mid, Spine-Base and Knee-Right joints, which form two vectors and cast an angle as shown in Figure 3.3. The angle

computation is given in Eq.3.6, which is used to distinguish sitting down from standing up and walking. For each subject, this angle α_{TUG} is computed per frame as shown in Eq.3.7.

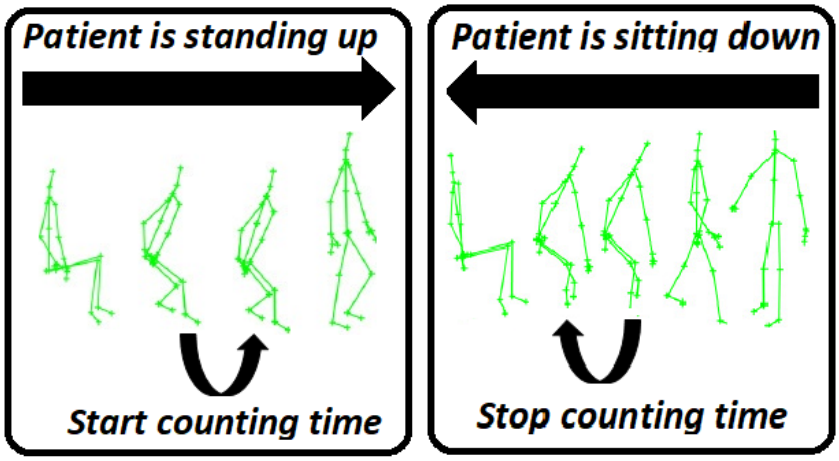
$$\alpha = \arccos \frac{v1.v2}{\|v1\|\|v2\|} \tag{3.6}$$

$$Ang_i = [\alpha_{f1}, \ldots, \alpha_{fmn}] \quad \rightarrow TAC_i = \alpha_{stop} - \alpha_{start} \tag{3.7}$$

where Ang_i shows the set of n angles for subject i during the action cycle. α_{f1} is the angle in frame 1. We use the angle to turn on a timer when the subject is standing up, and stop the timer when the subject is at the end position for sitting down. As you can see in the Figure 3.4, when the angle starts

Figure 3.3: The angle between two vectors : spine-base to spine-mid, and spine-base to knee which is used to determine the standing up and sitting down positions (α_{TUG}).

to increase, the subject is standing up and the timer can be started (α_{start}). When the angle is decreasing, the subject is back to sitting down and the timer can be stopped (α_{stop}). The time elapsed is taken as the TAC time in Eq.3.7. The total number of frames showing the TTC the task is also used as a feature for the TUG motion analysis.



(a) (b)

Figure 3.4: Time computation in TUG exercise: (a) α_{TUG} is increasing when the patient is standing up (α_{start}). (b) α_{TUG} is decreasing when the patient is sitting down.(α_{stop}).

Classification

The classification process is done using the features extracted from the kinect based motion data collected in the three exercises. The classification for tremor and SL are supervised because we have labeled samples as ground truth. However, there is no class label for the TUG experiment and thus it is an unsupervised classification. We use the RF classifier to assess gait data for tremor and SL. RF is an ensemble learning method, which can be used in classification, regression and other tasks. This method was presented by Breiman [24], who created a set of decision trees from a randomly selected subset of training data. The algorithm then aggregates the votes from different decision trees to assign a final class to the test object. Each tree in a random forest is a weak classifier. A large set of trees trained with randomly chosen data makes a single decision on a majority basis. In this work, we tested the accuracy of the random forests decision when the random trees are trained with the kinect-based motion data from PD patients.

Since there is no ground truth label for TUG data, we need to use an unsupervised clustering method. The k-means clustering algorithm is a widely used unsupervised learning method proposed by MacQueen in 1967 [63] and

later enhanced by Hartigan and Wong [132]. In this algorithm, a set of k points in the feature space are selected as the cluster centers. Then, the algorithm performs two steps: (1) Iteratively assign the feature samples to the cluster having the closest center in the feature space, based on the squared Euclidean distance. (2) Recompute the cluster centers in order to minimize the mean distance between data samples and the nearest cluster center. The goal of K-means goal is to keep a strong degree of association between data samples within the same cluster and a weaker degree of association between samples of different clusters [3].

3.4 Experimental Results and Discussion

The experiments were conducted at the University of Alberta Hospital Division of Neurology. The Tremor, TUG and SL exercises were performed by different groups of patients. In the SL exercise, there were 15 subjects in the control group and 15 in the PD group. The number of participants in the Tremor and TUG exercises were 43 and 20 respectively that all are PD subjects. There were 8 common subjects who did all the experiments. For the SL group, the ground truth was available for each subject with label (PD/HC) in our data. Also, the clinicians provided ground truth labels for the tremor group categorized under Mild, Moderate and Severe. However, there was no ground truth label for the patients who performed the TUG exercise.

The data distribution plots for different exercises based on the extracted features are presented in Figure 3.5. The tremor features can classify data into three separate groups based on the L1 and Euclidean distance displacement features (Figure 3.5(a)). The PD motion data are distinguishable based on the TAC and TTC features in TUG exercise (Figure 3.5(b)). Furthermore, the step length data distribution in terms of average step length and TTC, illustrate two groups of data representing PD and HC (Figure 3.5(c)). In addition to the RF classification, we also applied leave one out cross validation on the classification of SL and tremor data. The statistical analysis includes accuracy,

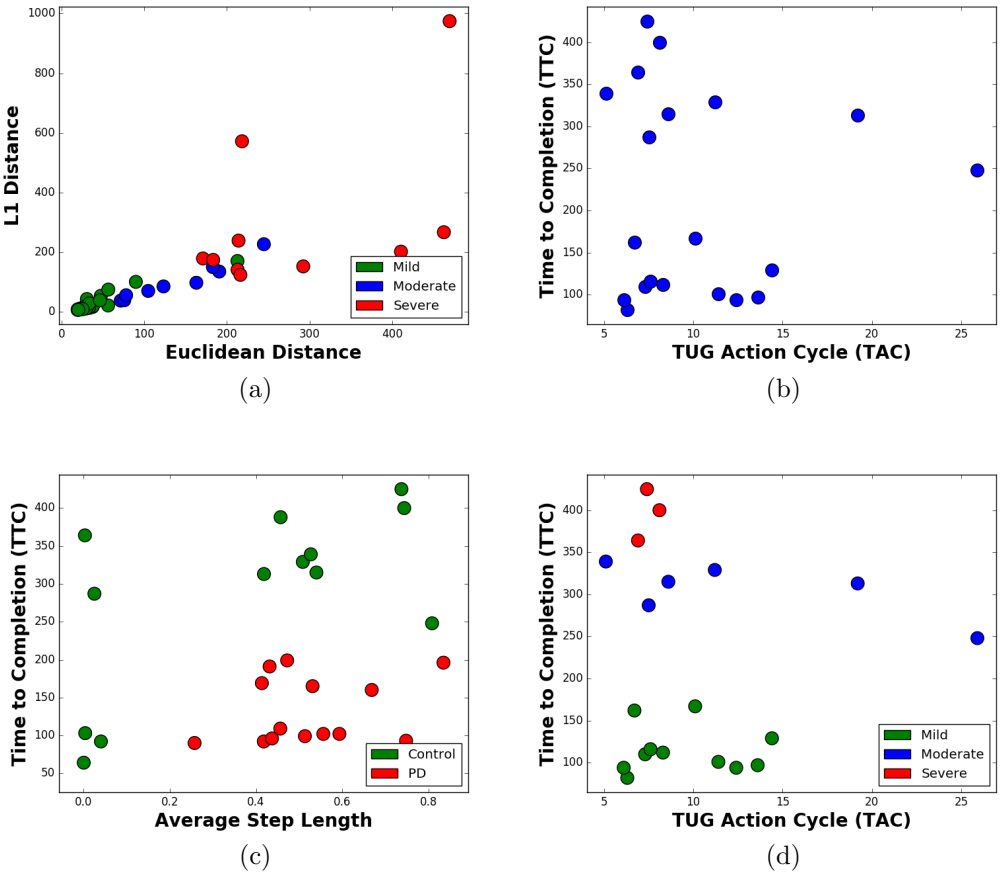


Figure 3.5: The data distribution based on the extracted features for: (a) Tremor, (b) TUG, (c) SL, (d) The prediction labels for data based on TUG features with K-means clustering method.

precision, recall, and F1-score which can be computed by Eq. 3.8.

$$\begin{aligned} Accuracy &= (TP + TN)/(TP + TN + FP + FN) \\ Precision &= TP/(TP + FP) \\ Recall &= TP/(TP + FN) \\ F1Score &= 2 * (Recall * Precision)/(Recall + Precision) \end{aligned} \tag{3.8}$$

Where True Positive (TP) is the number of cases correctly classified as PD, False Positive (FP) is the number of cases incorrectly classified as PD, True Negative (TN) is the number of cases correctly classified as healthy and False Negative (FN) is the number of cases incorrectly classified as healthy. The results are presented in Table 3.1.

	Accuracy	***Precision***	***Recall***	***F1-Score***
SL	93.3333	0.9411	0.9333	0.9330
Tremor	81.0697	0.8355	0.7376	0.7121

Table 3.1: Statistic analysis of RF classification results for tremor and SL exercises.

For the clustering of TUG data, K-means method is applied. The predicted labels are shown in the feature data distribution plot in (Figure 3.5(d)). Note that the TUG samples are distributed into three groups: Mild, Moderate, and Severe, as illustrated in green, blue, and red respectively.

There were 8 subjects who performed all the exercises including tremor, TUG and SL. It will be interesting to find out which exercise features are more influential in terms of classifying the patients based on their disease severity.

As mentioned before, for the SL, we have two groups of subjects (PD and HC), but we only need to consider the PD subjects in this analysis. When classifying these subjects in the three exercises, feature importance is computed to evaluate the significance of the extracted features. Each feature is identified by its index, name and the corresponding exercise: (0) Euclidean Distance Mean (Tremor); (1) Euclidean Distance STD (Tremor); (2) L1 Distance Mean (Tremor); (3) L1 Distance STD (Tremor); (4) TUG Time (TUG); (5) TTC (TUG); (6) Average Step Length (SL); (7) TTC (SL). Figure 3.6 shows the feature importance for all features extracted from the three exercises. Observe that the tremor features have higher importance comparing to other features.

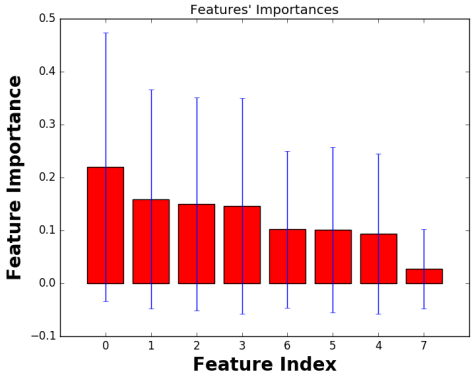


Figure 3.6: Comparing the extracted features importance for all the exercises.

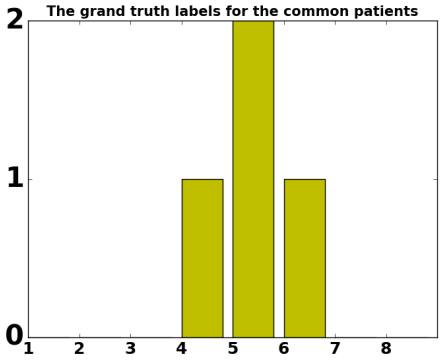


Figure 3.7: Ground truth labels for 8 common subjects.

We also plot the prediction labels for each of the 8 common subjects in different exercises (Figure 3.8 (a) to (h)). The red bar shows the prediction by SL, the green bar shows the tremor (Trem) prediction and the blue bar refers to the TUG exercise. The ground truth labels for these 8 patients are presented in Figure 3.7. The numbers $(0, 1, 2)$ on the y axis represent the three classes (Mild, Moderate and Severe). By comparing the predicted labels in different exercises with the ground truth labels, it is obvious that the tremor predication is the most accurate metric in the multi-motion analysis because it could classify 7 subjects correctly equivalent to 87.5% accuracy.

3.5 Conclusion and Future Work

In this study, Microsoft Kinect V2 is used to analyze three important characteristics of PD, namely TUG, SL and Tremor. The ground truth labels are provided by a clinician using the UPDRS for SL and tremor. Our assessment technique includes three major components: data acquisition, feature extraction, and classification. Representative features are extracted from the kinematic data collected in these exercises. RF is used to classify tremor and SL, and the K-means algorithm is used to cluster TUG data. The proposed method is able to achieve an accuracy of 93.33% for SL and around 80% for the analysis of tremor data. The prediction labels of K-means clustering illustrate the robustness of this method in categorizing the TUG motion data. In addition, multi-motion analysis is applied on the data obtained from the 8

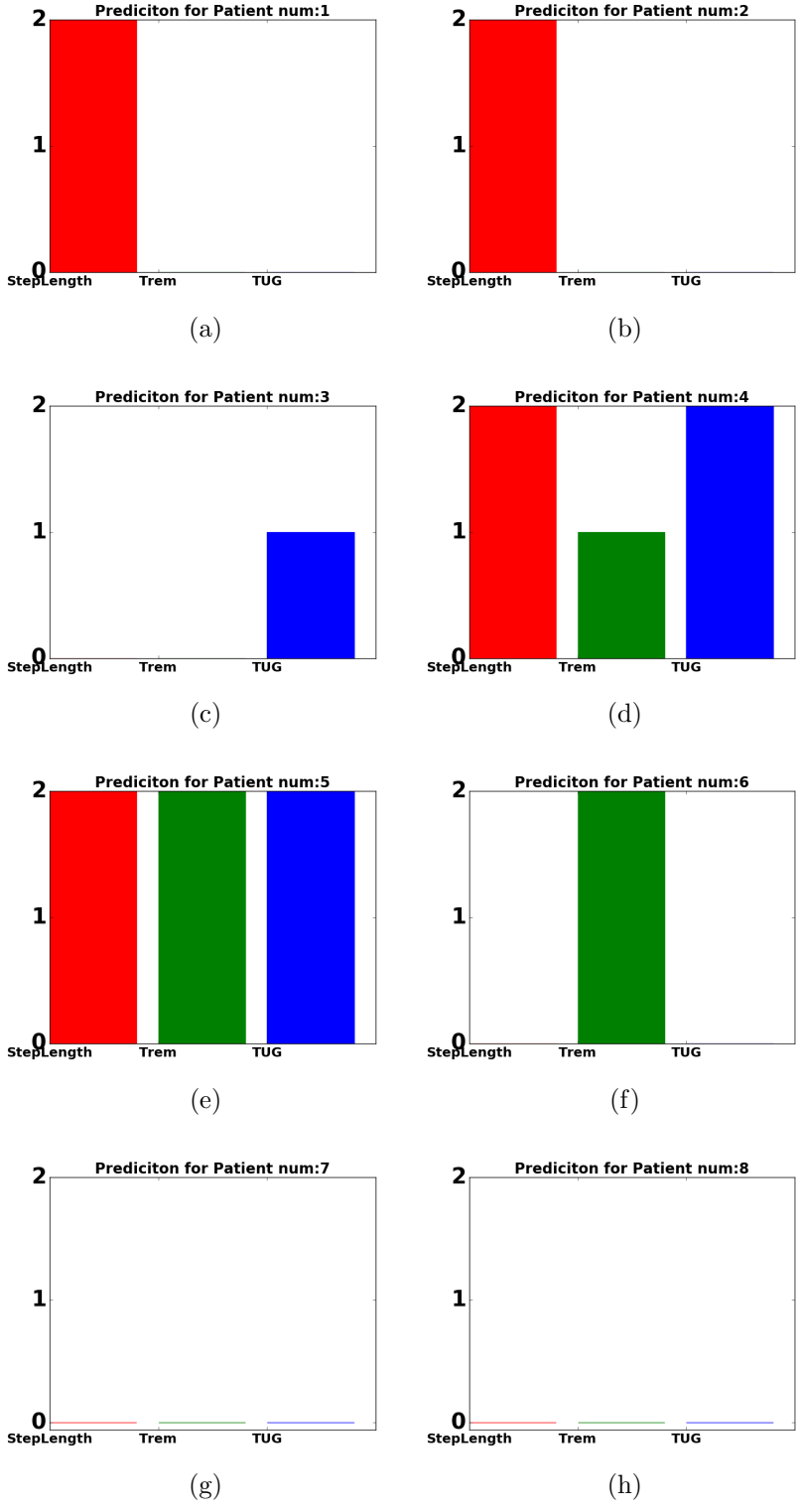


Figure 3.8: Prediction labels for the common subjects: (a) Patient 1, (b) Patient 2, (c) Patient 3, (d) Patient 4, (e) Patient 5, (f) Patient 6, (g) Patient 7, (h) Patient 8.

subjects who completed all three exercises. The results show the promising performance of our approach. The low cost, ease of use, accuracy and versatility of our system demonstrates its suitability in analyzing movement disorder patterns. In future work, we will further validate the benefits of using our Kinect or equivalent sensor system as a remote assessment tool to monitor the development of Parkinson's disease and other movement disorder, so that patients can be assessed remotely, e.g., at home, which will release the resources at rehabilitation clinics to individuals with urgent needs.

Chapter 4

Kin-FOG: Automatic Simulated Freezing of Gait (FOG) Assessment System for Parkinson's Disease

4.1 Introduction

PD is one of the most common progressive neuro-degenerative diseases with a worldwide prevalence of 22 per 100,000 person-years for all age groups, and up to 529 per $100,000$ person-years in older populations [81], [165]. The major cause for PD is loss of dopaminergic neurons in the part of the brain that is called the substantia nigra (SN) which is responsible for controlling the movement of different parts of the body. Therefore, the main symptoms of PD are related to motor disabilities including rigidity, bradykinesia, slowness, tremor, and FOG. However, it has other non-motor symptoms such as sleep disorder, cognitive changes, mood disorders, and fatigue. Estimates show that 60.5% of PD patients experience at least one fall and 39% of them have recurrent falls which can cause fractures [8]. Falls and fractures can cause disabilities, significant impairment in the quality of life, and death with a 10.6% rate [81]. Fall can be because of different PD symptoms of which the major one is FOG.

FOG is a brief episode of absence of forward progression of the feet despite the intention to walk. Patients have the impression that their feet are glued to

the ground [110] and they lose control over their gait. FOG is transient and lasts from a few seconds to up to 1–2 min [50], [110]. This phenomenon happens mainly during gait initiation, turning, performing dual tasks or approaching narrow spaces [43], [113].

Even though current medical drugs can successfully relieve most of the motor symptoms, FOG is one of the least responsive to medical treatments [50]. However, external cueing to induce an external auditory or visual stimulus are promising for tackling FOG and helps resume natural gait for patients. Currently, FOG assessment is done based on movement experiments in a lab environment, self-reported diaries from patients, manual video analysis by specialists and specific symptom questionnaires, namely FOG-questionnaire (FOG-Q) [137]. However, these strategies can provide biased information on a patient's daily experiences because of limitations, such as the ones listed below:

- The set up for experiments is different from a real environment at home resulting in different FOG patterns compared to a real one.
- Even though FOG-Q can provide relevant indicators for the identification and characterization of FOG, it is based on the patients' opinions which are subjective.
- Assessments are only done a few times a year. This is not suitable for FOG detection since we need continuous observations on patients.
- Self-assessment of FOG by PD patients is unreliable since most patients often experience memory loss and dementia [137].

On demand cueing is more efficient than continuous cueing in decreasing the duration of FOG episodes. Hence, automatic prediction of FOG is essential for generating the cueing only when a FOG event occurs [159]. FOG investigation is challenging because of its unpredictable and unreliable nature. The current treatments and cueing techniques are just temporary solutions for limiting a FOG event time. Moreover, there are many unknown aspects of the pathophysiology of FOG and its relation to PD. Furthermore, walking styles of PD

patients differ across subjects (including diverse motor anomalies) [99]. However, in this part of this thesis an automatic and accurate system is proposed which is necessary for neurologists to assess the FOG status of PD patients. The continuous assessment of FOG status for PD patients by neurologists using the proposed Kin-FOG system can control the FOG situation which can eventually prevent unexpected falls for patients.

4.2 Motivation

Conventional FOG assessment for patients is mainly conducted by a clinician visual inspection and observation which produce results that are subjective and rely on the observer's experiences [127]. Thus, it is crucial to get the knowledge of the rehabilitation process by quantitative gait analysis, and understand to what extent does a patient recover from the disease [127]. Therefore, accurate and automatic FOG assessment for PD patients is not only crucial for providing neurologists the disease status and progress, but it is also essential for earlier and more efficient use of cueing techniques for tackling FOG. The proposed research in this field is mostly for FOG detection for PD patients. Furthermore, the proposed system can be used remotely to send a patient's information to neurologists which can save significant resources for both patients and doctors. Therefore, this part of our research propose an automatic, accurate and fast FOG assessment system (Kin-FOG) which addresses these limitations by introducing the improvements listed below:

- The proposed Kin-FOG system uses an RGB sensor based on Microsoft Kinect which is more convenient than wearable sensors for elderly people. Moreover, it can capture video and motion data in parallel.

- The proposed system can present a patient's FOG conditions to specialists for disease stage determination and treatment. The FOG status includes number of FOGs, length of each FOG, and the time when a FOG happened. The continuous assessment of patients by doctors with proper treatment can control FOG and eventually prevent patients from

unexpected falls.

- The novel FOG algorithm can be used remotely by PD patients, at their home or other places such as senior homes, to send their FOG status to their doctors instead of going to a clinic. This will save significant time, money, and energy for patients and specialists.

- The proposed FOG detection technique can be applied on video captured by devices other than Kinect after extracting a subject's pose from a video.

Gait assessment tools in clinics may have limited applications because of three main reasons including cost, complexity and specific space requirements [127]. Our proposed Kin-FOG system is non-invasive, cheap, user friendly with easy and fast set up. However, our FOG assessment system can only be used in a lab, clinic, or home environment, but not outdoor.

4.3 Proposed Method

In this part of the research, an automatic FOG assessment system is proposed for helping specialists with automatic assessment of FOG for PD patients and provide efficient and on time treatment. The general framework of the proposed system is shown in Figure 4.1. It has three phases, including data captured by Kinect, data preprocessing, and FOG assessment. All the steps of the Kin-FOG system are explained in the next subsections.

4.3.1 Data Acquisition

The Microsoft Kinect V2 is a non-wearable sensor for capturing the 3D motion of a person. It was invented for Xbox console game devices which allow interactions with body movements, voice, and images. It has three main parts consisting of a 3D depth sensor, an RGB camera and multi-array microphones which are shown in Figure 4.2. Kinect V2 is cheap compared to other video capture devices and other motion capture sensors. It employs an RGB camera with a resolution of 1920×1080 at 25 fps and a depth sensor with an output

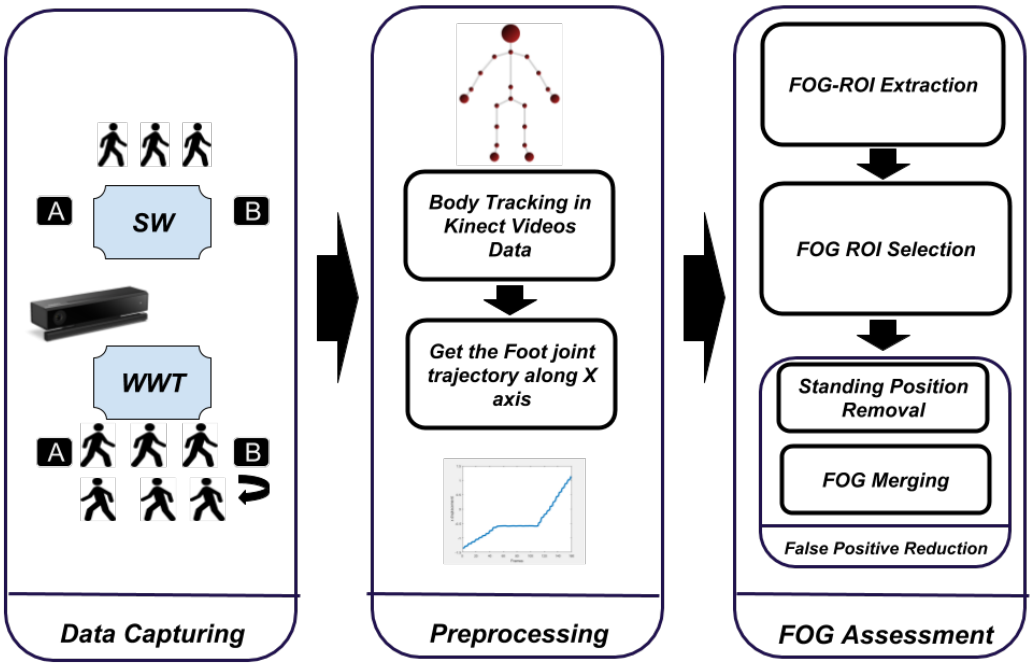


Figure 4.1: Kin-FOG framework for FOG assessment for PD patients.

resolution of 512×424 based on time of flight. The depth range is 0.5 to 4.5 m. Kinect V2 can detect up to 6 people, with each person represented by 25 joints. Figure 4.3 illustrates these joints of the body.

In our study, Kinect is used to capture data from 5 subjects. The subjects are healthy without any movement disability. The dataset consists of 1 female and 4 males. The average age of the subjects is 30.8. The subjects have different height and speed which is useful for having different types of data. Although the subjects are healthy, they are trained with real video on how FOG occurs for PD patients. The general information about the subjects is presented in Table 4.1. Two categories of experiments are done. The first experiment is in the Without Standing (WOST) mode and the second one is With Standing (WST). In both categories two types of tasks are used which are described as below:

- Simple Walking (SW):

  The subjects are asked to walk from a point (A) to another point (B). They do this experiment with two different number of FOGs (1 and 2) between the start and stop points. Figure 4.4(a) shows the SW experiment

scenario.

- Walking With Turning (WWT):
  The subjects are asked to walk from a point (A) to another point (B) and then turn and return to the starting point (A). They do this experiment with three different numbers of FOGs (3, 4, and 5). Figure 4.4(b) shows the WWT experiment. One of the FOG happens close to turning, and the rest occurs randomly between A to B.

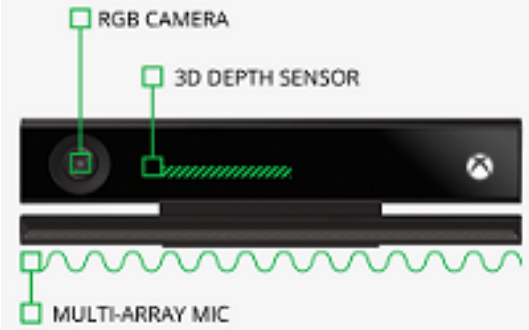


Figure 4.2: The components of Microsoft Kinect V2.

The Kinect sensor is set on a table with height 47 inches. The distance between points A and B is 104 inches and the subject walks in front of Kinect at a distance of 97 inches. Please note that Kinect can cover a longer range but because of the space limitation in our lab we only consider this set up. The video with motion data is captured for each subject with SW and WWT tasks for WOST and WST modes.

Subject Characteristics	***Range***	***Standard Deviation***
Age	26–33	1.87
Height (cm)	170–186	5.04
Weight (Kg)	55–85	9.09
BMI	17.8–24.8	2.44

Table 4.1: Subject demographics.

The number of FOGs in these experiments are presented in Table 4.2. To design a general and reliable system different numbers of FOGs are conducted

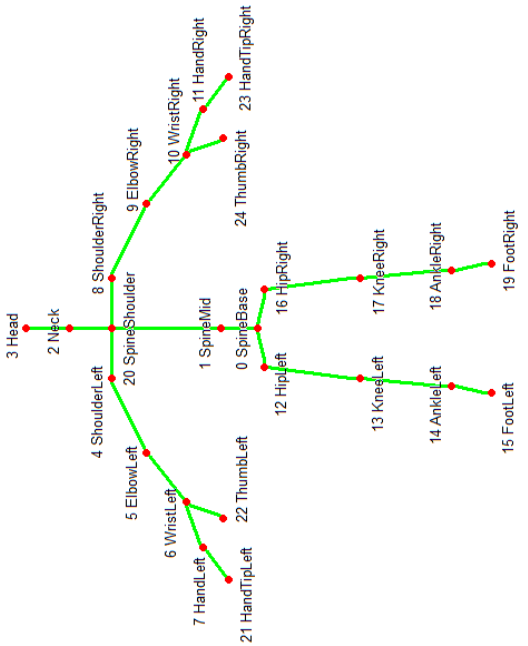
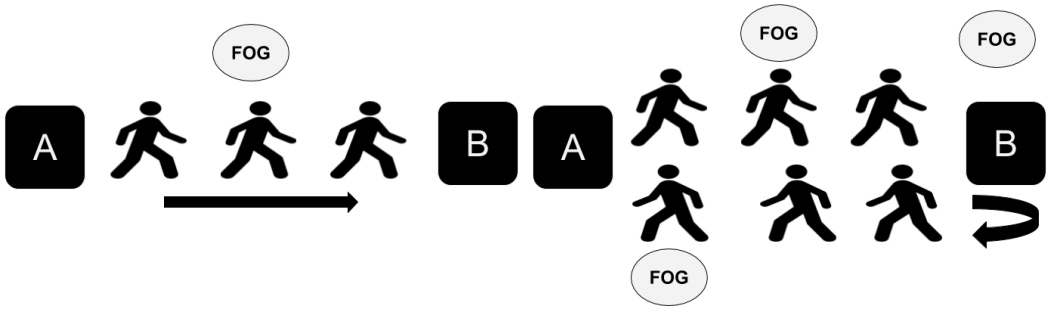


Figure 4.3: Kinect skeleton body joints with their indices. Kinect V2 can record motion data up to 25 joints.



(a) (b)

Figure 4.4: Different motion capture tasks in our experiments: (a) SW, (b) WWT.

in these experiments. Each subject is asked to repeat the experiments 3 times and the best experiment with the highest similarity to a real FOG phenomenon is selected to make sure the data is captured correctly.

Experiments/Modes	***SW***	***WWT***
WOST	1,2	3,4,5
WST	1	3

Table 4.2: The number of FOGs in each experiment and different modes.

4.3.2 Preprocessing

In this phase, the data captured by Kinect is processed to get the body joint's gait information and the proper motion signal for FOG assessment. In addition, the ground truth labels are extracted by checking the recorded video frames for final evaluation of the proposed method. The ground truth labels have three values including FOG, non-FOG, and non-related. The non-related labels are for the frames which are not related to the actual experiments.

Body Skeleton Tracking

The body skeleton and all the joints are extracted from the video data that are recorded by Kinect. Since Kinect records depth data, for each joint we have a list of $3D$ positions (X, Y, Z) and the length of the list is equal to the number of frames for the video. Figure 4.5 shows the coordinate system of Kinect when it is tracking the body joints. The Z axis data is the distance between the subject and the Kinect, which is almost stable in our experiments and thus provides no useful information. On the other hand, since the subject is walking along the X axis of the Kinect, the X trajectory data is selected for our FOG gait analysis.

Joint Trajectory Selection

As mentioned before, FOG happens when a person's foot is frozen and not able to move forward. Ahmadi et al. [4] proposed a low-cost system for human gait assessment which performs the following tasks: lower limb position estimation,

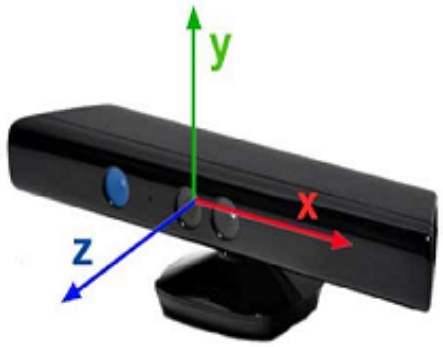


Figure 4.5: The coordinate system of the Microsoft Kinect V2 sensor.

orientation estimation, and 3D reconstruction of the estimated positions and orientations. They also propose a method for kinematic model adjustment at the end to cover the error from the wearable sensor's position uncertainty. The position estimation is done just by using the captured motion of Inertial Measurement Units (IMU) sensors on the feet. For orientation estimation and 3D reconstruction they used ankle and knee joints trajectories as well. FOG is mostly related to the leg joints which are hip, knee, ankle, and foot. Since, most of the papers in this field use sensors which are attached to these joints, we decided to select the foot and ankle as our FOG-related joints. The foot joints (left and right) are joint numbers 15 and 19 and the ankle joints (left and right) are joint numbers 14 and 18 as shown in Figure 4.3. The X trajectory of the right and left foot for all the subjects are collected for FOG assessment in the proposed system.

The SW experiment FOG assessment just needs the left foot X trajectory. However, the WWT experiment needs both left and right feet X trajectories to detect all the FOGs. In the WWT experiment the trajectory signal has two parts consisting of before turning and after turning. The left foot trajectory is used before turning and the right foot trajectory follows the same procedure for after turning. At the end, the analysis results of the left and right feet are fused to get the final FOG assessment for the WWT experiment. Figure 4.6 illustrates the process for the WWT experiment.

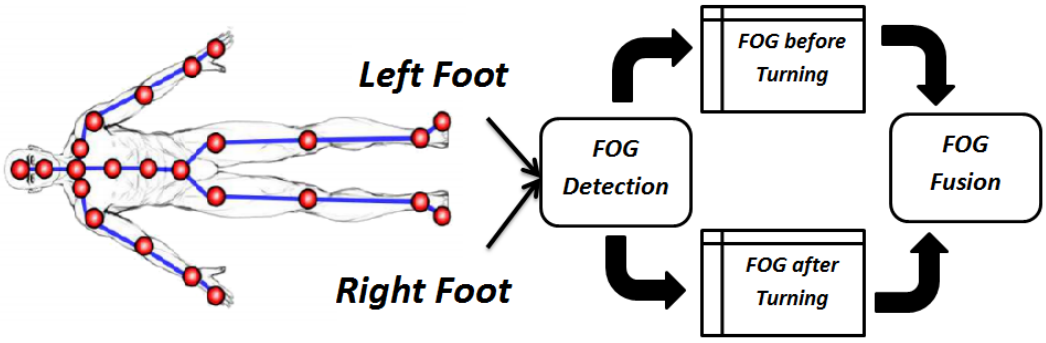


Figure 4.6: FOG assessment for the WWT experiment by our proposed Kin-FOG system using right and left feet. The left foot is used for 'before turning' and the right foot is used for 'after turning'.

4.3.3 FOG Assessment

After getting the X trajectory of the right and left feet, the goal is detecting the FOG episodes based on these signals. The Kin-FOG system follows three steps for FOG detection and monitoring which are described in the next sections.

FOG-ROI Extraction

As explained before, FOG episodes are the times when subjects are not able to move forward. Hence, FOG episodes are likely to be flat regions in the X trajectory of the foot joint. Figure 4.7 shows two samples of the X trajectory for the left foot of a subject for SW and WWT experiment tasks, in which the FOG regions are marked. Please note that the x-axis of the plots are frame IDs and the y-axis shows the X displacements. In Figure 4.7(a) the trajectory signal is related to the SW experiment with 1 FOG. The Figure 4.7(b) shows the X trajectory of the same subject in the WWT experiment with 3 FOGs which happen one before turning, and two after turning. The Turning Phase is marked with the orange array in this figure. Please note that the turning point is computed by finding the maximum point in the X trajectory plot. The X trajectory is defined as $f(x)$ for simplicity in clarifying the mathematics behind the proposed Kin-FOG assessment system.

For extracting the flat regions in the signal, we must compute the derivative and then find the regions equal or close to zero. Qiu et al. [127] used similar

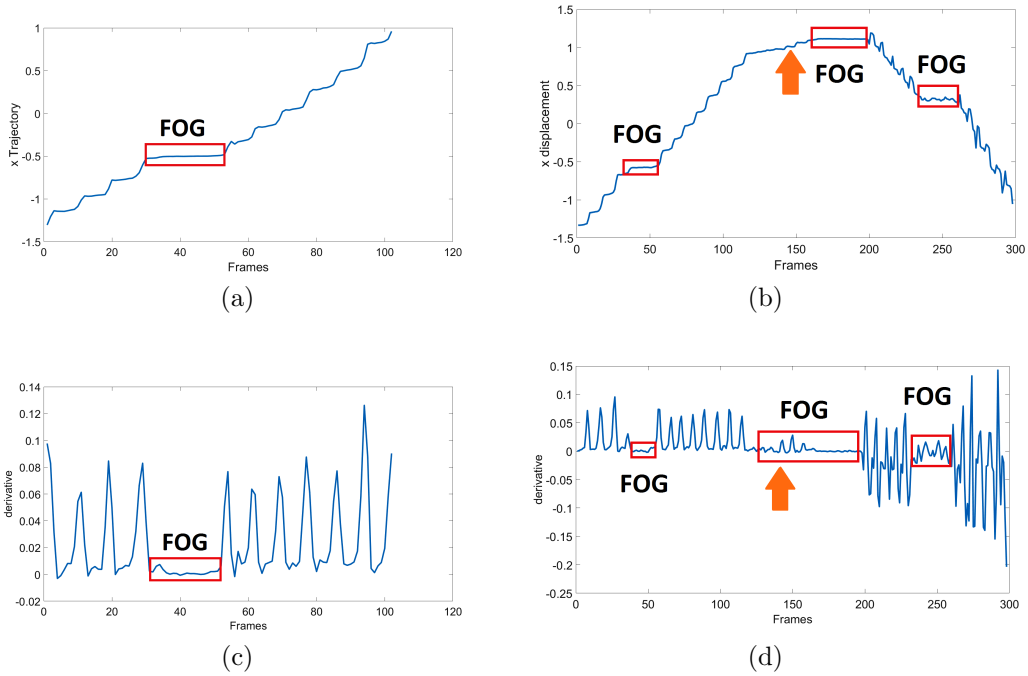


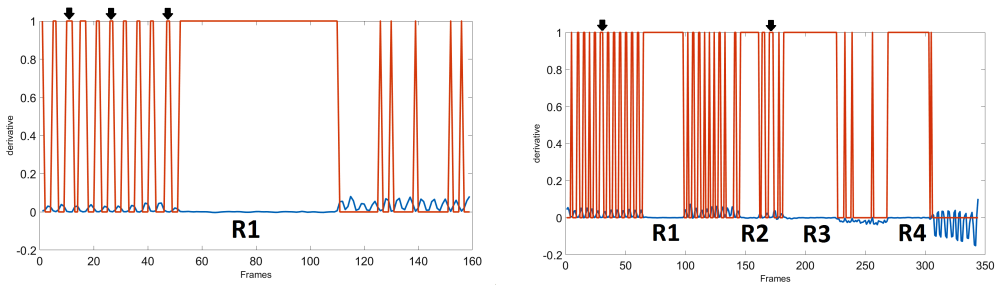
Figure 4.7: The X trajectory signals for: (a) SW with 1 FOG, (b) WWT with 3 FOGs. The turning point is marked with an arrow, and the derivative signals are: (c) SW with 1 FOG, (d) WWT with 3 FOGs.

techniques on foot joint trajectory for stance phase detection in the gait cycle. Figure 4.7 (c) and (d) illustrate the derivation of the previous X trajectory signals. The first plot, Figure 4.7(c), is related to the SW experiment and the second plot, Figure 4.7(d), shows the derivation signal of the WWT experiment. As shown, FOG episodes have derivatives close to zero. All these regions are going to be extracted as FOG (ROIs) and called FOG-ROIs. The derivative of the trajectory signal $f(x)$ is computed based on Equation (4.1).

$$g(x) = f^{'}(x_i) = \frac{f(x_i) - f(x_{i-1})}{f_i - f_{i-1}} \quad i = (1:n) \tag{4.1}$$

where $g(x)$ shows the derivative of the trajectory signal. The gradient computed by finding the distance between the displacement value of frame i, $f(x_i)$ and the displacement value of the previous frame $f(x_{i-1})$. Also, n is the total number of frames. We find all the areas with derivative close to zero based on Equation 4.2. These regions are the candidates of FOG.

$$h(x) = \begin{cases} 1, if|g(x)| \leq \epsilon \\ 0, otherwise \end{cases} \tag{4.2}$$



(a) (b)

Figure 4.8: The derivative plot in blue with the h(x) plot in red. Some samples of non-FOG regions with zero derivative are shown with an arrow. R1, R2, R3, and R4 are the real FOG regions. (a) SW with 1 FOG, (b) WWT with 4 FOGs.

FOG-ROI Selection

In the previous step, all the FOG-ROIs are extracted from the trajectory signal as candidates for FOG episodes. However, not all of them are real FOG episodes. For example, sometimes the derivative in just one frame is zero and the adjacent frames are not. This cannot be FOG since it is very short.

Figure 4.8 plots the $h(x)$ regions which is 1 for the area with derivative close to zero and is 0 for other areas. It can be seen that most of these regions are not related to FOG. Some samples of these non-FOG regions are shown with an arrow in Figure 4.8 for SW and WWT signals from the previous sections. Therefore, we must check $h(x)$ and find the places where there is a group of continuous 1 s which has "*uniform*" FOG patterns. The non-zero areas are found, and their lengths are computed. Then, in our Kin-FOG system the areas with length higher than a threshold ($FOGLen_{Thr}$) are selected as real FOG regions. By following this process, we ignore all the short zero derivative areas. The regions R1, R2, R3, and R4 are the real FOG regions which are selected after this process as shown in Figure 4.8.

FOG False Positive Reduction

After detecting the real FOG-ROIs, we must remove false positives areas. False positives (FP) can happen because of two reasons listed below:

- When a subject is in resting position or standing mode for a while during the experiment causing the foot to have a constant position. This FP reduction is called *Standing Position Removal* step.

- When we have a small movement in the middle of the FOG episode for which the system detects 2 FOGs instead of one FOG. This FP reduction is called *FOG merging* step.

The proposed Kin-FOG follows these two FP removal steps as explained in the next subsections.

Standing Position Removal The Kin-FOG system distinguishes the FOG episodes and the episodes in which a subject is just standing or resting in this phase. The similarity between both is that the foot has almost constant X position.

The leg joints, and hip, knee, ankle, and foot positions, are presented in Figure 4.9 in FOG and standing mode. Once subjects have FOG the angle between the ankle and foot joints and the ground will be changing; but when they are in the standing mode this angle is stable. Furthermore, the corresponding angle between the foot and ground is shown in Figure 4.9 with the red triangle. Observe that the angle is almost zero for the standing mode, but it changes in the FOG state. This angle is called α_{AG} for simplicity.

The α_{AG} is between two vectors in which the first one ($v1$) is between the ankle and foot joint and the second one ($v2$) is between the foot joint and the point that is the projection of the ankle joint on the ground line (*AG point*). The angle and the vectors are shown in Figure 4.10. α_{AG} can be computed by these two vectors based on Equation (4.3) for each frame which results in an angle list at the end ($AFG - list$). AFG stands for 'Ankle', 'Foot' and 'Ground.' The length of the $AFG - list$ is n which is the total number of frames.

$$\alpha_{AG} = \arccos \frac{v1.v2}{\|v1\|\|v2\|} \rightarrow AFG - list = \{\alpha_{AG1}, \alpha_{AG2}, ..., \alpha_{AGn}\} \qquad (4.3)$$

After computing α_{AG} for all the frames, the detected FOG episodes from the previous step are assessed in this angle data to remove the standing mode

areas. The displacement of the angle list ($AFG - list$) is taken to find out the amount of angle change over time for the current subject. The displacement can be computed by Equation (4.4).

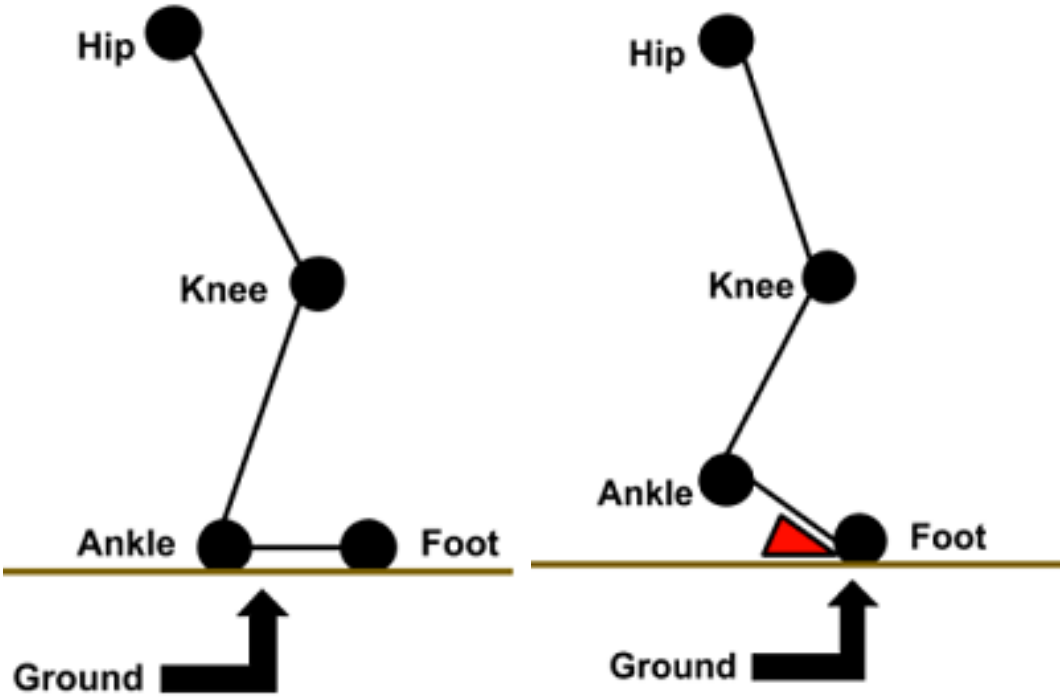
$$Disp_{AFGi} = \alpha_{AGi} - \alpha_{AG(i-1)} \quad i \in [1:n] \tag{4.4}$$

where n is the total number of frames. α_{AGi} is the angle in frame i and α_{AGi-1} is the angle in frame $i-1$. Figure 4.11(a) shows the current angle displacement in the SW experiment. The FOG episodes show higher changes compared to the standing episodes. The process of finding the resting mode regions is based on the peak values in that area. For finding the amount of change in this signal we get the gradient of the angle displacement data ($Disp - AFG$) using Equation (4.1). The gradient displacement signal is called ($GDisp - AFG$). Figure 4.11(b) shows the gradient of the current angle displacement. After detecting the FOG episodes ($FOG_{Candidates}$), they are evaluated in the ($GDisp_{AFG}$) signal. It can be seen that in Figure 4.11(b) the FOG regions have peaks with higher values compared to the resting area peak values. Therefore, the process of resting mode removal is based on finding the number of high-value peaks in the FOG candidate regions. The area with some high-value peaks is related to FOG and the area without the high-value peaks shows the resting mode. If the list of peaks in the current episode are $[p_1, p_2, ..., p_k]$ where k is the total number of peaks, the value for these peaks will be $[GDisp_{AFG}(p_1), GDisp_{AFG}(p_2), ..., GDisp_{AFG}(p_k)]$. After this the number of peaks with value higher than a simple threshold ($PksVal_{Thr}$) are computed, which is called $curFOG_{PksNum}$. The low number of high-value peaks results in a standing mode which needs to be removed from the FOG candidates list. The removal of resting area is done by making them zero since the FOG area will remain one in the FOG candidate list. The threshold for comparing the number of high-value peaks is $PksNum_{Thr}$.

Algorithm 1 Rest Removal Algorithm

procedure REST REMOVAL
    $AFG\text{-}list \leftarrow$ List of angles α_{AG}.
    $Disp_{AFG} \leftarrow$ Displacement of angle list α_{AG}.
    $GDisp_{AFG} \leftarrow$ Gradient of the $Disp_{AFG}$.
    $FOG_{Indxs} \leftarrow FOG_{Candidates}$ start and stop points.
    for *ii*,*jj* in FOG_{Indxs} **do**
        $curFOG_{PksVal} \leftarrow$ FindPeaks($GDisp_{AFG}$(ii to jj)).
        $curFOG_{PksNum} \leftarrow$ Length of the $curFOG_{PksVal} > PksVal_{Thr}$.
        if $curFOG_{PksNum} < PksNum_{Thr}$ **then**
            $FOG_{Candidates}$(ii to jj) $\leftarrow 0$
        end if
    end for
end procedure

The pseudo-code for rest removal is shown in Algorithm 1.



(a) (b)

Figure 4.9: The leg's joints position: (a) standing mode, (b) FOG mode.

In the SW experiment, only the left foot analysis would be enough to find the standing position and remove FP. However, in the WWT experiment as explained before (Figure 4.6) the left foot is going to be used for the

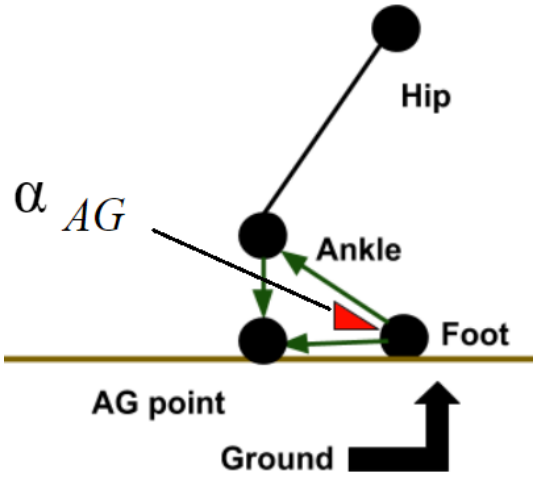
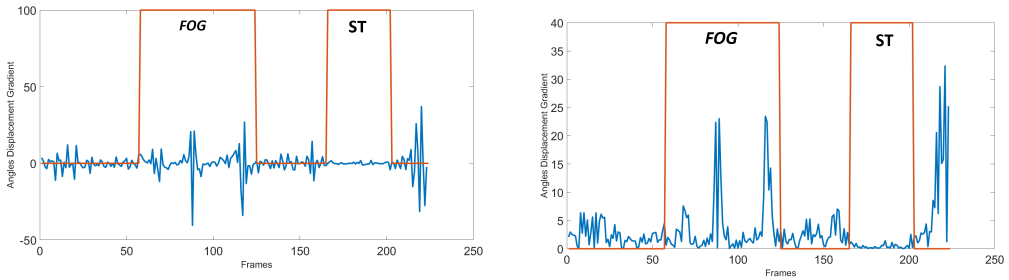


Figure 4.10: The angle between the ankle-foot joints and the foot-AG point is α_{AG}.



(a) (b)

Figure 4.11: The standing position removal plots for the SW experiment with 1 FOG and 1 ST mode: (a) angle displacement plot ($Disp_{AFG}$), (b) gradient angle displacement plot ($GDisp_{AFG}$). ST shows the standing mode.

analysis before turning and the right foot will be used for the analysis after turning. The angle displacement plot ($Disp_{AFG}$) for the left and right feet are shown in Figure 4.12(a) and (c), respectively. Furthermore, the gradient angle displacement plot ($GDisp_{AFG}$) for the left and right feet are presented in Figure 4.12(b) and (d), respectively. The FOG area and the standing mode area are determined in all the plots.

FOG Merging The proposed system finds adjacent FOGs with a small non-FOG area between them and combines these regions into one FOG to decrease the false positive rate of the proposed Kin-FOG system. Figure 4.13

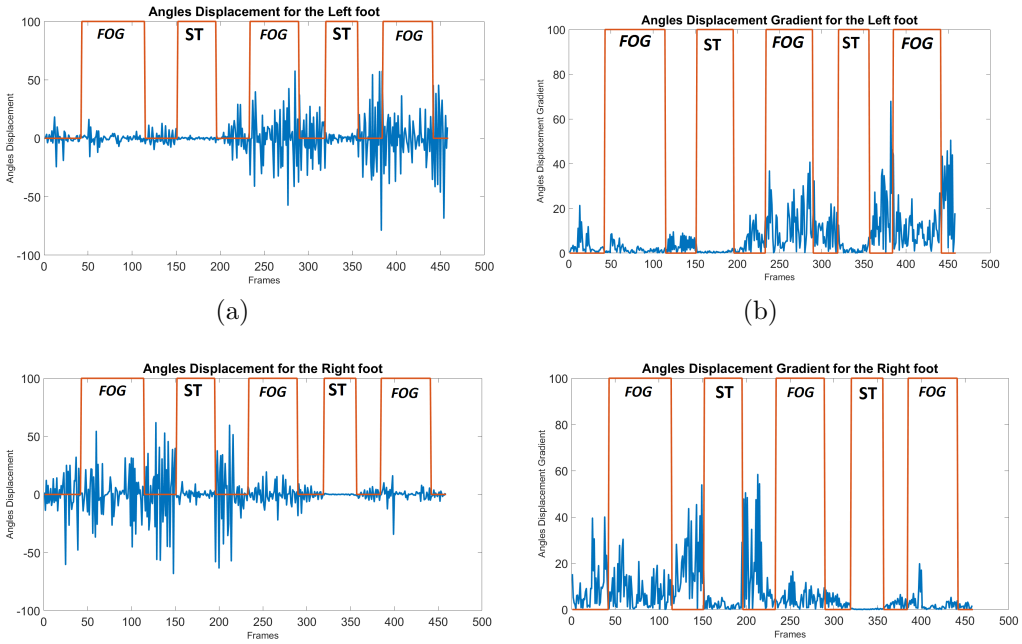


Figure 4.12: The standing position removal plots for the WWT experiment with 3 FOGs and 2 ST modes: (a) angle displacement plot for the left foot ($Disp_{AFG}$), (b) gradient angle displacement plot for the left foot ($GDisp_{AFG}$), (c) angle displacement plot for the right foot ($Disp_{AFG}$), (d) gradient angle displacement plot for the right foot ($GDisp_{AFG}$). ST shows the standing mode.

shows an example of this. The plot is for the SW experiment with 2 FOGs which are marked with labels. Observe that for the first FOG we have a small displacement (marked with the orange array) which causes the system to consider this as two FOGs and eventually report 3 FOGs instead of 2. This part of the system is for getting the right number of FOGs and their corresponding lengths which helps doctors have accurate information about their patients.

After finding the real FOG-ROIs, the time slot for each of them must be computed. Figure 4.14 presents the resulting FOG plots for the SW and WWT experiments of the previous signals (Figure 4.7). The number of FOGs and their lengths will be computed as well. The number of FOGs is the number of peaks in the FOG plot and the length of FOGs is computed based on Equation (4.5) for the final detected FOG-ROIs. $FOGLen_{fr}$ is the length of that particular

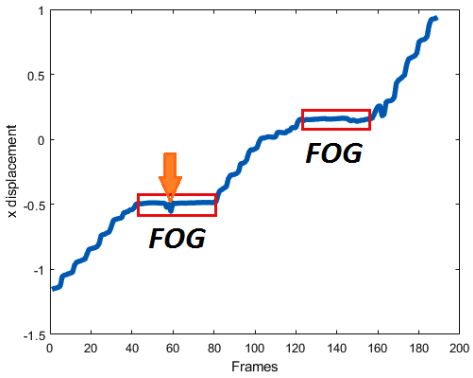
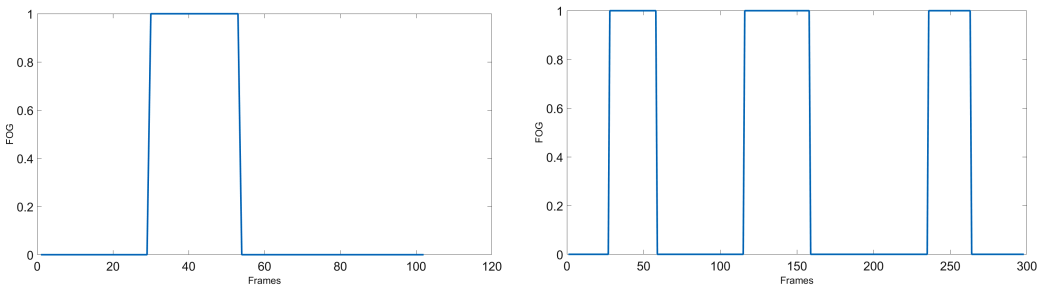


Figure 4.13: The X trajectory plot of a subject with 2 FOGs in the SW experiment. The arrow shows the part which may cause a false positive.



(a) (b)

Figure 4.14: The FOG plots for the subject with the trajectory in Figure 4.7 for: (a) the SW experiment with 1 FOG, (b) the WWT experiment with 3 FOGs.

FOG with the start frame ID (str_{FID}) and stop frame ID (sto_{FID}).

$$FOGLen_{fr} = sto_{FID} - str_{FID} \tag{4.5}$$

The computed length is based on the frame number but we can convert it back to the time space. Since Kinect captures data at 25 fps, the computed FOG length ($FOGLen_{fr}$) can be converted to a real-time space ($FOGLen_{tr}$) by Equation (4.6). The computed time unit is in milliseconds (ms).

$$FOGLen_{tr} = \frac{FOGLen_{fr}}{25} \tag{4.6}$$

In summary, the Kin-FOG system uses the foot and ankle joints motion data for FOG assessment for PD patients and outputs the detected FOGs, their lengths, and the time interval when they occurred.

4.4 Experimental Results and Discussions

The Kinect sensor is set up and subjects are asked to do two groups of experiments WOST mode and WST mode. In each of these experiments they perform two types of tasks including SW and WWT. The proposed Kin-FOG system uses the gradient displacement of the foot joint trajectory for finding the list of FOG candidates. It also uses the gradient displacement of the angle between the ankle and foot joints and the ground for false positive reduction from resting mode which has properties similar to FOG modes. Moreover, the merging algorithm at the end ensures that the system reports the correct number of FOGs to the users (doctors, specialists). The proposed system has three initial thresholds, which are $FOGlen_{Thr}$, $PKsVal_{Thr}$ and $PKsNum_{Thr}$. The values for these thresholds are $FOGlen_{Thr} = 10$, $PKsVal_{Thr} = 5$ and $PKsNum_{Thr} = 3$ which are constant for all the subjects in the SW and WWT experiments. Experimental results for each group are illustrated below.

The ground truth label for each subject is obtained based on the video data captured by Kinect. The video data is assessed and the label for each frame is determined. The ground truth label has three types of values consisting of *non-related*, *FOG* and *non-FOG*. The *non-related* frames are the ones that are not related to the experiments which are before point A and after point B (Figure 4.4). These frames are ignored for the evaluation part. The *FOG* frames are the ones for which a subject has FOG during the time interval. The rest of the frames are labeled as *non-FOG*. The evaluation of the proposed Kin-FOG frame is based on the FOG detection part for which quantitative results will be provided. After that, the number and length of the prediction will be evaluated by comparing to the real number and length of the FOG episodes. Other than quantitative results, qualitative assessment is also reported by the proposed FOG assessment system; such as the plots of FOG episodes for the subjects.

4.4.1 WOST Experiment

Video data is collected for each subject in WOST for different numbers of FOGs. As mentioned before for SW, the subjects perform 1 and 2 FOGs

and for WWT, 3, 4, and 5 FOGs. The FOG episodes happen randomly for each subject. The reason behind performing different numbers of FOGs in each experiment is having a more reliable and general system which works in different situations. The graphical user interface (GUI) is designed for our proposed Kin-FOG system, as shown in Figure 4.15. The proposed system has two modes including: (1) Offline FOG assessment, (2) Online FOG assessment, which both give useful FOG information on PD patients to the neurologists and care providers. The offline part gives all the required FOG information for the entire data captured by Kinect for the subjects. However, the online part is for visualization of the FOG status for each frame separately. The main GUI has two sections which are *'FOG Inputs'* (the inputs of the system) and *'FOG Output'* (the outputs of the system). The *'FOG Inputs'* part has options for choosing the patient, the task type (SW,WWT), the system's mode (offline, online) and the number of FOGs (for SW:$[1, 2]$ and for WWT: $[3, 4, 5]$). The *Num of FOG* parameter is for evaluation purpose since we need to get the right ground truth labels based on the selected criteria.

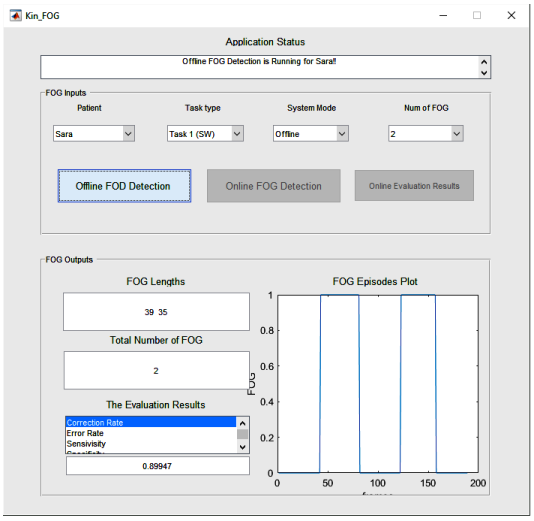


Figure 4.15: The main GUI of the proposed Kin-FOG system for FOG assessment of PD patients.

The offline FOG assessment uses the proposed FOG detection method and reports the *total number of FOGs*, the *FOG lengths*, and *FOG episodes plot*. The proposed system not only reports quantitative results but also provides

qualitative information, such as plots of FOG episodes. This information is useful for specialists to see in what situations patients have more FOGs. A patient might have a problem just when they turn but not when they walk straight. Therefore, they need to have treatment just for that special situation which cause FOG for them. This way doctors can provide the most efficient treatment, such as cueing to their patients. The FOG episodes plot illustrates the FOG time slots including the start frame IDs and the stop frame IDs. Furthermore, the *Evaluation Results* section presents different criteria to evaluate the Kin-FOG system to assess the current subject. These criteria are *correct rate*, *error rate*, *sensitivity* and *specificity*.

The online mode of the system has a separate GUI which is presented in Figure 4.16. The main GUI gets all the required information from the *FOG Inputs* part and shows the corresponding video frames for the particular subject in the online GUI. The online Kin-FOG presents the FOG status for each frame to the users (specialists or care providers) as shown in Figure 4.16. In

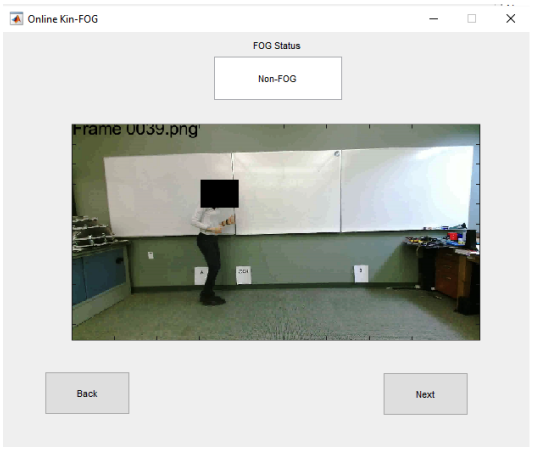


Figure 4.16: The online Kin-FOG GUI.

addition, the *'Online Evaluation Result'* button in the main GUI can show the evaluation results for the online mode of the system. Figure 4.17 shows the results of the online mode of the Kin-FOG for one of the subjects in the WWT experiment with 3 FOGs. The results show 3 FOGs with their lengths and the plot for showing the time slots. Also, the status of the Kin-FOG system is shown on top of the main GUI which presents exactly what the system is

doing. For instance, in this case it shows running for the subject with name *'Suba'* and the Kin-FOG system mode is online.

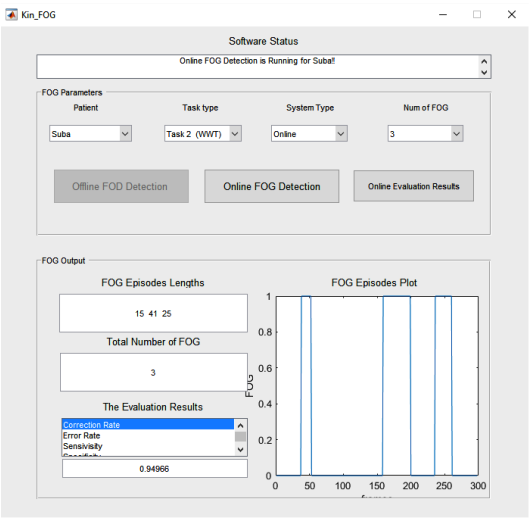


Figure 4.17: The Kin-FOG output for one of the subjects in the online mode, which consists of number of FOGs, their time slots, and their lengths. In addition, the performance of the Kin-FOG is also presented in the *'Evaluation Results'* section.

Quantitative results are shown for the proposed Kin-FOG system in two forms including: (1) local (subject-dependent), and, (2) general (subject-independent) similar to the evaluation in [137]. Table 4.3 shows the numerical evaluation results for different subjects using the above-mentioned criteria. The local evaluation for each subject is reported and the general evaluation is presented in the last row of the table. As the results show, the proposed Kin-FOG system has high accuracy in both experiments (SW and WWT).

The predicted number of FOGs and their lengths for each subject are presented in Table 4.4. *'PN'* means predicted FOG number and *'PL'* means predicted FOG lengths for each subject. The predicted number of FOGs are correct in all cases except 3 which are encircled in Table 4.4. The general evaluation is done for FOG number prediction in this part. Since we have 5 subjects and the number of evaluation tasks for each of them is 5 (2 FOGs types for the SW and 3 FOGs types for the WWT), the total number of experimental cases is 25. The general accuracy of the proposed Kin-FOG system in FOG

number prediction can be computed based on Equation (4.7).

$$PN_{accuracy} = ((25 - 3)/25) * 100 = 88 \tag{4.7}$$

The local accuracy for predicting the number of FOGs for each individual subject is listed in Table 4.4, last column. Each subject has 5 experimental cases (SW-FOG 1 and 2, WWT-FOG 3, 4, and 5). Thus, based on the number of correct predictions for FOG numbers in different cases the local accuracy for each subject is computed. In addition, FOG detection and FOG number prediction gives the corresponding FOG lengths. The predicted lengths are also presented in Table 4.4. The predicted lengths for the FOGs for all the subjects are compared with the real FOG lengths in each task for different experiments. The difference between the predicted FOG lengths and the real FOG lengths in most cases is less than 5 frames.

	SW				***WWT***			
Subjects\Evaluation	***Correction Rate***	***Error Rate***	***Sensitivity***	***Specificity***	***Correct Rate***	***Error Rate***	***Sensitivity***	***Specificity***
Subject 1	0.9525	0.0475	0.9268	0.9912	0.9387	0.06130	0.9112	0.9931
Subject 2	0.7979	0.2021	0.4726	1	0.9608	0.0392	0.9545	0.9611
Subject 3	0.9503	0.0497	0.9342	1	0.7896	0.2104	0.8499	0.6128
Subject 4	0.9360	0.064	0.9651	0.8919	0.8850	0.1150	0.8580	0.9644
Subject 5	0.9618	0.0382	0.9725	0.9306	0.9132	0.0868	0.8946	0.9755
General Evaluation	0.9197	0.08030	0.8542	0.9627	0.8960	0.1040	0.8910	0.9014

Table 4.3: Quantitative results of the proposed Kin-FOG system in the WOST experiment for PD patients. The local (subject-dependent) and the general (subject-independent) results are reported.

Task Type	***SW***				***WWT***						***Evaluation***
Task FOG Number	***FOG 1***		***FOG2***		***FOG3***		***FOG4***		***FOG5***		***Accuracy***
Prediction Number and Length ofFOGs	***PN***	***PL***	***PN***	***PL***	***PN***	***PL***	***PN***	***PL***	***PN***	***PL***	**PN**
Subject 1	1	"60"	2	"54,35"	④	"35,17,46,37"	4	"41,46,42,35"	5	"58,35,10,34,45"	80
Subject 2	1	"44"	2	"40,50"	3	"46,52,28"	4	"42,33,40,46"	5	"39,37,38,43,38"	100
Subject 3	1	"24"	2	"28,42"	3	"25,21,20"	4	"31,25,21,20"	⑥	"23,14,10,18,13,16"	80
Subject 4	1	"22"	2	"39,35"	3	"21,37,22"	4	"35,19,28,29"	5	"13,24,39,26,31"	100
Subject 5	1	"18"	2	"29,19"	3	"15,41,25"	⑤	"16,18,19,19,15"	5	"22,20,41,19,16"	80

Table 4.4: The predicted FOG numbers and lengths by the proposed Kin-FOG system in different tasks with different number of FOGs. *PN* is the predicted number of FOGs and *PL* is the predicted lengths of FOGs.

4.4.2 WST Experiment

Video data is collected for each subject in WST for SW and WWT tasks. In SW, the subjects perform 1 FOG and 1 standing mode and for WWT, they

perform 3 FOGs and 2 standing modes. The FOG and standing time slots are random for different subjects.

In this part of our research, we evaluate our system's ability to distinguish between the FOG and standing modes. The same GUIs are used for the FOG assessment in WST experiment for showing the detected FOGs, their lengths, and their plots. Table 4.5 presents the ability of Kin-FOG for detecting the real FOG episodes when we have the standing mode besides the FOG modes in our captured data with Kinect. The local evaluation for each subject is reported beside the general evaluation in the last row of Table 4.5 .

Our FOG assessment system is simpler than the proposed gait assessment system for stroke patients [127] and the one proposed for equestrian sport [163]; since it works based only on one individual sensor which provides the motion and video data together while others use fusion of data from different sensors. Furthermore, since the sensor is non-wearable, it is more comfortable for PD patients who are mostly elderly. Kinect provides video data besides the motion data; but, in the proposed gait analysis system by [127], [163], the authors use other optical sensors or cameras for collecting the video data in order to evaluate their proposed method at the end. Self-administration of Kinect compared to wearable sensors is simpler making it easier to set up. However, the proposed gait assessment system in [127] claimed that they solved the sensor drift problem by zero velocity update algorithm to address the challenge of using wearable sensors in an out-of-lab environment. It needs to be mentioned that the proposed system in [127], [163] is applicable in outdoor environments but our FOG assessment system only works inside a lab, a clinic or a patient's home. The proposed Kin- system is a useful tool for the neurologists and others who need to get the FOG information about patients.

Our proposed FOG assessment system is evaluated for healthy subjects performing simulations; however, in future work, it will be evaluated for real PD patients. The proposed foot-off detection algorithm by Amini et al. [10] was first evaluated for healthy subjects in [9] and later on PD patients in [10]. We have the same plan for our research. Even though the number of subjects in our experiments is limited, it needs to be noted that research with few subjects

can be faster for enlisting subjects, looking into their records, or performing biochemical investigations [62].

	SW				*WWT*			
Subjects\Evaluation	*Correction Rate*	*Error Rate*	*Sensitivity*	*Specificity*	*Correction Rate*	*Error Rate*	*Sensitivity*	*Specificity*
Subject 1	0.7640	0.2360	0.5581	1	0.8693	0.1307	0.8037	0.9913
Subject 2	0.9063	0.0938	0.8981	0.9254	0.8584	0.1416	0.9672	0.6973
Subject 3	0.9713	0.0287	0.9597	1	0.8610	0.1390	0.9174	0.6964
Subject 4	0.9792	0.0208	0.9664	1	0.9462	0.0538	0.9072	1
Subject 5	0.9628	0.0372	0.9479	1	0.9767	0.0233	0.9680	0.9946
General Evaluation	0.9167	0.0833	0.8661	0.9851	0.9023	0.0977	0.9127	0.8759

Table 4.5: The WST experiment's quantitative results for the proposed Kin-FOG system in FOG assessment of PD patients. The local (subject-dependent) and the general (subject-independent) results are reported.

4.5 Conclusions

FOG is one of the major motor symptoms of PD patients which can cause falls. Since, PD patients are mostly elderly, falling results in fractures and even death. Thus, it is critical to find a way to detect FOG and stop patients from falling. Neurologists require accurate information about the FOG status of patients for giving proper and effective treatments. In this research, an automatic, accurate and fast FOG assessment application (Kin-FOG) is designed, which works based on the video and motion data captured by a Microsoft Kinect sensor. The gradient of the displacement for foot joint trajectory is used for gait analysis and distinguishing between the FOG episodes. On the other hand, the angle between the foot and the ground is used for false positive reduction resulting from having resting modes in our gait analysis. Evaluation of the proposed method is done based on two types of experiments with and without resting modes (WST and WOST) and different modes (SW, WWT). Moreover, different number of FOGs are performed by the subjects to have a more reliable and general result of our assessment system. The outputs of the Kin-FOG system are the numbers and lengths of FOGs and the time interval for each of them. Kin-FOG is evaluated using data captured for 5 healthy subjects who are trained to imitate FOG. The experimental results show the local (user-dependent) and general (user-independent) evaluation for all the experiments. The overall accuracy rate for FOG prediction is around 90% for

both experiments which demonstrates the ability of the proposed system in FOG assessment. The proposed Kin-FOG system is low-cost, accurate, and easy to use for FOG assessment of PD patients. In addition, the proposed system can be used remotely at patients' homes for sending the FOG status to doctors. Even though the Kin-FOG is evaluated for a small dataset of healthy subjects performing simulations, the ability of the novel FOG assessment system will be evaluated for a larger number of real PD patients in our future work.

Chapter 5

Automatic Classification and Monitoring of Denovo Parkinson's Disease by Learning Demographic and Clinical Features

5.1 Introduction

PD is one of the main age-related neuro-degenerative disorders. This disease occurs because of a distortion of dopaminergic neurons in the brain, particularly in the region known as the SN. This disease affects around 10 million people around the world globally, with the prevalence rate increasing annually [98]. PD has a significant impact on the patients' quality of life and their care providers. Since the SN area is mostly responsible for controlling body movements, the major symptoms of PD are related to movement disability including rigidity, tremors, bradykinesia and postural instability [49]. However, it has non-motor symptoms as well, which are cognitive and behavioral abnormalities that can cause dementia in the later stages of the disease [2]. Although neuroimaging methods can provide complementary information in PD prognosis [140], in practice, the diagnosis mostly relies on the clinical judgment of an experienced neurologist. This is based on personal judgment, can be subjective and time consuming because of the high number of patients with a large amount of clinical information (demographic, imaging, laboratory, etc.) that the assessors

have to take into consideration [85].

5.2 Motivation

Earlier discovery of the disease increases the chances of more effective treatment. Therefore, it is beneficial to automatic assess the clinical information in order to help the neurologists and physiotherapists, who work with PD patients, so that they can deliver more accurate, easier and faster diagnosis and monitor progress.

In this research, we use machine learning techniques for automatic early PD classification and progress monitoring by making use of the clinical and demographic information. The demographic characteristics are available for every patient, and thus are useful to monitor PD patients at the early stages of the disease who are called *denovo PD*.

5.3 Dataset

The clinical data is downloaded from the Parkinson's Progression Markers Initiative, PPMI (www.ppmi-info.org/data) for research, which is a comprehensive set of clinical, imaging and bio-sample data defining PD progression and diagnosing biomarkers. The baseline summary in PPMI is obtained from the *denovo PD* and healthy subjects which include a set of clinical properties listed in Table 5.1. These clinical features are obtained using different methods such as questionnaire and imaging tools.

textbfFeatuers	***PD_range***	***PD_Mean***	***HC_range***	***HC_Mean***
Gender	0-1	-	0-1	-
Age (Years)	33-85	61.5	31-84	60.8
Number Of Relatives with PD	0-5	-	0-2	-
Years Of Education	5-26	15.5	8-24	16.0
MDS-UPDRS Total	6-72	31.9	0-20	4.5
MDS-UPDRS Part I	0-13	1.2	0-8	0.5
MDS-UPDRS Part I - Patient Questionnaire	0-20	4.3	0-11	2.4
MDS-UPDRS Part II - Patient Questionnaire	0-22	5.9	0-5	0.5
MDS-UPDRS Part III - Patient Questionnaire	4-49	20.4	0-13	1.2
UPSIT - Total Score	1-40	22.4	11-40	34
MoCA Score	17-30	27.1	26-30	28.2
GDS Score	0-14	2.3	0-15	1.3
SCOPA-AUT	0-39	9.5	0-20	5.9
SBR - Left Caudate	0.3-3.7	2.0	1.3-5.3	3.0
SBR - Right Caudate	0.4-4.0	2.0	1.3-5.1	3.0
SBR - Left Putamen	0.1-2.6	0.8	0.7-4.3	2.1
SBR - Right Putamen	0.1-2.5	0.8	0.5-3.8	2.2

Table 5.1: PPMI clinical and demographic data, the range and the mean value for PD and HC group are listed.

The feature range, mean value and their indices for the group of PD and the group of HC are shown in Table 5.1. The number of samples in the download PPMI clinical dataset is 564.

5.4 Proposed Method

The general framework of our proposed system is shown in Figure 5.1. The preprocessing step identifies the missing information and prepares the dataset for further analysis. The second step conducts features selection for the classification tasks based on the Mean Decrease Impurity (MDI) technique. In the last step, we use the RF classifier for two purposes, which include distinction PD from HC subjects and disease progress monitoring by predicting H&Y scores on identified PD subjects with values 0, 1 and 2. The proposed method handles both RF-based PD classification and stage prediction.

5.4.1 Preprocessing

In the preprocessing phase the subjects with missing values for the features listed in Table 5.1 will be excluded. After this step, the dataset has 17 features and 553 subjects composed of 402 PD subjects and 151 HC subjects.

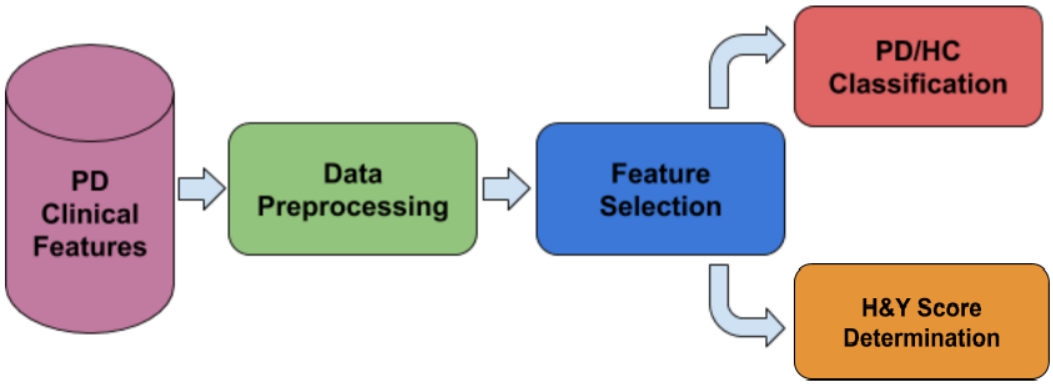


Figure 5.1: The proposed classification and progress monitoring RF-based model for PD.

5.4.2 Feature Selection

Using the preprocessed data, we evaluate the features ability in distinguishing between the PD and HC samples, and in monitoring PD severity by determination of the H&Y scores. RF is a popular machine learning technique because of its good accuracy, robustness and ease of use [23]. This method builds a set of decision trees for performing classification. Each node is a condition on a single feature, which decides how to divide the data into two similar sets. *Impurity* measures how well this condition is optimized to make a decision. In our method, a RF based feature selection method known as *Gini importance* or MDI is used. The MDI method calculates each feature's importance [72] by averaging over the impurity reductions in all the trees as shown in Eq. 5.1.

$$Imp(X_m) = (1/N_T)(1 - \sum_{k=1}^{N_T} Gini(X_m)^k) \tag{5.1}$$

where $Imp(X_m)$ is the MDI importance for variable X_m (m_{th} feature) and N_T is the total number of trees. It adds up the decrease of Gini index ($Gini(X_m)$) which measures the amount of impurity for variable X_m (current feature) in the k_{th} tree and gets the average over all of them. MDI has the advantage of easy and fast computation [72].

The features with their ranks and importance values are passed to the classification step for the two goals of PD diagnosis and PD progress monitoring. In accordance to the literature [157][144] for having a fair disease stage analysis,

only the scores from *MDS-UPDRS Part III* in H&Y determination are used and the other *UPDRS* score features are excluded. On the other hand, in the PD diagnosis process all the *UPDRS* related features are excluded in order to have a reliable and fair analysis.

5.4.3 Classification

This section has two parts: 1) PD diagnosis, and 2) PD progress monitoring using H&Y score determination. For these goals data normalization is performed for reducing data redundancy, increasing data integrity and making the classification more reliable. Min-Max normalization is the most popular method where the range values for the features are normalized to $[0, 1]$ using Eq. 5.2.

$$x_{norm} = (x - min(x))/(max(x) - min(x)) \tag{5.2}$$

After normalization, data division and classification are conducted. In this paper, 80% of the data is used for training and the rest 20% is the test set. The training process is done using a 10-fold cross validation technique. The classification algorithm used in this research is RF, which is also used for feature selection in the previous step. As explained earlier, RF is one of the strong ensemble learning classifiers and it uses a set of decisions trees [23]. It has many applications in classification and regression problems because of its important characteristics, such as handling high dimensional features, as well as noisy or imbalanced data and missing values [82]. Since the PPMI clinical data is imbalanced (the number of PD is more than the number of HC), RF is the appropriate classifier to apply on the most important features for PD diagnosis in the first part and for PD progress monitoring using the H&Y rating score in the second part of the proposed model.

5.5 Experimental Results and Discussion

In this part, the experimental results of the proposed model are presented for the two goals of this paper. For the first goal, we want to divide the subjects into two general classes: PD and HC, by using clinical features. After

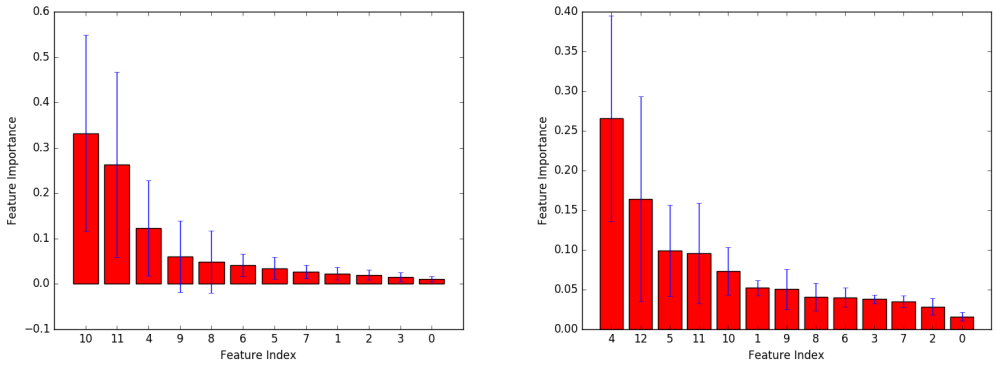
preprocessing, which makes the data more consistent and reliable, the total number of features for PD/HC classification is 12. These features are listed with their indices in Table 5.2.

Task	***Features***
PD Diagnosis	*0) Gender* *1) Age (Years)* *2) Relative Numbers with PD* *3) Years Of Education* *4) UPSIT - Total Score* *5) MoCA Score* *6) GDS Score* *7) SCOPA-AUT* *8) SBR - Left Caudate* *9) SBR - Right Caudate* *10) SBR - Left Putamen* *11) SBR - Right Putamen*
PD Progress Monitoring	*0) Gender* *1) Age (Years)* *2) Relative Numbers with PD* *3) Years Of Education* *4) MDS-UPDRS Part III - Patient Questionnaire* *5) UPSIT - Total Score* *6) MoCA Score* *7) GDS Score* *8) SCOPA-AUT* *9) SBR - Left Caudate* *10) SBR - Right Caudate* *11) SBR - Left Putamen* *12) SBR - Right Putamen*

Table 5.2: The list of the features for PD diagnosis and progress monitoring (H&Y score determination).

Figure 5.2 (a) shows the MDI feature importance plot for the PD/HC classification. The x-axis shows the feature indices based on Table 5.2 and the y-axis is the feature importance. The second goal of this paper is PD progress monitoring. As mentioned before, the H&Y rating scale is a popular method for monitoring PD progress. In our analysis, H&Y scale can have three values 0, 1 and 2, which are equivalent to mild, moderate and severe levels for PD. Thus, we have 3 classes. The list of features for H&Y score determination is the same as the features for PD diagnosis plus *MDS-UPDRS Part III - Patient Questionnaire* which results in 13 features. The list of these features

are presented in Table 5.2. The feature importance is computed based on the MDI algorithm. Figure 5.2 (b) presents the feature importance plot for the H&Y score determination. Based on the feature importance plots in Figure 5.2 demographic features such as *'Age'*, *Gender* and *Number of Relative with PD* have lower feature importance in PD diagnosis and PD progress determination.



(a) (b)

Figure 5.2: The feature importance plots based on the RF method for: (a) PD diagnosis (PD/HC classification), (b) PD progress monitoring (H&Y rating score determination).

Once the feature importance values are computed for PD diagnosis and progress monitoring, we need to select the best features for each task. The threshold for selecting the most important features is considered as the hyper parameter in our proposed method. The value list for this hyper-parameter is $[0.06, 0.08, 0.10, 0.12]$. After feature selection, RF is applied for classification using the 10-fold cross validation method. The number of elements in the training part of classification is 442 and in the testing part is 111.

The grid search method for tuning and finding the best parameter's value for the hyper-parameters is used for RF based classification. The average classification results are then reported.

In this work, the criteria for evaluating the proposed method are accuracy,

specificity and sensitivity, which are computed using Eq. 5.3 [149].

$$\begin{aligned} Accuracy &= (TP + TN)/(TP + TN + FP + FN) \\ Specificity &= TN/(TN + FP) \\ Sensitivity &= TP/(TP + FN) \end{aligned} \tag{5.3}$$

where, TP is the number of cases correctly classified as PD patients, FP is the number of cases incorrectly classified as patients, TN is the number of cases correctly classified as healthy and FN is the number of cases incorrectly classified as healthy.

In the second part of our research (H&Y score determination), we use precision, recall and f1 score to validate our results. These metrics are computed using Eq. 5.4 .

$$\begin{aligned} Precision &= TP/(TP + FP) \\ Recall &= TP/(TP + FN) \\ F1Score &= 2 * (Recall * Precision)/(Recall + Precision) \end{aligned} \tag{5.4}$$

Method	***Accuracy***	***Specificity***	***Sensitivity***
Proposed method in [126]	96.14	95.03	96.55
Proposed RF-based method	98.19	98.38	97.18

Table 5.3: Comparison of the PD/HC classification for the method in [126] versus our proposed RF-based method.

Class	***Accuracy***	***Specificity***	***Sensitivity***
0	0.97	0.96	1
1	0.85	0.91	0.70
2	0.85	0.86	0.86
Avg/Total	0.89	0.91	0.85

Table 5.4: The classification results of the proposed PD progress monitoring using H&Y rating score based on accuracy, specificity and sensitivity.

Class	***Precision***	***Recall***	***F1-score***	***Number of Samples***
0	0.91	1	0.95	30
1	0.75	0.70	0.72	30
2	0.84	0.84	0.84	50
Avg/Total	0.83	0.84	0.83	110

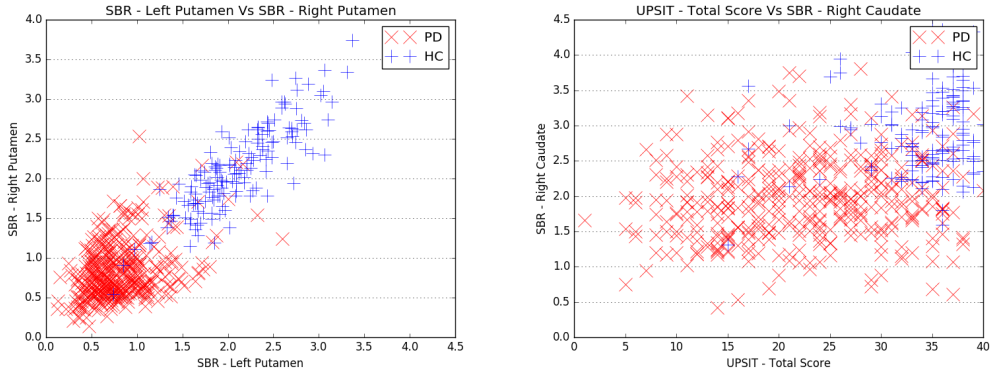
Table 5.5: The classification results of the the proposed PD progress monitoring using H&Y rating score based on the precision, recall and F1-score.

The quantitative results for PD diagnosis is presented in Table 5.3. The results are compared with the method in [126]. The proposed RF based classification model outperforms the method in [126]. The comparison demonstrates that the selected features in our method are more powerful than the features used in [126]. Additionally, the method in [126] does not account for the imbalance problem of their dataset. In contrast, the problem is addressed by our RF approach.

In PD progress monitoring using H&Y score determination, our proposed model evaluated results are presented in Table 5.4 and Table 5.5. Since there are multiple classes, the accuracy, specificity and sensitivity are computed individually per class and the average is presented in Table 5.4. The precision, recall and f1-score of the individual classes and the total are computed as reported in Table 5.5. The accuracy of our proposed progress monitoring model is 89.78% which is higher than the accuracy of the decision system in [157]. We should mention that the analysis in [157] is for 5 classes of H&Y score but in our proposed model, we have only 3 classes. Therefore, evaluation of the proposed system for the same number of classes will be conducted in future.

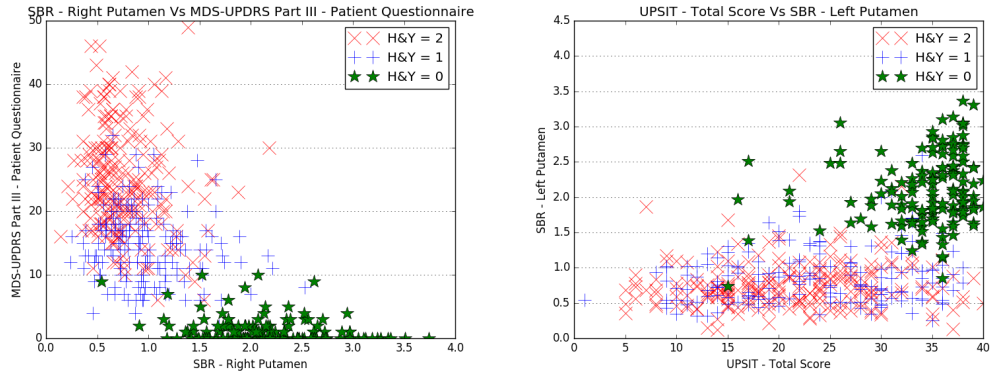
In order to demonstrate the ability of the selected features in distinguishing PD and HC, the data distribution is analyzed in Figure 5.3. In this plot, the ability of two pairs of selected features are illustrated in data classification, where the first pair is *SBR - Left Putamen* vs *SBR - Right Putamen* (Figure 5.3 (a)) and the second pair is *UPSIT - Total Score* vs *SBR - Right Caudate* (Figure 5.3 (b)). As the plots illustrate, the data are clustered in two groups perfectly by these two pairs of features. Similarly, for showing the ability of these selected features in determining the disease stage based on H&Y rating

scale, the data distribution is plotted in Figure 5.4. In this plot, the first pair of features is *SBR - Right Putamen* vs *MDS-UPDRS III* (Figure reffig:featdisthy (a)) and the second one is *UPSIT - Total Score* vs *SBR - Left Putamen* (Figure 5.4 (b)). As the plots show, the data are clustered well into three groups by these two pairs of features.



(a) (b)

Figure 5.3: The data distribution plots based on the two pairs of top features for PD/HC classification. (a) *SBR-Left Putamen* vs *SBR-Right Putamen*, (b) *UPSIT-Total Score* vs *SBR-Right Caudate*.



(a) (b)

Figure 5.4: The data distribution plots base on the two pairs of top features for H&Y score determination. (a) *SBR-Right Putamen* vs *MDS-UPDRS III*, (b) *UPSIT-Total Score* vs *SBR-Left Putamen*.

As has been noted, the *SBR* related features, the *UPSIT - Total Score* and the *MDS-UPDRS III* are the most important clinical information for

PD diagnosis and progress monitoring. Even though, it is commonly known that PD is an age-related disease, our analysis shows that age cannot be a good discriminator feature for the PD by itself. Because of the various aging scenarios that even a younger subject might have PD. Another key contribution of our work is that the proposed method for H&Y score determination can be used to supplement a clinical dataset that has missing H& Y values.

5.6 Conclusion

Early automatic detection and monitoring of PD is critical for on-time and more effective treatment of patients. The current diagnosis process is based on the demographic and clinical features evaluated by a specialist which can be subjective and time consuming. In this paper, an early automated detection and disease stage determination model based on the RF algorithm and machine learning techniques is proposed. The proposed system has three general phases consisting of preprocessing, features selection and classification. Feature selection is performed using the MDI method. Classification is composed of two parts: PD diagnosis and PD monitoring by H&Y rating score determination. The PPMI dataset is used as the PD demographic information for evaluation in the proposed method. Our experimental results show high accuracy in classifying PD and HC, and H&Y score determination, which outperforms state-of-the-art methods. Moreover, the proposed method can also be applied in clinical dataset, which has missing values for H& Y rating scales. In future work, we will further assess the ability of the proposed method as a supplementary tool for determination the H&Y rating scales in the dataset with missing values. Furthermore, we will evaluate the proposed model for determining the PD progress by determination of the UPDRS rating score.

Chapter 6

Comparing Subcortical Features Ability in Parkinson's Disease Assessment using only MRI T1 Images

6.1 Introduction

PD is the second most significant neurodegenerative disease after Alzheimer's Disease (AD). Middle-aged and elderly people are most vulnerable to PD. This disease is provoked by progressive loss of dopamine generating neurons in the brain which manifests as two types of symptoms: motor and non-motor. The motor symptoms are bradykinesia, muscle rigidity, tremor and abnormal gait [146], whereas non-motor symptoms are mental disorders, sleep problems and sensory disturbance [33]. Even though, there are some medical methods for PD diagnosing and progress determination, the results from these experiments are subjective and is highly dependent on the expertise of clinicians. On the other hand, most of these diagnosis techniques are expensive and time consuming for patients [130]. Neuroimaging techniques have remarkably improved the diagnosis of neurodegenerative disease. There are different types of neuroimaging techniques; MRI is among the most popular one because it is a cheap and non-invasive method and has the ability to produce high resolution images.

6.2 Motivation

PD patients show motor symptoms when they lose almost 80% of their brain's dopaminergic cells [1]. However, the non-motor symptoms might reveal at the earlier stages of the disease. The non-motor symptoms are associated with structural changes in the brain. MRI has the potential to detect the sub-cortical volume and shape variation in the brain [168]. These facts prove the urgency of having a CAD system for early and automatic detection of PD using MR images. In this research, the volume and surface-based features from the brain's subcortical regions are assessed and compared for PD diagnosis with machine learning techniques. T1 MR images have been used since they have the potential for showing brain structural changes, which are essential for early diagnosis of PD.

6.3 Proposed Method

Figure 6.1 shows the general framework of this research. In the first step, T1 images are preprocessed by neuroimaging analysis package, Freesurfer, to prepare the data for further analysis. After that, feature extraction is conducted to extract the volume and surface-based features of the subcortical regions of the brain. The extracted features are forwarded to the classification using: RF and SVM. Feature selection is done with PCA for choosing the most useful features and dimensionality reduction before classification.

6.3.1 Preprocessing

Preprocessing is an essential step in designing CAD systems. In this research, for extracting the brain subcortical characteristics using MR images, several preprocessing steps are performed. Freesurfer is used for preprocessing as a surface-based analysis package for the 3D MRI data. FreeSurfer is a package for the analysis and visualization of structural and functional neuroimaging data from cross-sectional or longitudinal studies [168]. The FreeSurfer library conducts cortical reconstruction, subcortical volumetric segmentation, and

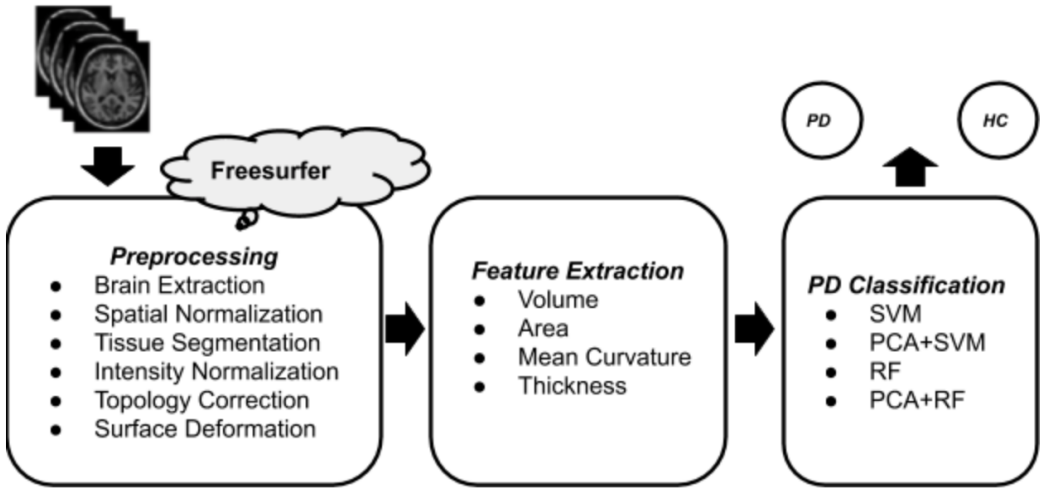


Figure 6.1: PD classification framework using machine learning and just with T1 MR images.

preprocessing [54].

Preprocessing process begins by converting DICOM, or other native scanner format, to the mgz format. In the next step, multiple scans from each subject are registered using the first scan as the template, and a single averaged, motion corrected volume for each subject is generated as output. The next few steps of volume processing for each subject begin with the output of motion correction. Several intensity normalization steps are next, along with a transformation to Talairach space. The intensity corrected T1 volume is fed into an watershed which strips out the skull and any remaining background noise. After creation of the skull scripted volume, the subcortical processing and segmentation occurs, yielding an automatic labeling of subcortical structures in the brain volume. It has to be noted that this is the most time-consuming phase in the processing. In this last portion of the volume processing pipeline, the input volume is normalized and segmented to generate volume containing only white matter. Subsequent automatic topology correction will be done for generating the WM volume. Finally the hemispheres will be separated from each other and from the brain stem, and creates a binary mask that distinguishes the two hemispheres for use in the surface processing pipeline. Figure 6.2 shows the MRI for a sample MR T1 image and the resulting image after preprocessing. The T1 image is presented in three different planes: axial, sagital, and coronal.

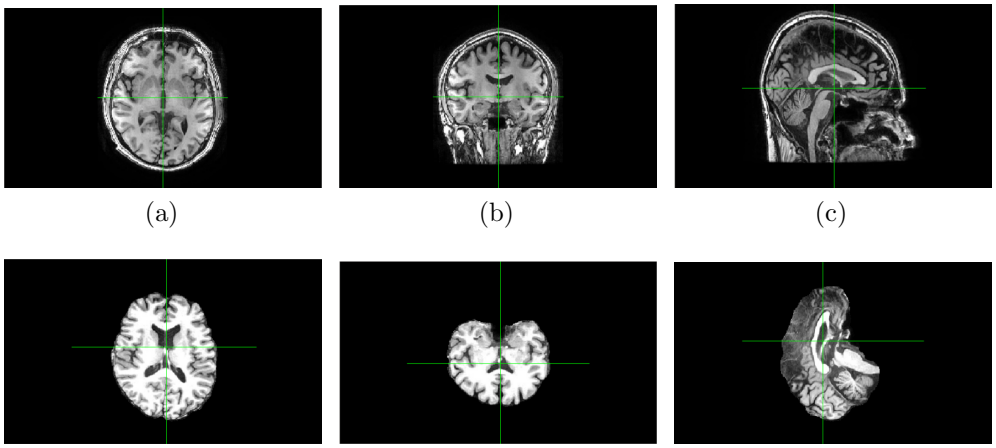


Figure 6.2: A MR T1 image in different planes: (a) Axial, (b) Coronal, (c) Sagital. The preproceesing results using Freesurfer for the T1 sample in different planes: (d) Axial, (e) Coronal, (f) Sagital.

6.3.2 Cortical Features Extraction

After preprocessing the MRI scans, volume and surface-based characteristics of the brain are extracted from the segmented subcortical regions. The output from the volume processing stream is used in surface creation; the entire surface processing stream is run twice, once for each hemisphere. In this stream, the research volume is first tessellated to create the orig surface. The orig surface is smoothed and inflated. Next, the topology correction is automatically run once. In the automatic topology correction steps, the inflated surface is transformed into spherical coordinates, corrected, and then smoothed and inflated again. Afterwards, the final surfaces are created. Visual checking of the final surfaces is necessary to check for geometric defects that may be present in the white and pial surfaces. Two versions of the surface parcellation are run, each using a different atlas. An output file is generated for each parcellation containing measurements of average thickness, surface area, mean curvature and more for each labeled area. In this research four features are considered for PD diagnosis: Volume, surface Area, Mean curvature and thickness. In the next part, the assessment over these four subcortical features beside their fusion are

explained. The number of volume, area, thickness and mean curvature features are 138, 74, 74 and 72 respectively.

6.3.3 Dimension Reduction and Classification

After the extraction of subcortical features from the different 3D MRI T1 images, the next step is classification. The goal is to use the extracted features for distinguishing between PD and HC subjects. The extracted features can contain redundant information, which may affect the classification result. Therefore, we perform feature reduction before classification. PCA is a popular feature reduction method in many applications such as neuroimaging.[129]. It makes a smaller number of uncorrelated variables (features) by linearly transforming correlated variables [131]. The resulting uncorrelated features are principal components which are capturing most of the variance in the data [129].

Classification is then performed using machine learning classifiers: RF and SVM. SVM [122] is a well-known supervised machine learning algorithm for classification and regression. It performs classification tasks by making optimal hyperplanes in a multidimensional space that distinguish different classes of data. This classification method is more popular because it's easier to use, has higher generalization performance and little tuning compared to other classifier. In our case, the kernel SVM is used. RF is an ensemble learning method for classification, regression and other tasks. This method is presented by Breiman [1] [91], which creates a set of decision trees (weak classifier) from randomly selected subset of training data. It then aggregates the votes from different decision trees to decide the final class of the test object. In the current stage of this research, we tested how accurate decisions can be made by RF with the data coming from PD's MRI volumes.

We apply four classification models on our four subcortical features, volume, area, mean curvature, and thickness. Moreover, the discriminating power of the fused features is also considered in this research. The four classification models are: 1) SVM, 2) PCA+SVM, 3) RF, and 4) PCA+RF. The general goal is classifying the T1 images using the mentioned features and classification models to two groups: PD and HC.

6.4 Experimental Results and Discussion

In this section, the dataset used in this research is explained and the results of different classification models are presented. We compare the effectiveness of the RF and SVM classifiers with and without using PCA. Our objective is to classify subjects T1 images to PD and HC using four important brain's subcortical features.

6.4.1 Dataset

The data used in the preparation of this article is the T1-weighted brain MR images which is obtained from the PPMI database `www.ppmi-info.org/data`. PPMI is a public and large-scale dataset to identify PD progression biomarkers [91]. The data that is used in our study contain the original T1 MR image of 598 samples with 411 PD and 187 Control. The summary of the data base demographic data are presented in Table 6.1.

Data Type	***Class***		***Sex***		***Age***		
	PD	***HC***	***F***	***M***	***(25-50)***	***(50-76)***	***(75-100)***
Number of Subjects	411	187	217	381	81	472	45

Table 6.1: Demographics information of PPMI dataset.

6.4.2 Results

Out of 598 subjects MRIs, 543 images were successfully preprocessed. Other images were excluded from the dataset due to poor quality of the original images or unknown parameters values during preprocessing pipeline. The number of PD and HC in the preprocessed MRI data are 369 and 174 respectively. The size of the MR images after preprocessing by Freesurfer package is $256*256*256$ because of using Talairach atlas. Since the number of PD is much more than the number of HC samples (more than 2 time), the dataset needs to be balanced. The method for balancing the dataset is over-sampling, which increases the number of samples in the minority group (HC) to make them equal or close to the number of samples in the majority class (PD). PCA used as dimensionality reduction beside RF and SVM as classifiers for classification of the MRI T1.

80% of the data is used for training and the rest (20%) for testing the classifier models. In this classification process, 10-fold cross validation and grid search with hyper parameters are used to get the most statistically accurate and reliable results. The data balancing process is done over the training data, which makes both the number of PD equal to the number of HC. There are two sets of experiments for RF and SVM, in which the first is without using PCA and the second one is with PCA. The evaluation metrics used in this research for comparing the results of the classification algorithms include specificity, sensitivity and accuracy, which are defined in Eq.6.1.

$$\begin{aligned} Accuracy &= (TP + TN)/(TP + TN + FP + FN) \\ Specificity &= TN/(TN + FP) \\ Sensitivity &= TP/(TP + FN) \end{aligned} \tag{6.1}$$

TP is the number of cases correctly classified as PD, FP is the number of cases incorrectly classified as PD, TN is the number of cases correctly classified as healthy and FN is the number of cases incorrectly classified as healthy. The four classification models are used to evaluate the discriminating abilities of the four brain subcortical features. In addition, fusion of these features are also evaluated for PD diagnosis using the classification models. For the fusion part, the subcortical features are concatenated for each subject in order to create the fused feature data as the input for the classification models.

Table.6.2 shows the general comparison between the calcification models. The experimental results are achieved by averaging the results of on 5 runs of 10-fold cross-validation for MR T1 data. Different criteria are reported in Table.6.2 by the mean and standard deviation values.

The quantitative results show that PCA+RF model with fusion of subcortical gives the best accuracy compared to other classification methods. However, the specificity values for RF is slightly lower than SVM whether with PCA or without PCA, which provides the capability of SVM in classification of Healthy subjects.

Most of the proposed studies use SPM and VBM toolbox for neuroimaging analysis. In this research, one of the important goals was to evaluate FreeSurfer

Class Methods	*SVM*	*RF*	*PCA+SVM*	*PCA+RF*
Features	*acc,spes,sens*	*acc,spes,sens*	*acc,spes,sens*	*acc,spes,sens*
vol	0.53 ± 0.04,0.31 ± 0.09,0.64 ± 0.03	0.63 ± 0.02,0.22 ± 0.03,0.82 ± 0.02	0.53 ± 0.03,0.27 ± 0.06,0.66 ± 0.08	0.65 ± 0.03,0.16 ± 0.06,0.85 ± 0.02
Area	0.60 ± 0.04,0.44 ± 0.11,0.68 ± 0.05	0.60 ± 0.03,0.21 ± 0.04,0.83 ± 0.03	0.55 ± 0.02,0.39 ± 0.05,0.65 ± 0.05	0.59 ± 0.04,0.17 ± 0.05,0.83 ± 0.05
Thickness	0.57 ± 0.02,0.41 ± 0.09,0.64 ± 0.06	0.55 ± 0.04,0.43 ± 0.06,0.62 ± 0.05	0.60 ± 0.04,0.41 ± 0.10,0.68 ± 0.07	0.58 ± 0.06,0.36 ± 0.07,0.70 ± 0.06
MC	0.60 ± 0.03,0.49 ± 0.04,0.65 ± 0.06	0.53 ± 0.06,0.31 ± 0.08,0.66 ± 0.04	0.58 ± 0.02,0.51 ± 0.10,0.61 ± 0.04	0.56 ± 0.02,0.45 ± 0.03,0.62 ± 0.03
GenFeature	0.57 ± 0.05,0.33 ± 0.07,0.67 ± 0.03	0.64 ± 0.02,0.21 ± 0.03,0.84 ± 0.021	0.52 ± 0.04,0.31 ± 0.07,0.65 ± 0.05	0.66 ± 0.04,0.11 ± 0.04,0.88 ± 0.04

Table 6.2: PD classification results using subcortical features.

package in terms of preprocessing and feature extraction for T1 MR images in order to PD classification using machine learning techniques. Generally, the experimental results show that the classification models need more information from the MR T1 data. In future work, the combination of high-level features and the low-level ones (subcortical features) will be assessed for PD diagnosis. Also, other classification models and dimensionality reduction techniques will be evaluated for PD classification.

6.5 Conclusion

We presented an automatic MRI based CAD system for diagnosing PD, the second most common neurological degenerative disease affecting elderly people. This disease is exposed by the loss of neurotransmitters that control body movements. Currently, there is no cure other than earlier diagnosis with better and more efficient treatment for patients. In this research, MR T1 images are used for assessment the ability of volume and surface-based features of the brain subcortical regions. The FreeSurfer package used for preprocessing of MRI data and also for extracting the subcortical features including volume, area, mean curvature and thickness. RF and SVM methods are used as decision models for conducting classification for T1 MR images with subcortical features and their fusion. In the experimental results, the ability of these classifiers for PD diagnosis are compared using different criteria: accuracy, specificity, and sensitivity.

In future work, the efficiency of the proposed method could be improved by adding high level features to the current ones. Moreover, other machine learning classifiers and dimension reduction techniques will be evaluated for PD classification using subcortical features.

Chapter 7

Parkinson's Disease Region-based Assessment with Gray/White Matter Volume of MR Images

7.1 Introduction

One of the main neurodegenerative diseases is PD caused by loss of dopaminergic neurons in the brain, especially in the BG region. PD primarily affects senior people (age >50 years) and incidence rates increase with age [116]. Two main symptoms of PD are motor and non-motor. The main motor symptoms are bradykinesia, tremor, rigidity, and slowness. Among the non-motor symptoms, cognitive impairment is one of the major one, which happens in $20\% - 50\%$ of PD patients [46]. The cognitive deficits may cause dementia at the later stages of the disease. Researchers illustrate cognitive impairment might develop at early stages of the disease [46]. Although there is currently no cure for PD, appropriate and on time treatment has been shown to have a beneficial effect on the disease progression [68], [116], [170].

7.2 Motivation

The cognitive impairments in PD have been related to GM, WM alterations, brain Functional Connectivity (FC), and brain activation alterations [46]. Thus, brain GM and WM assessment help to find the biomarker for cognitive deficits,

which will be useful for early PD diagnosis. Our proposed approach is based on medical image processing and machine learning techniques. Using SPM/CAT12 package, we pre processed MR images for meaningful analysis. The volume of GM and WM are selected as features for our further analysis. The selected features are then forwarded to machine learning algorithms, which classify MR images into PD and HC categories. Most of the proposed method for volume analysis for GM and WM of T1 MR images are done with SVM. However, in this research we compare the ability of two other classifiers beside SVM including RF and Gradient Boosting (GB). In addition, most of the research applied the proposed method on small datasets but in this research we evaluate our method with a big public dataset.

7.3 Proposed Method

In this part, the proposed idea for diagnosing PD using T1 MR images is explained. There are four general steps including preprocessing, feature extraction, feature selection and classification. Figure 7.1 shows the general framework of the proposed system.

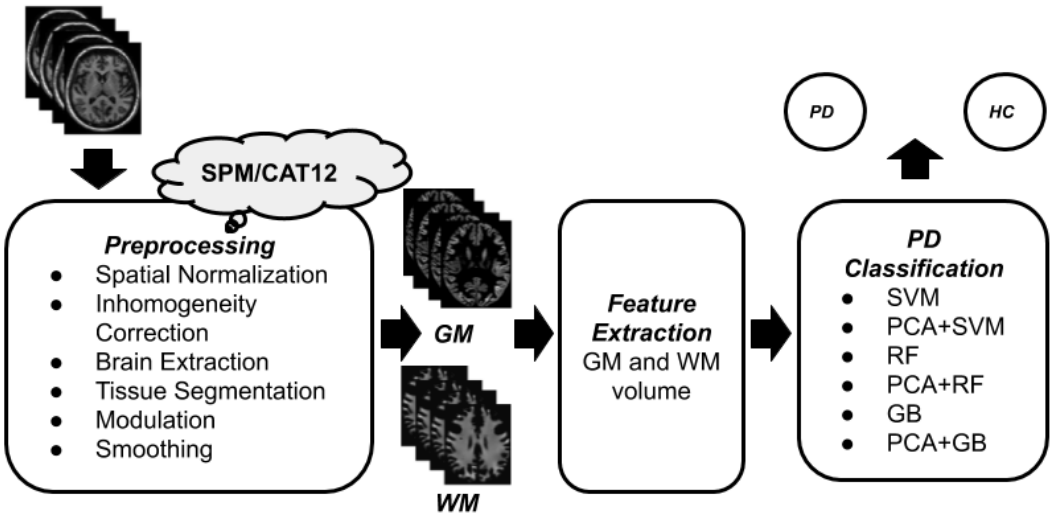


Figure 7.1: The pipeline of the proposed method for PD classification.

7.3.1 Dataset

The present study uses PPMI dataset which is the first substantial study for identifying the PD progression biomarkers [118]. The MR images are downloaded from their website beside the demographic information including age and sex for each subject. There exists a large class imbalance among the data, with 355 PD to 163 HC subjects. This class imbalance is resolved with the supplementation of 149 HC T1-weighted MRI scans from the publicly available IXI dataset [75], with a resulting total of 667 class-balanced patients. Demographic data is shown in Table 7.1. The number of PD samples in the balanced dataset is 355 and the number of HC samples is 312.

	PD	***HC***	***Average/Total***
Age	61.63 ± 9.6	49.88 ± 16.84	55.76 ± 17.99
Sex (F/M)	126/229	131/181	257/410

Table 7.1: Demographic information of the combination of IXI and PPMI dataset.

7.3.2 Preprocessing

Preprocessing is an essential part of any computer aided diagnosing system especially for neuroimaging data analysis. Since the neuroimaging data is comming from different scanners, they might have different intensity range. Thus, the data need to normalized in to the same space for having a comprehensive and reliable brain study. In addition, image processing such as increasing the image contrast and denoising are also need to be done. In this research, SPM12 [12], [152] and CAT12 [32] are used for preprocessing step. CAT12 is an extension to SPM12 to provide computational anatomy which is covering diverse morphometry methods such as VBM [13], Surface-Based Morphometry (SBM), Deformation-Based Morphometry (DBM) [14], and Region-Based Morphometry (RBM) [56]. Most of the presented papers in neuroimaging analysis use VBM8 for preprocessing. However, CAT12 is more robust and accurate than VBM8 [83] especially for detecting the small abnormalities in the brain images [38]. There are different steps in preprocessing including:

- Special Normalization:

  In this step, the goal is to alignment the images to a standard space which is called Montreal Neurological Institute (MNI). This work is done by registration that is adjusting the brain shape and positions with respect to the MNI standard. There are two registration methods for VBM analysis. The first one is affine registration which is aiming to get the global geometric transformation of the brain image. The second one is non-linear registration, where the goal is resolution matching between the MR images and the templates [96].

- Inhomogeneity Correction

  The intensity values in structural brain scans are not exclusively attributable to different tissue types, as an intensity-based tissue classification would assume [13]. Moreover, the magnetic field in image acquisition process will lead to some inhomogeneities which is called bias field in image as well especially with high-field scanners. Thus, bias field correction is a critical step in preprocessing.

- Brain Extraction

  In this step, the skull, fat, and background regions from the MRI scans are removed because they don't have any information for neuroimaging analysis.

- Tissue Segmentation

  The final goal of preprocessing is classifying the MR images to different tissues. In this phase, the brain MR images are segmented to three main parts including WM, GM and CSF.

- Modulation:

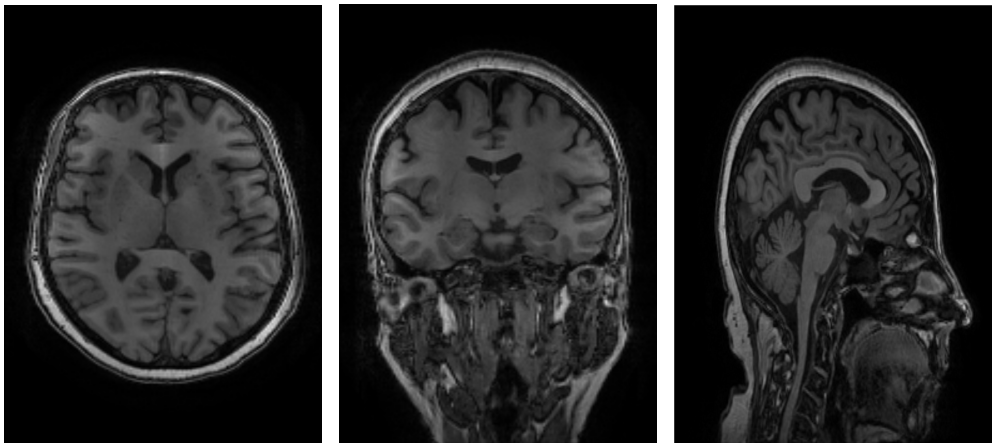
  This step causes correction of changes in the volume of the segmented images (to GM, WM) by applying a deformation method.

- Smoothing:

  Isotropic Gaussian kernel is used for smoothing in order to increase the

signal-to-noise ratio, reducing the impact of mis-registration between images and benefits on statistics normality. The size of the kernel is 8mm.

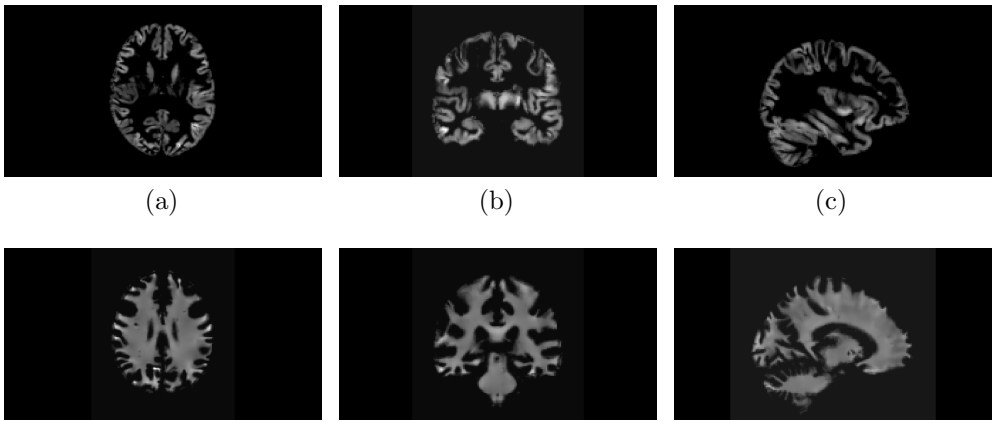
All the steps of preprocessing has been done by SPM/CAT12 package [47], [51]. A sample MRI scan in different planes including axial, coronal, and sagital



(a) (b) (c)

Figure 7.2: Selected original MRI T1 before preprocessing in three different planes: (a) Axial, (b) Coronal, (c) Sagital.

is presented in Figure 7.2. Figure 7.3 shows two samples of GM and WM in



(a) (b) (c)

(d) (e) (f)

Figure 7.3: A sample GM image in different planes: (a) Axial, (b) Coronal, (c) Sagital. A sample WM image in different planes: (d) Axial, (e) Coronal, (f) Sagital.

different planes: axial, coronal, and sagital.

7.3.3 Feature Extraction

The output of the preprocessing step is the smoothed 3D GM and WM for each subject. The goal of this research is an assessment of brain volume changes for PD patients. In our research, the GM Volume (GMV) and WM Volume (WMV) are used as discriminative features for our further analysis.

7.3.4 Dimensional Reduction and Classification

The GMV and WMV features might have irrelevant and redundant data which might affect the classification results at the end. PCA is one of the most popular feature reduction in neuroimaging analysis [108]. It finds relevant features by linearly transforming correlated variables into a smaller number of uncorrelated variables known as principal components [80]. The selected components are essentially linear combinations of the original data capturing most variance in the data [108]. In this research, 3 types of classifiers are used with and without PCA which are RF, SVM, and GB. SVM is proposed by Vapnic [41] in which find the optimum hyper planes for classifying the data. RF [24] is one of the major ensemble learning methods which consist of a group of weak decision trees. RF classify the data based on the majority vote of the weak classifiers. GB [55] is another popular machine learning classifier which is working based on combining many weak learning models together to create a strong predictive model.

For all the models, 80% of the data is used for training, and 20% for the testing. 10-fold cross validation is conducted. Furthermore, grid search with hyper parameters is performed for for training the models. For the classification part, six models are used and compared: SVM, RF, GB, PCA+SVM, PCA+RF, and PCA+GB.

7.4 Experimental Results

In this part, the classification results of using different models with GMV and WMV are reported. Our proposed system uses volume features to classify the T1 MR images into two groups: PD and HC. The extracted features were input

to the SVM, RF, and GB classifiers (with and without PCA). The criteria for evaluating the proposed system are accuracy, specificity, sensitivity which are computed based on Eq. 7.1.

$$\begin{aligned} Accuracy &= (TP + TN)/(TP + TN + FP + FN) \\ Specificity &= TN/(TN + FP) \\ Sensitivity &= TP/(TP + FN) \end{aligned} \tag{7.1}$$

TP as True Positive, is the number of samples correctly classified as PD, FP as False Positive is the number of HC samples incorrectly classified as PD, TN as True Negative is the number of samples correctly classified as healthy and FN as False Negative is the number of PD samples incorrectly classified as healthy. The six classification models described above are used to evaluate the descriptive abilities of the two volume feature descriptors GMV and WMV. The quantitative results from the GMV features are presented in Table.7.2, and for the WMV feature are shown in Table.7.3. The fusion of GMV and WMV classification results are shown in Table.7.4.

Criteria	***Accuracy***	***Specificity***	***Sensitivity***
RF	*0.79*	*0.60*	*0.93*
PCA+RF	*0.81*	*0.65*	*0.91*
SVM	*0.81*	*0.80*	*0.82*
PCA+SVM	*0.79*	*0.62*	*0.93*
GB	*0.76*	*0.53*	*0.98*
PCA+GB	*0.79*	*0.53*	*0.97*

Table 7.2: The classification results using different models for GMV.

Criteria	***Accuracy***	***Specificity***	***Sensitivity***
RF	*0.80*	*0.56*	*0.98*
PCA+RF	*0.79*	*0.61*	*0.94*
SVM	*0.79*	*0.52*	*1.00*
PCA+SVM	*0.79*	*0.76*	*0.83*
GB	*0.79*	*0.60*	*0.98*
PCA+GB	*0.80*	*0.53*	*1.00*

Table 7.3: The classification results using different models for WMV.

The experimental results illustrate that the maximum accuracy for PD classification is 0.81 which is achieved with GMV using two models: PCA+RF, and SVM. With considering other criteria (specificity and sensitivity), SVM is more reliable model compared to PCA+RF. The classification results of WMV in Table 7.3 show 0.80 accuracy which is very close to the maximum obtained accuracy for GMV. However, WMV highest accuracy is gained by RF and PCA+GB models.

Contrary to what we expected, the fusion of GMV and WMV did not represent higher performance in PD classification compared to each of these features. However, the results are close to what we achieved for GMV and WMV classification.

Comparing to the other proposed methods for PD classification using GMV and WMV [18], [53], [86], [141], our research used bigger dataset for evaluation which makes our results statistically significant. The highest reported accuracy in our method (0.81) is higher than the performance of the presented research in [86] which is 0.75. On the other hand, in our proposed method, we compared the ability of different machine learning classifiers. Since deep learning methods have high performance in many applications using MRI data [5]. For future work, the ability of some 3D and 2D networks will be assessed using GM and WM data.

Criteria	***Accuracy***	***Specificity***	***Sensitivity***
RF	*0.78*	*0.52*	*1.00*
PCA+RF	*0.79*	*0.56*	*0.98*
SVM	*0.79*	*0.53*	*1.00*
PCA+SVM	*0.79*	*0.77*	*0.80*
GB	*0.76*	*0.53*	*0.97*
PCA+GB	*0.77*	*0.54*	*0.98*

Table 7.4: The classification results using different models for the fusion of GMV and WMV.

7.5 Conclusion

Early diagnosis of PD is essential for having more effective treatment for the patients. The non-motor symptoms of PD especially cognitive impairment appears in the earlier stages of the disease. Brain structures including GM and WM are related to the cognitive functionality. In this research, the ability of GMV and WMV are evaluated for PD classification using MR T1 images. SPM/CAT12 is used for preprocessing and getting the GM and WM of T1 MR samples. These volume features are used with three important classifiers including SVM, RF, and GB. These classification models are used with and without PCA as dimension reduction technique which result to have six different models. The T1 MR data from IXI and PPMI dataset is used for evaluation purpose. IXI dataset is used to solve the imbalance problem of PPMI data. The experimental results show the highest accuracy for PD classification achieved by SVM model for GMV feature.

Chapter 8

Assessing the Capability of Deep-Learning Models in Parkinson's Disease Diagnosis

8.1 Introduction

PD is a progressive neurodegenerative disease that primarily affects the ability of an individual to perform basic motor movements. Common symptoms include bradykinesia (slow movement), difficulty in speaking and an unsteady gait. PD ranks as the second most common neurological disease after AD, with a prevalence of 6.2 million affected globally [161]. The disease is particularly prevalent in the aging community, affecting 1% of those above age 60 [134]. Despite this, the underlying pathology of the disease is not well understood [84].

In general, PD is diagnosed when motor symptoms begin to manifest. However, a patient may have lost $50 - 70\%$ of their dopaminergic neurons before these symptoms appear [34]. With this in mind, diagnostic methods such as MRI are of particular interest. When discussing diagnosis, it is important to consider the severity of the disease. A treatment that is helpful in early stage PD may become ineffective if applied only at a later stage. To classify severity, clinicians commonly use two metrics: HAY score [70] or the UPDRS [133]. These metrics provide a useful method for doctors to quantify and evaluate patient outcomes. Primarily, our work focuses on binary classification between PD and HC patients, but could be extended to fit these metrics.

8.2 Motivation

A pre-symptomatic diagnosis may prove crucial in stopping or slowing the disease before it progresses to a debilitating stage. Although PD is ostensibly incurable, promising new treatments such as Exenatide [16] may be effective if used at an early stage.

For the most part, structural sMRI is not currently used in PD diagnosis. This is due to a traditionally poor ability for sMRI to detect subtle physiological changes in areas associated with PD (i.e. BG) [176]. However, our approach seeks to circumvent this through the computational sensitivity of deep-learning based methods. As well, unlike CT or x-ray imaging, sMRI does not subject the patient to high levels of ionizing radiation [26].

Our work presents a non-invasive approach to diagnosis of PD based on deep-learning frameworks. Specifically, we construct multiple models to classify patients strictly from sMRI data. Similar approaches with computer-aided diagnosis have been taken in the past for conditions such as AD [162] and Attention Deficit Hyperactivity disorder [177].

8.3 Data and Preprocessing

8.3.1 Dataset

In our work, subject data is obtained from the PPMI public dataset. The dataset consists of T1-weighted sMRI scans for 568 PD and HC subjects. From this, 445 subjects are selected, with the rest omitted due to structural abnormalities during preprocessing. There exists a large class imbalance among the remaining data, with 299 PD to 146 HC subjects. This class imbalance is resolved with the supplementation of 153 HC T1-weighted sMRI scans from the publicly available IXI dataset, with a resulting total of 598 class-balanced patients. Demographic data is also collected and shown in Table 8.1.

	PD	***HC***	***Average/Total***
Age	62.0 ± 9.54	49.2 ± 16.09	55.6 ± 15.1
Sex (F/M)	189/110	172/127	361/237

Table 8.1: PPMI dataset demographic data.

8.3.2 Preprocessing

Due to morphological and dimensional differences between scans, the samples must be standardized to a common format so that they are comparable. We initially resize all scans to the same dimensions so that they will fit in our model. Then, we must perform an intensity normalization. MRI intensity is measured in arbitrary units, and as such, there is a large discrepancy in inter-subject intensity values. We correct this by standardizing intensity on a per-patient basis, fixing all values to the range $[0, 1]$. After this, more complex preprocessing operations are performed with the use of the FSL toolkit [167], as explained below.

A pipeline is constructed to preprocess the dataset using the Anatomical Processing Script (fsl_anat). Firstly, the data is reoriented to MNI orientation (fslreorient2std) such that all scans face the same direction. Then we perform a bias-field correction (FAST) [173] to remove general intensity non-uniformities. Following this, the brain is extracted from the scan (FNIRT / BET) [78] and linearly registered to the standard MNI152 format (FLIRT) [79][77]. Non-linear registration (FNIRT) was omitted because of a tendency to have unpredictable deformation corrections with regard to a standard template [7]. At this point, artifacts created in the preprocessing step, namely erroneous voxel intensity values higher than the global max of 1, are corrected to be within the normal range of $[0, 1]$. Note that scans are zero-padded from a size of $91 \times 109 \times 91$ to a size of $96 \times 112 \times 96$ when they are inputted into the model, since these dimensions are repeatedly divisible by 2 and thus lead to cleaner convolution.

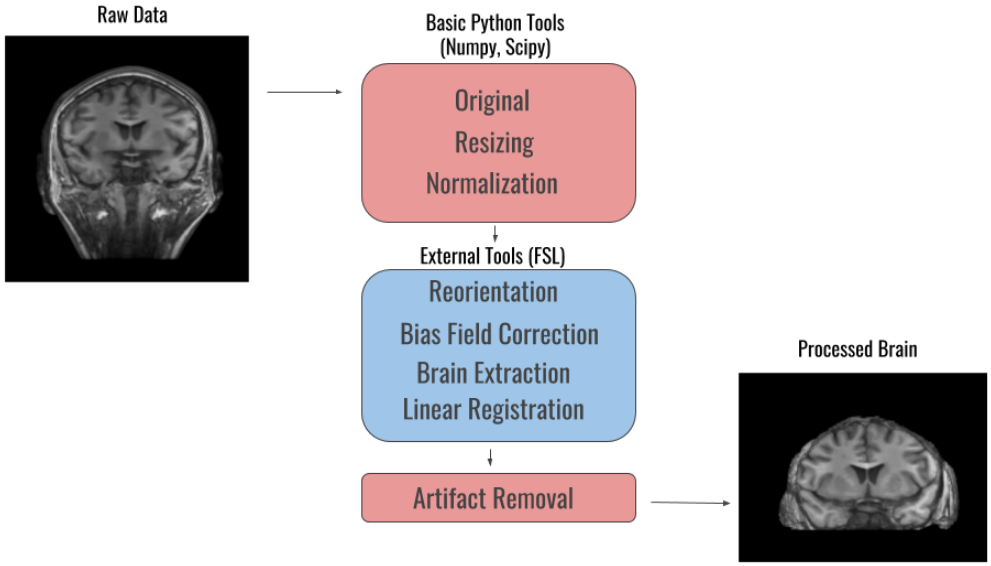


Figure 8.1: Preprocessing pipeline

8.4 Proposed Method

We take two distinct approaches to classification for deep-learning based models. The first approach is based on three-dimensional analysis, while the second is two-dimensional. All models are built using TensorFlow.

8.4.1 Three-Dimensional Models

(M1: 3D-CNN) In the three-dimensional approach, we test two models. The first of these is a generic $3D$ Convolutional Neural Network, with 3 sets of $3D$ Convolution + Max Pooling Layers. Kernel Size is $[3, 3, 3]$ and stride is $[2, 2, 2]$. Leaky ReLU units are used at each layer to introduce non-linearity to the data. We flatten the output of the final pooling layer, and use it as the input of a small 4 layer neural network. The flat layer is passed into the first dense layer of 512 neurons. Dropout with 30% loss is introduced after the first dense layer to minimize overfitting. Following this, we enter a 128 neuron layer, and then the final 2 class-logit layer. The output is the computed softmax of the two class probabilities. We train using cross-entropy loss and ADAM optimizer with a learning rate of 0.00001. Model architecture is shown in Figure 8.2 (a).

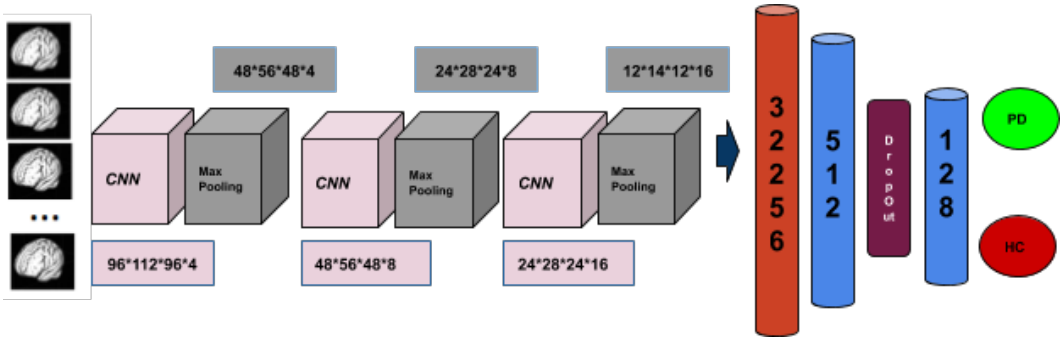
(M2: 3D-DAE) The second model is a $3D$ Denoising Convolutional Autoencoder followed by a classifying neural network. The autoencoder is undercomplete, and the input is purposely corrupted. The corruption is done by masking between $20-30\%$ of the input pixels to zeros before passing into the autoencoder. Both the corruption and under-completeness force the autoencoder to obtain an efficient latent-space-representation of the voxel data. We allow the autoencoder to train unsupervised for some time, using mean-squared-error and ADAM optimizer. Once it can recreate input scans with high accuracy, we take the compressed latent space representation of our "encoder" and pass it to the neural network. This is a form of feature reduction for our data. The classifying neural network has the same structure as the one used in the 3D-CNN, starting with a flat layer and ending in a softmax between the PD and HC logit. See Figure 8.2 (b) for the detailed structure. Note that we tried stacking autoencoders, but the results were generally poor and are not included in this research.

8.4.2 Two-Dimensional Models

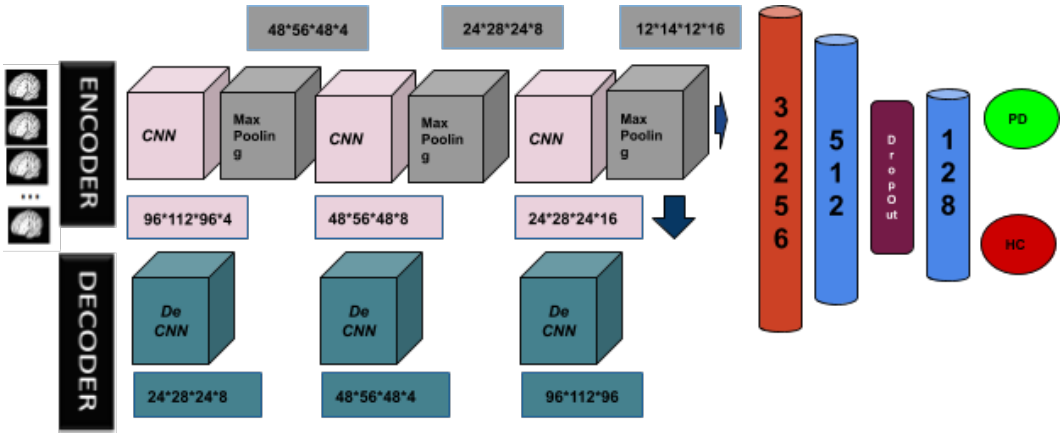
sMRI data is represented as a $3D$ matrix of voxel intensities, in which case it may be parsed into a series of $2D$ frames. In this sense, we can consider 3D-sMRI scans as a series of 96 2D-slices. This allows us to classify on a per-slice basis rather than from an entire 3D scan. This presents some unique facets to our task.

Classification on a per-slice basis means that for any given scan, some slices may be identified as Parkinsonian, while others are marked as healthy. Ultimately, a decision must be made for the entire 3D scan. In general, we simply take the mode of the class "votes" as the final decision, as seen in Figure 8.3. However, not all 96 slices are relevant to a diagnosis. If PD mostly affects the inner brain (BG), then it may be best to ignore slices on the boundary of the brain in our diagnosis.

(M3: 2D-CNN) The first $2D$ model we test is a basic two-dimensional CNN, which follows the same general structure as the $3D$-CNN. We stack



(a) 3D-CNN



(b) 3D-DAE

Figure 8.2: 3D model architecture

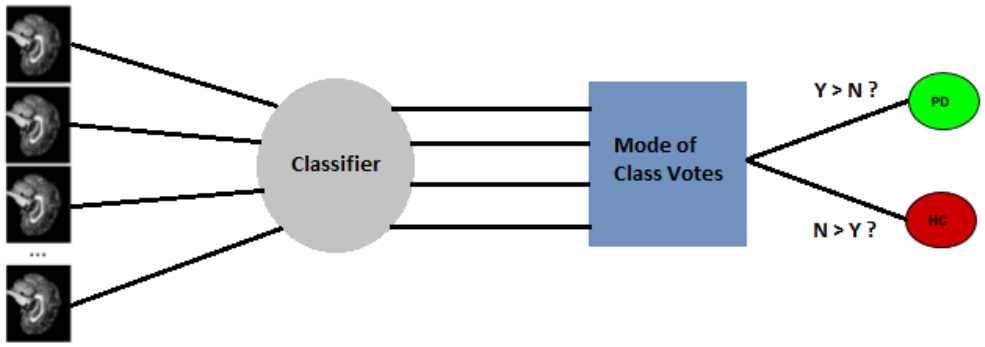


Figure 8.3: 2D voting system

three units one after another, in which one unit consists of a $2D$ convolutional layer followed by a max-pooling layer. Convolutional kernel size is $[5, 5]$ and stride is $[2, 2]$. ReLU is used at each step, and the final output is flattened and passed as the input to the neural network. This network follows the

same structure as the $3D$ neural network, and the resulting output is the softmax of the class logits. Note that we use a batch-size of 96 both because it allows our 2D model to train more efficiently, and also it is useful to calculate statistics with the same dimensions as our 3D models. We calculate loss with cross-entropy, and use ADAM optimizer on a per-slice basis with a learning rate of 0.0003. Final classification is done by taking the mode of the class votes for each slice. Detailed architecture is shown in Figure 8.4 (a).

(M4: 2D-DAE) The next model is a $2D$ Denoising Autoencoder and neural network classifier. This model follows the same basic structure as the $3D$ equivalent. Kernel and stride are the same as the $2D$-CNN. The autoencoder is trained with mean-squared error and ADAM optimizer with a learning rate of 0.00003. All other aspects are the same. See Figure 8.4 (b) for clarification.

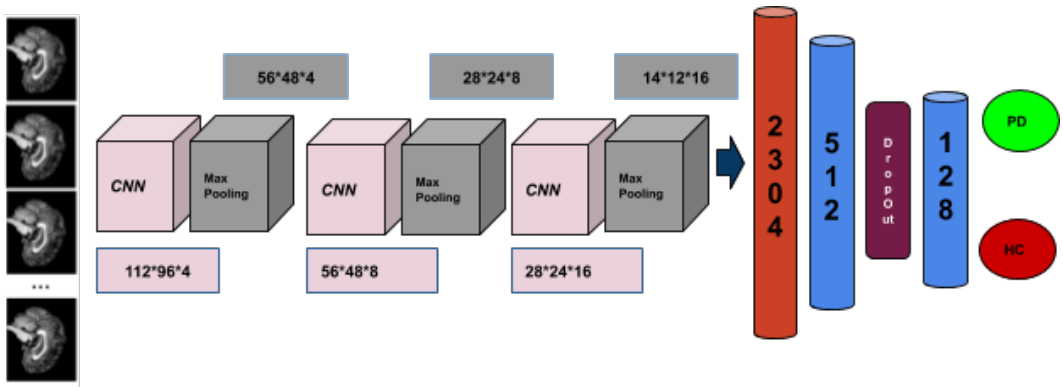
8.5 Results

The results of the 4 models are shown in Table 8.2. Although no model performs exceptionally well, all models perform better than the baseline class-distribution (50%). Interestingly, the $2D$ models generally have very high sensitivity but less than satisfactory specificity. This would suggest that they mainly accumulate error from false-positive diagnoses. In this sense, the $2D$ models over-diagnose PD. It's likely that this stems from the small filter-space of our 2D models, which must learn to detect very small features in only a few frames. Many of the frames we pass into the model would not seem to contain meaningful information in detecting PD.

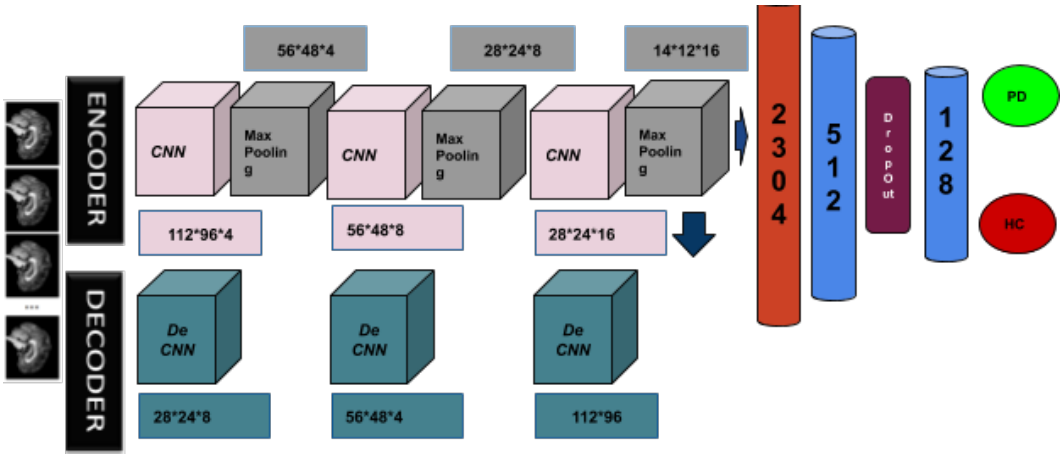
Criteria	***Sensitivity***	***Specificity***	***Precision***	***Accuracy***
M1: 3DCNN	0.76	0.74	0.74	*0.75*
M2: 3DDAE	0.72	0.57	0.62	*0.64*
M3: 2DCNN	1	0.54	0.58	*0.69*
M4: 2DDAE	0.92	0.53	0.65	*0.70*

Table 8.2: Deep learning model results.

The best performing $3D$ model was the 3D-CNN, with 75% accuracy and



(a) 2D-CNN



(b) 2D-DAE

Figure 8.4: 2D model architecture

balanced metrics in sensitivity, specificity and precision. It almost equally misdiagnoses with false-positives and false-negatives. This is promising because it suggests that the model learns meaningful features rather than preferentially guessing one class repeatedly. With this in mind, we consider the 3D-CNN useful for generalization. We perform an occlusion sensitivity analysis using our best model, the 3D-CNN. In the analysis, we allow the model to fully train, and then obtain the test sample which maximally activates the PD-positive logit. We occlude a small $4 \times 4 \times 4$ section of the sMRI sample by masking with zeros, and then obtain the new PD-positive logit value. This occlusion process is done for each block in the sample. We introduce a minimum threshold so that we can find the blocks which maximally activate the logit, which is

necessary because all blocks activate it to some degree. Those blocks which when occluded, lowered the logit value the most, are overlaid a sample brain as seen in Figure 8.5. The sample brain has the mean intensity values of the entire preprocessed dataset. In general terms, the red boxes shown in Figure 8.5 are the areas our model finds most important in making its diagnosis.

It follows from the occlusion sensitivity analysis that PD diagnosis is strongly correlated with voxel intensity in both the cerebellum and occipital lobe regions. As shown in the symmetry between the right and left sagittal (side) views in Figure 8.5, this relationship is bilateral. These findings are supported in literature, with significant atrophy in the cerebellum [22] and the occipital lobe [28] associated with Parkinson's disease.

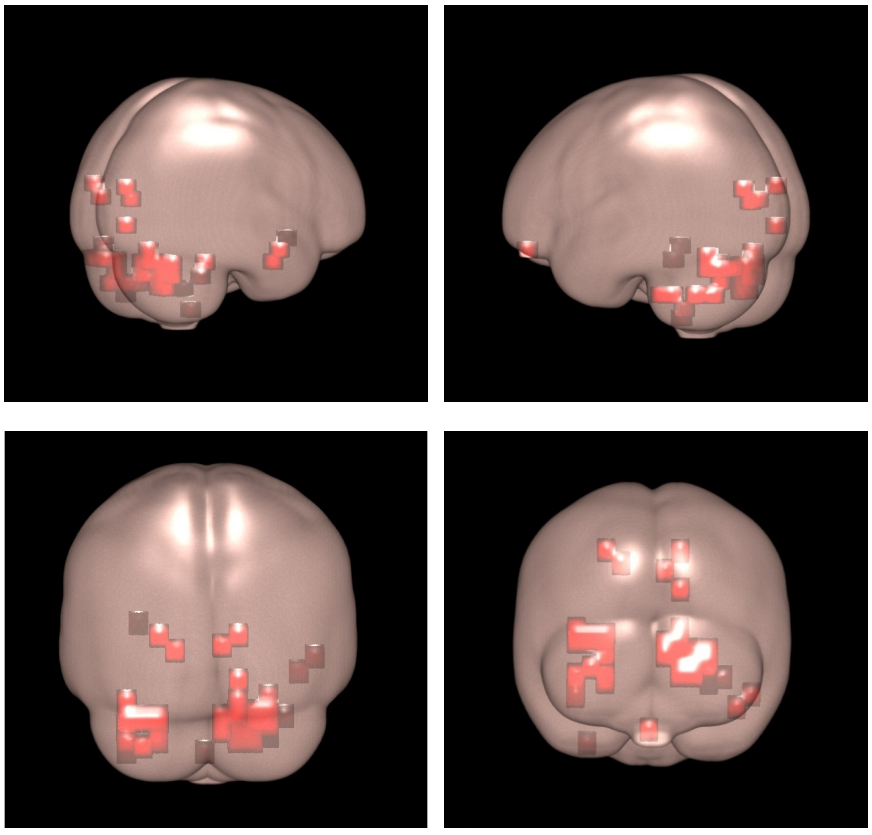


Figure 8.5: Occlusion sensitivity analysis

8.6 Discussion and Future Work

In general, no model tested is yet applicable in a clinical setting. Symptomatic diagnosis and medical history analysis are still the gold-standard in PD diagnosis for the time being. Despite this, our deep-learning model assessment demonstrates competency in the task of PD diagnosis, as well as highlighting potential biomarkers in PD progression. With improved accuracy, the proposed method could certainly be applied in a practical situation.

There is significant potential for future research in the task at hand. An area of consideration is deep-learning based classification with segmented subcortical structures. In the case of our research, classification models could be built based only on voxel intensities in the segmented cerebellum and occipital lobe regions. Furthermore, our models could be expanded to a categorical classification metric, such as UPDRS, rather than the current binary classification for more precise diagnosis.

8.7 Conclusion

We have demonstrated a deep-learning based approach to diagnose Parkinson's Disease from T1-weighted structural MRI images. The proposed method (3D-CNN) shows promising results in autonomous detection, which is an important field in modern medical applications. The work also highlights potential biomarkers of the disease in the cerebellum and occipital lobe. To summarize, the work suggests that structural MRI can provide useful and complementary information in the diagnosis of Parkinson's disease.

Chapter 9

Parkinson's Disease Mid-brain Assessment using MR T1 and T2 Images

9.1 Introduction

PD is the second most common movement disorder, which is a progressive neuro-degenerative disease. The incidence of PD, or the rate of newly diagnosed cases, generally increases with age, although it can stabilize in people who are older than 80. An estimated 4 percent of people with PD are diagnosed before the age of 50. Men are 1.5 times more likely to have PD than women [164]. This disease is caused by loss of dopaminergic neurons in the part of the brain known as the BG. The most affected part of BG is SN where the loss is typically at least 50% of the dopaminergic cells [25]. The cause of cell death is unknown. There are some other abnormal changes in SN such as excess iron accumulation, depletion of tyrosine hydroxylase and the presence of lewy bodies in the remaining nerve cells [125], [166]. SN is a dark region (the darkness is because of high content of neuromelanin which is a precursor of dopamine) presented in Figure. 9.1 with other regions of the BG which are located in midbrain [125]. The main components contain striatum; caudate nucleus, putamen, globus pallidus, SN, and subthalamic nucleus. The overall appearance of the normal SN on iron-sensitive MR sequences has been termed the 'swallow-tail-sign' or 'nigrosome-1', and it has been shown to be lost in almost all patients with PD studied so far [135], [145]. Histopathological studies

have reported that approximately 98% of dopaminergic neurons in nigrosome-1 are lost in patients with symptomatic PD [35]. Since, the BG is associated with movement related function, the most important symptoms, which people with PD manifest are motor symptoms including resting tremor, bradykinesia, postural impairments and rigidity [94]. In many patients, subsequent non-motor problems may arise such as cognitive, mood, sleep, and olfactory, with dementia commonly occurring in the advanced stages of the disease [61]. PD can be

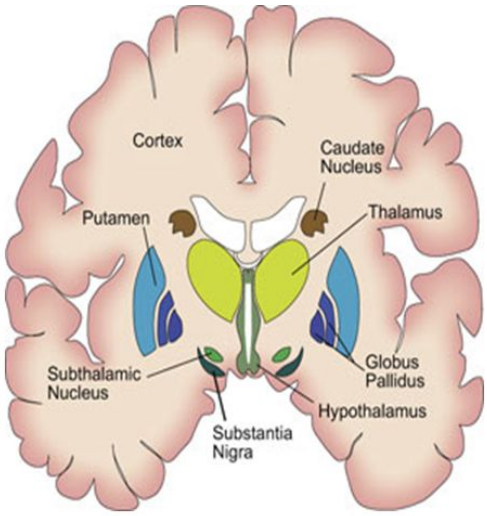


Figure 9.1: The structures of the basal ganglia.

difficult to diagnose accurately, particularly in the early stages of the disease when symptoms resemble other medical conditions, and misdiagnoses occurs occasionally [139]. There are currently no blood or laboratory tests that have been proven to help in diagnosing PD. The diagnosis is based on the medical demographic characteristics and neurological examination that is conducted by interviewing and observing the patient. In recent years, neuroimaging has been increasingly used as an objective method for the early diagnosis of PD and other neuro diseases [90]. There are some published research for diagnosing, progress monitoring and assessment of PD using different neuroimaging data such as MRI [130], PET [89], SPECT [36], and DTI [17]. Most of these neuroimaging techniques are helpful when 80% of the dopaminergic neurons are already lost [148]. MRI is one of the most commonly used tests in neurology and neurosurgery. Recently, promising results in providing better characterization

of early diagnosis of PD using MR images are reported and are expected to have better sensitivity than standard clinical measures [97], [148].

MRI is based on the magnetization properties of atomic nuclei. An external Radio Frequency (RFR) is perturbing the alignment of protons which are normally oriented within the water nuclei of the tissue. The emitted signals are measured after a certain period following the initial RFR. The frequency information contained in the signal from each location in the image plane will be converted by Fourier transform to corresponding intensity levels, which are then displayed as shades of gray in a matrix arrangement of pixels [93]. Different type of MRI image can be created using varying the sequence of RFR pulses applied and collected. Therefore, there are two parameters involved: Repetition Time (RT) that is the amount of time between successive pulse sequences applied to the same slice, and Time to Echo (TE) which is the time between the delivery of the RFR pulse and the receipt of the echo signal. Based on the mentioned parameters, two types of MRI exist. The first one is T1-weighted images which are produced by using short TE and RT times. The contrast and brightness of the image are predominantly determined by T1 properties of tissue. The second one is T2-weighted images which are produced by using longer TE and RT times. In these images, the contrast and brightness are predominantly determined by T2 properties of tissue [93].

9.2 Motivation

MRI methods are sensitive to detect and visualize tissue changes in three different planes: axial, sagittal and coronal. MRI can provide evidence of structural changes that occur as a result of the loss of dopamine neurons in the SN, as well as the loss of non-dopaminergic neurons in other brain regions. Because of its high resolution contrast, ready availability, non-invasive nature and no need for any pharmaceutical injections, many have used structural MRI methods in their pursuit of developing a biomarker for PD [39]. In this research, we analysis the SN and its neighborhood regions in the T1 and T2 MRI sequences just by focusing on the midbrain slices which SN is visually

existent. Two important computer vision techniques LBP and HOG are used for feature extraction to get textural, spatial, and shape information from the selected region in the MR images. Following that, extracted features used for the classification part with 4 models including: SVM, RF, PCA + SVM, PCA +RF. PCA has been used as feature selection or dimension reduction. The experimental results are conveyed with different number of slices from 3 to 8 to evaluate the relation between the classification accuracy and the required number of slices. This research is according to the listed objectives as below:

- Designing CAD system for PD classification using T1 and T2 MR images with computer vision techniques and using machine learning methods.
- Compare the ability of these two popular MR imaging techniques T1 and T2 MR for PD assessment.
- Evaluating the proposed CAD system based on different number of midbrain slices for T1 and T2 instead of using the whole 3D brain volume with a lot of useless information.

9.3 Proposed Method

In this research, we design and compare two CAD systems to distinguish PD and HC subjects using T1 and T2 MRI images. The proposed systems follow the same framework which have 4 main steps as shown in Figure. 9.2. We explain these steps, i.e., prep-processing, feature extraction, feature reduction, and classification, below.

9.3.1 Preprocessing

Preprocessing is one of the critical steps in most of the neuroimaging analysis. Not all the MRI slices are related to PD and have useful information for the classification task. For example the slices near the outer edge of the brain have no essential knowledge for our system. Hence, we need to find the Slice Of Interest (SOI) in the 3D MR MRI volume image for each subject which has information for PD classification. Since SN is more visible in axial view we

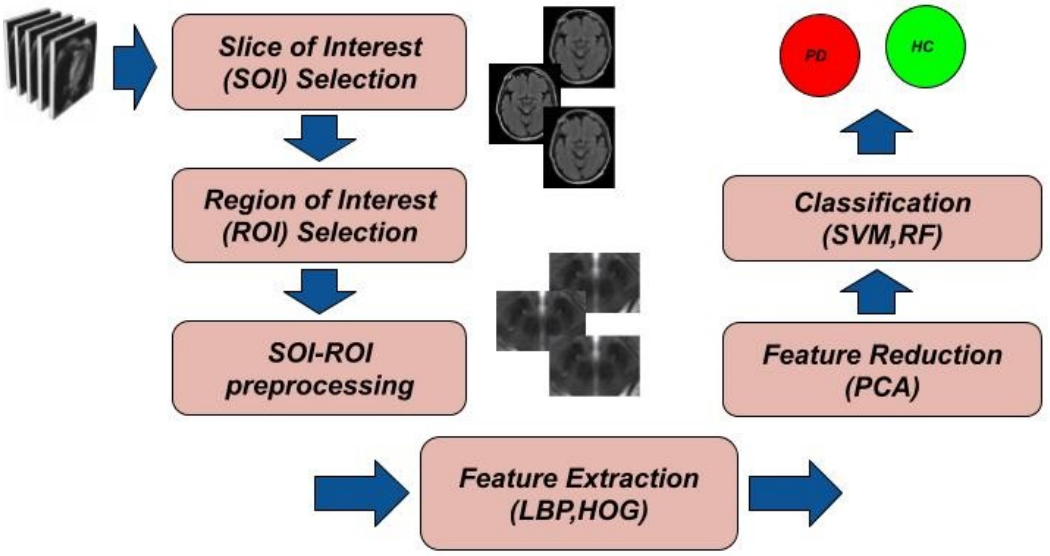


Figure 9.2: The proposed CAD system framework using T1 and T2 MRI scans for PD classification.

consider this plane for our 2D analysis. For each subject we selected between 3 to 8 slices around the center of the 3D MRI volume which have all sub-cortical regions including SN area. The SOI selection is conducted manually because the MR images are not preprocessed. For illustration purpose, we choose the same PD and HC subjects from T1 and T2 MR images. The SOI scans are shown in Figure. 9.4 and the selected ROI from these scans are presented in Figure. 9.5. Figure. 9.4(a) and (b) show two T1 SOI samples for PD and HC, respectively. Figure. 9.4(c) and (d) show two T2 SOI samples for PD and HC, respectively. For analysis the SN area which is the main affected region by PD, we choose the area inside those SOIs contains the SN region as ROI. These ROIs consist of SN and the neighbors sub-cortical regions such as Cerebral Palsy (CP) and Red Nucleus (RN) which are shown in Figure.9.3. The ROIs are automatically selected by choosing a square around the center of the MR image with the side length of 80. This part is usually done manually in other related works but our system is able to perform automatically. Figure. 9.5(a) and (b) show corresponding T1 ROI samples for PD and HC SOI, respectively. Figure. 9.5(c) and (d) show corresponding T2 ROI samples for PD and HC

SOI, respectively. The image quality of the selected SOI-ROIs needs to be

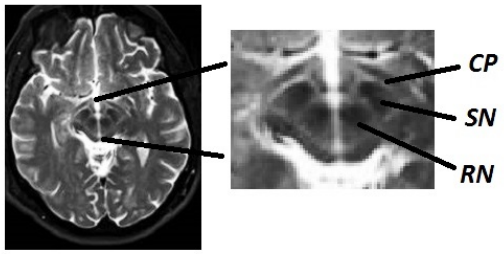
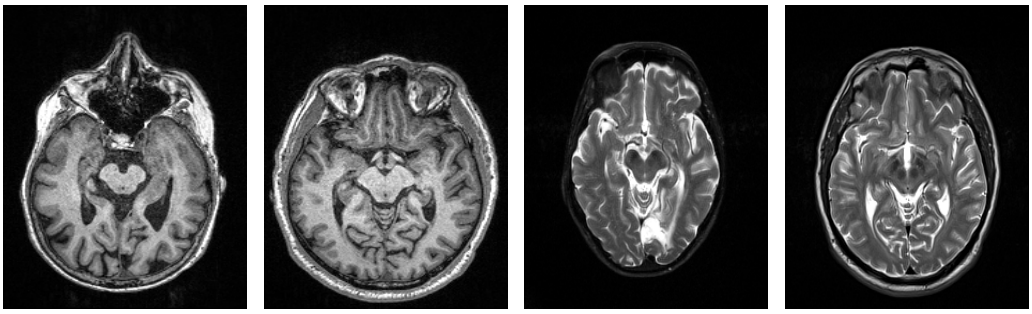
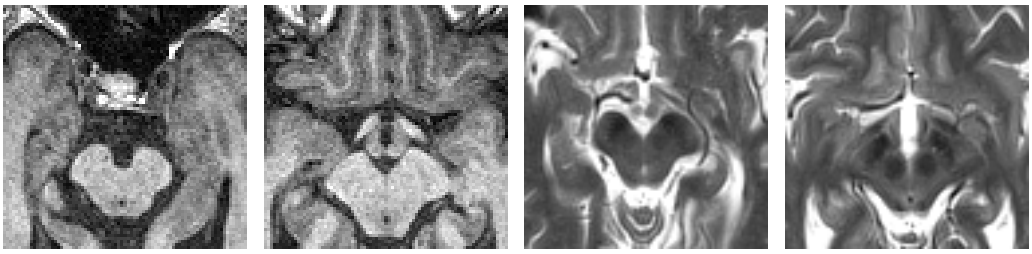


Figure 9.3: ROI in a random selected SOI, which contains the SN, CP and RN regions.



(a) (b) (c) (d)

Figure 9.4: The selected T1 SOIs for a HC subject in (a), and for PD subject in (b). The selected T2 SOIs for the same HC subject in (c), and for the same PD subject in (d).



(a) (b) (c) (d)

Figure 9.5: The ROI in the selected SOI for T1 image corresponding to Figure. 9.4: (a) PD subject, (b) HC subject. The ROI in the selected SOI for T2 image: (c) PD subject, (d) HC subject.

improved. Thus the preprocessing steps performs including noise reduction, and intensity non-uniformity correction. The denoising is done using the median

filter [6]. The intensity correction is done by adjusting the image intensity. Figure. 9.6 show two ROIs for PD sample in the first row and for HC sample in the second row for a T1 MR image. Figure. 9.7 show two ROIs for PD sample in the first row and for HC sample in the second row for a T2 MR image. Figure. 9.6 (a) and (d) are the original, Figure. 9.6 (b) and (e) are the denoised versions, and Figure. 9.6 (c) and (f) are the corrected non-uniformity of the T1 MR image. Figure. 9.7 (a) and (d) are the original, Figure. 9.7 (b) and (e) are the denoised, and Figure. 9.7 (c) and (f) are the corrected non-uniformity of the T2 MR image.

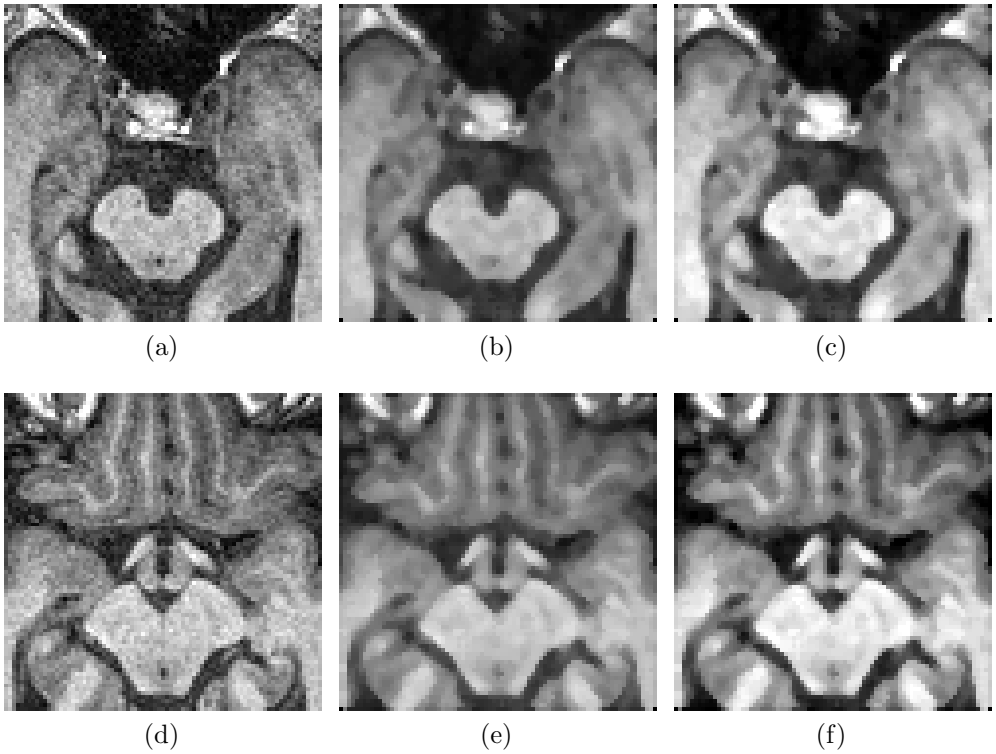


Figure 9.6: The selected T1 ROIs for a HC subject in (a), the denoised version in (b) and the enhanced version in (c), The selected ROI for a PD subject in (d), the denoised version in (e), and the enhanced version in (f).

9.3.2 Feature Extraction

Texture analysis in detection of neurological disorders with neuroimaging modalities are reviewed in [67] by Harrison. In their paper they claim that lesions and physiological changes either the visible or non-visible ones could

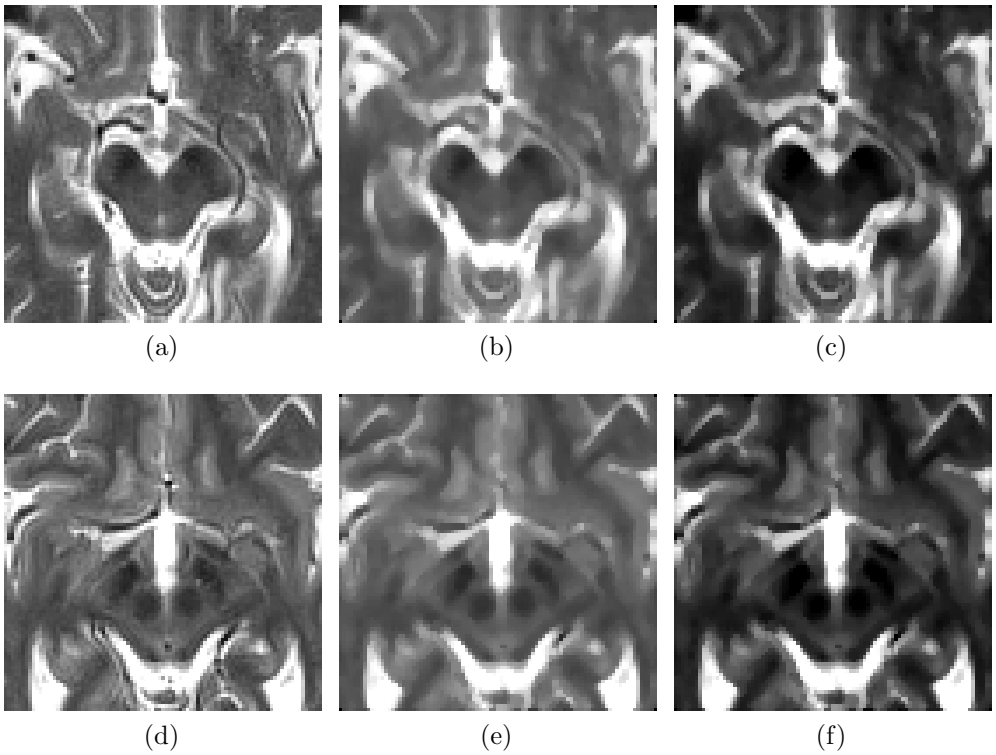


Figure 9.7: The selected T2 ROIs for a HC subject in (a), the denoised version in (b) and the enhanced version in (c), The selected ROI for a PD subject in (d), the denoised version in (e), and the enhanced version in (f).

be detected and characterized by texture analysis of routine clinical 1.5 Tesla scans [130][114]. LBP is one of the main texture feature descriptors in different applications of computer vision which is proposed by Ojala [111], [112]. After dividing the input images to a group of cells, it gets a histogram of the binary codes which are found by thresholding the neighborhood of each central pixel. The LBP operator is defined in Eq.9.1.

$$\begin{aligned} LBP_{P,R}(x_c, y_c) &= \sum_{i=0}^{i=P-1} s(g_i - g_c)2^i \\ s(x) &= \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{9.1}$$

where g_c is the central pixel with the coordinate (x_c, y_c), and the g_i are the neighborhood pixels as shown in Eq. 9.2. The sampling is done in a circular fashion around the central pixel [160]. The LBP operator uses either uniform patterns or non-uniform patterns [160] which the first is used in this research.

For a uniform pattern, the number of bit-wise transitions is no more than two. The LBP features are extracted for all the MRI SOI-ROIs images in this step.

$$\begin{bmatrix} g_3 & g_2 & g_1 \\ g_4 & g_c & g_8 \\ g_5 & g_6 & g_7 \end{bmatrix} \tag{9.2}$$

The other feature descriptor that is used in this research is HOG. This method is used for object detection in different applications. The algorithm first divides the image to a group of cells and then for each cell the direction and the magnitude of gradient are computed. Based on the gradient direction and magnitude the HOG histogram is derived. The concatenation of different normalized histograms create the HOG feature for each of the SOI-ROI of the MRI images. The direction and magnitude of the gradient in each cell is computed based on Eq.9.3.

$$\begin{aligned} G_x(x,y) &= I(x+1,y) - I(x-1,y) \\ G_y(x,y) &= I(x,y+1) - I(x,y-1) \\ G_{Mag} &= \sqrt{G_x^2 + G_y^2} \\ G_{Dir} &= \frac{G_y}{G_x} \end{aligned} \tag{9.3}$$

$I(x,y)$ is the central pixel in the cell of the input image and the others are the neighborhood pixels which are shown in Eq. 9.4. G_{Mag} is the gradient magnitude and G_{Dir} it the gradient direction.

$$\begin{bmatrix} & I(x,y-1) & \\ I(x-1,y) & I(x,y) & I(x+1,y) \\ & I(x,y+1) & \end{bmatrix} \tag{9.4}$$

Two SOI-ROIs HOG images are shown in Figure. 9.8 for T1 and T2 subjects. Figure. 9.8(a) and (c) are visualized HOG images for T1 and T2 PD samples, respectively. Figure. 9.8(b) and (d) are HOG images for T1 and T2 HC subjects, respectively. Two SOI-ROIs LBP images are shown in Figure. 9.9 for T1 and T2. Figure. 9.9(a) and (c) are visualized LBP images for PD samples of T1 and T2. In addition, Figure. 9.9(b) and (d) are LBP images for T1 and T2 HC subjects, respectively.

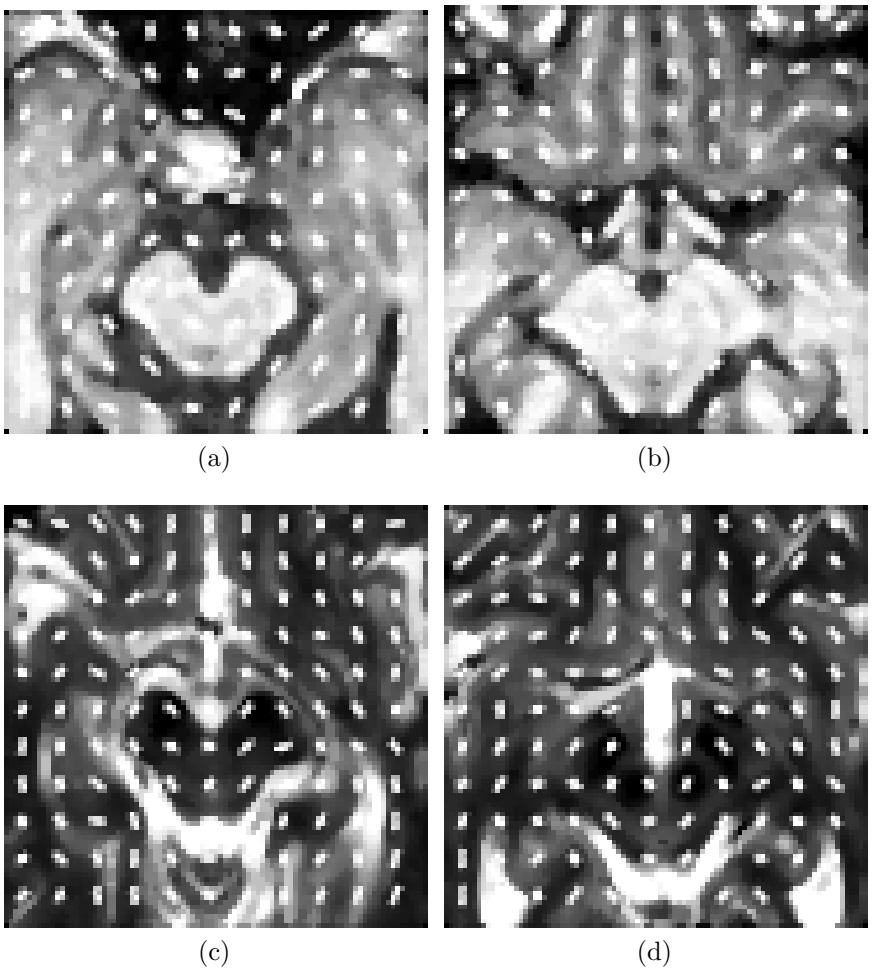


Figure 9.8: The HOG images of two SOI-ROIs for: (a) PD, (b) HC T1 subject, and (c) PD, (d) HC T2 subject.

9.3.3 Feature Reduction and Classification

After extraction of LBP and HOG features from SOI-ROIs of different subject's MRI T2 and T1 images, the next part is classification. In this part the goal is using the extracted features for distinguishing between the PD and HC subjects. The extracted features can have redundant information which may affect the classification results. Therefore, we perform feature reduction before doing the classification to eliminate redundant information. One of the popular feature reduction in neuroimaging analysis reviewed in Benson et al.[108] is PCA. PCA finds relevant features by linearly transforming correlated variables into a smaller number of uncorrelated variables which are known as principal

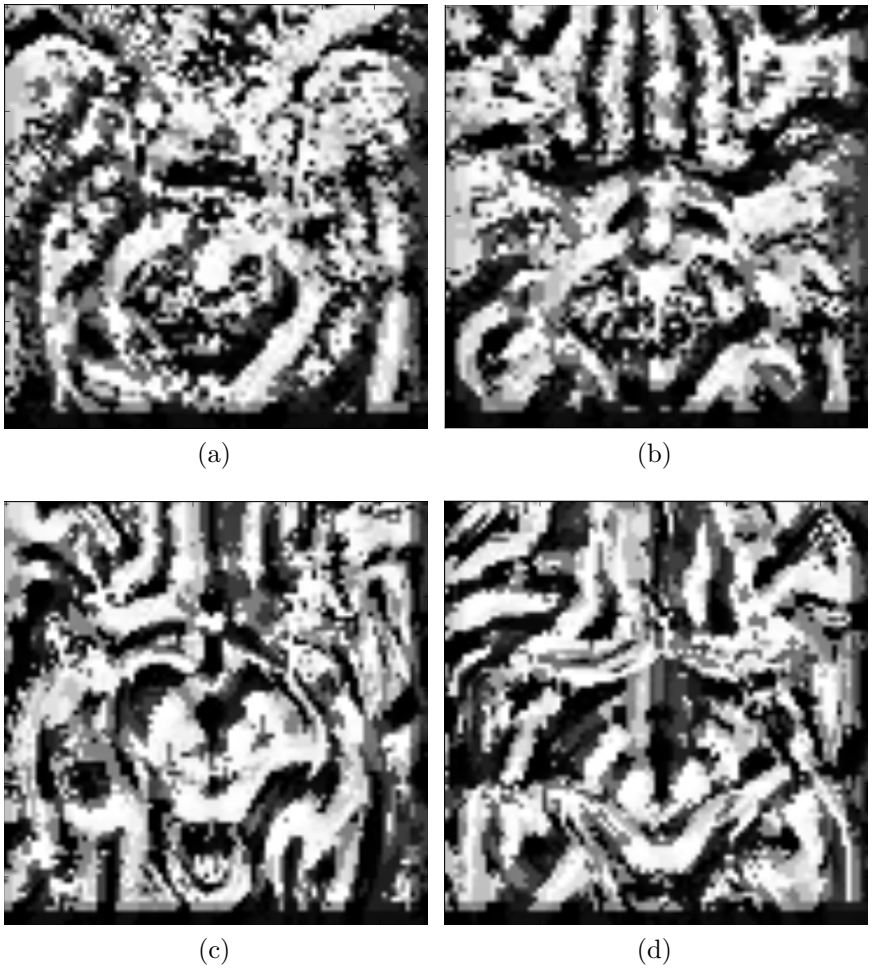


Figure 9.9: The LBP images of two SOI-ROIs for: (a) PD, (b) HC T1 subject, and (c) PD, (d) HC T2 subject.

components [80]. The resulting principal components are essentially linear combinations of the original data capturing most of the variance in the data [108].

The classification is done by two types of machine learning classifiers including RF and SVM. SVM is one of the main classification algorithms in different application proposed by Vapnik in 1995 [41]. This algorithm finds the optimal hyperplane that has the maximum margin and minimum distance to the data points of different classes. RF is one of the important ensemble learning method (Breiman et al.[24], 1984) trained on data sets of the same size as training sets parallel [143]. The final decision is going to be taken based on the majority votes of all the classification trees in the forest. We use four types

	PD	***HC***	***Average/Total***
Age	62.04 ± 9.71	61.95 ± 10.41	61.99 ± 10.06
Sex (F/M)	58/41	26/15	84/56

Table 9.1: Demographic information of T2 and T1 MRI data obtained from PPMI data set.

of classification models for the evaluation of LBP, HOG and the combination of these two feature descriptors. The classification models are: 1) RF, 2) PCA +RF, 3) SVM, and 4) PCA +SVM. The abilities of these four models will be compared for our PD classification goal.

9.4 Experimental Results and Discussion

In this section, the data set used in this research is explained and the results of different classification models are presented. The effectiveness of the RF and SVM classifiers with and without using PCA for the two texture features (LBP, HOG, and the fusion of those) are evaluated. Furthermore, the performance of these models is compared using different number of midbrain MR scans. As mentioned in the previous sections, our objective is to classify PD and HC based on T1 and T2 MR data.

9.4.1 Data Set

In our work, we used MRI T2 and T1 images, which are obtained from the PPMI database (www.ppmi-info.org/data). PPMI is the biggest international study for determining PD progression biomarkers. It consists of different types of data including neuroimaging, demographic and clinical information. We downloaded MR images from PPMI. We use 99 PD and 41 HC subjects who have both MR T1 and T2 images. The demographic information of this data is presented in Table 9.1.

9.4.2 Results

The proposed system in this research uses LBP and HOG features with 4 types of classification models including: SVM, RF, PCA +SVM and PCA +RF for

PD classification with a group of subjects having both T1 and T2 data. In addition, combinations of the texture feature descriptors are also evaluated for PD classification. LBP and HOG features are concatenated for each subject in order to create the fused features for the classification models.

80% of the data were used for training and the rest (20%) were used for testing. Cross validation is used with different numbers including $k = [3, 5, 10]$ in the classification process. The best classification result is defined based on grid search with the hyper parameter values for different models (SVM, PCA +SVM, RF, PCA +RF).

As you can see in Table 9.1, the number of PD is a lot more than the number of HC samples. Many problems in data mining and machine learning involve the analysis of rare patterns of occurrence. This imbalance between the rare category and the common category can cause significant bias toward the common category in resulting models. Balancing data sets is necessary to correct the bias in the learning process [74]. There are several options for learning unbalanced data which can be ccarried out using down-sampling or up-sampling. We used down-sampling for handling imbalance problem. In our proposed method, the same number of the minority class (HC group) are randomly selected from the majority class (PD group). As shown in Table 9.1, the number of HC samples is 41. Thus, 41 samples are randomly selected from the PD group to make a balanced data set for the classification process. There are practical benefits and constraints for using randomness. Randomness may impact the model's convergence rate, the stability of the results, and the final quality of the neural network. In order to tackle these issues, a solution is to have full visibility into the data, parameters, and details in the environment that led to a specific result [37]. Therefore, in our system we defined a single variable that contains a static random seed and use it for down-sampling, where we want to randomly select samples from the majority (PD) class. Seed values are used to make sure that the data are selected from the data distribution. We use 5 randomly seed values for this purpose.

The criteria for evaluating the proposed system are accuracy, specificity, and sensitivity, which are defined below (Eq. 9.5). Different criteria are used

Criteria	*Accuracy*						*Specificity*						*Sensivisity*					
Slice #	*3*	*4*	*5*	*6*	*7*	*8*	*3*	*4*	*5*	*6*	*7*	*8*	*3*	*4*	*5*	*6*	*7*	*8*
RF	*0.65*	*0.65*	*0.88*	*0.82*	*0.82*	*0.71*	*0.70*	*0.50*	*0.90*	*1.00*	*0.90*	*0.80*	*0.57*	*0.86*	*0.86*	*0.70*	*0.71*	*0.57*
PCA+RF	***0.94***	*0.82*	*0.71*	*0.82*	*0.82*	*0.82*	*0.85*	*0.70*	*0.70*	*0.90*	*0.70*	*0.90*	*1*	*1.00*	*0.71*	*0.71*	*1.00*	*0.71*
SVM	*0.76*	*0.82*	*0.82*	***0.94***	*0.82*	*0.88*	*0.90*	*0.90*	*0.80*	*1.00*	*0.90*	*0.80*	*0.57*	*0.71*	*0.86*	*0.86*	*0.71*	*1.00*
PCA+SVM	*0.71*	*0.82*	*0.71*	*0.76*	*0.88*	***0.94***	*0.90*	*0.90*	*0.70*	*0.80*	*0.71*	*0.90*	*0.43*	*0.71*	*0.71*	*0.71*	*1.00*	*1.00*

Table 9.2: Classification results using LBP feature for T1 MR images.

Criteria	*Accuracy*						*Specificity*						*Sensivisity*					
Slice #	*3*	*4*	*5*	*6*	*7*	*8*	*3*	*4*	*5*	*6*	*7*	*8*	*3*	*4*	*5*	*6*	*7*	*8*
RF	*0.71*	*0.76*	*0.71*	*0.71*	*0.76*	*0.82*	*0.70*	*0.57*	*0.50*	*0.71*	*0.70*	*0.90*	*0.71*	*0.90*	*1.00*	*0.70*	*0.86*	*0.71*
PCA+RF	*0.76*	*0.76*	*0.65*	*0.71*	*0.65*	*0.71*	*1.00*	*1.00*	*0.80*	*0.90*	*0.86*	*0.80*	*0.60*	*0.60*	*0.43*	*0.43*	*0.50*	*0.57*
SVM	*0.71*	*0.76*	*0.71*	*0.65*	*0.71*	*0.71*	*0.71*	*0.60*	*1.00*	*0.70*	*0.70*	*0.90*	*0.70*	*1.00*	*0.50*	*0.57*	*0.71*	*0.43*
PCA+SVM	*0.76*	*0.88*	*0.59*	*0.76*	*0.65*	*0.82*	*0.90*	*0.90*	*0.50*	*0.90*	*0.80*	*0.71*	*0.57*	*0.86*	*0.71*	*0.57*	*0.43*	*0.90*

Table 9.3: Classification results using HOG feature for T1 MR images.

to confirm the reliability and correctness of our results.

$$\begin{aligned} Accuracy &= (TP + TN)/(TP + TN + FP + FN) \\ Specificity &= TN/(TN + FP) \\ Sensitivity &= TP/(TP + FN) \end{aligned} \tag{9.5}$$

TP is the number of cases correctly classified as PD, FP is the number of cases incorrectly classified as PD, TN is the number of cases correctly classified as healthy and FN is the number of cases incorrectly classified as healthy. The four classification models described above are used to evaluate the descriptive abilities of the two feature descriptors LBP and HOG.

In our method, we only use the mid-brain slices from T1 and T2 MR data instead of the whole 3D volume. We compared the performance by using different numbers of mid-brain slices, i.e., from 3 to 8, and select the optimal number. This is one of the many contributions of this research. The classification results for T1 MR images are presented in Table 9.2, Table 9.3, and Table 9.4 for LBP, HOG and fusion of LBP and HOG, respectively. The results for different number of slices, features and the four classifier models are reported.

The experimental results for T2 MR images using LBP, HOG, and their fusion are presented in Table 9.5, Table 9.6, and Table 9.7, respectively.

Criteria	*Accuracy*						*Specificity*						*Sensivisity*					
Slice #	*3*	*4*	*5*	*6*	*7*	*8*	*3*	*4*	*5*	*6*	*7*	*8*	*3*	*4*	*5*	*6*	*7*	*8*
RF	0.59	0.71	0.76	0.82	0.65	0.71	0.8	0.7	0.7	0.8	0.5	0.7	0.29	0.71	0.86	0.86	0.86	0.71
PCA+RF	0.71	0.59	0.76	0.82	0.65	0.65	0.90	0.86	0.80	0.80	0.60	0.60	0.43	0.40	0.71	0.86	0.71	0.71
SVM	0.76	0.59	0.71	0.76	0.59	0.76	0.6	0.4	0.6	0.86	0.6	0.9	1	0.86	0.86	0.7	0.57	0.57
PCA+SVM	0.76	0.65	0.71	0.82	0.76	0.71	0.7	0.7	1	0.8	0.8	0.5	0.86	0.57	0.5	0.86	0.71	1

Table 9.4: Classification results using fusion of LBP and HOG features for T1 MR images.

Criteria	*Accuracy*						*Specificity*						*Sensivisity*					
Slice #	*3*	*4*	*5*	*6*	*7*	*8*	*3*	*4*	*5*	*6*	*7*	*8*	*3*	*4*	*5*	*6*	*7*	*8*
RF	*0.82*	*0.88*	*0.82*	*0.82*	*0.76*	*0.76*	*0.9*	*0.9*	*0.86*	*0.8*	*0.86*	*0.7*	*0.71*	*0.86*	*0.8*	*0.86*	*0.7*	*0.86*
PCA+RF	*0.64*	*0.64*	*0.70*	*0.70*	*0.82*	*0.82*	*0.90*	*0.50*	*0.50*	*0.70*	*0.90*	*0.70*	*0.28*	*0.85*	*1*	*0.71*	*0.71*	*1*
SVM	*0.76*	*0.82*	*0.76*	*0.88*	***0.94***	*0.76*	*0.86*	*0.57*	*0.8*	*0.9*	*0.9*	*0.6*	*0.7*	*1*	*0.71*	*0.86*	*1*	*1*
PCA+SVM	*0.82*	*0.76*	*0.76*	*0.82*	*0.88*	*0.88*	*0.90*	*0.70*	*0.71*	*0.80*	*0.90*	*0.80*	*0.71*	*0.86*	*0.80*	*0.86*	*0.86*	*1.00*

Table 9.5: Classification results using LBP feature for T2 MR images.

9.5 Discussion and Future Works

Our proposed CAD system only needs a few SOI-ROI for each subject for PD classification using either T1 or T2 MR images. Thus, it has low computational complexity and fast processing. The analysis in this research proves that not all slices in the DICOM volume have useful information for PD classification. Furthermore, in each slice (2D) not all the regions are essential for analysis. Compared to other proposed papers in this field, we did not consider just the SN part, but we consider SN and the area around it as our ROI. By taking consideration of neighborhood information, better validation can be performed on extracted features.

The results of classification T1 illustrate that the LBP feature descriptor is more powerful in discriminating PD and HC. The highest accuracy for T1 is achieved for LBP feature. It is worth noting that each model achieves its high accuracy with a different number of slices. For example, the number of slices for getting maximum accuracy for SVM is 6. Whereas, for getting the same accuracy with PCA +RF model, we need just 3 SOIs and with PCA +SVM model, 8 SOIs are needed. Figure. 9.10 shows the relation between accuracy and number of SOI for T1 MR image classification using LBP and the three models with higher accuracy. As the first plot in Figure. 9.10(a) shows, the accuracy is increasing the highest when the number of $SOI = 3$. The PCA +RF model in Figure. 9.10(b), has the highest accuracy at $SOI = 6$.The

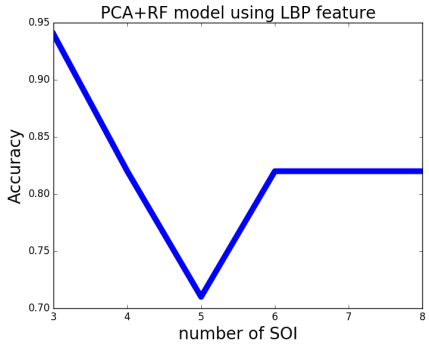
Criteria	*Accuracy*						*Specificity*						*Sensivisity*					
Slice #	*3*	*4*	*5*	*6*	*7*	*8*	*3*	*4*	*5*	*6*	*7*	*8*	*3*	*4*	*5*	*6*	*7*	*8*
RF	*0.76*	*0.71*	*0.82*	*0.65*	*0.88*	*0.82*	*0.7*	*0.8*	*0.8*	*0.6*	*1*	*0.9*	*0.86*	*0.57*	*0.86*	*0.71*	*0.71*	*0.71*
PCA+RF	*0.82*	*0.76*	*0.82*	*0.71*	*0.71*	*0.71*	*0.90*	*0.80*	*0.70*	*0.80*	*0.60*	*0.50*	*0.71*	*0.71*	*1.00*	*0.57*	*0.86*	*1.00*
SVM	*0.65*	*0.88*	*0.88*	*0.88*	*0.76*	*0.88*	*0.8*	*0.8*	*0.9*	*0.9*	*0.7*	*0.8*	*0.43*	*1*	*0.86*	*0.86*	*0.86*	*1*
PCA+SVM	*0.88*	*0.76*	*0.82*	*0.76*	*0.71*	*0.88*	*0.86*	*0.60*	*0.70*	*0.43*	*0.60*	*0.90*	*0.90*	*1.00*	*1.00*	*1.00*	*0.86*	*0.86*

Table 9.6: Classification results using HOG feature for T2 MR images.

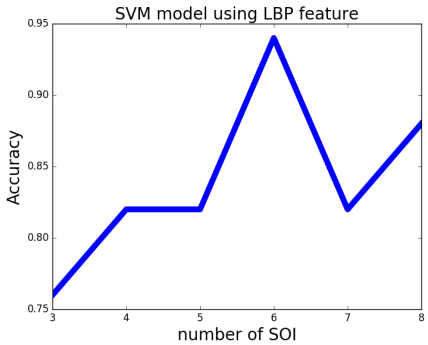
Criteria	*Accuracy*						*Specificity*						*Sensivisity*					
Slice #	*3*	*4*	*5*	*6*	*7*	*8*	*3*	*4*	*5*	*6*	*7*	*8*	*3*	*4*	*5*	*6*	*7*	*8*
RF	*0.71*	*0.65*	*0.71*	***0.94***	*0.65*	*0.76*	*0.70*	*0.60*	*0.70*	*1.00*	*0.60*	*0.70*	*0.71*	*0.71*	*0.71*	*0.90*	*0.71*	*0.86*
PCA+RF	*0.64*	*0.64*	*0.70*	*0.70*	*0.82*	*0.82*	*0.85*	*0.5*	*0.5*	*0.7*	*0.9*	*0.7*	*0.5*	*0.85*	*1*	*0.71*	*0.71*	*1*
SVM	*0.76*	*0.76*	*0.88*	*0.71*	*0.65*	*0.71*	*0.71*	*0.70*	*0.80*	*0.60*	*0.40*	*0.71*	*0.80*	*0.86*	*1.00*	*0.86*	*1.00*	*0.70*
PCA+SVM	*0.71*	*0.65*	*0.82*	*0.76*	*0.65*	*0.71*	*0.80*	*0.50*	*0.86*	*0.60*	*0.40*	*0.71*	*0.57*	*0.86*	*0.80*	*1.00*	*1.00*	*0.70*

Table 9.7: Classification results using fusion of LBP and HOG features for T2 MR images.

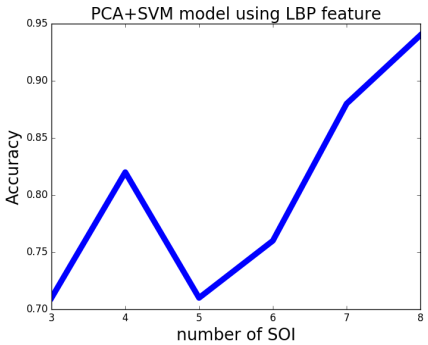
results of PCA +SVM model (Figure. 9.10(c)) shows maximum accuracy at 8 SOIs.For specificity and sensitivity values, PCA +RF model has higher values comparing to the two other winner models.



(a)



(b)



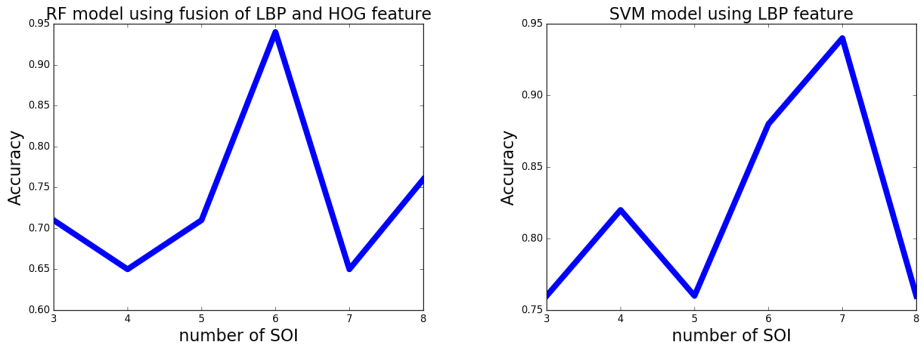
(c)

Figure 9.10: The plots of the relation between the accuracy and number of slices for the best models in T1 classification.

Based on the classification results for T1 with HOG feature, the maximum

accuracy is given by PCA +SVM model with 4 SOIs which is 0.88. The other models report similar accuracies between 0.65 to 0.76. Contrary to what was expected, the fusion of LBP and HOG features does not report any higher accuracy. In most cases, the accuracies are similar to either LBP or HOG results.

For the T2 MR images, the results illustrate the highest accuracy is obtained using SVM model with LBP feature descriptor. The maximum accuracy for T2 classification is 0.94 which is same as the highest classification accuracy for T1 data. The number of slices used for getting this amount of accuracy is 7. Plots in Figure. 9.11 show the relation between accuracy and number of SOI for T2 MR image classification using LBP and the models with highest accuracy. As the first plot shows (Figure. 9.11(a)), the accuracy is almost stable or increasing with higher number of SOI till $SOI = 7$ and after that it's decreasing. In the fusion case (Figure. 9.11(b)) with RF model has fluctuation with a big peak at $SOI = 6$ with the maximum accuracy value. The highest



(a) (b)

Figure 9.11: The plots of the relation between the accuracy and number of slices for the best models in T2 classification.

accuracy given by other classification models is 0.88 which is reported with different number of SOI. The results for HOG feature with different models represent similar outcomes. The highest accuracy with HOG feature for T2 MR image calcification is 0.88 which can be gained with different models and various number of SOIs. Same as the reported results for T1, the feature fusion for T2 has not higher accuracy than each of these features. However, the

maximum accuracy still can be obtained with fused features using RF model for T2.

Note that using PCA for feature dimension reduction is not always effective on our results. In the T1 classification using LBP feature descriptor, the results with PCA are smaller than the results without PCA using SVM model. However, RF with PCA gives higher accuracy than the same model without PCA. On the contrary, RF with PCA giving lower results compared to the same model without PCA when HOG feature is used. However, PCA does have positive impacts on SVM classifier with HOG feature.

Our research proves the classification results of T1 and T2 MR images have similar outcomes for PD. However, the highest accuracy for each data type are obtained with different models and different number of SOIs.

For future work, more MR images for T1 and T2 will be considered. In this research, the preprocessing was not done on the MR images. In the future work, the preprocessed MR T1 and T2 will be passed to our CAD system which cause to make the SOI selection automatically and even the ROI extraction will be faster and more accurate. However, the preprocessed data might not guarantee the higher accuracy for PD classification using T1 and T2 MR images. Moreover, localization the most discriminative pats of the ROI in either T1 or T2 images should be added to our proposed CAD system.

9.6 Conclusion

The reduction of dopaminergic neurons in the SN region of the brain is the main reason behind PD. In this research, we propose a method using MR T1 and T2 midbrain scans for PD classification. The system's main components are preprocessing, feature extraction, feature reduction and classification. When preprocessing each subject, 3 to 8 midbrain slices are selected as SOI. In each SOI, SN and its neighbor are chosen as ROI. We use the LBP and HOG techniques in the feature extraction. PCA is used for feature reduction to remove the redundant information irrelevant to PD. Finally, four classification models using RF and SVM, with or without PCA, are evaluated. The PPMI

data set is used for the evaluation of the proposed CAD system. MRI T1 and T2 images of 91 PD and 41 Health Control subjects are used for analysis. Seed-based random under sampling is conducted for making the data set balanced. The classification models using LBP feature, HOG feature and the fusion of these features are evaluated and compared. Experimental findings present and compare the ability of the proposed system for T1 and T2 MR images. Furthermore, the classification ability of the CAD system is evaluated using different number of midbrain MR images from T1 and T2. The experimental results illustrate the maximum accuracy for T1 and T2 MR image classification are similar (0.94). The highest accuracy is obtained with different classification models using LBP features for both MR imaging samples (T1 and T2). Depends on the classification model, different number of scans are used for discriminating between PD and HC subjects. In the nutshell, PD classification can be done with either T1 or T2 and with just a few numbers of midbrain slices with high performance.

Chapter 10

Conclusions and Future Directions

The current thesis is proposing automatic CAD systems with non-invasive techniques for diagnosing and monitoring of PD, the second-most prevalent neurodegenerative age-related disease. This disease is characterized by the loss of dopamine neurotransmitters in the brain region known as SN. There is no established treatment and diagnostic method for PD but earlier diagnosis results in more successful and efficient treatment for patients. The main signs of PD are motor symptoms such as bradykinesia, rigidity, slowness, tremor and FOG, which results in falls that are particularly dangerous for elderly people. In addition, PD has non-motor symptoms such as cognitive impairment, excessive sleeping, and olfactory problems. The motor-symptoms usually reveal themselves at later stages of the disease when most of the dopaminergic cells are lost. However, the non-motor signs might appear at earlier stages of the disease, which makes them more useful for possible early diagnosis of PD. In this research, a list of contributions have been made using three types of data: kinematic motion data, demographic and clinical information, and neuroimaging (MR T1 and T2 image) data.

The gait analysis section consists of two parts. The data in the first part is captured using a non-invasive and low-cost mocap device known as Kinect, with the collaboration of a specialist at the University of Alberta neuroscience department. The gait data is collected using three experiments, which are normally performed by specialists to assess PD patients (SL, Tremor, TUG).

In this part, gait-related features are extracted from the motion data for PD and HC subjects. Thereafter, RF and K-means are used for PD diagnosis and progress monitoring. The experimental results demonstrate the high accuracy of our proposed approach to assess PD using kinematic motion data. The limitation of the current study arises from the small group of subjects for evaluation purpose. However, gait analysis will be done for PD patients as future work using bigger dataset which is provided by Arizona State University. In the second part of the motion analysis section, an automatic FOG assessment system ("Kin-FOG") is designed for PD patients to provide objective information to neurologists about the FOG condition and the symptoms' characteristics. The proposed FOG assessment system uses Kinect for capturing data of healthy subjects who are trained to imitate the FOG phenomenon. The gradient of displacement for foot joint trajectory is used for distinguishing the FOG episodes. Furthermore, the angle between the foot and the ground is used for false positive reduction, resulting from having resting modes in our gait analysis. The proposed Kin-FOG system is low-cost, accurate, and easy to use in FOG assessment for PD patients. In addition, the proposed system can be used remotely at patients' homes for sending the FOG status to doctors and other healthcare providers. The Kin-FOG system evaluation part is done using healthy subjects, which is a shortcoming. However, In future work, we will reevaluate using PD subjects.

Clinical properties are one of the primary sources in PD diagnosis and monitoring. Assessing the large amount of clinical data by specialists is time consuming, difficult, and also subjective. Therefore, in the second section of our research, an automatic system is designed using clinical characteristics, with machine learning methods (MDI for feature selection and RF for classification). The clinical data used is taken from PPMI. The experimental results show high accuracy in classifying PD and HC, and progress monitoring using HAY score, which outperforms current state-of-the-art methods. In this part, 13 clinical features for PD patients are used. However, there are more clinical properties for PD which need to be assessed and ranked based on their relative impacts on classification and progress monitoring.

In the neuroimaging analysis section, non-invasive MRI imaging is used for PD classification. The neuroimaging analysis is conducted using two popular types of MR images: T1 and T2. Three contributions are presented for T1 MR images. The first contribution uses FreeSurfer for preprocessing the MRI T1 data. Sub-cortical features, including volume, area, mean-curvature, and thickness are extracted and used for PD classification. The decision models for classification are based on RF and SVM methods. In the second contribution, SPM/CAT12 is used for preprocessing and parcellation of T1 MR images to GM and WM regions. SVM, RF, and GB are used with GM and WM volumes for PD diagnosis. Moreover, in both contributions PCA has been used for feature selection and dimensionality reduction. In the third contribution for T1, different DL models (2D and 3D) are evaluated and compared based on the ability to distinguish PD and HC. The dataset for T1 analysis is downloaded from PPMI and IXI datasets. IXI dataset has been used for fixing the imbalance issue in the PPMI dataset. Furthermore, there is a contribution which is performed for T1 and T2 MR images. The proposed method designs a PD assessment system solely based on the midbrain scans. Moreover, the analysis is based on the SN region as ROI. In the proposed system, LBP and HOG are used as feature descriptors, which are two powerful computer vision techniques. The classification models are SVM and RF with or without PCA as the dimensionality reduction method. The proposed classification system's results are compared for T1 and T2. The relation between the performance of the proposed CAD system and the number of midbrain scans are evaluated. The goal was PD classification using just MRI. Also, finding the MRI biomarkers for PD was another goal. Since, the dataset is not specifically for early stage of PD, the proposed CAD systems using MRI might not be applicable for early PD classification. However, the proposed method can be evaluated on the dataset of MRI for a group of PD subjects whose data is collected at different time starting from when they just got sick till the later stages of their disease.

The dataset for evaluation of FOG detection system were simulated by healthy subjects. Therefore, more gait analyses need to be done in order to use FOG detection for real PD patients. In this regard, we have started a

collaboration with Arizona State University, who have provided motion data for a group of PD patients who may have FOG or not. The first goal in this project is utilizing machine learning techniques for FOG classification using provided motion data. Additionally, The proposed CAD systems in neuroimaging analysis can be applied for other neuroimaging such as PET, DTI, and fMRI data. Since there is no single biomarker can accurately detect PD, we can see researchers analyzing different features in the literature including MRI and gait. In this thesis by analyzing gait and probably the correlation between gait and neural features, future research can focus on the detected significant feature(s) and monitor how a feature evolves over time in a time series. My thesis will eventually lead to a longer term goal on detecting PD at an early stage.

Physicians intuitively apply pattern recognition when evaluating a patient. However, there are different data characteristics which need to be considered, which makes the evaluation process complex, leading to high rates of deadly and costly diagnostic errors. Therefore, computerized diagnosis support systems are necessary to make the diagnosis process easier, faster, and much more accurate. The reason why these systems are still not widely accepted by clinicians is because of their current limitations. First is the lacking of having a simple and user-friendly graphic user interface such as the one this thesis proposes for Kin-FOG system. Secondly, clinicians require time to learn how to use the CAD system, which they often avoid due to their busy schedules and lack of trust in the system. Note that, crowd-sourcing can cause inaccurate or biased training data when machine learning is adopted [29]. In addition, utilizing CAD systems require real-time data sharing between different hospitals and clinics which sometimes is not possible. Addressing these limitations will be a good direction for future research.

References

[1] E. Adeli, F. Shi, L. An, C.-Y. Wee, G. Wu, T. Wang, and D. Shen, "Joint feature-sample selection and robust diagnosis of parkinson's disease from MRI data," *NeuroImage*, vol. 141, pp. 206–219, 2016.

[2] C. H. Adler, "Premotor symptoms and early diagnosis of parkinson's disease," *International Journal of Neuroscience*, vol. 121, no. sup2, pp. 3–8, 2011.

[3] V. Aharonson, I. Schlesinger, A. M. McDonald, S. Dubowsky, and A. D. Korczyn, "A practical measurement of parkinson's patients gait using simple walker-based motion sensing and data analysis," *Journal of Medical Devices*, vol. 12, no. 1, p. 011 012, 2018.

[4] A. Ahmadi, F. Destelle, L. Unzueta, D. S. Monaghan, M. T. Linaza, K. Moran, and N. E. O'Connor, "3d human gait reconstruction and monitoring using body-worn inertial sensors and kinematic modeling," *IEEE Sensors Journal*, vol. 16, no. 24, pp. 8823–8831, 2016.

[5] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, "Deep learning for brain mri segmentation: State of the art and future directions," *Journal of digital imaging*, vol. 30, no. 4, pp. 449–459, 2017.

[6] H. M. Ali, "Mri medical image denoising by fundamental filters," in *High-Resolution Neuroimaging-Basic Physical Principles and Clinical Applications*, InTech, 2018.

[7] J. S. Allen, J. Bruss, S. Mehta, T. Grabowski, C. K. Brown, and H. Damasio, "Effects of spatial transformation on regional brain volume estimates," *Neuroimage*, vol. 42, no. 2, pp. 535–547, 2008.

[8] N. E. Allen, A. K. Schwarzel, and C. G. Canning, "Recurrent falls in parkinson's disease: A systematic review," *Parkinson's disease*, vol. 2013, 2013.

[9] A. Amini, K. Banitsas, and S. Hosseinzadeh, "A new technique for foot-off and foot contact detection in a gait cycle based on the knee joint angle using microsoft kinect v2," in *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, IEEE, 2017, pp. 153–156.

[10] A. Amini, K. Banitsas, and W. R. Young, "Kinect4fog: Monitoring and improving mobility in people with parkinson's using a novel system incorporating the microsoft kinect v2," *Disability and Rehabilitation: Assistive Technology*, pp. 1–8, 2018.

[11] V. F. Annese and D. De Venuto, "Gait analysis for fall prediction using emg triggered movement related potentials," in *2015 10th International Conference on Design & Technology of Integrated Systems in Nanoscale Era (DTIS)*, IEEE, 2015, pp. 1–6.

[12] J. Ashburner, G. Barnes, C. Chen, J. Daunizeau, G. Flandin, K. Friston, S. Kiebel, J. Kilner, V. Litvak, R. Moran, *et al.*, "Spm12 manual," *Wellcome Trust Centre for Neuroimaging, London, UK*, 2014.

[13] J. Ashburner and K. J. Friston, "Voxel-based morphometry—the methods," *Neuroimage*, vol. 11, no. 6, pp. 805–821, 2000.

[14] J. Ashburner, C. Hutton, R. Frackowiak, I. Johnsrude, C. Price, and K. Friston, "Identifying global anatomical differences: Deformation based morphometry," *Human brain mapping*, vol. 6, no. 5-6, pp. 348–357, 1998.

[15] H. T. Atasoy, O. Nuyan, T. Tunc, M. Yorubulut, A. E. Unal, and L. E. Inan, "T2-weighted mri in parkinson's disease; substantia nigra pars compacta hypointensity correlates with the clinical scores," *Neurology India*, vol. 52, no. 3, p. 332, 2004.

[16] D. Athauda, K. Maclagan, S. S. Skene, M. Bajwa-Joseph, D. Letchford, K. Chowdhury, S. Hibbert, N. Budnik, L. Zampedri, J. Dickson, *et al.*, "Exenatide once weekly versus placebo in parkinson's disease: A randomised, double-blind, placebo-controlled trial," *The Lancet*, vol. 390, no. 10103, pp. 1664–1675, 2017.

[17] C. Atkinson-Clement, S. Pinto, A. Eusebio, and O. Coulon, "Diffusion tensor imaging in parkinson's disease: Review and meta-analysis," *NeuroImage: Clinical*, vol. 16, pp. 98–110, 2017.

[18] G. S. Babu, S. Suresh, and B. S. Mahanand, "A novel PBL-McRBFN-RFE approach for identification of critical brain regions responsible for parkinson's disease," *Expert Systems with Applications*, vol. 41, no. 2, pp. 478–488, 2014.

[19] M. Bachlin, M. Plotnik, D. Roggen, I. Maidan, J. M. Hausdorff, N. Giladi, and G. Troster, "Wearable assistant for parkinson's disease patients with the freezing of gait symptom," *IEEE Trans. Information Technology in Biomedicine*, vol. 14, no. 2, pp. 436–446, 2010.

[20] G. Bartzokis, J. L. Cummings, C. H. Markham, P. Z. Marmarelis, L. J. Treciokas, T. A. Tishler, S. R. Marder, and J. Mintz, "Mri evaluation of brain iron in earlier-and later-onset parkinson's disease and normal subjects," *Magnetic resonance imaging*, vol. 17, no. 2, pp. 213–222, 1999.

[21] A. A. M. Bigy, K. Banitsas, A. Badii, and J. Cosmas, "Recognition of postures and freezing of gait in parkinson's disease patients using microsoft kinect sensor," in *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, IEEE, 2015, pp. 731–734.

[22] P. Borghammer, K. Ostergaard, P. Cumming, A. Gjedde, A. Rodell, N. Hall, and M. Chakravarty, "A deformation-based morphometry study of patients with early-stage parkinson's disease," *European journal of neurology*, vol. 17, no. 2, pp. 314–320, 2010.

[23] L. Breiman, "Out-of-bag estimation," *University of Brekeley*, 1996.

[24] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[25] J. Brown, "An atlas of parkinson's disease and related disorders," *Journal of neurology, neurosurgery, and psychiatry*, vol. 66, no. 2, p. 257, 1999.

[26] N. Brown and L. Jones, "Knowledge of medical imaging radiation dose and risk among doctors," *Journal of medical imaging and radiation oncology*, vol. 57, no. 1, pp. 8–14, 2013.

[27] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks*, vol. 106, pp. 249–259, 2018.

[28] E. J. Burton, I. G. McKeith, D. J. Burn, E. D. Williams, and J. T. O'Brien, "Cerebral atrophy in parkinson's disease with and without dementia: A comparison with alzheimer's disease, dementia with lewy bodies and controls," *Brain*, vol. 127, no. 4, pp. 791–800, 2004.

[29] A. Cahan and J. J. Cimino, "A learning health care system using computer-aided diagnosis," *Journal of medical Internet research*, vol. 19, no. 3, e54, 2017.

[30] J. Camps, A. Sama, M. Martin, D. Rodriguez-Martin, C. Perez-Lopez, S. Alcaine, B. Mestre, A. Prats, M. C. Crespo, J. Cabestany, *et al.*, "Deep learning for detecting freezing of gait episodes in parkinson's disease based on accelerometers," in *International Work-Conference on Artificial Neural Networks*, Springer, 2017, pp. 344–355.

[31] J. Camps, A. Sama, M. Martin, D. Rodriguez-Martin, C. Perez-Lopez, J. M. M. Arostegui, J. Cabestany, A. Catala, S. Alcaine, B. Mestre, *et al.*, "Deep learning for freezing of gait detection in parkinson's disease patients in their homes using a waist-worn inertial measurement unit," *Knowledge-Based Systems*, vol. 139, pp. 119–131, 2018.

[32] *CAT A Computational Anatomy Toolbox for SPM*, `http://www.neuro.uni-jena.de/`, Online; accessed 30 September 2019.

[33] K. R. Chaudhuri and A. H. Schapira, "Non-motor symptoms of parkinson's disease: Dopaminergic pathophysiology and treatment," *The Lancet Neurology*, vol. 8, no. 5, pp. 464–474, 2009.

[34] H.-C. Cheng, C. M. Ulane, and R. E. Burke, "Clinical progression in parkinson disease and the neurobiology of axons," *Annals of neurology*, vol. 67, no. 6, pp. 715–725, 2010.

[35] Z. Cheng, J. Zhang, N. He, Y. Li, Y. Wen, H. Xu, R. Tang, Z. Jin, M. Haacke, F. Yan, *et al.*, "Radiomic features of the nigrosome-1 region of the substantia nigra: Using quantitative susceptibility mapping to assist the diagnosis of idiopathic parkinson's disease," *Frontiers in aging neuroscience*, vol. 11, p. 167, 2019.

[36] H. Choi, S. Ha, H. J. Im, S. H. Paek, and D. S. Lee, "Refining diagnosis of parkinson's disease with deep learning-based interpretation of dopamine transporter imaging," *NeuroImage: Clinical*, vol. 16, pp. 586–594, 2017.

[37] *Choosing the correct seed for reproducible research/results*, `https://stats.stackexchange.com/questions/335936/`, Accessed: 2019-08-08.

[38] O. Cigdem, I. Beheshti, and H. Demirel, "Effects of different covariates and contrasts on classification of parkinson's disease using structural mri," *Computers in biology and medicine*, vol. 99, pp. 173–181, 2018.

[39] O. Cigdem, A. Yilmaz, I. Beheshti, and H. Demirel, "Comparing the performances of PDF and PCA on parkinson's disease classification using structural MRI images," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, IEEE, 2018.

[40] W. W. Cohen, "Fast effective rule induction," in *Machine learning proceedings 1995*, Elsevier, 1995, pp. 115–123.

[41] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[42] K. Cresswell, Y. Shin, and S. Chen, "Quantifying variation in gait features from wearable inertial sensors using mixed effects models," *Sensors*, vol. 17, no. 3, p. 466, 2017.

[43] A. L. S. De Lima, L. J. Evers, T. Hahn, L. Bataille, J. L. Hamilton, M. A. Little, Y. Okuma, B. R. Bloem, and M. J. Faber, "Freezing of gait and fall detection in parkinson's disease using wearable sensors: A systematic review," *Journal of neurology*, vol. 264, no. 8, pp. 1642–1654, 2017.

[44] A. Delval, A. H. Snijders, V. Weerdesteyn, J. E. Duysens, L. Defebvre, N. Giladi, and B. R. Bloem, "Objective detection of subtle freezing of gait episodes in parkinson's disease," *Movement Disorders*, vol. 25, no. 11, pp. 1684–1693, 2010.

[45] *Diagnostic imaging dataset statistical release*, https://www.england.nhs.uk/statistics, [Online; Accessed Aug. 6, 2018], 2017.

[46] M. Diez-Cirarda, N. Ibarretxe-Bilbao, J. Pena, and N. Ojeda, "Neurorehabilitation in parkinson's disease: A critical review of cognitive rehabilitation effects on cognition and brain," *Neural plasticity*, vol. 2018, 2018.

[47] P. Dirac, "The lorentz transformation and absolute time," *Physica*, vol. 19, no. 1-–12, pp. 888–896, 1953.

[48] B. P. Drayer, "Magnetic resonance imaging and extrapyramidal movement disorders," *European neurology*, vol. 29, no. Suppl. 1, pp. 9–12, 1989.

[49] S. Fahn *et al.*, "Description of parkinson's disease as a clinical syndrome," *ANNALS-NEW YORK ACADEMY OF SCIENCES*, vol. 991, pp. 1–14, 2003.

[50] M. L. Ferster, S. Mazilu, and G. Troster, "Gait parameters change prior to freezing in parkinson's disease: A data-driven study with wearable inertial units," in *Proceedings of the 10th EAI International Conference on Body Area Networks*, ICST, 2015, pp. 159–166.

[51] R. Feynman and F. Vernon Jr., "The theory of a general quantum system interacting with a linear dissipative system," *Annals of Physics*, vol. 24, pp. 118–173, 1963.

[52] M. M. Flores-Barranco, M.-A. Ibarra-Mazano, and I. Cheng, "Accidental fall detection based on skeleton joint correlation and activity boundary," in *International Symposium on Visual Computing*, Springer, 2015, pp. 489–498.

[53] N. K. Focke, G. Helms, S. Scheewe, P. M. Pantel, C. G. Bachmann, P. Dechent, J. Ebentheuer, A. Mohr, W. Paulus, and C. Trenkwalder, "Individual voxel-based subtype prediction can differentiate progressive supranuclear palsy from idiopathic parkinson syndrome and healthy controls," *Human Brain Mapping*, vol. 32, no. 11, pp. 1905–1915, 2011.

[54] FreeSurferWiki, *Freesurfer*, https://surfer.nmr.mgh.harvard.edu/fswiki, [Online; Accessed Aug. 6, 2019].

[55] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[56] C. Gaser and R. Dahnke, "Cat-a computational anatomy toolbox for the analysis of structural mri data," *HBM*, vol. 2016, pp. 336–348, 2016.

[57] H. Gharagyozyan, *A practical application of machine learning in medicine*, https://www.macadamian.com/learn/a-practical-application-of-machine-learning-in-medicine/, 2019.

[58] F. S. Gharehchopogh and P. Mohammadi, "A case study of parkinson's disease diagnosis using artificial neural networks," *International Journal of Computer Applications*, vol. 73, no. 19, 2013.

[59] W. Gibb and A. Lees, "The relevance of the lewy body to the pathogenesis of idiopathic parkinson's disease," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 51, no. 6, pp. 745–752, 1988.

[60] C. G. Goetz, B. C. Tilley, S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M. B. Stern, R. Dodel, *et al.*, "Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): Scale presentation and clinimetric testing results," *Movement disorders: official journal of the Movement Disorder Society*, vol. 23, no. 15, pp. 2129–2170, 2008.

[61] J. G. Goldman and R. Postuma, "Premotor and non-motor features of parkinson's disease," *Current opinion in neurology*, vol. 27, no. 4, p. 434, 2014.

[62] A. Hackshaw, *Small studies: Strengths and limitations*, 2008.

[63] I. Halperin, M. Morelli, A. D. Korczyn, M. B. Youdim, and S. A. Mandel, "Biomarkers for evaluation of clinical efficacy of multipotential neuroprotective drugs for alzheimer's and parkinson's diseases," *Neurotherapeutics*, vol. 6, no. 1, pp. 128–140, 2009.

[64] J. H. Han, W. J. Lee, T. B. Ahn, B. S. Jeon, and K. S. Park, "Gait analysis for freezing detection in patients with movement disorder using three dimensional acceleration system," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439)*, IEEE, vol. 2, 2003, pp. 1863–1865.

[65] A. A. Handojoseno, J. M. Shine, T. N. Nguyen, Y. Tran, S. J. Lewis, and H. T. Nguyen, "The detection of freezing of gait in parkinson's disease patients using eeg signals based on wavelet decomposition," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, 2012, pp. 69–72.

[66] A. A. Handojoseno, J. M. Shine, T. N. Nguyen, Y. Tran, S. J. Lewis, and H. T. Nguyen, "Analysis and prediction of the freezing of gait using eeg brain dynamics," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 23, no. 5, pp. 887–896, 2014.

[67] L. Harrison, K. Holli, A. Laaperi, S. Soimakallio, P. Dastidar, and H. Eskola, "Clinical applicability of mri texture analysis," *Computational Vision and Medical Image Processing: VipIMAGE 2009*, p. 409, 2009.

[68] R. A. Hauser, "New considerations in the medical management of early parkinson's disease: Impact of recent clinical trials on treatment strategy," *Parkinsonism & related disorders*, vol. 15, S17–S21, 2009.

[69] T. Herman, N. Giladi, and J. M. Hausdorff, "Properties of the 'timed up and go' test: More than meets the eye," *Gerontology*, vol. 57, no. 3, pp. 203–210, 2011.

[70] M. M. Hoehn and M. D. Yahr, "Parkinsonism: Onset, progression, and mortality," *Neurology*, vol. 17, no. 5, pp. 427–427, 1967.

[71] S. K. Holden, T. Finseth, S. H. Sillau, and B. D. Berman, "Progression of mds-updrs scores over five years in de novo parkinson disease from the parkinson's progression markers initiative cohort," *Movement disorders clinical practice*, vol. 5, no. 1, pp. 47–53, 2018.

[72] J.-H. Hur, S.-Y. Ihm, and Y.-H. Park, "A variable impacts measurement in random forest for mobile cloud computing," *Wireless Communications and Mobile Computing*, vol. 2017, 2017.

[73] *International Parkinson and Movement Disorder Society*, `https://www.movementdisorders.org/MDS.htm`, Online; accessed 30 September 2017.

[74] S. Iram, F.-B. Vialatte, and M. I. Qamar, "Early diagnosis of neurodegenerative diseases from gait discrimination to neural synchronization," in *Applied Computing in Medicine and Health*, Elsevier, 2016, pp. 1–26.

[75] *IXI Dataset*, `https://brain-development.org/ixi-dataset/`, Online; accessed 30 September 2019.

[76] J. Jankovic, "Parkinson's disease: Clinical features and diagnosis," *Journal of neurology, neurosurgery & psychiatry*, vol. 79, no. 4, pp. 368–376, 2008.

[77] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, "Improved optimization for the robust and accurate linear registration and motion correction of brain images," *Neuroimage*, vol. 17, no. 2, pp. 825–841, 2002.

[78] M. Jenkinson, M. Pechaud, S. Smith, *et al.*, "Bet2: Mr-based estimation of brain, skull and scalp surfaces," in *Eleventh annual meeting of the organization for human brain mapping*, Toronto., vol. 17, 2005, p. 167.

[79] M. Jenkinson and S. Smith, "A global optimisation method for robust affine registration of brain images," *Medical image analysis*, vol. 5, no. 2, pp. 143–156, 2001.

[80] I. Jolliffe, *Principal component analysis*. Springer, 2011.

[81] L. Kalilani, M. Asgharnejad, T. Palokangas, and T. Durgin, "Comparing the incidence of falls/fractures in parkinson's disease patients in the us population," *PloS one*, vol. 11, no. 9, e0161689, 2016.

[82] M. Khalilia, S. Chakraborty, and M. Popescu, "Predicting disease risks from highly imbalanced data using random forest," *BMC medical informatics and decision making*, vol. 11, no. 1, p. 51, 2011.

[83] F. Kurth, E. Luders, and C. Gaser, "Vbm8 toolbox manual," *Jena: University of Jena*, 2010.

[84] A. E. Lang and A. J. Espay, "Disease modification in parkinson's disease: Current approaches, challenges, and future considerations," *Movement Disorders*, vol. 33, no. 5, pp. 660–677, 2018.

[85] J. W. Langston, J. C. Wiley, and M. Tagliati, "Optimizing parkinson's disease diagnosis: The role of a dual nuclear imaging algorithm," *NPJ Parkinson's disease*, vol. 4, no. 1, p. 5, 2018.

[86] P.-L. Lee, K.-H. Chou, C.-H. Lu, H.-L. Chen, N.-W. Tsai, A.-L. Hsu, M.-H. Chen, W.-C. Lin, and C.-P. Lin, "Extraction of large-scale structural covariance networks from grey matter volume for parkinson's disease classification," *European radiology*, vol. 28, no. 8, pp. 3296–3305, 2018.

[87] H. Lei, Y. Zhao, Y. Wen, Q. Luo, Y. Cai, G. Liu, and B. Lei, "Sparse feature learning for multi-class parkinson's disease classification," *Technology and Health Care*, vol. 26, no. Papers from the 6th International Conference on Biomedical Engineering and Biotechnology (iCBEB2017), 17–20 October 2017, Guangzhou, China, pp. 193–203, 2018, ISSN: 0928-7329.

[88] G. N. Lewis, W. D. Byblow, and S. E. Walt, "Stride length regulation in parkinson's disease: The use of extrinsic, visual cues," *Brain*, vol. 123, no. 10, pp. 2077–2090, 2000.

[89] F. Liu, X. Xu, H. Zhu, Y. Zhang, J. Yang, L. Zhang, N. Li, L. Zhu, H. F. Kung, and Z. Yang, "PET imaging of 18f-(2s, 4r)4-fluoroglutamine accumulation in breast cancer: From xenografts to patients," *Molecular Pharmaceutics*, vol. 15, no. 8, pp. 3448–3455, 2018.

[90] 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," *PLoS ONE*, vol. 7, no. 11, e47714, 2012.

[91] LONI, *Neuroscience Datasets*, `https://ida.loni.usc.edu/login.jsp`, [Online; Accessed Aug. 6, 2019].

[92] E. D. Louis and J. J. Ferreira, "How common is the most common adult movement disorder? update on the worldwide prevalence of essential tremor," *Movement Disorders*, vol. 25, no. 5, pp. 534–541, 2010.

[93] *Magnetic resonance imaging (mri) of the brain and spine: Basics*, `https://casemed.case.edu/`, Accessed: 2019-08-08.

[94] F. Magrinelli, A. Picelli, P. Tocco, A. Federico, L. Roncari, N. Smania, G. Zanette, and S. Tamburin, "Pathophysiology of motor dysfunction in parkinson's disease as the rationale for drug treatment and rehabilitation," *Parkinson's disease*, vol. 2016, 2016.

[95] P. Mahlknecht, F. Krismer, W. Poewe, and K. Seppi, "Meta-analysis of dorsolateral nigral hyperintensity on magnetic resonance imaging as a marker for parkinson's disease," *Movement Disorders*, vol. 32, no. 4, pp. 619–623, 2017.

[96] J. A. Maintz and M. A. Viergever, "An overview of medical image registration methods," in *Symposium of the Belgian hospital physicists association (SBPH/BVZF)*, Citeseer, vol. 12, 1996, pp. 1–22.

[97] E. Mak, L. Su, G. B. Williams, M. J. Firbank, R. A. Lawson, A. J. Yarnall, G. W. Duncan, B. Mollenhauer, A. M. Owen, T. K. Khoo, *et al.*, "Longitudinal whole-brain atrophy and ventricular enlargement in nondemented parkinson's disease," *Neurobiology of aging*, vol. 55, pp. 78–90, 2017.

[98] C. Marras, J. Beck, J. Bower, E. Roberts, B. Ritz, G. Ross, R. Abbott, R. Savica, S. Van Den Eeden, A. Willis, *et al.*, "Prevalence of parkinson's disease across north america," *NPJ Parkinson's disease*, vol. 4, no. 1, p. 21, 2018.

[99] S. Mazilu, A. Calatroni, E. Gazit, D. Roggen, J. M. Hausdorff, and G. Troster, "Feature learning for detection and prediction of freezing of gait in parkinson's disease," in *International workshop on machine learning and data mining in pattern Recognition*, Springer, 2013, pp. 144–158.

[100] S. Mazilu, M. Hardegger, Z. Zhu, D. Roggen, G. Troster, M. Plotnik, and J. M. Hausdorff, "Online detection of freezing of gait with smartphones and machine learning techniques," in *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, IEEE, 2012, pp. 123–130.

[101] A. Mobini, S. Behzadipour, and M. Saadat, "Test-retest reliability of kinect's measurements for the evaluation of upper body recovery of stroke patients," *Biomedical engineering online*, vol. 14, no. 1, p. 75, 2015.

[102] S. T. Moore, H. G. MacDougall, and W. G. Ondo, "Ambulatory monitoring of freezing of gait in parkinson's disease," *Journal of neuroscience methods*, vol. 167, no. 2, pp. 340–348, 2008.

[103] S. T. Moore, D. A. Yungher, T. R. Morris, V. Dilda, H. G. MacDougall, J. M. Shine, S. L. Naismith, and S. J. Lewis, "Autonomous identification of freezing of gait in parkinson's disease from lower-body segmental accelerometry," *Journal of neuroengineering and rehabilitation*, vol. 10, no. 1, p. 19, 2013.

[104] S. Morris, M. E. Morris, and R. Iansek, "Reliability of measurements obtained with the timed "up & go" test in people with parkinson disease," *Physical therapy*, vol. 81, no. 2, pp. 810–818, 2001.

[105] H. Mousavi Hondori and M. Khademi, "A review on technical and clinical impact of microsoft kinect on physical therapy and rehabilitation," *Journal of medical engineering*, vol. 2014, 2014.

[106] *movement disorders assessment and treatment*, `https://jfkmc.org/`, Online; accessed 30 September 2018.

[107] A. Muro-De-La-Herran, B. Garcia-Zapirain, and A. Mendez-Zorrilla, "Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications," *Sensors*, vol. 14, no. 2, pp. 3362–3394, 2014.

[108] B. Mwangi, T. S. Tian, and J. C. Soares, "A review of feature reduction techniques in neuroimaging," *Neuroinformatics*, vol. 12, no. 2, pp. 229–244, 2014.

[109] J. R. Nocera, E. L. Stegemoller, I. A. Malaty, M. S. Okun, M. Marsiske, C. J. Hass, N. P. F. Q. I. I. Investigators, *et al.*, "Using the timed up & go test in a clinical setting to predict falling in parkinson's disease," *Archives of physical medicine and rehabilitation*, vol. 94, no. 7, pp. 1300–1305, 2013.

[110] J. G. Nutt, B. R. Bloem, N. Giladi, M. Hallett, F. B. Horak, and A. Nieuwboer, "Freezing of gait: Moving forward on a mysterious clinical phenomenon," *The Lancet Neurology*, vol. 10, no. 8, pp. 734–744, 2011.

[111] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," in *Proceedings of 12th International Conference on Pattern Recognition*, IEEE, vol. 1, 1994, pp. 582–585.

[112] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[113] Y. Okuma, "Freezing of gait and falls in parkinson's disease," *Journal of Parkinson's disease*, vol. 4, no. 2, pp. 255–260, 2014.

[114] K. Oppedal, T. Eftestol, K. Engan, M. K. Beyer, and D. Aarsland, "Classifying dementia using local binary patterns from different regions in magnetic resonance images," *Journal of Biomedical Imaging*, vol. 2015, p. 5, 2015.

[115] H. Ozkan, "A comparison of classification methods for telediagnosis of parkinson's disease," *Entropy*, vol. 18, no. 4, p. 115, 2016.

[116] R. Pahwa and K. E. Lyons, "Early diagnosis of parkinson's disease: Recommendations from diagnostic clinical guidelines," *The American Journal of Managed Care*, vol. 16, S94–9, 2010.

[117] S. Pan, S. Iplikci, K. Warwick, and T. Z. Aziz, "Parkinson's disease tremor classification–a comparison between support vector machines and neural networks," *Expert Systems with Applications*, vol. 39, no. 12, pp. 10 764–10 771, 2012.

[118] *Parkinson's Progression Markers Initiative (PPMI)*, `https://www.ppmi-info.org/`, Online; accessed 30 September 2019.

[119] S. Patel, B.-r. Chen, C. Mancinelli, S. Paganoni, L. Shih, M. Welsh, J. Dy, and P. Bonato, "Longitudinal monitoring of patients with parkinson's disease via wearable sensor technology in the home setting," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, 2011, pp. 1552–1555.

[120] M. Peker, B. Sen, and D. Delen, "Computer-aided diagnosis of parkinson's disease using complex-valued neural networks and mrmr feature selection algorithm," *Journal of healthcare engineering*, vol. 6, no. 3, pp. 281–302, 2015.

[121] E. Pelosin, L. Avanzino, M. Bove, P. Stramesi, A. Nieuwboer, and G. Abbruzzese, "Action observation improves freezing of gait in patients with parkinson's disease," *Neurorehabilitation and neural repair*, vol. 24, no. 8, pp. 746–752, 2010.

[122] B. Peng, S. Wang, Z. Zhou, Y. Liu, B. Tong, T. Zhang, and Y. Dai, "A multilevel-ROI-features-based machine learning method for detection of morphometric biomarkers in parkinson's disease," *Neuroscience Letters*, vol. 651, pp. 88–94, 2017.

[123] C. R. Pereira, S. A. Weber, C. Hook, G. H. Rosa, and J. P. Papa, "Deep learning-aided parkinson's disease diagnosis from handwritten dynamics," in *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Ieee, 2016, pp. 340–346.

[124] S. V. Perumal and R. Sankar, "Gait and tremor assessment for patients with parkinson's disease using wearable sensors," *Ict Express*, vol. 2, no. 4, pp. 168–174, 2016.

[125] J. A. K. Polonen, "Mri shape analysis of midbrain in parkinson's disease," Master's thesis, 2016.

[126] R. Prashanth, S. D. Roy, P. K. Mandal, and S. Ghosh, "Automatic classification and prediction models for early parkinson's disease diagnosis from spect imaging," *Expert Systems with Applications*, vol. 41, no. 7, pp. 3333–3342, 2014.

[127] S. Qiu, Z. Wang, H. Zhao, L. Liu, and Y. Jiang, "Using body-worn sensors for preliminary rehabilitation assessment in stroke victims with gait impairment," *IEEE Access*, vol. 6, pp. 31 249–31 258, 2018.

[128] C. Ramaker, J. Marinus, A. M. Stiggelbout, and B. J. Van Hilten, "Systematic evaluation of rating scales for impairment and disability in parkinson's disease," *Movement disorders: official journal of the Movement Disorder Society*, vol. 17, no. 5, pp. 867–876, 2002.

[129] B. Rana, A. Juneja, M. Saxena, S. Gudwani, S. S. Kumaran, R. K. Agrawal, and M. Behari, "Voxel-based morphometry and minimum redundancy maximum relevance method for classification of parkinson's disease and controls from t1-weighted mri," in *Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing*, ser. ICVGIP '16, Guwahati, Assam, India: ACM, 2016, 22:1–22:6, ISBN: 978-1-4503-4753-2.

[130] B. Rana, A. Juneja, M. Saxena, S. Gudwani, S. S. Kumaran, M. Behari, and R. Agrawal, "Relevant 3d local binary pattern based features from fused feature descriptor for differential diagnosis of parkinson's disease using structural mri," *Biomedical Signal Processing and Control*, vol. 34, pp. 134–143, 2017.

[131] B. Rana, A. Juneja, M. Saxena, S. Gudwani, S. S. Kumaran, M. Behari, and R. K. Agrawal, "Graph-theory-based spectral feature selection for computer aided diagnosis of parkinson's disease using t1-weighted MRI," *International Journal of Imaging Systems and Technology*, vol. 25, no. 3, pp. 245–255, 2015.

[132] O. Rascol, P. Payoux, F. Ory, J. J. Ferreira, C. Brefel-Courbon, and J.-L. Montastruc, "Limitations of current parkinson's disease therapy," *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, vol. 53, no. S3, S3–S15, 2003.

[133] M. D. S. T. F. on Rating Scales for Parkinson's Disease, "The unified parkinson's disease rating scale (updrs): Status and recommendations," *Movement Disorders*, vol. 18, no. 7, pp. 738–750, 2003.

[134] A. Reeve, E. Simcox, and D. Turnbull, "Ageing and parkinson's disease: Why is advancing age the biggest risk factor?" *Ageing research reviews*, vol. 14, pp. 19–30, 2014.

[135] E. Reiter, C. Mueller, B. Pinter, F. Krismer, C. Scherfler, R. Esterhammer, C. Kremser, M. Schocke, G. K. Wenning, W. Poewe, *et al.*, "Dorsolateral nigral hyperintensity on 3.0 t susceptibility-weighted imaging in neurodegenerative parkinsonism," *Movement Disorders*, vol. 30, no. 8, pp. 1068–1076, 2015.

[136] A. P. Rocha, H. Choupina, J. M. Fernandes, M. J. Rosas, R. Vaz, and J. P. S. Cunha, "Kinect v2 based system for parkinson's disease assessment," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2015, pp. 1279–1282.

[137] D. Rodriguez-Martin, A. Sama, C. Perez-Lopez, A. Catala, J. M. M. Arostegui, J. Cabestany, A. Bayes, S. Alcaine, B. Mestre, A. Prats, *et al.*, "Home detection of freezing of gait using support vector machines through a single waist-worn triaxial accelerometer," *PloS one*, vol. 12, no. 2, e0171764, 2017.

[138] I. Rustempasic and M. Can, "Diagnosis of parkinson's disease using fuzzy c-means clustering and pattern recognition," *Southeast Europe Journal of Soft Computing*, vol. 2, no. 1, 2013.

[139] I. Rustempasic and M. Can, "Diagnosis of parkinson's disease using principal component analysis and boosting committee machines," *Southeast Europe journal of soft computing*, vol. 2, no. 1, 2013.

[140] U. Saeed, J. Compagnone, R. I. Aviv, A. P. Strafella, S. E. Black, A. E. Lang, and M. Masellis, "Imaging biomarkers in parkinson's disease and parkinsonian syndromes: Current and emerging concepts," *Translational neurodegeneration*, vol. 6, no. 1, p. 8, 2017.

[141] C. Salvatore, A. Cerasa, I. Castiglioni, F. Gallivanone, A. Augimeri, M. Lopez, G. Arabia, M. Morelli, M. Gilardi, and A. Quattrone, "Machine learning on brain MRI data for differential diagnosis of parkinson's disease and progressive supranuclear palsy," *Journal of Neuroscience Methods*, vol. 222, pp. 230–237, 2014.

[142] A. Samà, D. Rodrıguez-Martın, C. Pérez-López, A. Català, S. Alcaine, B. Mestre, A. Prats, M. C. Crespo, and A. Bayes, "Determining the optimal features in freezing of gait detection through a single waist accelerometer in home environments," *Pattern Recognition Letters*, vol. 105, pp. 135–143, 2018.

[143] A. Sarica, A. Cerasa, and A. Quattrone, "Random forest algorithm for the classification of neuroimaging data in alzheimer's disease: A systematic review," *Frontiers in aging neuroscience*, vol. 9, p. 329, 2017.

[144] B. K. Scanlon, H. L. Katzen, B. E. Levin, C. Singer, and S. Papapetropoulos, "A formula for the conversion of updrs-iii scores to hoehn and yahr stage," *Parkinsonism & related disorders*, vol. 14, no. 4, pp. 379–380, 2008.

[145] S. T. Schwarz, M. Afzal, P. S. Morgan, N. Bajaj, P. A. Gowland, and D. P. Auer, "The 'swallow tail' appearance of the healthy nigrosome–a new accurate test of parkinson's disease: A case-control and retrospective cross-sectional mri study at 3t," *PloS one*, vol. 9, no. 4, e93814, 2014.

[146] K. Seppi, D. Weintraub, M. Coelho, S. Perez-Lloret, S. H. Fox, R. Katzenschlager, E.-M. Hametner, W. Poewe, O. Rascol, C. G. Goetz, *et al.*, "The movement disorder society evidence-based medicine review update: Treatments for the non-motor symptoms of parkinson's disease," *Movement Disorders*, vol. 26, no. S3, S42–S80, 2011.

[147] S. Shinde, S. Prasad, Y. Saboo, R. Kaushick, J. Saini, P. K. Pal, and M. Ingalhalikar, "Predictive markers for parkinson's disease using deep neural nets on neuromelanin sensitive mri," *NeuroImage: Clinical*, vol. 22, p. 101 748, 2019.

[148] S. Sivaranjini and C. Sujatha, "Deep learning based diagnosis of parkinson's disease using convolutional neural network," *Multimedia Tools and Applications*, pp. 1–13, 2019.

[149] S. Soltaninejad, I. Cheng, and A. Basu, "Robust lung segmentation combining adaptive concave hulls with active contours," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE, 2016, pp. 004 775–004 780.

[150] S. Soltaninejad, I. Cheng, and A. Basu, "Towards the identification of parkinson's disease using only t1 mr images," in *International Conference on Smart Multimedia*, Springer, 2018, pp. 145–156.

[151] S. Soltaninejad, A. Rosales-Castellanos, F. Ba, M. A. Ibarra-Manzano, and I. Cheng, "Body movement monitoring for parkinson's disease patients using a smart sensor based non-invasive technique," in *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*, IEEE, 2018, pp. 1–6.

[152] *SPM*, `https://www.fil.ion.ucl.ac.uk/spm/software/`, Online; accessed 30 September 2019.

[153] M. Sreejith, S. Rakesh, S. Gupta, S. Biswas, and P. P. Das, "Real-time hands-free immersive image navigation system using microsoft kinect 2.0 and leap motion controller," in *2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, IEEE, 2015, pp. 1–4.

[154] A. Stezin, R. M. Naduthota, R. Botta, S. Varadharajan, A. Lenka, J. Saini, R. Yadav, and P. K. Pal, "Clinical utility of visualisation of nigrosome-1 in patients with parkinson's disease," *European radiology*, vol. 28, no. 2, pp. 718–726, 2018.

[155] A. Tay, S. Yen, P. Lee, C. Wang, A. Neo, S. Phan, K. Yogaprakash, S. Liew, and W. Au, "Freezing of gait (fog) detection for parkinson disease," in *2015 10th Asian Control Conference (ASCC)*, IEEE, 2015, pp. 1–6.

[156] E. E. Tripoliti, A. T. Tzallas, M. G. Tsipouras, G. Rigas, P. Bougia, M. Leontiou, S. Konitsiotis, M. Chondrogiorgi, S. Tsouli, and D. I. Fotiadis, "Automatic detection of freezing of gait events in patients with parkinson's disease," *Computer methods and programs in biomedicine*, vol. 110, no. 1, pp. 12–26, 2013.

[157] K. M. Tsiouris, G. Rigas, A. Antonini, D. Gatsios, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "Mining motor symptoms updrs data of parkinson's disease patients for the development of hoehn and yahr estimation decision support system," in *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, IEEE, 2017, pp. 445–448.

[158] P. Tuite, "Brain magnetic resonance imaging (MRI) as a potential biomarker for parkinson's disease (PD)," *Brain Sciences*, vol. 7, no. 12, p. 68, 2017.

[159] R. Velik, U. Hoffmann, H. Zabaleta, J. F. M. Masso, and T. Keller, "The effect of visual cues on the number and duration of freezing episodes in parkinson's patients," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, 2012, pp. 4656–4659.

[160] B. S. Vidya and E. Chandra, "Entropy based local binary pattern (elbp) feature extraction technique of multimodal biometrics as defence mechanism for cloud storage," *Alexandria Engineering Journal*, vol. 58, no. 1, pp. 103–114, 2019.

[161] T. Vos, C. Allen, M. Arora, R. M. Barber, Z. A. Bhutta, A. Brown, A. Carter, D. C. Casey, F. J. Charlson, A. Z. Chen, *et al.*, "Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: A systematic analysis for the global burden of disease study 2015," *The Lancet*, vol. 388, no. 10053, pp. 1545–1602, 2016.

[162] S.-H. Wang, P. Phillips, Y. Sui, B. Liu, M. Yang, and H. Cheng, "Classification of alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling," *Journal of medical systems*, vol. 42, no. 5, p. 85, 2018.

[163] Z. Wang, J. Li, J. Wang, H. Zhao, S. Qiu, N. Yang, and X. Shi, "Inertial sensor-based analysis of equestrian sports between beginner and professional riders under different horse gaits," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 11, pp. 2692–2704, 2018.

[164] *Who has parkinson's?* `https://www.parkinson.org/Understanding-Parkinsons/Statistics`, Accessed: 2019-08-15.

[165] K. Wirdefeldt, H.-O. Adami, P. Cole, D. Trichopoulos, and J. Mandel, "Epidemiology and etiology of parkinson's disease: A review of the evidence," *European journal of epidemiology*, vol. 26, no. 1, p. 1, 2011.

[166] J. B. Wood, "Mr imaging of parkinsonism," in *Parkinson's Disease*, CRC Press, 2012, pp. 762–777.

[167] M. W. Woolrich, S. Jbabdi, B. Patenaude, M. Chappell, S. Makni, T. Behrens, C. Beckmann, M. Jenkinson, and S. M. Smith, "Bayesian analysis of neuroimaging data in fsl," *Neuroimage*, vol. 45, no. 1, S173–S186, 2009.

[168] A. Worker, C. Blain, J. Jarosz, K. R. Chaudhuri, G. J. Barker, S. C. Williams, R. Brown, P. N. Leigh, and A. Simmons, "Cortical thickness, surface area and volume measures in parkinson's disease, multiple system atrophy and progressive supranuclear palsy," *PloS one*, vol. 9, no. 12, e114167, 2014.

[169] Y. Xia, J. Zhang, Q. Ye, N. Cheng, Y. Lu, and D. Zhang, "Evaluation of deep convolutional neural networks for detection of freezing of gait in parkinson's disease patients," *Biomedical Signal Processing and Control*, vol. 46, pp. 221–230, 2018.

[170] A. J. Yarnall, D. P. Breen, G. W. Duncan, T. K. Khoo, S. Y. Coleman, M. J. Firbank, C. Nombela, S. Winder-Rhodes, J. R. Evans, J. B. Rowe, *et al.*, "Characterizing mild cognitive impairment in incident parkinson disease: The icicle-pd study," *Neurology*, vol. 82, no. 4, pp. 308–316, 2014.

[171] H. Zach, A. M. Janssen, A. H. Snijders, A. Delval, M. U. Ferraye, E. Auff, V. Weerdesteyn, B. R. Bloem, and J. Nonnekes, "Identifying freezing of gait in parkinson's disease during freezing provoking tasks using waist-mounted accelerometry," *Parkinsonism & related disorders*, vol. 21, no. 11, pp. 1362–1366, 2015.

[172] J. Zhang, W. Xu, Q. Zhang, B. Jin, and X. Wei, "Exploring risk factors and predicting updrs score based on parkinson's speech signals," in *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*, IEEE, 2017, pp. 1–6.

[173] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm," *IEEE transactions on medical imaging*, vol. 20, no. 1, pp. 45–57, 2001.

[174] Y. Zhao, K. Tonn, K. Niazmand, U. M. Fietzek, L. T. D'Angelo, A. Ceballos-Baumann, and T. C. Lueth, "Online fog identification in parkinson's disease with a time-frequency combined algorithm," in *Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics*, IEEE, 2012, pp. 192–195.

[175] Y. Zhao, Z. S.-Y. Wong, and K. L. Tsui, "A framework of rebalancing imbalanced healthcare data for rare events' classification: A case of look-alike sound-alike mix-up incident detection," *Journal of healthcare engineering*, vol. 2018, 2018.

[176] D. A. Ziegler and S. Corkin, "New magnetic resonance imaging biomarkers advance the characterization of parkinson disease.," *US Neurology*, vol. 9, no. 1, pp. 8–12, 2013.

[177] L. Zou, J. Zheng, C. Miao, M. J. Mckeown, and Z. J. Wang, "3d cnn based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural mri," *IEEE Access*, vol. 5, pp. 23 626–23 636, 2017.