Parkinson's Disease Detection Using Structural MRI

Özkan Çiğdem

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Electrical and Electronic Engineering

Eastern Mediterranean University
September 2018
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Assoc. Prof. Dr. Ali Hakan Ulusoy
Acting Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy in Electrical and Electronic Engineering.

Prof. Dr. Hasan Demirel
Chair, Department of Electrical and Electronic Engineering

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Doctor of Philosophy in Electrical and Electronic Engineering.

Prof. Dr. Hasan Demirel
Supervisor

Examining Committee

1. Prof. Dr. Aydın Akan
2. Prof. Dr. Hasan Demirel
3. Prof. Dr. Osman Eroğul
4. Prof. Dr. Erhan A. İnce
5. Assoc. Prof. Dr. Önsen Toygar

ABSTRACT

Parkinson's Disease (PD) is the second most encountered neurodegenerative disorder, second only to Alzheimer's Disease (AD), and the most common movement disorder affecting 1% of people over the age of 60. PD is characterized by progressive loss of muscle control that causes trembling of the limbs and head at rest position, rigidity, slowness, impaired balance, and later on a shuffling gait. As the disease is progressed, difficulties in walking, talking, and completing basic tasks might occur. The causes of PD are unknown, yet it is believed that both the environmental and genetic factors might lead to PD. High-quality images obtained using neuroimaging methods could give beneficial support to the clinicians for evaluating the treatments. Three-dimensional magnetic resonance imaging (3D-MRI) has been effectively utilized in the detection of progressive neurodegenerative diseases including PD. Therefore, using neuroimaging techniques with Computer-Aided Diagnosis (CAD) has gained increasing attention in the early and accurate diagnosis of PD. In this thesis, the extensive reviews on the studies of PD detection using MRI data and CAD methods since 2008 are studied. Furthermore, the affected brain regions owing to PD are obtained by using the 3D Volume of Interests (VOIs) and the captured affected brain regions are compared with the regions reported in the-state-of-the-art studies in the last decade. The obtained affected brain regions might shed light on the existing literature on PD diagnosis.

In order to build an automatic method for PD detection, machine learning algorithms are applied to the processed neuroimaging data. It is obvious that pre-processing of 3D-MRI scans plays an important role for post-processing. In this thesis, the

preprocessing of 3D-MRI data has been performed by using a Voxel-Based Morphometry (VBM) technique which evaluates the whole brain morphology with voxel-by-voxel comparisons. In VBM, some parameters such as covariates need to be defined to build a model for Gray Matter (GM) and White Matter (WM) volumes of Structural MRI (sMRI) datasets. In this thesis, the effects of using different covariates (i.e. total intracranial volume, age, sex and combination of them) on the classification of PD groups from Healthy Controls (HCs) have been studied. Additionally, in order to determine the 3D VOIs, the significant local alterations in GM and WM volumes of PD groups and HCs, a hypothesis either f-contrast or t-contrast need to be defined. In this thesis, the effects of two different hypotheses on PD detection have been investigated. Furthermore, a feature-level fusion technique in which the 3D GM and WM VOIs are combined considering the effects of both GM and WM volumes in PD diagnosis.

The voxels extracted from 3D GM, WM, and the combination of GM and WM VOIs are considered as raw features. Even though the feature extraction decreases the number of features in raw data, using an automatic feature selection method from high dimensional feature space is an asset in PD classification. In this thesis, to select the most discriminative attributes from high-dimensional data, all raw features are ranked by using various feature ranking methods such as minimum redundancy maximum relevance, Relief-F, unsupervised feature selection for multi-cluster data, Laplacian score, regularized discriminative feature selection for unsupervised learning, correlation- based feature selection, and feature selection and kernel learning for local learning-based clustering. In order to select the optimal number of top-ranked discriminative features, a Fisher Criterion (FC) is calculated for different sizes of

feature vectors and the optimal number of top-ranked features is selected when the vector size maximizes the FC. In order to classify the PD and HC, five different classification algorithms, namely k- nearest neighbor, naive Bayes, ensemble bagged trees, ensemble subspace discriminant, and support vector machines are used. Moreover, a decision fusion technique which combines the binary outputs of all five classifiers by using a majority voting method is investigated to achieve higher performance in PD diagnosis. The experimental results indicate that the proposed methods are reliable approaches that are highly competitive with the state-of-the-art methods in PD classification.

Keywords: Parkinson's disease, structural MRI, covariates, f-contrast, voxel-based morphometry.

ÖZ

Parkinson Hastalığı (PH), Alzheimer Hastalığından (AH) sonra en çok karşılaşılan ikinci norodejeneratif hastalıktır ve 60 yaş üstü insanların %1'ini etkileyen en yaygın hareket bozukluğudur. PH, dinlenme pozisyonundayken bacaklarda ve kafada titremeye, sertliğe, hareketlerde yavaşlamaya, denge bozukluğuna ve daha sonra ayakları yere sürüyerek yürümeye sebep olan zamanla ilerleyen kas kontrolünün kaybıyla tanımlanmaktadır. Hastalık ilerledikçe, yürüme, konuşma ve temel ihtiyaçları giderme zorlukları ortaya çıkabilir. PH'nin sebepleri bilinmemektedir, ancak hem çevresel hem de genetik faktörlerin hastalığa sebep olabileceğine inanılmaktadır. Nöro-görüntüleme yöntemleri kullanılarak elde edilen yüksek kaliteli görüntüler, klinisyenlere tedaviyi değerlendirirken yararlı desteği sunabilirler. Üç boyutlu manyetik rezonans görüntüleme (3B-MRG), PH'yi de içeren zamanla ilerleyen nörodejeneratif hastalıkların teşhisinde etkili bir şekilde kullanılmaktadır. Bu nedenle, PH'nin erken ve kesin teşhisinde, nöro-görüntüleme tekniklerinin bilgisayar destekli tanı ile kullanılması giderek dikkat çekmeye başlamıştır. Bu tezde, PH'nin sebep olduğu etkilenmiş beyin kısımları, 3B ilgili vokseller kullanılarak elde edilmiştir ve elde edilen bu etkilenmiş beyin bölgeleri son on yılda çalışılmış modern yöntemlerden çıkarılan bölgelerle karşılaştırılmıştır. Ayrıca, 2008'den bugüne MRG verileri ve bilgisayar destekli tanı yöntemleri kullanarak, PH teşhisi üzerine yapılan çalışmalar detaylı bir şekilde incelenmiştir. Elde edilen PH tarafından etkilenmiş beyin bölgeleri PH teşhisinde mevcut literatüre ışık tutabilir.

PH teşhisinde otomatik bir metot oluşturmak için, otomatik öğrenme algoritmaları işlenmiş nöro-görüntü verilerine uygulanır. 3B-MRG taramalarının on işlemesinin,

son işleme üzerinde önemli bir rol oynadığı açıkça görülmektedir. Bu tezde, 3B-MRG verilerinin ön işlemesi, tüm beyin morfolojisini voksel-voksel karşılaştırarak değerlendiren Voksel Tabanlı Morfometri (VTM) tekniği kullanılarak yapılmıştır. VTM tekniğinde, yapısal MRG verisetlerinin gri madde (GM) ve beyaz madde (BM) hacimleri için model oluşturulmasında, kovaryant gibi bazı parametrelerin tanımlanması gerekmektedir. Bu tezde, PH'ye sahip hastalar ile Sağlıklı Bireylerin (SB'lerin) sınıflandırılmasında, farklı kovaryantların (toplam intrakraniyal hacim, yaş, cinsiyet ve bunların kombinasyonu) kullanımının etkileri çalışılmıştır. Ayrıca, PH hastaları ile SB'lerin, GM ve BM hacimleri arasındaki önemli lokal farklılıklar olarak da bilinen 3B ilgili hacimleri belirlemek için, t-kontrast veya f-kontrast olabilen bir hipotez tanımlamak gerekmektedir. Bu tezde, bahsedilen iki farklı hipotezin, PH teşhisi üzerindeki etkileri incelenmiştir. Buna ek olarak, PH teşhisinde GM ve BM hacimlerini birlikte değerlendiren, 3B GM ve BM birleşimi olarak tanımlanan bir kaynak birleştirme tekniği kullanılmıştır.

3B GM, BM ve bu iki hacmin kombinasyonundan elde edilmiş ilgili vokseller, işlenmemiş özellikler olarak ele alınmıştır. Öznitelik bulma yöntemi işlenmemiş verideki özellik sayısını azaltıyor olsa bile yüksek boyutlu öznitelik uzayından, öznitelikleri otomatik olarak seçme yöntemi kullanmak PH teşhisinde olması gereken bir süreçtir. Bu tezde, yüksek boyutlu veriden, en ayırt edici özellikleri seçmek için işlenmemiş tüm özellikler en az artıklık en çok ilgililer, Relief-F, çoklu-kümeli veri için denetlenmemiş öznitelik seçme, Laplas skoru, denetlenmemiş öznitelik öğrenme için düzenlenmiş ayırt edici öznitelik seçme, ilinti tabanlı öznitelik seçme ve lokal öğrenme tabanlı kümeleme için (öznitelik seçme ve çekirdek öğrenme gibi farklı öznitelik sıralama yöntemleri kullanılarak sıralanmıştır. En ayırt edici özellikleri

seçmek için, tüm farklı boyutlardaki özellik vektörlerinin bir Fisher kriter değeri hesaplanır ve vektör boyutunun Fisher kriterini maksimum yaptığı anda optimal sayıdaki en ayırt edici özellikler seçilmiş olur. PH grubu ile SB grubunu sınıflandırmak icin beş farklı sınıflandırma algoritması kullanıldı. Kullanılan bu algoritmalar k-en yakın komşu, Naïve Bayes, topluluk torbalı ağaç, topluluk altuzay ayırtaç ve destek vektor makineleridir. Buna ek olarak, PH teşhisinde daha iyi bir performans elde etmek amacıyla bahsedilen beş farklı sınıflandırma algoritmalarının ikili çıktılarını çoğunluk onayı yöntemi kullanılarak birleştiren bir karar birleştirme tekniği kullanılmıştır. Deneysel sonuçlar, PH sınıflandırması için bu tezde önerilen yöntemlerin modern yöntemlerle ciddi oranda rekabet edebildiğini göstermiştir.

Anahtar Kelimeler: Parkinson hastalığı, yapısal MRI, kovaryant, f- kontrast, voksel tabanlı morfometri.

DEDICATION

To all kids in the world.

ACKNOWLEDGMENT

I would like to express my heartfelt appreciation to my supervisor Prof. Hasan Demirel who gave me the opportunity to work with him. I have learned a lot from him and he has had a great influence on me. I would like to extend my gratitude to the committee members for their academic guidance. Additionally, I would like to thank Eastern Mediterranean University, KU Leuven, and Dokuz Eylül University for giving me the opportunity to conduct this research.

I thank Dr. Iman Beheshti, Dr. Faezeh Yeganli, Dr. Faegeh Yeganli, Arif Yılmaz, Ozan Özgür Özgün, and Bermal Harmancı who have provided me with irreplaceable guidance and friendship during the composition of my thesis.

I would also like to thank Sadık Ulusoy and Ada Mina Sarı for doing a lot for me in both my life and my thesis by supporting and encouraging me at any time, and to my precious nieces and nephews Toprak Çağan, Irmak, Elifsu, Nujen, Aysu, and Serok who have caused nothing but joy in my life.

Last but not least, I would like to express my utter indebtedness to my family, who has given me endless love and guidance, throughout the entire course of my academic career. There are no words with which I could properly express my appreciation of their steadfast support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

3D-MRI	Three-Dimensional Magnetic Resonance Imaging
β	Estimated Parameter
λ	Contrast
γ	Kernel Width
μ	Mean
σ	Standard Deviation
C	Regularization
e	Additive Noise
K	Fold
P_x	Projector onto X
R_x	Orthogonal Projector onto X
s	Standard Deviation
S_B	Between-Class Scatter Matrix
S_W	Within-Class Scatter Matrix
X	Design Matrix
var	Variance
AAL	Automated Anatomical Labeling
AC	Anterior Commissural
ACC	Accuracy
AD	Alzheimer's Disease
ARS	Age-Related Subgroups
AUC	Area Under the Curve
AUG	Age-Unrelated Subgroups

BG	Basal Ganglia
CAD	Computer Aided Diagnosis/Detection
CAT	Computational Anatomy Toolbox
CCA	Canonical Correlation Analysis
CDT	Cauchy Deformation Tensor
CFS	Correlation-Based Feature Selection
CSD	Constrained Spherical Deconvolution
CSF	Cerebra Spinal Fluid
CT	Computed Tomography
CV	Cross Validation
DAT	Dopamine Transporter
DBM	Deformation Based Morphometry
DICOM	Digital Imaging and Communications in Medicine
dMRI	Diffusion MRI
DT	Decision Trees
DTI	Diffusion Tensor Imaging
EAP	Ensemble Average Propagator
EBT	Ensemble-Bagged Trees
ELM	Extreme Learning Machine
EPI	Echo-Planar Imaging
ESD	Ensemble-Subspace Discriminant
F	Female
FA	Fractional Anisotropy
FBM	Feature-Based Morphometry
FC	Fisher Criterion

FDR	Fisher Discriminant Ratio
FFD	Fused Feature Descriptor
FL	Frontal Lobe
fMRI	Functional MRI
FN	False Negative
FP	False Positive
FS	Feature Selection
FSL	FMRIB Software Library
FSS	Feature Subset Selection
FWHM	Full-Width-Half-Maximum
GA	Genetic Algorithm
GLM	General Linear Model
GM	Grey Matter
HC	Healthy Control
HO	HouldOut
HOG	Histogram of Oriented Gradient
ICA	Independent Component Analysis
ICC	Intraclass Correlation Coefficient
ICCA	Iterative Canonical Correlation Analysis
IPD	Idiopathic Parkinson's Disease
IPS	Idiopathic Parkinson Syndrome
JFSS	Joint Feature-Sample Selection
JFSS-C	Joint Feature-Sample Selection as a Classifier
LDA	Linear Discriminant Analysis
LLCFS	Kernel Learning For Local Learning-Based Clustering

LOOCV	Leave-One-Out Cross-Validation
LS	Laplacian Score
LSSVM	Least Squares Support Vector Machine
M	Male
MCFS	Unsupervised Feature Selection For Multi-Cluster Data
McRBFN	Meta-Cognitive Radial Basis Function Network
MNI	Montreal Neurological Institute
MRI	Magnetic Resonance Imaging
mRMR	Minimum Redundancy Maximum Relevance
MSA	Multiple Systems Atrophy
MSA-C	Cerebellar Type Multiple System Atrophy
MSA-P	Parkinsonian Multiple Systems Atrophy
MSMT	Multi-Shell, Multi-Tissue
NB	Naive Bayes
NN	Nearest Neighbor
NoFS	No Feature Selection
noFSS	No Feature Sample Selection
PBL	Projection Based Learning
PC	Principal Component
PCA	Principal Component Analysis
PD	Parkinson's Disease
PDCI	Cognitively İntact Patients With Parkinson’s Disease
PDD	Parkinson’s Disease Dementia
PDF	Probability Distribution Function
PDMCI	Parkinson’s Disease With Mild Cognitive Impairment

PD-NC	Parkinson’s Disease Patients With Normal Cognition
PET	Positron Emission Tomography
PPMI	Parkinson's Progression Markers Initiative
PS	Parkinsonian Syndromes
PSP	Progressive Supranuclear Palsy
RBFNN	Radial Basis Function Neural Network
RBF-SVM	Radial Basis Function Kernel Support Vector Machine
RBM	Region/Label-Based Morphometry
rET	Essential Tremor With Rest Tremor
RF	Random Forest
RFE	Recursive Feature Elimination
RFS-LDA	Robust Feature-Sample Linear Discriminant Analysis
RLDA	Regularized Linear Discriminant Analysis
ROI	Region Of Interest
RPCA	Robust Principle Component Analysis
rsfMRI	Resting-State fMRI
SBM	Surface-Based Morphometry
SEN	Sensitivity
SFS	Sparse Feature Selection
SIFT	Scale Invariant Feature Transform
sMRI	Structural Magnetic Resonance Imaging
SN	Substantia Nigra
SOM	Self Organizing Map
SPE	Specificity
SPECT	Single Photon Emission Computed Tomography

SPM	Statistical Parameter Mapping
SR	Sparse Regression
SSNMV	Skull-Stripped and Normalized MRI Volumes
SSST	Single-Shell, Single Tissue
SVC	Support Vector Classification
SVM	Support Vector Machine
SWEDD	Scans Without Evidence Of Dopamine Deficit
SWI	Susceptibility-Weighted Imaging
TE	Echo Time
TI	Inversion Time
TIV	Total Intracranial Volume
TN	True Negative
TP	True Positive
tPD	Tremor-Dominant Parkinson’s Disease
TR	Repetition Time
UDFS	Unsupervised Discriminative Feature Selection
UPDRS	Unified Parkinson’s Disease Rating Scale
VBM	Voxel-Based Morphometry
VOI	Volume of Interest
VP	Vascular Parkinsonism
WM	White Matter

Chapter 1

INTRODUCTION

1.1 Introduction

Neurodegenerative diseases affect the neurons in the human brain. The peripheral nervous system or the central nervous system, affected due to neurodegeneration, progressively lose their structures and functions over time and finally the neuron cells die. Since the neuron cells are not regenerable, when they are damaged or ultimately die, they leave permanent damages on the human body. Even though current treatments might help to relieve the symptoms of neurodegenerative diseases, they are still incurable and no effective methods are available to slow the progression of the diseases. There are several neurodegenerative diseases studied, namely Alzheimer's Disease (AD) and other dementias, Parkinson's Disease (PD) and PD-related disorders, Huntington's disease, spinocerebellar ataxia, spinal muscular atrophy, and motor neuron diseases [3, 4, 5, 6, 7, 8]. Since the early diagnosis of the diseases is helpful to take precautions and develop required treatments, a Computer Aided Diagnosis (CAD) method has been progressively used in neurodegenerative disorder imaging [9, 10, 11]. In this thesis, the detection of PD, the second widely encountered neurodegenerative disorder following the AD, is studied by using the machine learning methods [9, 10, 11].

1.2 Parkinson's Disease

The diseases related to parkinsonism have been studied for years. Some ancient Indian medical notes as well in Western medical literature contribute descriptions of PD [12]. The symptoms of the disease were studied by various researchers, namely the rest tremor was studied by Sylvius de la Boe, festination was represented by Sauvages, and physician Galen described shaking palsy in AD175 [12]. PD was first described by James Parkinson in 1817 [13] and later refined and expanded by Jean-Martin Charcot in the mid-1800s.

For many decades, the PD was treated effectively and has been seen as a terminal illness. In order to overcome the symptoms of the disease, different types of medication have been used such as arsenic, morphine, hemlock, and cannabis [12]. In the early 1940s, neurosurgeons began to operate surgery on the part of the brains mostly on the Basal Ganglia (BG) one. However, owing to the manual risky operations performed on the brain, approximately 10% of patients passed away [14]. In the 1960s, studies identified chemical differences on the brain parts of PD patients and Healthy Control (HC). They observed that the level of dopamine was lower in PDs comparing to the HCs which cause the degeneration of nerve cells in Substantia Nigra (SN), the part of the brain as seen in Fig. 1.1. Currently, PD is the second most common neurodegenerative disorder, second only to AD, affecting 1% of people over the age of 60 [15]. The PD has mainly four symptoms, namely bradykinesia, resting tremor, rigidity, and postural instability [16]. Depression, anxiety, sensory, sleep, thinking, and behavioral problems might also be seen. These overall motor symptoms are named as parkinsonism [17, 18].

The motor symptoms of the PD lead to the death of dopamine-generating cells in the SN which locates in the BG of the cerebral, and the reasons for cell death are not clearly understandable [19].

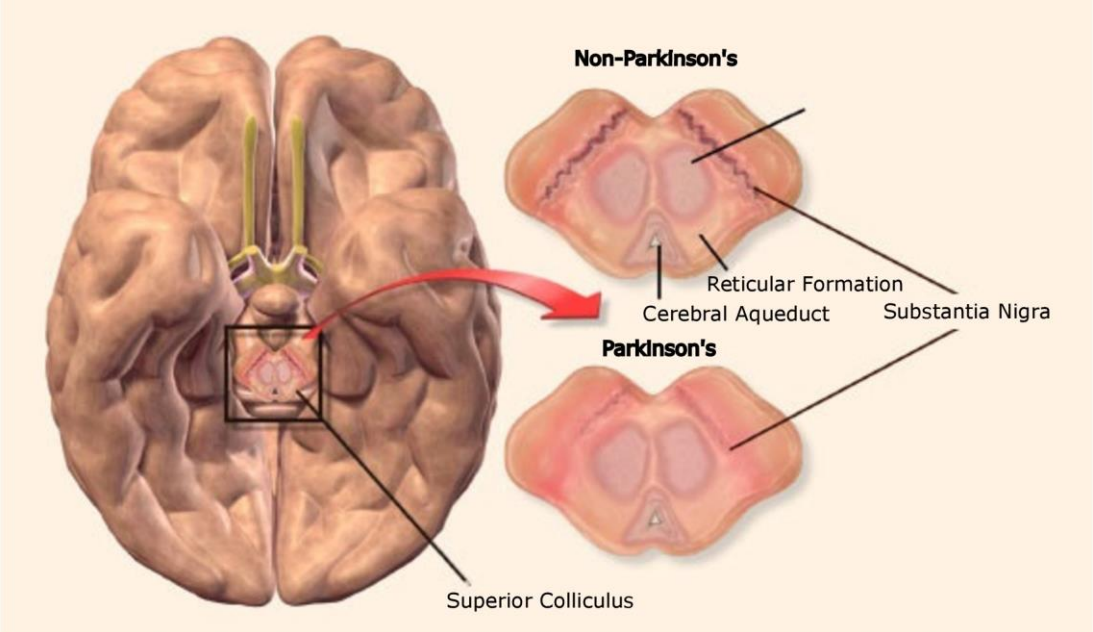

Figure 1.1: The SN affected owing to the PD [1].

PD also involves many non-motor symptoms and a group of them precede the motor dysfunction by more than a decade. The causes of having PD are unknown, yet it is believed that genetics and environmental conditions increase the risk. The PD might begin years before the diagnosis and some neuroanatomical regions of the brain might have already been damaged at the time of detection [19]. The motor impairments and the results of treatment are examined by clinicians using a standard Unified Parkinson's Disease Rating Scale (UPDRS) [16, 19]. Evaluating this scale is clinician-dependent and there is no certain method for analyzing the progression and the efficiency of the treatment. Therefore, in this thesis, the early diagnosis of PD by using neuroimaging data together with machine learning methods has been studied.

1.3 NeuroImaging Techniques

Neurodegenerative disorders characterized by the progressive deterioration of brain neurons have been extensively reported in the literature over the past decade. The aim of a neurodisease classification is to generate a model which is trained on a large pool of diseased and HCs and to predict the future progression of the disease at earlier stages by using a single neuroimaging data [20, 21]. In literature, there are different neuroimaging techniques such as Structural Magnetic Resonance Imaging (sMRI) [2], Functional MRI (fMRI) [22], Positron Emission Tomography (PET) [23], and Single Photon Emission Computed Tomography (SPECT) [24], and X-ray Computed Tomography (CT) [25]. Among these techniques, the MRI is more common than other neuroimaging techniques, since it has good contrast, high spatial resolution neuroanatomy, and no need for any pharmaceutical injections [26, 27]. In Fig. 1.2, an example of MRI data is provided. In sMRI, atrophies and physical differences are examined among different tissue types, while in fMRI, hemodynamic responses of the brain regarding neural activities have been investigated [22]. In this thesis, the sMRI data has been used for classification of PD and HC.

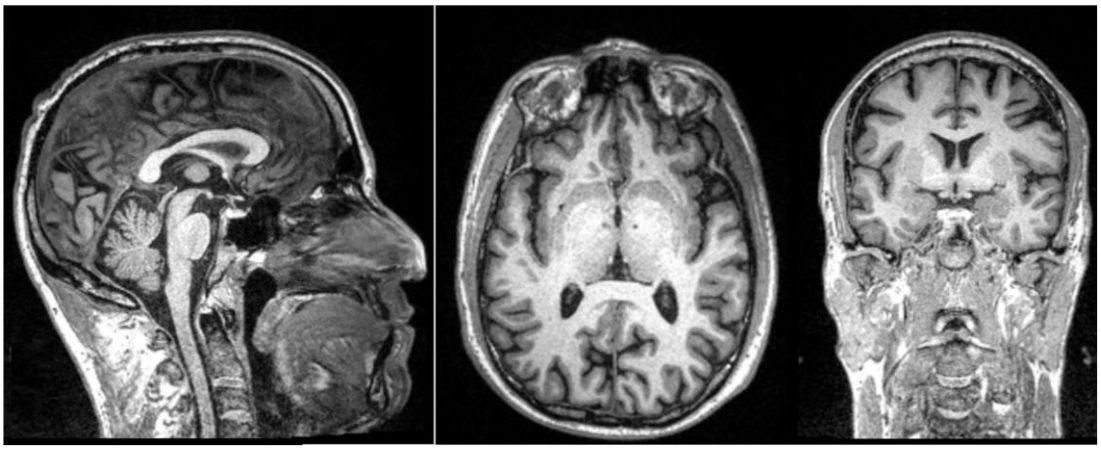

Figure 1.2: The sagittal, axial, and coronal view of an MRI from the left to the right.

1.4 Computer-Aided Detection

In neurodisease diagnosis, in order to detect the disorder by using only one MRI scan, there is a need to have a model generated from a large collection of diseased and HCs datasets [20, 21]. Even though in the AD the atrophies in the brain are clearly visible from the sMRI and might be sufficient to decide the level of the disease, in PD, the atrophies might not be sufficient. Hence, in addition to sMRI data, the clinical examinations and the medical histories of the patients are required [27, 28]. To decrease the comprehensive evaluation time, improve the accuracy of clinical tests on PD identification, and make it robust, a CAD has been progressively used in neurodegenerative disease detection [9, 10, 11]. The aim of using CAD is to diagnose the earliest indications of abnormalities in patients that clinicians may not detect. In this thesis, an automatic CAD tool is used to detect PD by using sMRI data.

1.5 Motivation

The neurodegenerative disorders are characterized by the progressive deterioration of the brain neurons [29, 30]. PD is the second common encountered disease which affects 1% of people over 60 years old [15]. Since the clinical examinations do not have any certain methods to assess the disease and have instabilities on evaluating and observing the progressive of the treatment, using a CAD system is an asset in the early and accurate diagnosis of PD. Additionally, the PD might not indicate its symptoms as it begins to occur in the brain. When the symptoms of the disease are observed by the professionals, the important parts of the nerve cells related to PD might already be damaged. Therefore, early diagnosis of PD is extremely crucial. In the AD, early diagnosis of the disease by using neuroimaging data and machine learning methods has been achieved importantly in the last decade [2, 7, 31]. Even though the atrophies in the brain of AD patients are clearly visible from the sMRI data and might be

sufficient to decide the level of the disease, in PD, the atrophies might not be sufficient. Hence, in addition to sMRI data, the clinical examinations and the medical histories of the patients are still required [27, 28]. In order to make the diagnosis of PD more robust than the state-of-the-art studies, the novel preprocessing parameters, image processing techniques, and pattern recognition methods are applied to the Three-Dimensional (3D) neuroimaging datasets of PDs and HCs.

1.6 Thesis Objectives

In this thesis, the PD detection by using sMRI data is proposed. The main objectives of the thesis are:

- Reviewing the existing literature in the last decade using MRI datasets and CAD methods for the diagnosis of PD and reporting the obtained affected brain regions due to the PD. The pipeline of each reviewed study on PD detection is provided individually.

- Using 3D-MRI with Voxel-Based Morphometry (VBM) method. Comparing the effects of different covariates (i.e. Total Intracranial Volume (TIV), age, sex and combination of them) and two different hypotheses, t-contrast and f-contrast, on the classification of PD from HC. Obtaining the 3D Volumes of Interests Volume of Interests (VOIs) which encapsulate the most discriminative voxels between PD and HC for GM and WM tissues. Multiplying the obtained 3D masks or VOIs with the processed GM and WM tissues and extracting the differentiated features.

- Combining the GM and WM data by concatenating them and obtaining the fused GM+WM dataset. To analyze the effect of both GM and WM tissues on PD detection.

- Using various feature ranking methods (i.e. Minimum Redundancy Maximum Relevance (mRMR), Relief-F, Unsupervised Discriminative Feature Selection (UDFS), Laplacian Score (LS), Unsupervised Feature Selection For Multi-Cluster Data (MCFS), Correlation-Based Feature Selection (CFS), and Feature Selection and Kernel Learning For Local Learning-Based Clustering (LLCFS)) to rank the features and an adaptive Fisher Criterion (FC) method to select the optimal number of top-ranked discriminative features.

- Using different classification methods which are Support Vector Machine (SVM), k Nearest Neighbor (kNN), Naïve Bayes (NB), Ensemble-Bagged Trees (EBT), and Ensemble-Subspace Discriminant (ESD) and comparing their performances on PD detection. Combining the binary outputs of the used classification methods by using Majority Voting (MV) technique.

1.7 Thesis Contributions

In PD detection, most studies using sMRI datasets with VBM technique consider t-contrast for obtaining 3D masks (the GM volume of PD is lower than the one of HC or vice versa). Contrary to existing literature, in this thesis, in addition to t-contrast, f-contrast hypothesis which assesses the differences between two groups without any direction limitations has also been investigated [9, 16, 27, 32, 33]. Additionally, after preprocessing the data, various classification methods with different feature ranking methods have been studied. The main contributions of this thesis are given as follows:

- The-state-of-the-art studies on PD detection since 2008 by using MRI data and CAD methods and obtained affected brain regions due to PD have been reported in a review article.

- The VBM technique is used to compare morphological differences between the brain volumes of PD patients and HCs. Since the pre-processing of 3D-MRI scans plays an important role for post-processing, different combinations of covariates (i.e. TIV, age, and sex) in model building and two hypotheses, t-contrast and f-contrast, in 3D mask building are studied. Using f-contrast for the first time in PD detection increased the classification accuracy of PD detection, significantly.

- To analyze the morphological differences in both GM and WM brain volumes of PD patients and HCs, a feature-level fusion technique, concatenating the GM and WM data, is used and it is obtained that the classification performance of using combined GM+WM data outperforms that of using GM and WM data individually.

- An automatic approach by comparing the performances of five classification algorithms with five feature ranking approaches using an automatic FC method as selecting the top-ranked number of features for PD detection is proposed. Different feature raking methods are applied to the 3D sMRI data of PD for the first time. Using a decision fusion technique, combining the outputs of five classification methods by using the MV technique, enhances the classification performance of PD detection.

1.8 Thesis Outline

In Chapter 2, the extensive literature review on PD detection using MRI data and CAD methods in the last decade is given. In Chapter 3, the methodology of the thesis is provided. It includes the database, image acquisition, pre-processing steps, feature selection, and classification methods. Chapter 4 describes the effects of different covariates and contrasts on the classification of PD using sMRI. In Chapter 5, performance analysis of different classification algorithms using different feature

selection methods on PD Detection is studied. Chapter 6 analyzes the comparison of obtained affected brain regions with the ones in state-of-the-art studies since 2008. Finally, in Chapter 7, the conclusions of the thesis, scopes for improvement, and future work are given.

Chapter 2

LITERATURE REVIEW

2.1 Introduction

Neurodegenerative diseases have been investigated recently to build an automatic CAD model in order to discriminate PD apart from HC by using MRI neuroimaging technique [8, 26, 27]. Due to the fact that the early detection of the disease is helpful to take precautions and develop required treatments and currently, the detection of the disease is highly clinician-dependent, many researchers have focused on the diagnosis of PD using machine learning methods. In this thesis, the state-of-the-art studies in the last decade investigating CAD of parkinsonism by utilizing MRI data have been reviewed.

2.2 Computer-Aided Diagnosis of PD Using MRI

The diagnosis of PD has been studied by using neuroimaging techniques together with machine learning methods. Since MRI has excellent spatial resolution with good tissue contrast and no need for any pharmaceutical injections, it has been widely used in diagnosing of various diseases and conditions. It yields better soft tissue contrast than CT and may discriminate better fat, water, muscle, and other soft tissue than CT [7]. The local and global differences in the brain are detected by using MRI and these measurements are powerful biomarkers of the stage and intensity of the neurodegenerative view of PD pathology [2]. In this chapter, several studies using MRI with different classification algorithms using different feature selection methods are reviewed. In most of the publications, SVM classification algorithm with Leave-One-

Out Cross-Validation (LOOCV) scheme and MRI data is utilized [16, 27, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45]. In some other studies, the SVM classification algorithm with K-fold CV and MRI data is used [8, 46, 47, 48, 49, 50, 51, 52, 53]. In [54, 55], SVM classification algorithm with HO CV and MRI data is considered. In other studies, instead of SVM classification algorithm, Linear Discriminant Analysis (LDA) and it is derivatives with K-fold CV and MRI data are used [56, 57, 58, 59, 60]. In other publications, various classification algorithms including the proposed novel ones with different CV schemes and MRI data are utilized [9, 32, 61, 62, 63, 64, 65]. Finally, in some studies, instead of using only MRI data, other neuroimaging data are also taken into account [66, 67, 68, 69, 70, 71, 72, 73]. The state-of-the-art studies using MRI data and CAD methods are reviewed individually and the results of them are provided in Table 2.1.

In [59], a method in which features and samples are selected jointly is proposed in order to develop new approaches for the diagnosis of PD. All MRI images are preprocessed by skull stripping, cerebellum removing, and tissue segmentation into the GM, WM, and CSF volumes. Automated Anatomical Labeling (AAL) Atlas brain mapping is used for brain parcellation after segmentation of MRI data of 374PDs and 169HCs obtained from PPMI database. The features extracted by using AAL mapping are eliminated with a Joint Feature-Sample Selection (JFSS) method. The aim of the proposed JFSS approach is to select the most salient and optimal features as well as samples. The Regularized Linear Discriminant Analysis (RLDA) method is used both for de-noising and classification. The proposed scheme is evaluated with both synthetic and real-world data from PPMI. In order to show the superiority of the

proposed method which includes Robust LDA using JFSS approach over the LDA, SVM, sparse SVM, matrix completion, Joint Feature-Sample Selection as a Classifier (JFSS-C), and Sparse Regression (SR) classification algorithms using various feature selection techniques, namely JFSS, Feature Subset Selection (FSS), Sparse Feature Selection (SFS), No Feature Sample Selection (noFSS), mRMR, Principal Component Analysis (PCA), Robust Principle Component Analysis (RPCA), autoencoder-restricted Boltzmann machine, non-negative matrix factorization, and random sample consensus. The experimental results for the proposed method given in Table 2.1 indicate that the accuracy of 96.10% is obtained while using 100 synthetic data and the accuracy of 81.90% is obtained while using the PPMI data. Furthermore, the results represent that increasing the number of synthetic data decreases the classification accuracy. It is also reported that left and right red nucleus, left and right SN, pons, left superior temporal gyrus and both left and right middle frontal gyrus are the regions that are the most associated with PD detection.

In [58], several classification methods are compared with different feature selection processes in order to achieve a better accuracy for PD detection. The proposed method in [59] has experimented with 56PDs and 56HCs. RLDA classifier combined with JFSS feature selection outperforms other classifiers and RLDA classifier with other feature selection methods. In order to show the superiority of the proposed method which includes RLDA using JFSS approach over the LDA, SVM, and SR classification algorithms using different feature selection techniques, namely JFSS, FSS, SFS, noFSS, mRMR, and PCA. The experimental results given in Table 2.1 provide that the classification accuracy of 96.10% is obtained while using 100 synthetic data and that of 82.50% is obtained while using the PPMI data.

In [71], a kernel-based feature selection scheme, in which the features and kernels are selected in a way that they induce the best classification performance in the kernel space is studied. The method proposed in the study uses a Joint Kernel-based feature selection which selects the features that benefit the classification method in the kernel space. Furthermore, the kernel functions are investigated, specifically for designing the used non-negative feature types. As a classification method, a max-margin classification with *l*1 regression function is proposed. The experiments are performed for the MRI and SPECT data of 538 subjects from PPMI database. Considering only the MRI data of the subjects, the diagnosis accuracy of 70.50% is obtained. However, as given in Table 2.1, for using both the SPECT and MRI data with the proposed method, the 97.50% of classification accuracy is achieved in PD diagnosis.

In [60], a novel semi-supervised discriminative classification approach which is robust against both sample-outliers and feature-noises is proposed. The least-square formulation of LDA forms the proposed robust discriminative classification method. Furthermore, all the labeled training and the unlabeled testing data are utilized to identify the outliers and are denoised simultaneously. The MRI of 374PDs and 169HCs are processed to test the proposed method for classifying PD. The number of features is reduced by RPCA feature selection and then selected features are given to Robust Feature-Sample Linear Discriminant Analysis (RFS-LDA) classifier. The classifier is trained via 10-Fold CV method. The experimental results indicate that the proposed method performs accurately at classifying AD from HCs with an accuracy of 92.10%, whereas the classification accuracy of the method for PD vs HC classification is 79.80% as given in Table 2.1. Brain regions which are the most relevant to PD are found to be middle frontal gyrus, pons, left and right SN, left red

nucleus, left pallidum, left putamen, right caudate, left inferior temporal and right superior temporal gyrus.

In [66], a novel method based on complex networks is proposed to contribute the early diagnosis of PD by using MRI data and the most affected brain regions owing to the PD are investigated. A network model of brain regions is identified for the MRI of 374PDs and 169HCs from PPMI database and proper connectivity measures are assigned to each region. Hence, each brain is characterized by a feature vector encoding the local relationships brain regions interweave. The Random Forest (RF) feature selection algorithm is applied to the extracted feature vectors and SVM approach with 10-fold CV scheme is utilized in order to combine complex network features with clinical scores typical of PD prodromal phase and to obtain a diagnostic index. The experimental results indicate that the proposed method achieves 93.00% of classification accuracy in PD diagnosis as seen in Table 2.1. Hence, it is concluded in the study that the connectivity of several brain regions is remarkably related to PD. Additionally, the method also offers a ranking of brain regions based on the effects of the disease in each brain region. The middle temporal gyrus, superior temporal gyrus, sub-gyral, superior occipital gyrus, middle frontal gyrus, culmen, medial frontal gyrus, precentral gyrus, cingulate gyrus, and precuneus form the top discriminative brain regions affected due to the PD.

In [32], a new meta-cognitive radial basis function network classifier with recursive feature elimination approach is proposed in order to detect the critical brain regions affected due to the PD. The VBM toolbox in SPM package is utilized to segment the MRI data of 112PDs and 112HCs obtained from PPMI database and to extract the differentiated features among the segmented GM volumes of the PDs and HCs. The

Meta-Cognitive Radial Basis Function Network (McRBFN) classifiers utilize the extracted features and a projection based learning approach. The experiments are performed with holdout-75% CV scheme and the classification accuracy of 87.21% is obtained for the proposed method as given in Table 2.1. The performance of Projection Based Learning (PBL) McRBFN classifier is compared with the SVM classification approach and it is reported in the study that the former one produces better generalization performance on PD detection. Furthermore, in order to identify the significantly affected brain regions owing to the PD, a recursive feature elimination algorithm is also proposed. The results of PBL-McRBFN Recursive Feature Elimination (RFE) selected features indicate that the GM loss in the superior temporal gyrus may lead to the PD.

In [61], an ensemble average propagator which is a Probability Distribution Function (PDF) that fully describes the diffusion locally at a voxel level, is used as a structural biomarker and an unbiased template brain map is constructed from the Ensemble Average Propagator (EAP) field of HCs in order to observe the changes in brain structure of 46PDs compared to 22HCs. Using EAP provides to capture both the orientation and shape information of the diffusion process at each voxel in Diffusion MRI (dMRI). After constructing the template brain map, in order to evaluate the variations between the patients and HCs, the dMRI brain data of the patients is compared with the template map and the number of elastic deformations are obtained by using a nonrigid warping algorithm. In order to simplify capturing the morphometric differences, a manifold-valued feature named as the Cauchy deformation tensor (CDT) is taken into account.

A principal geodesic analysis which is a nonlinear generalization of PCA algorithm is applied to CDT fields for reducing the dimensionality of the data. Then for classification, the Radial Basis Function Kernel Support Vector Machine (RBF-SVM) method with LOOCV scheme is employed. The same method was applied to another experimental set-up in which Fractional Anisotropy (FA) is selected instead of CDT. The proposed method using CDT which provides 98.53% of classification accuracy is more accurate than that using FA which gives 76.47% of classification accuracy.

In [35], a novel local feature based SVM method is developed in order to identify structural differences between the MRI of ADs and HCs as well as PDs and HCs as potential biomarkers. Considering the aim of this review article, only the tests related to PDs are reviewed here. The morphological structures are characterized by using Scale Invariant Feature Transform (SIFT) algorithm and the classification of patients, HCs, and noise is performed by using the likelihood scores obtained utilizing the feature based-morphometry. The noise features in the extracted ones are eliminated by using FBM feature selection method and the selected features are classified by using the SVM algorithm with LOOCV scheme. The experiments are performed for two datasets, namely PPMI-15 and PPMI- 37 which include 15 and 37 subjects respectively. The findings of the study state that the classification accuracy of the proposed method using PPMI-15 dataset is 80.00%, while that using PPMI-37 dataset is only 68.00% as given in Table 2.1. The classification accuracy rate decreases by the increasing number of samples. The regions of the limbic lobe, frontal lobe, sub-lobar, midbrain, pons, posterior lobe and occipital lobe are reported to be highly predictive at PD detection.

In [74], the whole-brain resting-state functional connectivity patterns of PD is studied to build an automated PD detection approach. The functional connectivities among the predefined 116 ROIs obtained by AAL Atlas brain map are constructed and the images of all Resting-State fMRI (rsfMRI) of 47 subjects, including 21PDs are preprocessed by using SPM8 package. After preprocessing the data, first the Kendall tau correlation coefficient is used to select the most informative features and then these features are given to SVM with LOOCV scheme for classification. The experimental results indicate that the classification accuracy reaches up to 93.62% as given in Table 2.1. The most discriminative functional connections were placed within or across the default mode, cingulo-opercular and frontal-parietal networks, and the cerebellum.

In [70], a combination of whole-brain, VBM, and Diffusion Tensor Imaging (DTI) analyses is used in order to differentiate 15 Tremor-Dominant Parkinson's Disease (tPD) from 15 Essential Tremor With Rest Tremor (rET) patients. The GM and WM volumes derived by using VBM and the mean diffusivity and FA derived by using DTI are considered together. The Dopamine Transporter (DAT)-SPECT data is used as a ground truth. The preprocessing is performed by using FMRIB Software Library (FSL) VBM and FSLDTIfit. In order to select the significant features out of all features, an f-test is used as a feature selection method. The four features, namely GM, WM, mean diffusivity, and FA of each voxel are given to SVM with LOOCV scheme both separately and combined. The classification accuracy reaches up to 100% when combined features are taken into account.

In [8], the importance of the preprocessing step of PD detection has been emphasized. The VBM technique is used to compare morphological differences of PDs and HCs in GM and WM. The effects of different combinations of three covariates (i.e. TIV, age,

and sex) and two different contrasts, t-contrast and f-contrast, on PD classification have been studied. The 3D masks are obtained by using two-sample t-test statistical method for GM and WM. The 3D GM and WM masks are concatenated and the fused GM+WM data is obtained. The PCA is applied to the extracted features in order to reduce the dimensionality and RBF-SVM is utilized for classification. The proposed method is evaluated on 40PDs and 40HCs obtained from the PPMI dataset. The experimental results indicate that compared to using t-contrast, using f-contrast shows a superior performance for GM, WM, and the combination of GM as well as WM. Comparing to other combinations of covariates, using only TIV gives more robust results for PD identification. As seen in Table 2.1, the highest classification accuracies of 73.75%, 72.50%, and 93.7% are obtained when TIV is used as a covariate and f-contrast is used for model building for GM, WM, and the combination of them, respectively. The most important ROIs detected using different combinations of covariates and contrasts belong to the regions of the left superior frontal gyrus, right middle temporal, left anterior cingulate gyrus, right anterior insula, right angular gyrus, left middle temporal gyrus, left inferior temporal, and right putamen.

In [46], the performance of RBF-SVM using two different feature selection methods, namely PCA and PDF-based feature selection, on PD detection are compared. The VBM technique with DARTEL toolbox in SPM12 package is used to segment, normalize, modulate and smooth the MRI images of 40PDs and 40HCs obtained from PPMI database. In the statistical model building, TIV is used as a covariate and in obtaining the 3D mask of GM and WM volumes the t-contrast is utilized where it is assumed that PD leads to structural atrophies in brain parts. The experimental results indicate that meaningful distributions may not be achieved by taking the pdf of the 3D

masked GM and WM datasets, since the number of obtained VOIs are not too much. Hence, using PCA for dimensionality reduction outperforms using PDF-based feature selection method as given in Table 2.1.

In [36], structural and cross-sectional T1-weighted MRI of 181 subjects are processed in order to evaluate an automated classification which tries to classify 16 Idiopathic Parkinson's Disease (IPD) patients from 149HCs, 8 Progressive Supranuclear Palsy (PSP), and 8 Multiple Systems Atrophy (MSA) patients. The deformation features and tissue map of the hindbrain region are automatically extracted from the images. The most salient features are selected by using the PCA feature selection method and then provided to SVM with least-square optimization within a multidimensional composition/deformation feature space constructed from the data of HCs. LOOCV scheme is used to avoid over-determination. As provided in Table 2.1, the mean classification accuracy, sensitivity, and specificity obtained by using the proposed method are 90.60%, 93.32%, and 88.20% respectively.

In [75], it is aimed to detect morphometric markers among the brains of 14 vascular parkinsonism patients, 32PDs, and 29HCs. Volumetric morphometry between PDs and HCs is identified by utilizing the Morphobox software and the AUC is considered in PD diagnosis. The AUC is computed as 0.86 when WM hyper intense lesions are considered for comparing Vascular Parkinsonism (VP) from HCs. The volumes of the caudate nucleus and WM hyper intense lesions are observed to be affected by VP. The GM volume in the caudate nucleus of patients with VP is lower compared to the GM volume in the same region of HCs.

In [37], MRI of 21 Idiopathic Parkinson Syndrome (IPS), 11 Parkinsonian Multiple Systems Atrophy (MSA-P), 10PSP patients, and 22HCs are processed using VBM with DARTEL toolbox in SPM8 package in order to achieve an automated discrimination between the disease groups and HCs. The most discriminative VOIs among all the voxels are determined by using f-contrast in 3D mask building. It is stated in the study that the advantage of the f-contrast over t-contrast is that it has no negative values and increases as well as decreases are treated equally. The extracted features are classified by using SVM with LOOCV scheme classification method. Besides defining WM and GM losses in patients, SVM also tries to classify the subjects based on the selected features. The results of the study show that the proposed method is not able to find a significant difference between IPS and HCs, only 41.86% as seen in Table 2.1. However, the method manages to discriminate PSP patients from HCs with an accuracy of 93.75% when the WM volumes are considered. Patients with PSP experience WM volume loss in the brainstem and in basal ganglia and also GM loss in the cerebellum.

In [47], DTI of 131PDs and 58HCs are processed in order to test a method for diagnosing PD. To check the test-retest reliability, a second image is taken from each subject just after the first one. A total of 68 Region Of Interest (ROI) is predetermined by using FreeSurfer software. The ROI signal correlations features are extracted from the images and selected based on Intraclass Correlation Coefficient (ICC) and t-test methods in order to eliminate irrelevant and non-significant features.

In order to check if Echo-Planar Imaging (EPI) correction affects the classification accuracy, in addition to no correction, two correction methods, namely 3D inverse-consistent mutual information elastic registration and custom variational image-based

elastic registration are applied to the data. SVM classifier with 12 regularization is used as the classification method and is tested via 10-Fold CV scheme. The results indicate that 3DMI EPI correction increases the classification accuracy to 60.10% as given in Table 2.1. Second images taken from the subjects' record slightly lower accuracy, 59.50%, for 3DMI correction.

In [76], it is aimed to detect and evaluate cerebral GM volume and WM density changes in Parkinsonian diseases in order to understand how these diseases affect the brain structure. The MRI of 23 Parkinson's Disease With Mild Cognitive Impairment (PDMCI) patients, 23PD-HC patients, and 21HCs are processed and changes in the structural images are obtained by using VBM8 toolbox in SPM8 package. The findings of the study state that GM atrophy is observed at both PDMCIs and PN-HCs compared to HCs. Cerebellum posterior lobe, cingulate gyrus, middle temporal gyrus and lentiform nucleus are the affected regions. The results also show that WM density of healthy cases is higher than patients in putamen, caudate and lentiform nucleus. It is noted in the study that the regions with observed GM atrophy and WM density decrease may provide significant information for diagnostic classification studies when deciding feature selection and ROI.

In [77], falls in PD patients are investigated by applying a multidisciplinary approach which benefits from clinical, demographic and neuroimaging data of subjects. In addition to clinical and demographic data gathered from patients, features obtained from MRI are also evaluated. The relevant and discriminative features of the neuroimaging data are selected by employing Knockoff filtering and RF method. Six features which are chosen among the top discriminative features by both methods are selected as the significant features. These features are then given to different

classifiers, namely logistic regression, RF, SVM, Adaboost, SuperLearner, and XGboost. The performances of those classifiers are evaluated by external out-of bag validation and also by 5-Fold CV techniques. The tests are conducted for both selected six features and all features obtained before the feature selection process. It is noted in the study that while some classification approaches using feature selection method provide higher classification accuracies, some others using all the features without the selection process give better classification accuracies. As given in Table 2.1, the classification accuracy of 77.70% is achieved when SVM classification method with the selected features is used and the classification accuracy of 73.70% is obtained when RF classification method with all features are considered.

In [53], functional connectivity patterns in fMRI data is used for PD detection. The fMRI data is divided into five filtered frequency ranges. The features from each frequency range are selected by using the mRMR method and a proximal SVM classification algorithm with 5-fold CV scheme is used. Then the outputs of all five classification methods using five filtered frequency ranges are combined by using the MV technique. As provided in Table 2.1, the classification accuracy of 84.00% is obtained.

In [78], a machine learning-based automatic classification is studied in order to separate patients with the postural instability and gait difficulty subtype of PD from those with the non-postural instability and gait difficulty subtypes. The self-acquired sMRI, fMRI, and DTI data of PD patients have been used. The SVM classification algorithm with RFE method and LOOCV scheme is utilized to identify the postural instability and gait difficulty by using multi-modal MRI datasets. The classification accuracy of 92.31% and AUCmax = 0.9585 are obtained as given in Table 2.1.

In [73], the group-level comparison of Susceptibility-Weighted Imaging (SWI) values in the brain of PD patients is studied. The experimental results show that at the group level, the simple visual analysis gives no differences between groups, yet in the bilateral right-dominant thalamus and dentate nucleus of PD patients compared to other Parkinsonism the SWI values are increased. In addition, SWI value decreases in the left SN, left putamen, and right putamen of PD patients. Visual SWI analysis might not discriminate idiopathic from atypical PD. However, at the individual level, the SVM method is used and the classification accuracy of 86.92% is obtained as given in Table 2.1.

In [79], a method is proposed for automatic tissue segmentation of the marmoset monkey brain using a 7-T animal scanner and to evaluate the dopaminergic degeneration in a PD model. The experiments indicate that the volumes of the bilateral SN of 1- methyl-4-phenyl-1,2,3,6-tetrahydropyridine-treated marmoset is decreased remarkably. Furthermore, the volume decreases in the locus coeruleus and lateral hypothalamus are also obtained.

In [38], an atlas-based brain volumetry is used to determine the structural alterations emerged in the brain of PD patients. The different MRI data of patients with PD, PSP with Richardson's syndrome phenotype, MSA of the cerebellar type, and MSA of the Parkinsonian type are analyzed. The 3D masked feature extraction and SVM classification algorithm are used. The experiments are performed for each Parkinsonian Syndromes (PS) versus NCs, each PS versus all other PS, and each combination of two single PS. Furthermore, the multiclass comparisons of all PS together with NCs, all PS without NCs, and all PS with Cerebellar Type Multiple System Atrophy (MSA-C) and MSA-P combined in one MSA group are done.

However, since the scope of this review study is to separate PDs apart from NCs, the obtained 66.20% balanced binary classification accuracy of PD versus HC is provided. Cerebrum, frontal lobe, temporal lobe, parietal lobe, and occipital lobe are reported to have at least 5% GM loss. In addition, it is found out that there is a total volume decrease in the hippocampus, amygdala, and caudate while WM volume increases by 5% in the cerebellum.

In [48], the structural connectome of PD patients is mapped with probabilistic Multi-Shell, Multi-Tissue Constrained Spherical Deconvolution (MSMT- CSD), deterministic Single-Shell, Single Tissue (SSST) CSD, and probabilistic SSST-CSD tracking methods by diffusion-weighted MRI tractography. The five global network metrics are computed: global path length, clustering coefficient, global efficiency, mean strength, and the small-worldness ratio. Additionally, four local network metrics are computed: nodal strength, betweenness centrality, local clustering, and local efficiency. A graph-based theory and the SVM algorithm to predict the diagnosis based on a linear combination of graph metrics are used. The experimental results indicate that the probabilistic MSMT-CSD could identify remarkably decreased global strength, efficiency, clustering, and small-worldness, and increased global path length in PDs comparing to HCs. On the other hand, the probabilistic SSST-CSD could only identify the differences in global strength and small-worldness. The obtained 78.33% classification performance of probabilistic MSMT-CSD method outperforms the classification performances of probabilistic SSST- CSD and deterministic SSST-CSD tracking ones. The study reports a significant decrease in local efficiency at right putamen, right thalamus, right postcentral gyrus, and left putamen while significant

decrease in local clustering at right supramarginal gyrus, right insula, right postcentral gyrus, right thalamus, and left pars triangularis.

In [39], a graph theoretical analysis for automated diagnosis of PD by using ROI is evaluated. A brain network graph is constructed using the regions as nodes and the Pearson correlation values between their average time series as edge weights. For each subject, the metrics of integration (characteristic path length and efficiency), segregation (clustering coefficient and transitivity), centrality, and nodal degree are obtained as features. A floating forward automatic feature selection method is used to select the most discriminative features from the extracted features and SVM is used as a classification method. The classification accuracy of 94.59% is achieved as seen in Table 2.1. The regions of right cuneus, left precuneus, left and right middle frontal gyrus, right superior frontal gyrus are found to be the most discriminative brain regions at discriminating patients from HCs.

In [67], a new feature selection method based on an improved loss function is proposed. Four different feature sets which are GM and CSF tissue maps of MRI, FA coefficient of DTI, and the combinations of all these three sets are analyzed. The Support Vector Classification (SVC) with sigmoid kernel method is used for classification. As given in Table 2.1, the highest classification accuracy of 83.28% and AUC of 84.17% are obtained when the combination of GM, CSF, and DTI feature sets are considered. The regions of left hippocampus formation, right superior occipital gyrus, left amygdala, right calcarine, right lingual, left postcentral, left middle occipital gyrus, left thalamus, left putamen, and right middle frontal gyrus are reported to be the most discriminative brain regions according to the proposed method.

In [68], a feature selection approach via relational learning in a unified multi-task feature selection method is proposed. Four different types of features, namely FA coefficient of DTI, CSF and GM of MRI, and CSF biomarkers. The GM, CSF, and FA coefficients of DTI feature sets are investigated separately. Furthermore, the combination of these four feature sets are studied. The SVC with sigmoid kernel approach is used for classification. The experiments indicate that the performance of using multi-modal data in PD identification outperforms that of using single modal data. As given in Table 2.1, the highest classification accuracy of 84.40% is obtained when the combination of GM, CSF, DTI, and CSF biomarkers features are considered.

In [69], an adaptive sparse learning framework in which feature selection and subspace learning are integrated is used. The idea of Fisher's LDA and locality preserving projection are combined in order to use both the local and global information of the raw data. To improve the classification performance, a generalized norm to the loss function which aims to regulate the sparseness degree is proposed for feature selection. After using feature selection, by using a p-norm regularization technique, additional features are added to the selected features to boost the classification performance. The SVC approach is used for classification. The experiments are performed for three different feature subsets which are the combinations of i) GM, CSF, and DTI, ii) WM, CSF, and DTI, and iii) GM, WM, CSF, and DTI. Furthermore, the three CSF biomarkers and Montreal cognitive assessment scores feature sets are combined with combinations of GM, CSF, and DTI, the feature sets which gives the highest classification accuracy. As given in Table 2.1, the experimental results indicate that the best classification accuracy of 79.25% is obtained when the combination of GM,

CSF, DTI, CSF biomarkers and Montreal cognitive assessment scores feature sets are used.

In [40], a combination of SIFT and Histogram of Oriented Gradient (HOG) is presented for feature extraction and selection. The study also investigates three different cell block sizes, namely 4x4, 8x8 and 16x16 for feature extraction. Extracted features are submitted to SVM classifier in order to differentiate PD patients from HCs and also AD patients from HCs. Because of the extent of this review, only the classification of PDs and HCs are examined. Two datasets are used which contain 46 (26PD and 20NC) and 212 (145PD and 67NC) MRI data available at PPMI database. The experimental runs compare different local feature selection methods (gray value, gray-level co-occurrence matrix, and HOG). Additionally, the impacts of SIFT features and cell block size on accuracy are examined. 16x16 cell block size design is reported to outperform the 4x4 and 8x8 designs. In addition, utilizing SIFT features increases the accuracy rates. It is noteworthy that larger dataset decreases the accuracy rate; when the design is applied for classification of the dataset consisting of 46 subjects an accuracy rate of 78.26% is achieved whereas the accuracy rate is only 66.98% for the dataset with 212 subjects as provided in Table 2.1. Diagnosing PD is most probable when HOG and SIFT features are combined with 16x16 cell block size according to the findings of the mentioned study. The results also indicate that the regions of the thalamus, temporal lobe, calcarine, and cingulum are the most remarkable features.

In [62], a hybrid classifier which is the combination of the basal ganglion prior classifier and a data-driven classifier is proposed for PD detection. A LOOCV scheme selected the most predictive imaging features and repeated for all study participants of

145 subjects of which 72 are PD patients. As seen in Table 2.1, 74.00% of accuracy, 76.00% of sensitivity and 73.00% of specificity are achieved from the classification experiments.

In [56], a novel feature selection method called Iterative Canonical Correlation Analysis (ICCA) is proposed for exploring the roles of different brain regions in PD. In the first experiment, features in GM/WM feature space are selected in order to test the proposed ICCA feature selection method. The second experiment is implemented with features that are selected in a canonical feature space. RLDA classification algorithm is used for classifying the data gathered from the PPMI database (56PDs and 56HCs). As it is seen in Table 2.1, a classification accuracy of 70.50% is achieved with the features selected in GM/WM feature space outperforming the features from canonical feature space which reach 68.80% of classification accuracy.

In [57], WM and GM features are extracted from 112 subjects half of which are PD patients. 90 regions pre-defined by AAL Atlas are considered as ROI. Among all extracted features more relevant and significant features are selected by ICCA and cascaded Canonical Correlation Analysis (CCA) feature selection methods. Selected features are classified by using the RLDA classification method with 10-Fold CV scheme. ICCA feature selection records 70.50% of classification accuracy, whereas cascaded CCA gives 68.80% of classification accuracy. Both feature selection methods result in higher accuracy rates compared to other feature selection approaches, namely No Feature Selection (NoFS), ElasticNet, SFS, PCA, FSASL, CCA, and mRMR. The study indicates that the postcentral gyrus, precuneus, thalamus, middle and medial frontal gyrus, hippocampus, and superior temporal pole are the

most discriminative brain regions that are available in at least 60% of feature subset selected.

In [41], resting-state functional MRI of 46 subjects are collected and the characteristics obtained from the data are extracted in three different levels, the amplitude of low-frequency fluctuations, regional homogeneity, and regional functional connectivity strength. The volume characteristics from GM, WM, and CSF are extracted for the structural images. In order to reduce the number of features, a two-sample t-test is used. The selected features are classified by using the SVM classification method with LOOCV scheme. The experiment results indicate that the proposed method provides 86.96% of classification accuracy as given in Table 2.1.

In [80], it is aimed to classify clinically unclassifiable PD into subgroups of PSP, PD, and MSA. In order to construct a model, midbrain diameter and area, pontine diameter and area, middle and superior cerebellar peduncles, midbrain to pontine diameter and area ratios are taken as features of interest. DT classifier is trained with data from 55PSPs, 194PDs, and 63MSAs and tested with 84 clinically unclassifiable PD patients. The decision algorithm can discriminate PD from other groups with a classification accuracy of 86.50% according to the results of the study.

In [81], in order to evaluate the importance of anatomical patterns obtained from a structural neuroimaging dataset, a multi-class pattern recognition is supported instead of pair-wise categorical classification as it tries to distinguish various diseases simultaneously. In the study structural MRI of 17PSPs, 14IPDs, and 19MSAs along with 19HCs are scanned and processed. The predefined anatomical patterns, namely subcortical motor network, its component regions, and whole brain are evaluated via

probabilistic pattern recognition. During classification process, four classifying sets are defined as (i) PSP vs IPD vs MSA, (ii) PSP vs IPD vs MSA-P vs MSA-C, (iii) HC added to classifying set (i) and (iv) HC added to classifying set (ii). The results state that midbrain/brainstem region is the most helpful region for classification with accuracies of 91.70%, 73.60%, 84.50%, and 66.20% for classification sets (i), (ii), (iii), and (iv) respectively.

In [64], the performances of four classification models, namely NB, multivariate filter-based NB, filter selective NB, and SVM on discrimination of 16 Cognitively İntact Patients With Parkinson's Disease (PDCI), 15PDMCI, and 14 Parkinson's Disease Dementia (PDD) patients are investigated. The Freesurfer software which is used to quantify the volume of multiple subcortical structures is utilized in order to extract the anatomical neurological information automatically (volume of subcortical structures and cortical thickness). The relevancies of these measurements are considered simultaneously in early diagnosis of PDMCI and PDD. Owing to a large number of variables, the most discriminative cerebral regions are determined automatically by using the FSS approach together with NB classifiers. The Bayesian classifiers as an accurate automatic decision support system are applied to the different PD patient groups, namely PDMCI vs. PDCI, PDMCI vs. PDD, PDD vs. PDCI, and among all three groups. Multivariate filter-based NB model is reported to achieve the highest classification accuracy, sensitivity, and specificity of classifying different types of parkinsonian diseases. Additionally, it is reported that the volumetric measures, specifically the enlargement of the lateral inferior ventricles together with hippocampus and reduction of WM are the most discriminative neuroanatomical biomarkers for predicting dementia in PD.

In [43], pattern recognition techniques are studied in order to classify patients with different types of parkinsonian disorders. The DTI, proton spectroscopy, and morphometric- volumetric data are used together in order to acquire MR quantitative markers. A two-step feature selection approach is utilized. First, the most discriminative features are detected based on relative entropy criterion and then the feature subsets of MR markers with different sizes are tested for generalizability. Additionally, a graph-based method on the set of quantitative markers is also studied in order to extract additional features from the dataset and enhance the classification accuracy. The SVM approach with LOOCV scheme is utilized to implement the multi-class classification. Integrating selected features with graph-based features increases the classification accuracy as the results indicate.

In [42], a method for an automated Parkinsonian disorders classification utilizing SVM is investigated. The MR quantitative markers are taken into account as features. Due to the large of number of features, a feature selection method based on relative entropy criterion is used. The feature selection is reported to suit with the opinions of clinicians, since the selected features are described as the most important features by the clinicians as well. The classification accuracy of PD vs all with 40 selected features is 90.00%, as seen in Table 2.1.

In [82], the classification results are inspected in order to evaluate the effect of CV approach on the results. The binomial distribution and the permutation test are compared and it is noted that distributions built from classification of random data with CV do not follow a binomial distribution. On the other hand, the CV scheme does not affect the permutation test. When a real dataset from a brain-computer interface experiment is used to analyze the effect of CV scheme, it is obtained that the three of

16 patients have significant accuracy on binomial testing, yet just one patient shows significant accuracy on permutation test. Furthermore, when the real fMRI dataset is used to analyze the effect of CV scheme, it is observed that the mental imagery of gait might classify IPD patients apart from HCs significantly according to the permutation test, yet not according to the binomial test. As a result, it is recommended that the permutation testing should be selected for clinical classification with CV, since the binomial distribution might estimate the significance level incorrectly.

In [65], a novel approach utilizing the MRI data of 30PDs and 30HCs from PPMI database is proposed for identification of PD. The VBM toolbox in SPM8 package is used for segmentation, normalization, and smoothing the MRI data. Out of six segmented volumes, only the GM one is taken into account. The two-sample t-test statistical method is used in model building and t-contrast, the atrophies in the brain of PD is considered. The significant VOIs captured by 3D GM mask are extracted and the genetic algorithm feature selection method is applied to the extracted features. The performances of two different classification algorithms, Extreme Learning Machine (ELM), and SVM, are compared and it is noted that the ELM method outperforms the SVM one. As seen in Table 2.1, the classification accuracy of ELM with Genetic Algorithm (GA) method is reported as 94.87%.

In [83], classification of PDs apart from HCs by using T2-weighted (spin-spin relaxation time) MRI data is studied. The bar graph options and gray level co-occurrence matrix options are extracted in order to classify the MRI of 78PDs and 32HCs obtained from PPMI database. The extracted features are given to the SVM for classification. Gaussian, polynomial and sigmoidal kernels are also compared in the study and it is indicated that Gaussian one gives higher accuracy rates. The results

show that classification accuracy for gray level co-occurrence features are higher than that for bar graph options features. However, combining those features enhances the classification accuracy up to 87.00%.

In [84], an automated classification method is used to first detect the patients with PD and then classify them into sub-groups based on the severity of the disease. The histogram features and the gray level co-occurrence matrix features are extracted from the MRI of 60PDs and 26HCs. The extracted features are provided to two different classification techniques, namely SVM and Radial Basis Function Neural Network (RBFNN). The aim of the classifiers is first to differentiate PDs from HCs and then to classify them into three severity-based subgroups; mild, moderate, and advanced. The results show that the highest classification accuracy is obtained when the histogram features and gray level co-occurrence matrix features are combined and given to SVM, rather than RBFNN, with Gaussian filtering.

In [49], a multilevel-ROI-features-based machine learning method which is a combination of low-level ROI features and high-level correlative features is proposed in order to identify the sensitive morphometric biomarkers in PD. The multilevel ROI features are captured by combining particularly the low-level ROI features (gray matter volume, cortical thickness, etc.) and high-level correlative features (connectivity between ROIs). The GM volume, cortex thickness and cortex surface area of ROI are measured by using BrainLab. The filter and wrapper types of feature selection techniques are used to select the discriminative features captured by using MRI of 69PDs and 103HCs obtained from PPMI database. The multi-kernel SVM algorithm with 10-fold CV scheme is employed to perform the classification of PDs and HCs. The method proposed in the study provided a classification accuracy of

85.78%. The results indicate that proposed scheme which benefits the multi-level ROI features is more promising than the methods with single-level ROI features in PD detection. The most sensitive biomarkers between PD patients and HCs are detected mostly in the frontal lobe, parental lobe, limbic lobe, temporal lobe, and central region.

In [50], applying the machine learning methods to the multilevel ROI based features in order to enhance the accurate detection of mild cognitive disease in parkinsonism is investigated. The MRI data of 77PDs and 32HCs obtained from PPMI database is used to evaluate the performance of the proposed method. The GM volume, cortex thickness and cortex surface area at each ROI are measured by using the BrainLab software. The t-test and mRMR feature selection methods are used in order to select the most discriminative features and the SVM algorithm with 10-fold CV scheme is utilized for classification PD apart from NC. The method manages to classify 92.35% of PDs and HCs correctly according to the test results as given in Table 2.1.

In [85], a multi-modal MRI approach which is a combination of MRI parameters sensitive to different tissue characteristics such as volume atrophy, iron deposition, and microstructural damage is proposed for detecting the differences between PD, MSA patients, and HCs. In order to discriminate 26PDs, 29MSA patients (16MSA-P and 13MSA-C), and 26HCs, the GM density, T2 relaxation rates, and DTI scalars are combined by using the whole-brain voxel-based MRI. Self-organizing map method is computed for unsupervised classification. 10-Fold CV method is used for calculating the area under the ROC curve. Combinations with AUC greater than 95% are reported in the study. Additionally, it is noted that the proposed method explains the multiparametric changes within the cerebellum and putamen in both MSA-C and

MSA-P patients compared to PD patients and specific single multimodal MRI markers may discriminate MSA-P and MSA-C patients from PDs.

In [72], the rapid eye movement, sleep behavior disorder, olfactory loss, CSF measurements, and dopaminergic imaging markers from 401PDs and 183HCs obtained from PPMI database are studied in order to increase the chance of early PD diagnosis. Four different classifiers, namely NB, SVM, BT, and RF are used to classify early PD patients apart from HCs. The experiment results indicate that SVM classifier which provides the highest classification accuracy of 96.40% outperforms the other three classifiers as provided in Table 2.1. It is stated in the study that using the non-motor, CSF, and imaging markers together may help in the preclinical detection of PD.

In [44], to determine features of normalized and skull-stripped volumes obtained from 3D T1-weighted MRI dataset, a graph-theory-based spectral feature selection method is analyzed. A decision model is built using SVM with LOOCV scheme. The experiments are performed for both self-acquired and PPMI datasets. As given in Table 2.1, the classification accuracy of 88.89% for PPMI datasets, and 86.67% for self-acquired datasets are obtained.

In [27], an ROI based CAD technique using GM, WM, and CSF volumes of sMRI dataset for PD detection is studied. The predefined five ROI, namely SN, thalamus, hippocampus, FL, and mid-brain are studied both individually and in combinations of pairs and triplets. The mutual information technique with unpaired two-tailed two-sample t-test based ranking is used for feature selection. The SVM method with LOOCV algorithm is used for classification. As given in Table 2.1, the classification accuracy of 86.67% with SN for GM and combination of SN and Frontal Lobe (FL)

for WM. The ROI based classification accuracy for PD detection outperforms the one of VBM technique. Additionally, while the volumes of GM and WM tissue maps decrease, the one of CSF tissue map increases in PD patients comparing to HCs.

In [33], VBM technique is applied to the self-acquired PDs and HCs sMRI datasets. A spatial clustering is performed to obtain clusters/groups of spatially contiguous voxels. The clusters of spatially contiguous voxels in the statistical map are constructed and three statistical features per cluster are extracted. After the 3D masks of GM, WM, and CSF tissue maps are extracted separately, an mRMR feature extraction method is used to reduce the number of extracted features from the different combinations of GM, WM, and CSF tissue maps. The SVM algorithm with LOOCV is used for classification. As given in Table 2.1, the experimental results indicate that the best classification accuracy of 88.33% is obtained both for the combination of GM and WM as well as the combination of GM, WM, and CSF tissue maps.

In [16], an Fused Feature Descriptor (FFD) method is defined in which the information and inter-correlation of all GM, WM, and CSF tissue maps are obtained simultaneously, hence the PD detection performance of the CAD system is improved. The obtained three brain volumes are represented in terms of the FFD and each fused volume is segmented into 118 brain regions by using WFU Pickatlas map and 2 regions covering the left and the right SN region are added manually. The 3D LBP is used to extract the features of each brain region and the t-test with mRMR feature selection method is applied to the extracted features. The SVM algorithm with LOOCV and 10-fold CV is used for classification. As provided in Table 2.1, the highest classification accuracy of 95.00% is obtained for LOOCV.

In [9], a supervised machine learning is used to identify PD by using self-acquired sMRI dataset. The BET tool of the FSL software is used to obtain Skull-Stripped and Normalized MRI Volumes (SSNMV) data and the data is normalized to Montreal Neurological Institute (MNI) space. The PCA is used to reduce the dimensionality of the normalized SSNMV data. The SVM algorithm with LOOCV is used for classification. Even though the experiments are performed for PD vs HC, PSP vs HC, and PSP vs PD, since the scope of this study is to focus on PD vs HC, the classification accuracy of 85.80% is obtained for PD vs HC case.

In [86], different variations of tasks and masks are examined in order to develop a method for diagnosing PD. The fMRI of 29 subjects, including 14PDs and 15HCs, are obtained while each subject standing, STAND, walking at a comfortable pace, COMF, and walking briskly, BRISK along a path. All seven fMRI task combinations, namely STAND+COMF, STAND+BRISK, COMF+BRISK, STAND+COMF+BRISK and each task separately are processed. In addition to different tasks, three different masks which are, whole brain mask, motor mask and mesencephalic locomotor region mask are used in the experiment. A total of 21 models that are produced by using all 7 task combinations with all three masks are tested by a binary SVM classifier. The results indicate that mesencephalic locomotor region mask combined with COMF task is the most accurate model with 76.00% of classification accuracy. It is reported in the study that using a whole brain mask dramatically decreases the diagnosis accuracy which is also stated [81].

In [54], a Kohonen self-organizing map-based feature extraction is proposed for extracting the GM and WM features from MRI of 831 subjects including 518PDs, 68 Scans Without Evidence Of Dopamine Deficits (SWEDDs), and 245HCs. The

extracted features are classified by using SVM classifier with holdout- 80% of CV scheme. When the classification accuracies are compared, it is indicated that the WM features enhance the classification accuracy up to 96.84%, while the GM features help the classifier predict the 93.25% of the classification tests correctly as seen in Table 2.1.

In [55], a novel pattern recognition based automated individual-level clinical diagnosis of PD is proposed. The features obtained from the same dataset. In [54] are extracted by using Self Organizing Map (SOM) method. The study offers two different experimental designs. The first design divides subjects into 6 age unrelated groups while the second design divides them into 32 age related subgroups in order to compare the classification accuracies depending on the age and progress of the disease. In brief, the preprocessed MRI are modeled by utilizing SOM for capturing the features, the Fisher Discriminant Ratio (FDR) technique is used to detect the discriminative features, and the Least Squares Support Vector Machine (LSSVM) algorithm with holdout-80% CV scheme is used for classification of PDs. The proposed feature extraction and adopted classification algorithm are reported to suit best for diagnosing patients in the early stages of the disease. The results show that the classification accuracy of Age-Related Subgroups (ARS) design which is 97.22% is higher than that of Age-Unrelated Subgroups (AUG) one which is 87.42%.

In [51], a framework which used PCA for feature extraction, FDR for selecting the number of top-ranked features, and SVM algorithm with 10-Fold CV scheme for classifying the MRI of 50PDs, 50HCs, and 50SWEDDs obtained from PPMI database is investigated. FDR feature selection method is used after the feature extraction process in order to limit the number of features to be taken into account. The

experiment is performed for both GM and WM volumes. The experiment results, as given in Table 2.1, indicate that the highest classification accuracy of 99.00% is obtained when features from only GM volume is considered.

In [87], the analysis of DaTscan images using 3D voxel-based logistic lasso model is studied. The analysis indicates that the sub-regional voxels in the caudate, the putamen, and in the globus pallidus includes discriminative information in PD detection. A logistic component analysis technique, explaining two uncorrelated sources which reveal the most of the variance of the logistic feature is proposed. The most important factors affecting the logistic analysis are the intra-population variations in dopamine transporter concentration and the imperfect normalization. The interactions with handedness, sex, and age are studied and strong interaction of the logistic feature with sex (for controls) and with age are obtained. However, the interaction of the logistic feature with handedness is not that significant.

In [45], an image-based classification method of patients with PD and PSP is studied. The mean apparent diffusion coefficient parameter maps are calculated based on diffusion-tensor MRI datasets and mean values of them are calculated for ROI which are determined by using the AAL Atlas map. These extracted values are used as features and the linear SVM algorithm with LOOCV scheme is used for classification. As given in Table 2.1, the obtained classification accuracy of separating PD apart from PSP patients is 94.80%.

In [63], the PD detection by using the combinations of the GM, WM, and CSF volumes are studied. The VBM method is used to extract the 3D masks of all tissue maps separately and the different combinations of the extracted features are performed. The

PCA dimensionality reduction technique is used to reduce the dimensions and a relevance vector machine-based classifier is used for classification. Different total variance rates to select the number of PCs are performed and the highest classification accuracy of 89.13% is obtained for 0.95 PCA variance rate and the combination of GM, WM, and CSF tissue maps.

In [88], PD related atrophies are estimated by using Deformation Based Morphometry (DBM) of T1 weighted MRI data. A network of cortical and subcortical regions in which the atrophies are related with the clinical observations are determined by using PLS method. The results indicate that the atrophies occur in lower brainstem, substantia nigra, basal ganglia, and cortical areas, specifically inferior frontal gyrus, fusiform gyrus, putamen, cingulate gyrus, nucleus accumbens, cerebellum, SN, thalamus, and caudet are the brain regions with the most significant atrophies. Furthermore, individual changes in this network estimated the longitudinal clinical progression in both motor and non-motor symptoms of de novo PD patients.

In [89], the structural changes in cerebellar brain part of PD patients are analyzed. The GM volume of sMRI data is preprocessed by using VBM technique and the RFE feature selection method is used. The selected features are classified by using the SVM classification approach with four different CV techniques, namely LOOCV, 2-fold, 5-fold, and 632-fold CVs. The results indicate that the GM volume changes in the brain of PD patients are primarily located within the cerebellar Crus I. Furthermore, 97.70% of the classification accuracy using the GM volume changes in cerebellum for PD detection is obtained with 632-fold CV.

In Table 2, the abbreviations of NI, FS, CV, AUC, ACC, SEN, and SPE stand for NeuroImagin, Feature Selection, Cross-Validation, Area Under Curve (%), Accuracy (%), Sensitivity (%), and Specificity (%), respectively.

Table 2.1: The state-of-the-art studies on PD detection using MRI data.

Study	Database	NI	FS	Classification	CV	AUC	ACC	SEN	SPE
Chen et al.[35]	PPMI-15(9PD 6HC)	MRI	FBM	SVM	LOO	N/A	80.00	90.70	N/A
	PPMI-37(18PD 21HC)						68.00	78.00	N/A
Chen et al.[74]	Self-Acquired(21PD 26HC)	MRI	Kendall tau RCC	SVM	LOO	N/A	93.62	90.47	96.15
Duchesne et al.[36]	Self-Acquired+ICBM(16IPD 8MSA 8PSP 149HC)	MRI	PCA	SVM	LOO	N/A	90.60	93.32	88.20
Focke et al.[37]	Self-Acquired(21IPS 11MSA 10PSP 22HC)	MRI	f-contrast	SVM (WM)	LOO	N/A	41.86	38.10	45.45
				SVM (GM)			39.53	28.57	50.00
Huppertz et al.[38]	Self-Acquired(204PD 106PSP 21MSA-C 60MSA-P 73HC)	MRI	Results of ABV	SVM	LOO	N/A	66.20	65.20	67.10
kazeminejad et al.[39]	PPMI(19PD 18HC)	MRI	Floating Forward	SVM	LOO	N/A	94.59	N/A	N/A
Li et al.[40]	PPMI(26PD 20HC)	MRI	HOG+SIFT	SVM	LOO	N/A	78.26	76.92	80.00
	PPMI(145PD 67HC)						66.98	70.15	65.52
Long et al.[41]	Self-Acquired(19PD 27HC)	MRI	Two Sample t-test	SVM	LOO	N/A	86.96	78.95	92.59
Morisi et al.[42]	Self-Acquired(26PD 17PSP 8MSA-C 7MSA-P)	MRI	REC	SVM	LOO	N/A	90.00	N/A	N/A
Morisi et al.[43]	Self-Acquired(47PD 22PSP 9MSA-C 7MSA-P)	MRI	REC	SVM	LOO	N/A	87.00	90.00	84.00
Rana et al.[44]	Self-Acquired(30PD 30HC)	MRI	Graph-Theory-Based SFS	SVM	LOO	N/A	86.67	90.00	83.33
	PPMI(27PD 27HC)						88.89	96.29	81.48

Table 2.1: The state-of-the-art studies on PD detection using MRI data (continued).

Study	Database	NI	FS	Classification	CV	AUC	ACC	SEN	SPE
Rana et al.[27]	Self-Acquired(30PD 30HC)	MRI	MI-based FFS	SVM (GM)	LOO	N/A	86.67	90.00	83.33
				(WM)			86.67	86.67	86.67
				(GM+WM)			83.33	76.67	90.00
Rana et al.[33]	Self-Acquired(30PD 30HC)	MRI	mRMR	SVM	LOO	N/A	88.33	90.00	86.67
Rana et al.[16]	Self-Acquired(30PD 30HC)	MRI	FFD-LBP	SVM	LOO	N/A	95.00	93.33	96.67
					10-Fold	N/A	89.67	86.67	92.67
Talai et al.[45]	Self-Acquired(56PD21 PSP)	MRI	SVM based FS	SVM	LOO	94.90	94.80	94.60	98.10
Cigdem et al.[8]	**PPMI(40PD 40HC)**	**MRI**	**PCA**	**SVM**	**10-Fold**	**N/A**	**93.75**	**95.00**	**92.50**
Cigdem et al.[46]	**PPMI(40PD 40HC)**	**MRI**	**PCA (GM)**	**SVM**	**10-Fold**	**N/A**	**86.25**	**82.50**	**90.00**
Galvis et al.[47]	PPMI(131PD 58HC)	MRI	ICC+t-test	SVM	10-Fold	N/A	60.10	N/A	N/A
Kamagata et al.[48]	Self-Acquired(21PD 21HC)	MRI	Graph-based Theory	SVM	10-Fold	85.28	78.33	85.00	81.67
Peng et al.[49]	PPMI(69PD 103HC)	MRI	mRMR + SVM-RFE	SVM	10-Fold	83.63	85.78	87.64	87.79
Peng et al.[50]	PPMI(77PD 32HC)	MRI	mRMR + t-test + SVM	SVM	10-Fold	97.44	92.35	90.35	94.31
Singh et al.[51]	PPMI(50PD 50HC 50SWEDD)	MRI	FDR	SVM (GM)	10-Fold	N/A	99.00	98.00	100.00
				SVM (WM)	10-Fold	N/A	88.00	86.00	90.00
Gellerup et al.[53]	Self-Acquired(24PD 21HC)	MRI	mRMR	PSVM	5-Fold	N/A	84.00	74.00	93.00

Table 2.1: The state-of-the-art studies on PD detection using MRI data (continued).

Study	Database	NI	FS	Classification	CV	AUC	ACC	SEN	SPE
Zeng et al.[89]	Self-Acquired(45PD 40HC)	MRI	RFE	SVM	632-Fold	N/A	97.70	97.80	97.50
					LOO	N/A	96.90	98.10	95.50
					5-Fold	N/A	98.10	97.50	96.90
					2-Fold	N/A	95.60	99.60	93.80
Fu et al.[90]	N/A	MRI	mRMR,IG,Relief,t-test	BFO-SVM	10-Fold	N/A	96.90	98.75	90.83
				PSO-SVM			94.89	N/A	N/A
				GRID-SVM			93.87	N/A	N/A
Gao et al.[77]	Self-Acquired(251PD)	MRI	KO+RF	LR	5-Fold	81.70	77.30	43.00	95.20
				RF		77.40	70.50	45.30	83.60
				AdaBoost		76.50	71.70	55.80	80.00
				XGBoost		78.10	74.50	54.70	84.80
				SVM		78.50	77.70	51.20	91.50
				Neural Network		N/A	66.10	51.20	73.90
				SuperLearner		N/A	72.90	45.30	87.30
Liu et al.[56]	PPMI(56PD 56HC)	MRI	ICCA	RLDA	10-Fold	71.10	70.50	N/A	N/A
Liu et al.[57]	PPMI(56PD 56HC)	MRI	ICCA	RLDA	10-Fold	71.10	70.50	62.50	78.60
Adeli et al.[59]	PPMI(374PD 169HC)+Synthetic(200)	MRI	JFSS	LS-LDA	10-Fold	N/A	91.90	N/A	N/A
Adeli et al.[60]	PPMI(374PD 169HC)	MRI	RPCA	RFS-LDA	10-Fold	N/A	79.80	N/A	N/A
Adeli et al.[58]	PPMI(56PD 56HC)+Synthetic (200)	MRI	JFSS	LS-LDA	10-Fold	N/A	82.50	N/A	N/A
Banerjee et al.[61]	Self-Acquired(46PD 22HC)	MR	EAP	CDT	LO	N/A	98.53	98.00	100.00

Table 2.1: The state-of-the-art studies on PD detection using MRI data (continued).

Study	Database	NI	FS	Classification	CV	AUC	ACC	SEN	SPE
Lin et al.[62]	Self-Acquired(72PD 73HC)	MRI	LOOCV based	BGPC+DDC	LOO	N/A	74.00	76.00	73.00
Salvatore et al.[9]	Self-Acquired(28PD 28HC)	MRI	PCA	LR	LOO	N/A	85.80	86.00	86.00
Wang et al.[63]	Self-Acquired(19PD 27HC)	MRI	PCA	RVM	LOO	N/A	89.13	78.95	96.30
Morales et al.[64]	Self-Acquired(16PDCI 15PDMCI 14PDD)	MRI	Filter-based	CFS-NB	5-Fold	N/A	97.00	93.33	100.00
Babu et al.[32]	PPMI(127PD 112HC)	MRI	PBL-McRBFN-RFE	PBL-McRBFN	HO-75%	N/A	87.21	87.39	87.00
Pahuja et al.[65]	PPMI(30PD 30HC)	MRI	GA	GA-ELM	HO-80%	N/A	94.87	92.45	97.30
Amoroso et al.[66]	PPMI(374PD 169HC)	MRI	RFE+CS	SVM	10-Fold	97.00	93.00	93.00	92.00
Gu et al.[78]	Self-Acquired(52PD 45HC)	MRI	RFE+DTI	SVM	LOO	95,85	92.31	84.21	96.97
Lei et al.[67]	PPMI(123PD 29SWEDD 56HC)	MRI	ILFFS+DTI	SVM	10-Fold	84.17	83.28	N/A	N/A
Lei et al.[68]	PPMI(123PD 29SWEDD 56HC)	MRI	RRFS+DTI	SVM	10-Fold	84.40	84.40	75.70	83.10
Lei et al.[69]	PPMI(123PD 29SWEDD 56HC)	MRI	LDA+LPP+p-norm	SVM (CSF)	10-Fold	81.62	67.58	58.48	54.83
				SVM (WCF)		80.24	66.35	58.05	55.80
				SVM (GWCF)		82.34	67.37	61.22	54.90
				SVM (GCF + 3CFSs)		86.63	65.34	41.94	60.48
				SVM (GCF + DSSM)		91.22	78.21	84.39	56.23
				SVM (GCF + 3CDs)		94.79	79.25	85.45	66.32

Table 2.1: The state-of-the-art studies on PD detection using MRI data (continued).

Study	Database	NI	FS	Classification	CV	AUC	ACC	SEN	SPE
Cherubini et al.[70]	Self-Acquired(15PD 15rET)	MRI+SPECT	f-test	SVM	LOO	N/A	100.00	100.00	100.00
Adeli et al.[71]	PPMI(369PD 169HC)+Synthetic(200)	MRI+SPECT	Joint Kernel-Based	Max-Margin	10-Fold	N/A	97.50	N/A	N/A
Prashant et al.[72]	PPMI(401PD 183HC)	MRI+UPSIT	Wilcoxon Ranksum Statitical Test	SVM	10-Fold	98.88	96.40	97.03	95.01
				RF		98.40	96.18	97.55	93.15
				Boosted Trees		98.23	95.08	96.07	92.90
				Logistic Regression		98.66	95.63	96.78	93.26
				NB		96.77	93.12	92.67	93.52
Haller et al.[73]	Self-Acquired(16PD 20OtherPDs)	SWI	Relief FS	SVM	10-Fold	N/A	86.92	87.00	87.00
Singh et al.[54]	PPMI(518PD 245HC 68SWEDD)	MRI	FDR+WAT	LS-SVM	HO-80%	N/A	93.25	97.12	85.03
Singh et al.[55]	PPMI(518PD 245HC 68SWEDD)	MRI	SOM	LS-SVM	HO-80%	N/A	7.42	71.90	94.70

Chapter 3

METHODOLOGY

3.1 Introduction

In this chapter, an automatic CAD method for PD classification using sMRI data has been studied. The scheme of the method includes databases, image acquisition, pre-processing, feature selection, and classification methods.

3.2 Database and Image Acquisition

The sMRI data utilized in this thesis are obtained from PPMI datasets (www.ppmi-info.org/data). 50 PDs and 50 HCs data are downloaded, however among 100 persons, 10 HCs and 10 PD patients' MR data were excluded due to mismatch X/Y/Z matrix size, hence the failure of the segmentation method. For the present study, 40 PD patients (mean age±standard deviation = 60.37±8.63 years, range: 40.0-75.2 years, gender: 19M-21F) and 40 HCs (mean age ± standard deviation = 60.09 ±10.35 years, range: 32.5-78.9 years, gender: 27M-13F) are used. The data selection parameters are provided in Table 3.1.

Table 3.1: Data selection parameters.

Acquisition Type = 3D	TE = 2.98 ms	Matrix X = 240 pixels
Acquisition Plane = SAGITTAL	TR = 2300.0 ms	Matrix Y = 256 pixels
Manufacturer = SIEMENS	TI = 900.0 ms	Matrix Z = 176 pixels
Slice Thickness = 1 mm	Flip angle = 9.0	Degree Weighting = T1

3.3 MRI Data Pre-processing and Statistical Analysis

The preprocessing plays a crucial role in CAD. In this thesis, the 3D Tl-weighted sMRI images from PPMI database are pre-processed by using an SPM package [129] and its extension called CAT [130]. The SPM software package is designed to investigate the brain neuroimaging data sequences, namely sMRI, fMRI, MEG, SPECT, and PET. It constructs and evaluates the spatially extended processes which are utilized to test hypotheses about neuroimaging data. The CAT toolbox, the extension of SPM and newer version of VBM, provides the computational anatomy and it includes different morphometric techniques, namely VBM, Region/Label-Based Morphometry (RBM), Surface-Based Morphometry (SBM), and DBM. The CAT toolbox is more robust and accurate than VBM package for detecting the small morphological abnormalities [92]. In this thesis, the version 12 of CAT toolbox with version 12 of SPM software are used for pre-processing the data.

The 3D high-resolution image data are downloaded from PPMI database in Digital Imaging and Communications in Medicine (DICOM) formats. By using SPM12, all DICOM files are converted into NifTi formats as seen in Fig. 3.1. First of all, the AC points of all subjects are co-registered to the central point spaces to have every image with the same central locations. A VBM technique is an automated method to analyze tissue volumes between different subject groups. It takes into account the whole brain structure by comparing it voxel-by-voxel and discriminates the degenerated tissue concentrations by referencing the brain of HC as a template.

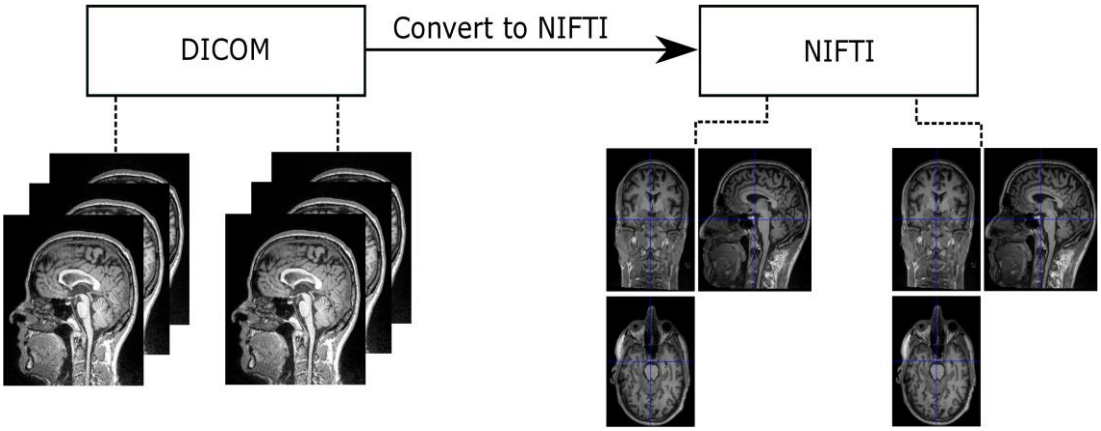

Figure 3.1: Converting the DICOM format to the NifTi one by using SPM12.

Then by using the SPM12 together with the CAT12 toolbox, the data are segmented into six tissue probability maps regarding to the existing templates for each of six modalities which are GM, WM, CSF, skull, scalp and air cavities as seen in Fig. 3.2. In this study, only the GM and WM tissues have been analyzed.

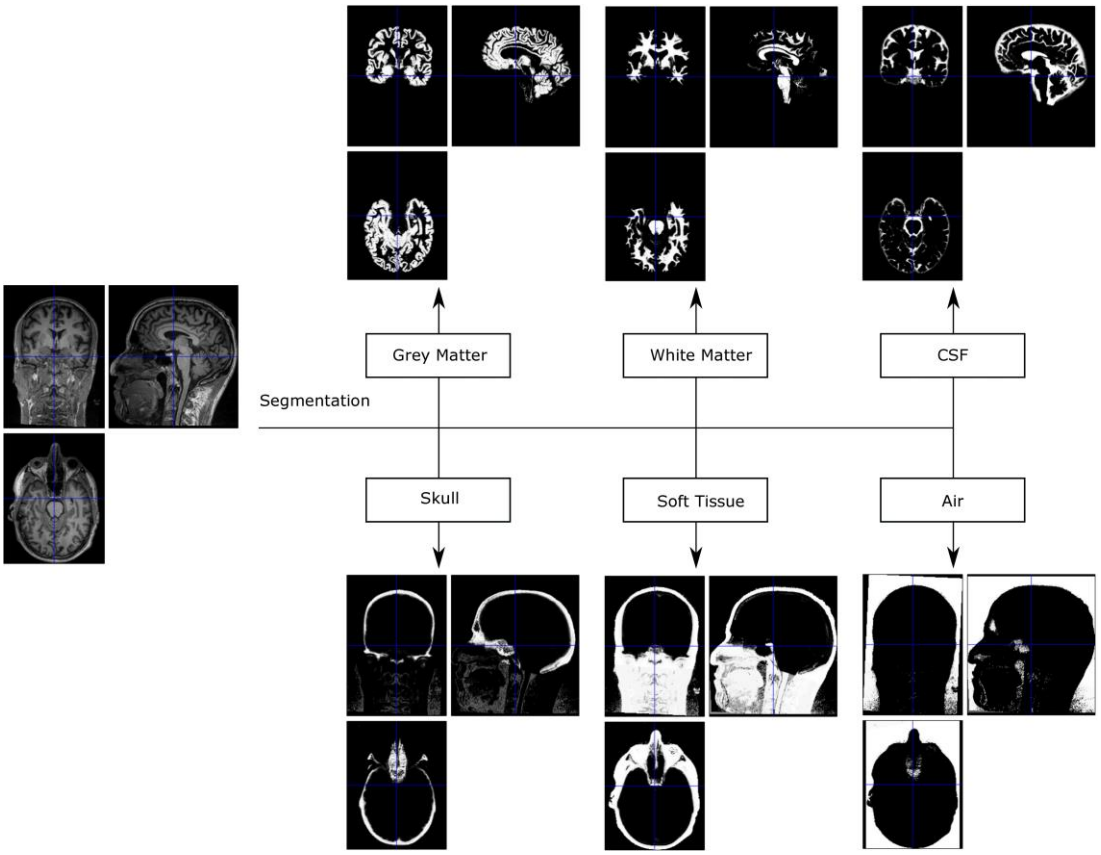

Figure 3.2: Segmentation the sMRI image into six modalities which are GM, WM, CSF, skull, scalp and air cavities.

Using the segmented GM and WM tissue probability maps together with the existing DARTEL template captured from 555 HCs, the flow field of each subject for GM modality is determined as given in Fig.3.3

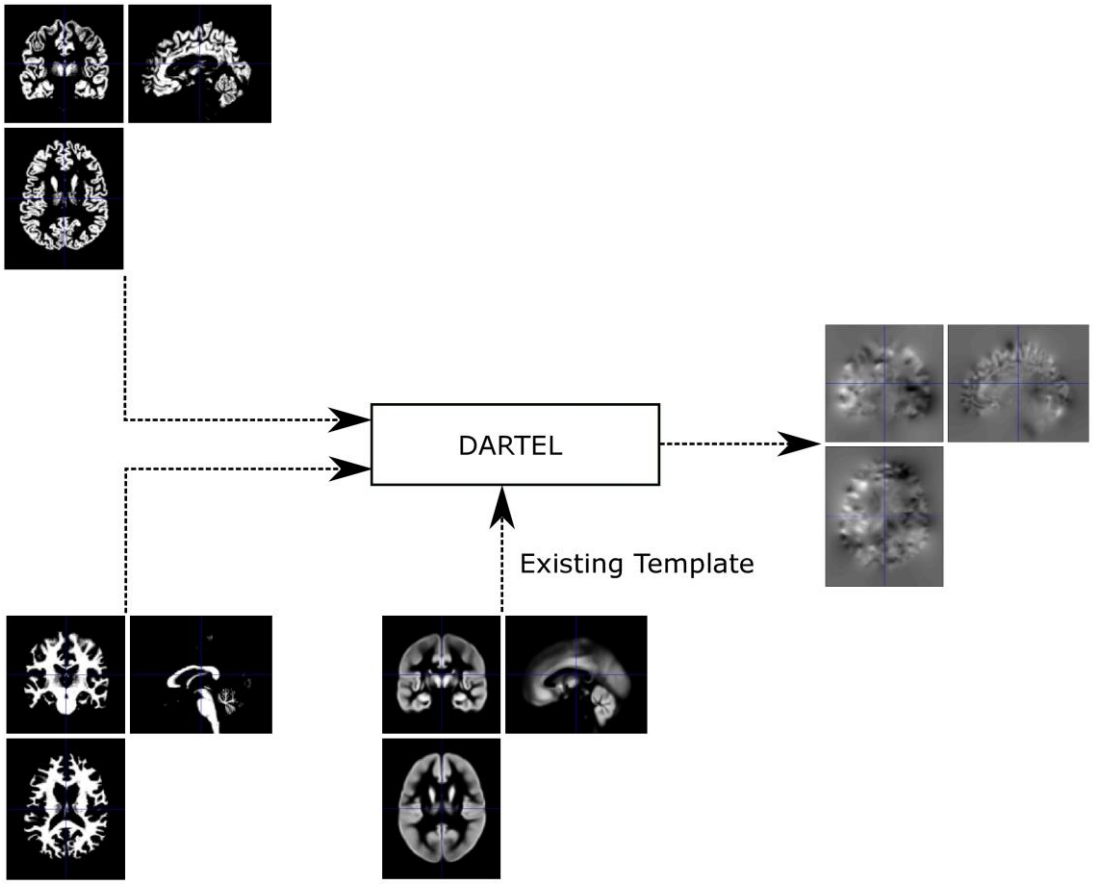

Figure 3.3: Obtaining flow fields of each subject by using the segmented GM (left top) and WM (left bottom) images together with the existing DARTEL template image.

After GM and WM tissues are obtained, they are normalized with a DARTEL toolbox [2, 93]. It has the idea of registering images by computing a flow field for each subject. Firstly, images are imported and the parameter files are extracted which are obtained in the segmentation step. Then it writes the transferred versions of the tissue class images out in a way that they are aligned with the tissue probability maps as close as possible [91]. The mean of all images are used as an initial template and deformations of each image from this template is calculated. Then the template is updated by

applying the inverse of deformations to each of the image and taking the mean again. This pipeline is performed many times and finally, the warped images are obtained [91]. Therefore, using DARTEL has more precise inter-subject alignment of MRI images than VBM [2, 8]. Additionally, it gives precise and accurate localization of structural deformation on the MRI images. In this thesis, in order to create deformation fields of each data, instead of creating a template for the studied data, an existed DARTEL template which has been created from 555 healthy controls is utilized. By using DARTEL normalization, all data are registered to standard MNI space that includes both affine transformation and nonlinear deformation. An example of normalized to MNI space process is given in Fig. 3.4. Later on, the segmented, normalized images are modulated by the preserving amount which preserves total amounts of tissue corrected for individual differences in brain size, and finally spatially smoothed with an 8 mm Full-Width-Half-Maximum (FWHM) Gaussian kernel. All other parameters of SPM12 are kept as default settings.

Finally, in order to design a General Linear Model (GLM) for the segmented, DARTEL-warped, modulated, and smoothed images by using the statistical analysis for the GM and WM modalities individually, the parameter of covariate needs to be defined. In literature, different covariates such as TIV, age, and sex have been taken into account [7, 8, 33, 33]. Since these covariates might have impact on creating the model, different combinations of them are considered. In this thesis, after analyzing different combinations of TIV, age, and sex covariates, only the TIV one are provided in model building since it gives consistent results for both the GM and WM datasets [8]. A voxel-wise two sample t-test is used to set up the model of the GM and WM

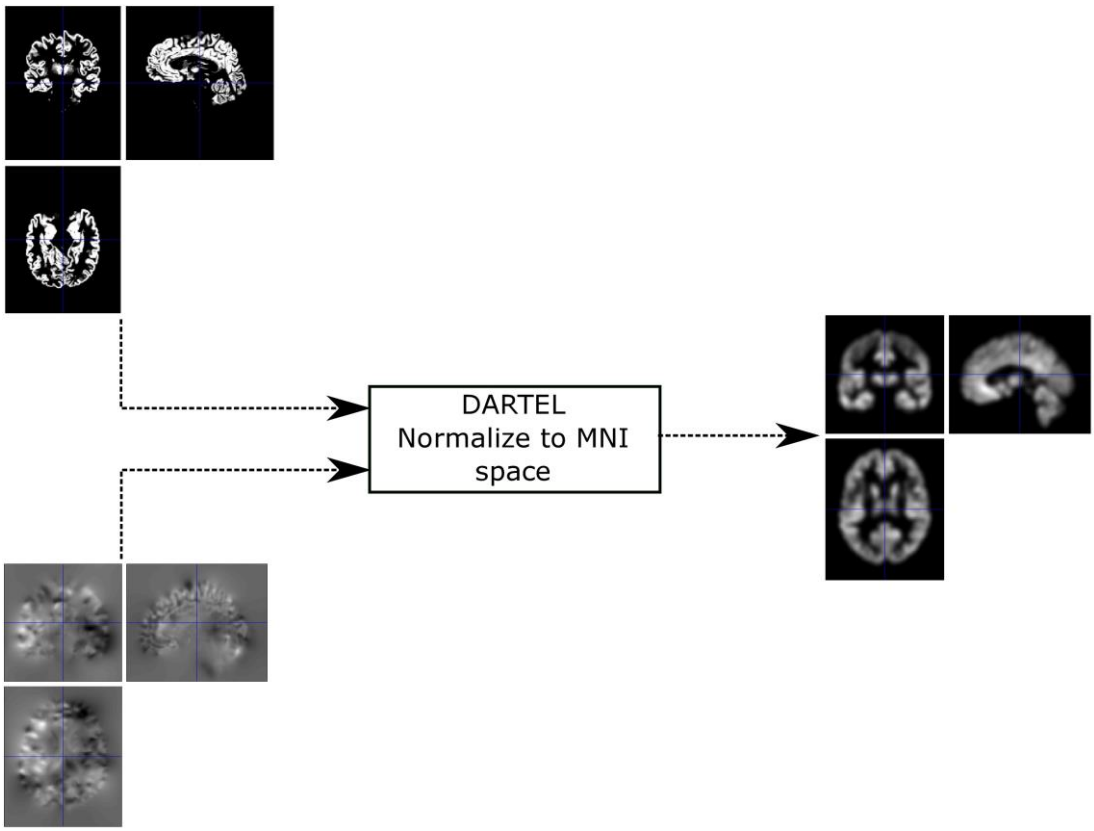

Figure 3.4: Obtaining the segmented GM, DARTEL normalized, modulated, and smoothed data (right) by using the flow field of each image (left bottom) together with the segmented GM image (left top).

datasets individually. After the GLM is specified, the parameters of the model given in Eq. 3.1 are needed to be estimated.

$$Y = X\beta + e \tag{3.1}$$

where Y is the data, X is the design matrix, β is the estimated parameter, and e is the additive noise. As it is known, all parameters might not be estimable by taking rather unlikely model which includes the same regressor twice, such that x_1 and $x_2 = x_1$ have the corresponding β_1 and β_2 parameters. Hence, the information on how to select the β_1 compared to β_2 is not included in the data. In this case, any solution of $\hat{\beta}_1 + \hat{\beta}_2 = constant$ gives the same fitted data and residuals, yet an infinity of possible $\hat{\beta}_1$ and

$\hat{\beta}_2$ [94]. To overcome this, linear functions of the parameter estimates are considered as in Eq. 3.2.

$$\lambda_1\hat{\beta}_1 + \lambda_2\hat{\beta}_2 + \ldots = \lambda^T\hat{\beta}. \quad (3.2)$$

The constants λ_i are the coefficients of a function that contrasts the parameter estimates. The contrast vector is given as $\lambda^T = [\lambda_1\lambda_2\ldots\lambda_p]$ where p denotes the number of parameters in design matrix X. Therefore, a contrast is a linear combination of the parameters $\lambda^T\hat{\beta}$.

After building and estimating the GLM, in order to obtain the 3D masks, namely the differentiated VOIs, a hypothesis needs to be provided to the software by using contrast manager from the SPM Stats menu. The contrast evaluates whether significant differences are captured over canonical regressors and generates the 3D masks which encapsulate the most discriminative voxels between two data groups. There are two alternatives hypotheses embedded in the software, a t-contrast and an f-contrast. The t-contrast in SPM is the simple combinations of the β_i, they are either positive or negative ($\beta_1 - \beta_2$ is different from $\beta_2 - \beta_1$). However, the f-contrast in SPM is the sum of the squares of one or several combinations of the β_i [94].

3.3.1 The t-contrast

In order to generate a 3D mask by using SPM12, the contrast type, t-contrast or f-contrast, is required to be defined. A simple contrast for an SPM{T} tests the null hypothesis, $H_0: \lambda^T\hat{\beta} = 0$, against the 1-tailed alternative hypothesis, positive-tailed $\lambda^T\hat{\beta} > 0$ or negative-tailed $\lambda^T\hat{\beta} < 0$. The t-contrast considers only one direction either

positive-tailed (e.g., the mean of GM volume in HC is higher than that in PD) or negative-tailed (e.g., the mean of GM volume in PD is higher than that in HC).

The value of t statistics is calculated as the contrast of estimated parameters divided by the square root of variance estimate. A large t-value indicates that the sample estimate is less precise, since it has more random error. The general equation of t-contrast is given in Eq. 3.3.

$$t = \frac{\lambda^T \hat{\beta}}{\sqrt{var(\lambda^T \hat{\beta})}} \tag{3.3}$$

where

$$var(\lambda^T \hat{\beta}) = \hat{\sigma}^2 \lambda^T (X^T X)^- X^T \sum_i X (X^T X)^- \lambda \tag{3.4}$$

The t-contrast may also given as in Eq. 3.5.

$$t = \frac{\bar{\mu}_1 - \bar{\mu}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{3.5}$$

where $\bar{\mu}_1$ and $\bar{\mu}_2$ are the mean of first and second set of values respectively.s_1 and s_2 are the standard deviation of first and second set of values, respectively. n_1 and n_2 are the total number of elements in the first and second set of values, respectively.

3.3.2 The f-contrast

As in t-contrast, f-contrast is also evaluated through considering null and alternative hypotheses. The null hypothesis $\lambda^T \hat{\beta} = 0$ might be seen as a linear constraint on the full model under consideration, giving a reduced model [94]. For the full model which

is an extension of the reduced model, the null hypothesis states that the additional terms in the full model are redundant. The f-contrast compares the variances of the groups and rejects the null hypothesis if the variances of them are not equal without considering any tailed limitations. While t-contrast is 1-tailed hypothesis which is said to have directionality, f-contrast takes into account effects of interests [95, 96, 97].

In statistical inference, the value of f statistic is the ratio of explained variability and unexplained variability (error). In other words, the f value is calculated as additional variance accounted for by tested effect divided by error variance estimate. The f-statistic simply computes the extra sum of squares for each voxel. The high values of f statistic represent evidence against the null hypothesis $\lambda^T\hat{\beta} = 0$ in favor of the 2-tailed alternative $\lambda^T\hat{\beta} \neq 0$. The general equation of f-contrast is given in Eq. 3.6.

$$F = \frac{(Y^T(I-P_{X_0})Y - Y^T(I-P_X)Y)/v_1}{Y^T(I-P_X)Y)/v_2} \tag{3.6}$$

The v_1 and v_2 are defined as:

$$v_1 = tr((R_0 - R)\sum_i) \tag{3.7}$$

$$v_2 = tr(R\sum_i) \tag{3.8}$$

where R_0 is the projector onto the residual space of X_0, reduced design matrix, and P_X is the orthogonal projector onto X, full model design matrix [94].

The f-contrast follows the F-distribution with degrees of freedom $d_1 = K - 1$ and $d_2 = N - K$ under the null hypothesis. The value of f is large if the between-group variability is large relative to the within-group variability. The general representation of f-contrast is given in Eq. 3.9

$$F = \frac{\text{explainedvariance}}{\text{unexplainedvariance}} = \frac{\text{between-groupvariability}}{\text{within-groupvariability}} \tag{3.9}$$

The explained variance or between-group variability is given as:

$$\sum_{i=1}^{K} n_i(\bar{Y}_i - \bar{Y})^2/(K-1) \tag{3.10}$$

where $\bar{Y}_i$ represents the sample mean in the i^{th} group, n_i is the number of observations in the i^{th} group, $\bar{Y}$ represents the overall mean of the data, and K represents the number of groups.

The unexplained variance or within-group variability is given as:

$$\sum_{i=1}^{K} \sum_{j=1}^{n_i} \left(Y_{ij} - \bar{Y}_{i\cdot}\right)^2/(N-K), \tag{3.11}$$

where Y_{ij} denotes the j^{th} observation in the i^{th} out of K groups and N denotes the overall sample size.

3.4 Feature Selection

The generated 3D masks for GM and WM tissues are multiplied with the DARTEL normalized, modulated, and smoothed GM and WM tissues, respectively. Therefore, only the voxels clustered in the VOIs are extracted from the whole GM and WM tissue maps. The dimensions of the extracted GM and WM data are reduced significantly

through feature extraction. In this study, in addition to the two extracted GM and WM datasets, another dataset is created by concatenating the extracted GM and WM VOIs.

3.4.1 PCA Dimensionality Reduction

In order to eliminate irrelevant and redundant data without losing relevant information, feature selection is necessary. PCA is a widely used dimensionality reduction method [9]. It searches for the most informative components which are orthogonal to each other in the original data distribution and uses these components as PCs. In [9], PCA projects the higher number of original data into a much lower number of PCA coefficients which are used for classifications. However, in this study, PCA is applied to reduce the dimension of the 3D masked data instead of the original data. The number of components for all experiments are determined by using 95% of total variance.

3.4.2 Feature Ranking

In order to alleviate the effect of the curse of dimensionality, improve the performance of the designed model, reduce the learning process time, and enhance data understanding, an FS method is required. There are mainly three FS techniques, namely filters, wrappers, and embedded methods. In filter methods, irrelevant features are filtered out independently from the learning algorithm. The classifier is ignored and only the intrinsic properties of the data are examined. First, the features are taken into account exclusively and ranked according to their importance level. Then a feature subset is obtained from the ranked ones. In wrappers methods, classifiers are considered in order to score a given subset of features. Hence, in the wrapper approach, there is need to evaluate a specific learning algorithm and to determine which features are selected. In embedded ones, a selection process is injected into the learning of the classifier [98]. With wrappers and embedded methods, first, the feature subsets are sampled and evaluated. Then they are kept as the final outputs [98]. The wrappers and

embedded methods tend to find the more optimal features for the specific learning algorithms, hence the computational cost is high. However, the filter methods are fast and easy to implement. In this study, filter feature ranking methods including both supervised and unsupervised ones are studied for detection of PD. The FS method, learning type, and computational complexity of the used five feature ranking approaches are given in Table 3.2.

Table 3.2: The class (Cl.) and the computational (Com.) complexity of the five filter methods based feature selection approaches are given. T, for the number of samples; n, for the number of initial features; C, for the number of classes; d, for the number of selected features; i, for the number of iterations; s, for supervised learning; u, for unsupervised learning are used.

FS	Cl.	Com. Complexity
Relief-F	s	O(iTnC)
Laplacian	u	N/A
MCFS	u	O($nT^2 + Cd^3 + TCd^2 + n\log n$)
UDFS	u	N/A
CFS	u	O($\frac{n^2}{2}$T)

3.4.2.1 mRMR: minimum Redundancy Maximum Relevance

mRMR is a forward feature selection method which aims to select the features mutually different from each other while still having a high correlation to the classes [33]. For a dataset of $X = \{x_1, x_2, x_3, K, x_d\}$ given that $x_i \varepsilon R^M$, where M indicates the number of samples and d indicates the number of features, mRMR selects a set **S** of k lower than d features which maximizes the following criterion:

$$J(S) = \frac{1}{|S|}\sum_{x_i \varepsilon S} I(x_i, c) - \frac{1}{|S|^2}\sum_{x_i, x_j \varepsilon S} I(x_i, x_j) \qquad (3.12)$$

where $I(x, y)$ indicates mutual information between the two random variables x, and y. The mutual information is defined as follows:

$$I(x, y) = \sum_{i,j} p(x_i, y_j) log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \quad (3.13)$$

where $p(x_i)$ and $p(x_y)$ are marginal probability distributions of x and y, respectively and $p(x_i, y_j)$ is the joint probability distribution of x and y. The higher the mutual information score, the higher dependencies between the random variables [33].

3.4.2.2 Relief-F

Relief-F is an iterative, randomized, and supervised approach in which the quality of the features are estimated according to how well the values of the features separate close data samples. Discrimination is not done among redundant features, hence using few data might decrease the performance of the algorithm [99]. In the Relief-F algorithm, the weight of a feature vector is identified regarding the feature relevance. The feature weights are obtained by solving a convex optimization problem. However, these feature weights may fluctuate with the instances and due to the uncertainty in sampling frequency. Therefore, the Relief-F algorithm is unstable and might reduce the expected accuracies [98].

3.4.2.3 LS:Laplacian Score

LS is an unsupervised method, in which the significance of a feature is evaluated based on how powerfully it preserves the locality of the data manifold structure. The nearest neighbor graph is designed to model the local geometric structure. LS method searches for the features that preserve the existing structure of the graph.

3.4.2.4 MCFS: Unsupervised Feature Selection for Multi-Cluster Data

MCFS is an unsupervised filter method which selects the relevant features such that the multi-cluster structure of the data might be preserved the best [100]. The MCFS

method is inspired from the manifold learning and the L1-regularized models for the subset selection.

3.4.2.5 UDFS: Unsupervised Discriminative Feature Selection

In UDFS, the most discriminative feature subsets are ranked in batch mode, namely by taking into account the structure of the manifold [100]. It exploits local discriminative information and feature correlations simultaneously [102].

3.4.2.6 CFS: Correlation-Based Feature Selection

CFS proposes a metric to assess a feature subset. The main purpose of CFS is to reduce feature-to-feature correlation (r_{ff}) and increase feature-to-class correlation (r_{fo}). In other words, according to CFS, good feature sets include features which have a high correlation with the class, but uncorrelated with each other. In 3.14, higher ratio represents a better subset.

$$\frac{r_{fo}}{r_{ff}} \tag{3.14}$$

3.4.2.7 LLCFS: Feature Selection and Kernel Learning for Local Learning Based Clustering

In LLCFS, the goal is to obtain an appropriate data representation through FS or kernel learning within the framework of the LLC approach which can outperform the global learning-based ones when dealing with the high-dimensional data lying on manifold. A weight is assigned to each feature or kernel as well as it is added into the built-in regularization of the LLC algorithm to consider the relevance of each feature or kernel for the clustering [52].

3.4.3 Feature Selection Based on Fisher Criterion Methods

The aim of FS is to find the feature subset of a certain size which causes the largest possible generalization or minimal risk [103]. In order to determine the optimal feature subset which has the number of top discriminative features, an automatic approach based on the FC, J(w), given in Eq. 3.15 is used [31].

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \tag{3.15}$$

where S_B and S_W indicate the determinant of between class and within class scatter matrices, respectively. The between class scatter and within class scatter matrixes for y_1 and y_2 classes are defined as follows:

$$S_B = (\mu_{y_1} - \mu_{y_2})(\mu_{y_1} - \mu_{y_2})^T$$
$$S_W = \sum_{x_i \varepsilon y_1} (x_i - \mu_{y_1})(x_i - \mu_{y_1})^T + \sum_{x_i \varepsilon y_2} (x_i - \mu_{y_2})(x_i - \mu_{y_2})^T m \tag{3.16}$$

where $w = {S_w}^{-1}(\mu_{y_1} - \mu_{y_2})$ and μ_{y_i} is the mean on the data in each class. The number of top discriminative features from the ranked datasets is selected adaptively in each set of training data instead of a fixed number of features. The number of top-ranked features increases iteratively and the respective FC value is calculated for each iteration. The iterations are performed until a maximum FC value is obtained and the optimal number of top features is determined by taking the number of top-ranked features maximizing the FC.

3.5 Classification Methods

In order to separate the PD patients apart from the HCs, classification algorithms need to be used. In literature, many different classifiers are used to train a model. In this

thesis, five different classification methods are used for training the model, namely k nearest neighbor classifier, naive Bayes, ensemble-subspace discriminant, ensemble-bagged trees, and support vector machines.

3.5.1 K Nearest Neighbor

In kNN classification, an object is classified using majority vote of its neighbors, namely the object is simply labeled by the most common class label among its k nearest neighbors measured by a distance function. Hence, the output in this classification algorithm is a class membership. In this thesis, a Euclidean distance with a default k value, and a squared inverse distance weight are selected and 10-fold CV is used. The drawback of k-NN is that equal weight is intrinsically assigned to each feature by comparing the distances. Hence, for irrelevant and noisy feature datasets, the accuracy results of the classification might be insufficient [87].

3.5.2 Naive Bayes

NB is a classification algorithm which assumes the independence of the predictors. In other words, in NB classifier, it is assumed that the effect of a feature in a given class is unrelated to the many other existing features [104]. By doing so, the computational cost is reduced. Since there is no exhaustive iterative parameter estimation, an NB algorithm performs well, especially on large datasets. The main idea behind the NB classification is to try and classify data by maximizing the $P(C_i|X)$.

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \tag{3.17}$$

where for a class index, i, for measurements, X, and for the classes, C are used. In this thesis, the hyperparameters of NB are optimized to minimize the CV loss in an NB classifier.

3.5.3 Ensemble Subspace Discriminant

The error in machine learning is mostly occurred owing to the noise, the bias, and the variance. Ensemble learning combines multiple learners in order to decrease the variance and improve the predictive performance of the machine learning models compared to a single classifier. A subspace ensemble method is used to improve the accuracy of discriminant analysis classifier which is a Gaussian mixture model for data generation. It has the advantage of overcoming the missing values (NaNs) problem and using less memory than ensembles with all predictors [105]. In this thesis, 30 for the number of learner and 175 for subspace dimension are selected and 10-fold CV is used.

3.5.4 Ensemble Bagged Trees

In bagging mate-algorithm, also known as bootstrap aggregation, each model in the ensemble is voted with equal weight and each model is built independently. In order to obtain the data subsets for training base learners, a bagging method takes into account a bootstrap sampling. Each model in the ensemble is trained by using a randomly decided subset of training sets. Decision trees might suffer from the variance and taking the average of multiple estimates of an estimate by using bagging algorithm is a way of reducing the variance. As a result, the higher accuracy is achieved by ensemble bagged trees method [98]. In this thesis, 30 for the number of learners, 79 for the maximum number of splits are selected and 10-fold CV is used.

3.5.5 Support Vector Machines

SVM classification algorithm is used in order to classify PD patients apart from HCs. It is a powerful classifier based on principle of structural risk minimization. SVM classification has been used in many different machine learning application successfully [2, 27]. The main idea behind SVM is to search an optimal class-

separation hyperplane in the maximal margin. The training is performed by using classification learner application [131]. In this application, there are several algorithms to train and validate classification models for binary or multiclass problems. It provides comparing the validation errors side-by-side after obtaining multiple trained models and allows to extract the highest accuracy model. There are various kernel types of SVM such as linear, quadratic, cubic, and gaussian. In this study, RBF-SVM classifier has been used. The RBF model has two parameters that need to be selected, C (regularization) and γ (controls the kernel width), in which the performance of the classifier depends on these parameters.

3.5.6 Validation Process

To evaluate the performance of the classifier, a procedure of two CVs was combined with a grid search. This was done to elude unwarp bias in the estimation of accuracies produced by the CV procedure. This procedure consisted of two nested loops. In the outer loop, the data were split into K_1 folds. At each step, one fold was used as a test and the remaining $K_1 - 1$ folds were used for training and validation. In the inner loop, the training data ($K_1 - 1$ folds) were further divided into K_2 folds. For each combination of C and γ, the classifier was trained using the training data and its performance was assessed using the fold left for validation by estimating the classification accuracy. One fold was left for validation and the remaining $K_2 - 1$ fold for training was combined with the grid search to determine the optimal parameters. In the grid search, the values of C and γ varied logarithmically from 2^{-5} to 2^{20} and from 2^{-15} to 2^{15}, respectively. The inner loop was repeated K_2 times and the accuracy of the classifier was obtained across the K_2 folds for every combination of C and γ. Optimal parameters were selected that produced maximum average accuracy across the K_2 folds. Then, the class label of the test data was predicted, which was left out in

the outer loop using selected optimal parameters. The above procedure was repeated K_1 times by leaving a different fold as test data, which was used to compute the classification accuracy[2, 8]. In this work, $K_1 = 10$ and $K_2 = 10$ were used. The pipeline of the 10-fold CV procedure is given in Fig. 3.5.

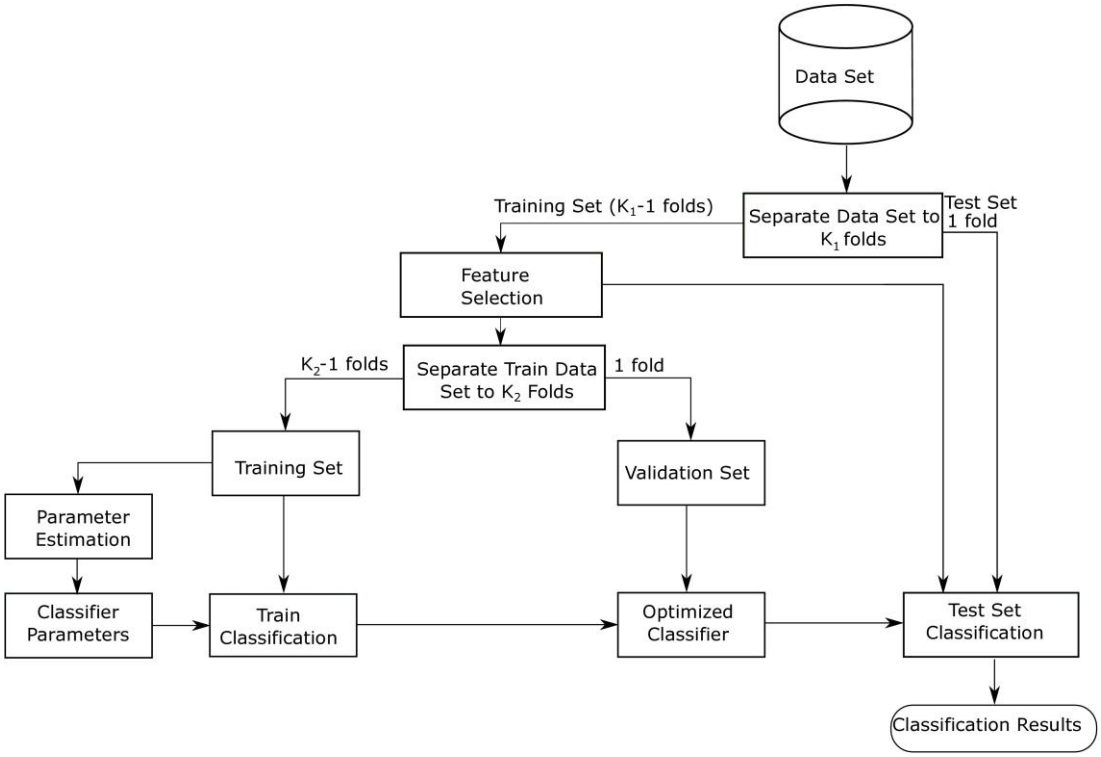

Figure 3.5: The pipeline of the 10-fold cross validation procedure [2].

3.5.7 Performance Evaluation

The classification performances are assessed as a result of ACC, SEN, and SPE. They are identified as follows:

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \tag{3.18}$$

$$SEN = \frac{TP}{TP+FN} \tag{3.19}$$

$$SPE = \frac{TN}{TN+FP} \tag{3.20}$$

where TP, TN, FN, and FP are the number of true positives, true negatives, false negatives, and false positives, respectively. TP, TN, FN, and FP are determined as follows:

TP: The number of patients with PD correctly identified as PDs.

TN: The number of HCs correctly identified as HCs.

FN: The number of patients with PD incorrectly identified as HCs.

FP: By counting the number of HCs incorrectly identified as PD.

Chapter 4

EFFECTS OF DIFFERENT COVARIATES AND CONTRASTS ON CLASSIFICATION OF PARKINSON'S DISEASE USING STRUCTURAL MRI

4.1 Introduction

In this chapter, sMRI data have been used for classification of PD and HC. Differentiated voxels which are considered as VOIs are obtained by using the VBM method which discriminates PD from HC by analyzing group-wise comparisons of cross-sectional sMRI scans [32]. In order to improve inter-group registration and provide more precise and accurate localization of structural differences of sMRI data, a DARTEL approach has been used with combination of the VBM [106]. After preprocessing of the data, statistical analysis is studied for extracting VOIs through voxel-by-voxel comparison of the brains of PD and HC. In statistical analysis, there are some parameters that need to be provided such as covariates [107]. To detect tissue abnormalities optimally, including the appropriate covariates among TIV, age, and sex is necessary [108]. In various studies, different covariates have been considered. In [109], the effects of TIV for AD dataset have been studied and it is observed that in neurodegenerative diseases, TIV is an important covariate for volumetric analyses of the brain. In [110], all combinations of different covariates have been studied with different versions of gradient drivers, therefore their individual effects are not considered. Also, instead of taking TIV as a nuisance variable, the log(TIV) is taken

with the different version of gradient drivers [110]. In [108], the different covariates have been taken into account for GM and WM tissues, separately. For GM, it is shown that TIV, by itself has a significant impact on explaining structural variances, hence adjustment of only TIV has a remarkable impact on GM volume changes. However for WM, in [108] it is found that all covariates, namely age, sex, and TIV need to be considered. Including age as a covariate increases WM tissue area of PD patients compared to HCs. Furthermore, since there is a large degree of multi-collinearity between TIV and gender, its alteration was found to greatly overlap with TIV-dependent variation. As mentioned in [108], the results for WM are inconsistent and might be found specifically for the analyzed data set. Moreover, while in some studies the effects of age on WM volume have no significant impact [111, 112], in some of them, the volume of WM is increased until middle age followed by a decrease [113, 114], and in some others it is decreased [110, 115]. Having said this, influences of age on WM tissue varies, that is why the use of age as a nuisance covariate is inconsistent. In this study, the effects of covariates on PD have been studied and the experimental results indicate that using TIV as a single covariate in the classification of PD apart from HC is optimal as well as consistent for GM, WM, and combination of them.

In a statistical model building, in order to obtain a 3D mask including only the differentiated voxels, there is a need to define a contrast type either t-contrast or f-contrast. While t-contrast compares the mean of two groups, f-contrast compares variances of the two groups. Therefore, t-contrast is directional (either mean of group one is larger than group two or vice versa), f-contrast checks for the differences between the groups. In literature, for AD studies, t-contrast has been used widely [2, 116]. Since AD makes structural degenerations in the GM brain segment of patients,

it is meaningful to use a 1-tailed t-contrast, where total GM volume of an AD is lower than HC, focusing on atrophies. However, it might not be reasonable to search for only atrophies in PD brain segments. In most PD detection studies, the directional t-contrast hypothesis has been studied [32, 33, 117]. However, in some studies, the effects of either tail of t-contrast have also been examined. In [106], differences in brain volume have been investigated by comparing the performance of using VBM and VBM with DARTEL. By modeling with 2-tailed t-contrast, it is noted that PD leads to specific patterns of alterations in the examined brains. The thalamus (ventral posterior lateral nucleus, VPL) of PD patients is larger than that of the HCs. In [118], a 2-tailed t-contrast has been used for determining the pattern of GM changes between the PD and the HC dataset. It is noted that there are important amounts of voxels determined for both directions of t-contrast. As a result, in many studies that use sMRI data, the classification of PD and HC groups are analyzed by using t-contrast (GM volume of PD is lower than the one of HC or vice versa). Contrary to existing literature, in this chapter, in addition to t-contrast, f-contrast hypothesis which assesses the differences between two groups without any direction limitations has also been investigated. The experimental results indicate that using f-contrast improves the classification accuracy significantly.

In neurodegenerative disorder diagnosis, mostly the alterations in GM using VBM have been considered [9, 32, 117]. Other research work has been done on WM analysis using VBM [16, 119]. In [119], SVM classification has been used for GM as well as WM separately and the obtained classification results were 39.53% and 41.86%, respectively. In [33], different combinations of GM, WM, and CSF have been studied with VBM findings in conjunction with mRMR to consider only relevant features. The

classification accuracy of 88.33% was obtained for both GM+WM and GM+WM+CSF combinations. The experiments are done with self-acquired 30 PD and 30 HC datasets which are not available online. Contrary to these studies, in this chapter, 40 PD and 40 HC data from PPMI datasets (www.ppmi-info.org/data) have been used with different combinations of three covariates as well as two different contrasts and a 93.75% classification accuracy has been obtained. By doing statistical analysis, various numbers of clusters are being generated using VOIs which are the voxels differentiated between brain parts of PD and HC. The obtained clusters are used to generate 3D masks and for both GM and WM, two different 3D masks have been created [2]. Masking the data decreases the amount of raw data significantly. Hence, only the differentiated voxels clustered in VOIs are taken into account. In [2] and other existing AD studies, the number of features after masking is extremely high, since voxel changes between AD and HC are more than the ones between PD and HC. In this study, the generated 3D masks are applied to the GM and WM tissues and only differentiated voxels are extracted. In order to reduce the dimension of the extracted raw data, PCA dimensionality reduction method has been used. The reduced 3D masked data of GM and WM have been classified for PD detection by using SVM classification.

In this chapter, several aspects in PD are provided, such as i) searching for optimal covariates in the VBM design for PD, ii) using t-contrast or f-contrast for model building, and iii) fusing GM and WM for performance improvement. Therefore, different combinations of covariates (TIV, age, TIV+age, TIV+age+sex) with two different hypotheses (t-contrast and f-contrast) are analyzed for all GM, WM, and combination of GM as well as WM tissues. The SVM classification results

demonstrate that among all these different experiments, using only TIV as a covariate and f-contrast gives the highest accuracy of 73.75% for GM, 72.50% for WM and 93.75% for fusion of extracted GM as well as WM VOIs dataset.

4.2 Materials

The materials used in this chapter are explained as MRI acquisition and subjects.

4.2.1 MRI Acquisition

The data used in this research are obtained from the PPMI dataset (www.ppmi-info.org/data). The protocol included T1-weighted MRI images based on a scanner by Siemens with acquisition plane=sagittal, acquisition type=3D, coil=Body, flip angle=9.0 degrees, matrix X/Y/Z=240.0/256/176 pixel, mfg model=TrioTim, pixel spacing X/Y=1.0/1.0 mm, pulse sequence=GR/IR, slice thickness=1 mm, and TE/TI/TR=2.98/900/2300 ms.

4.2.2 Subjects

50 PDs and 50 HCs data are downloaded, yet among 100 persons, 10 HCs and 10 PD patients' MR data were excluded due to mismatch X/Y/Z matrix size, hence the failure of the segmentation method. For the present study, 40 PD patients (mean age$\pm$standard deviation = 60.37$\pm$8.63 years, range: 40.0-75.2 years, gender: 19M-21F) and 40 HCs (mean age $\pm$ standard deviation = 60.09 $\pm$10.35 years, range: 32.5-78.9 years, gender: 27M-13F) are used.

4.3 Methodology of the CAD System

In this section, the methodology is given for three different covariates and two contrasts to design an automatic CAD system for PD detection. First, the preprocessing of 3D sMRI data by using the VBM plus the DARTEL approach is employed. Second, 3D masks are generated for different combinations of three covariates, two contrasts, and different tissue maps, namely GM, WM, and the combination of them. Third, the

dimensions of the 3D masks are reduced by using the PCA method. Finally, the SVM algorithm is used for classification. The framework is given in Fig. 4.1.

4.3.1 MRI Data Pre-processing and Statistical Analysis

The 3D T1-weighted sMRI images are preprocessed by taking advantage of the SPM12 package and its extension called CAT12 implemented in Matlab 2017b. CAT12 is the newer version of VBM8 and it covers diverse morphometric methods such as VBM, SBM, DBM, and RBM. CAT12 is more robust and accurate than VBM8 for detecting the small morphological abnormalities [92].

The 3D high-resolution image data are downloaded from PPMI in DICOM formats. By using SPM12, all DICOM files are converted into NifTi formats. First of all, the AC points of all subjects are co-registered to the central point spaces to have every image with the same central locations. A VBM technique is an automated method to analyze tissue volumes between different subject groups. It takes into account the whole brain structure by comparing it voxel-by-voxel and discriminates the degenerated tissue concentrations by referencing the brain of HC as a template. Then by using the SPM12 together with the CAT12 toolbox, the data are segmented into six tissue probability maps regarding to the existing templates for each of six modalities which are GM, WM, CSF, skull, scalp and air cavities. In this study, only the GM and WM tissues have been examined. After GM and WM tissues are obtained, they are normalized with the DARTEL approach, since it has more precise inter-subject alignment of MRI images than VBM [2]. Additionally, it gives precise and accurate localization of structural deformation on the MRI images. In order to create deformation fields of each data, instead of creating a template for studied data, an existed DARTEL template created from 555 healthy controls is used. By using

DARTEL normalization, all data are registered to standard MNI space that includes both affine transformation and nonlinear deformation. Later on, the segmented, normalized images are modulated by the preserving amount which preserves total amounts of tissue corrected for individual differences in brain size, and finally spatially smoothed with an 8 mm FWHM Gaussian kernel. All other parameters of SPM12 are kept as default settings. The framework is given in Fig. 4.1. Finally, the GLM is described for the segmented, DARTEL-warped, modulated, and smoothed images by using statistical analysis for GM and WM modalities individually. The voxel-wise two sample t-test is used to set up the model of 3D masks of GM and WM individually. In order to configure a design matrix, describing the GLM, data specification, and other parameters are necessary for the statistical analysis. Different combinations of mentioned covariates are assigned for model designing. The design parameters are estimated and the inference on these estimated parameters are handled by using SPM12 Results section. This is done in order to find the volume changes of GM and WM tissues among PD and HC. The analysis is performed with a threshold of uncorrected $p<.001$ and none extend threshold voxels. These experiments are done with both t-contrast and f-contrast for GM as well as WM.

After pre-processing the data and statistical analysis, feature extraction, dimensionality reduction, and classification steps are performed as seen in Fig. 4.1.

4.3.2 Feature Extraction

The generated 3D masks for GM and WM tissues are multiplied with the DARTEL normalized, modulated, and smoothed GM and WM tissues, respectively. Therefore, only the voxels clustered in the VOIs are extracted from the whole GM and WM tissue maps. The dimensions of the extracted GM and WM data are reduced significantly

through feature extraction. In this study, in addition to the two extracted GM and WM datasets, another dataset is created by concatenating the extracted GM and WM VOIs. The pipeline of the procedure is given in Fig. 4.1. The performance of GM, WM, and combination of them are studied, separately.

4.3.3 Dimensionality Reduction

In order to eliminate irrelevant and redundant data without losing relevant information, feature selection is necessary. PCA is a widely used dimensionality reduction method [9]. It searches for the most informative components which are orthogonal to each other in the original data distribution and uses these components as PCs. In [9], PCA projects the higher number of original data into a much lower number of PCA coefficients which are used for classifications. However, in this study, PCA is applied to reduce the dimension of the 3D masked data instead of the original data. The number of components for all experiments are determined by using 95% of total variance.

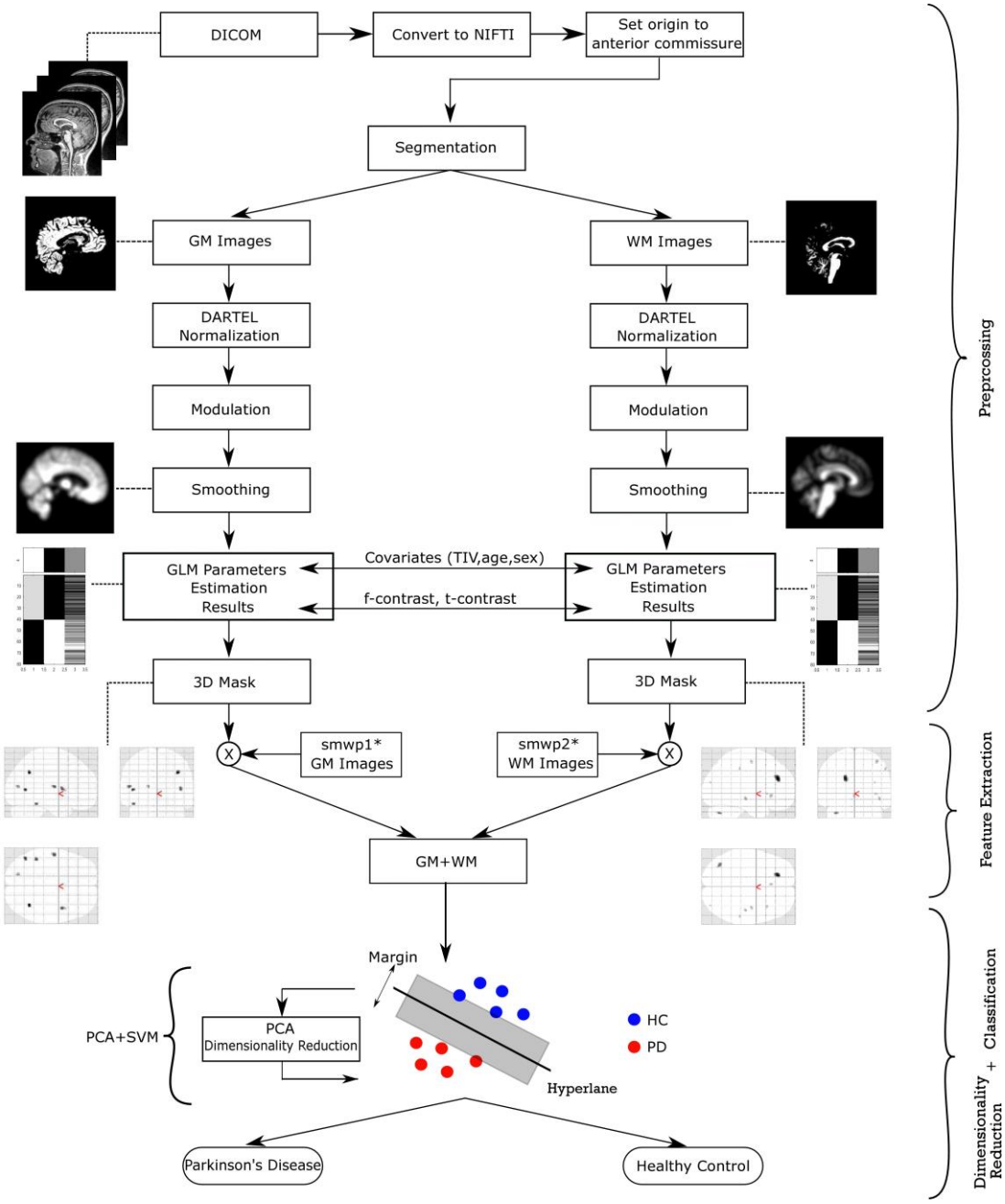

Figure 4.1: The framework of VBM plus DARTEL processing pipeline and classifying PD apart from HC.

smwp1*=DARTEL warped, modulated, smoothed GM tissue. smwp2*=DARTEL warped, modulated, smoothed WM tissue

4.3.4 The SVM Classifier

An SVM classification algorithm is used in order to classify PD patients apart from HCs. The main idea behind the SVM is to search for an optimal class-separation hyperplane in the maximal margin [120]. In this chapter, the training is performed by using a classification learner application. In this application, there are several algorithms to train and validate classification models for binary or multiclass problems. It provides a comparison of the validation errors side-by-side after obtaining multiple trained models and allows to extract the highest accuracy model. In this study, an RBF-SVM classifier has been used. To evaluate the performance of the classifier, a procedure of two CVs was combined with a grid search [2]. In this work, $K_1 = 10$ and $K_2 = 10$ were used. The pipeline of the 10-fold cross-validation procedure is given in Fig. 5.3. As shown in Fig. 5.3, the feature selection by using the PCA is performed through outer loop cross-validation in the SVM classification algorithm.

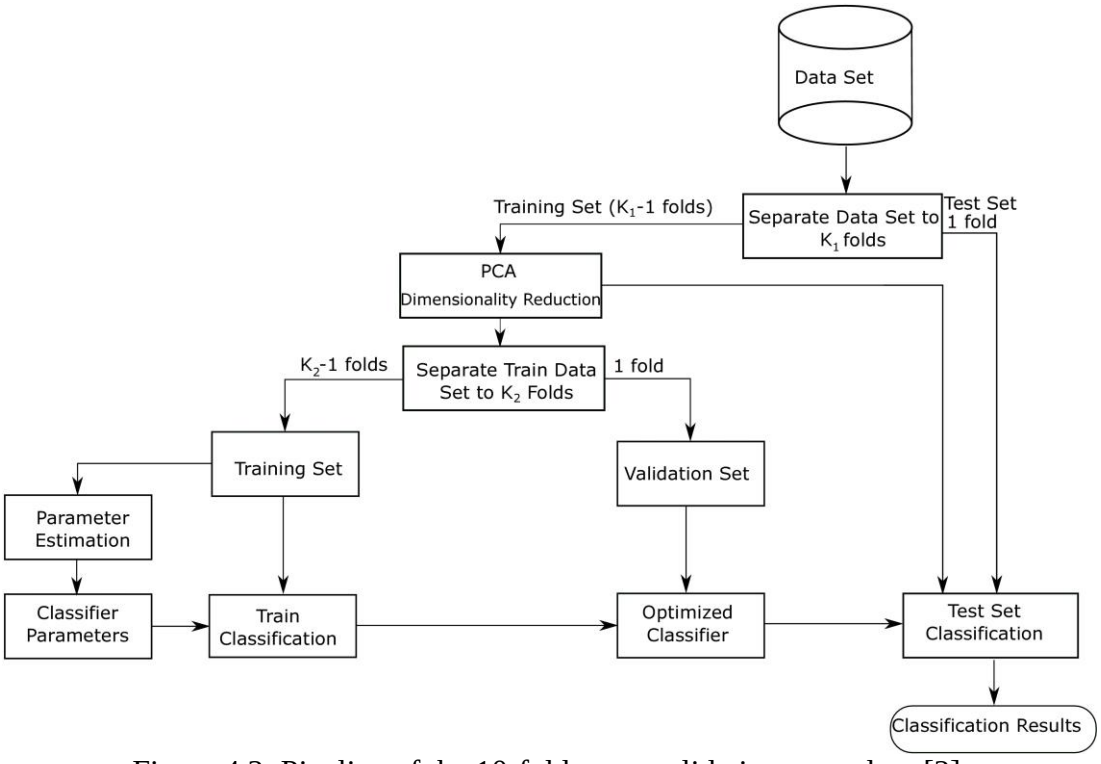

Figure 4.2: Pipeline of the 10-fold cross validation procedure [2].

4.3.5 Experimental Results

The experimental results obtained from 3D T1-weighted sMRI data by using the SPM12 and the CAT12 toolbox with DARTEL analysis have been studied. The segmented, normalized, modulated and smoothed data are used in order to generate a 3D mask by using four combinations of covariates and two different contrasts for GM as well as WM tissues. The 3D masks are the VOIs which encapsulate the most discriminative voxels between PD and HC. For each different combination of covariates and contrasts, 3D masks are generated for both GM and WM tissues, separately.

4.3.6 GM Analysis

The 3D masks obtained for GM tissue with only TIV as a covariate is given in Fig. 9. The positive-tailed t-contrast is selected as $\lambda^T = [-1 \quad 1]$ in which the data located as $[PD \quad HC]$. The generated 3D masks for GM tissue are given in Fig. 9a. The negative-tailed t-contrast is selected as $\lambda^T = [1 \quad -1]$ in which the data located as $[PD \quad HC]$. The results are shown in Fig. 9b. Since f-contrast for two groups include both sides, defining it as $\lambda^T = [-1 \quad 1]$ or $\lambda^T = [1 \quad -1]$ gives the same 3D masks. Therefore, generated 3D masks through f-contrast without direction limitation are given in Fig. 9c for GM tissue.

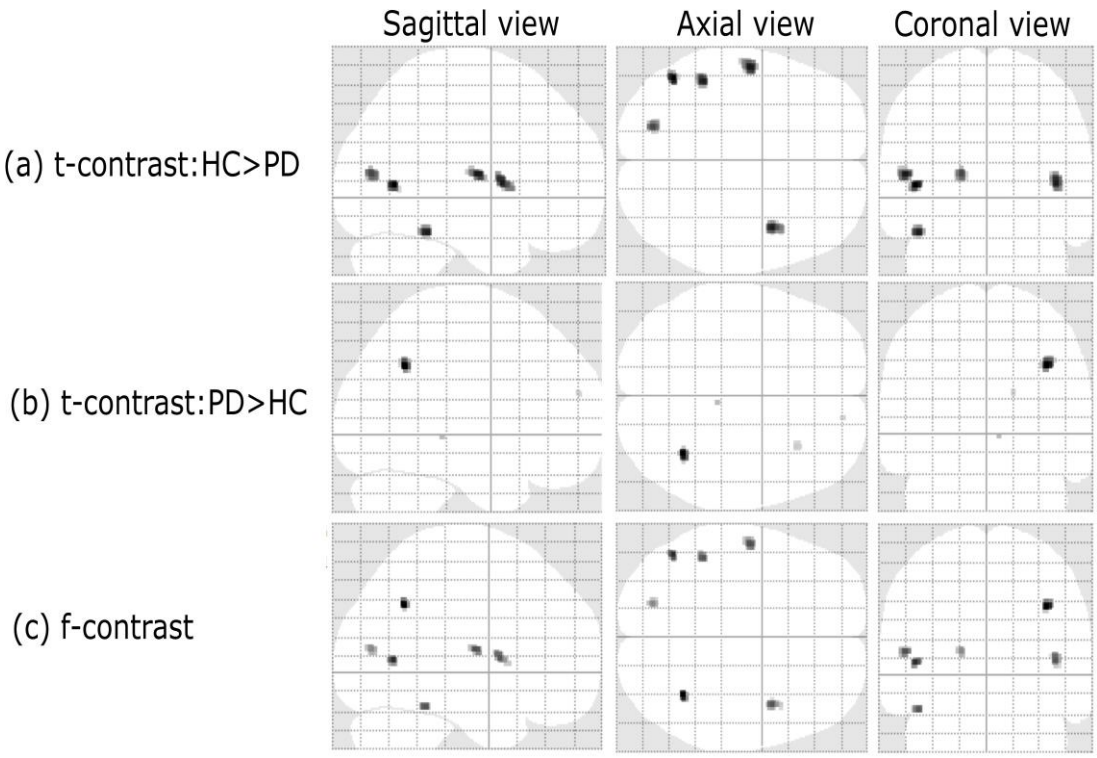

Figure 4.3: 3D masks for GM tissue with TIV as a covariate and different contrasts.

After multiplying the obtained 3D masks with the whole GM tissue, the extracted GM data is obtained. The experiments are performed for the different combinations of TIV, age, and sex covariates and two different contrasts. The number of features in the extracted 3D data and the number of obtained PCs explaining 95% of total variance after applying PCA to the extracted 3D data are given in Table 17. As it is seen, the numbers of features are reduced drastically through using the PCA method. The results indicated in Table 18 show that among TIV, age, TIV+age, and TIV+age+sex covariates, the highest accuracy, 73,75% with f-contrast and 63.75% with t-contrast, are obtained by using only TIV as a covariate. What is more, the experiments represent that using f-contrast outperforms using t-contrast importantly. As a result, using TIV as a covariate and f-contrast gives the highest accuracy of 73.75%.

4.3.7 WM Analysis

The 3D masks obtained for WM tissue with only TIV as a covariate are given in Fig. 4.2. The positive-tailed t-contrast is selected as $\lambda^T = [-1 \quad 1]$ in which the data located as $[PD \quad HC]$. The generated 3D masks for WM tissues are given in Fig. 4.2a. The negative-tailed t-contrast is selected as $\lambda^T = [1 \quad -1]$ in which the data located as $[PD \quad HC]$. The result is shown in Fig. 4.2b. Since f-contrast for two groups include both sides, defining it as $\lambda^T = [-1 \quad 1]$ or $\lambda^T = [1 \quad -1]$ gives the same 3D masks. Therefore, generated 3D masks through f-contrast without direction limitation are given in Fig. 4.4c for WM tissue.

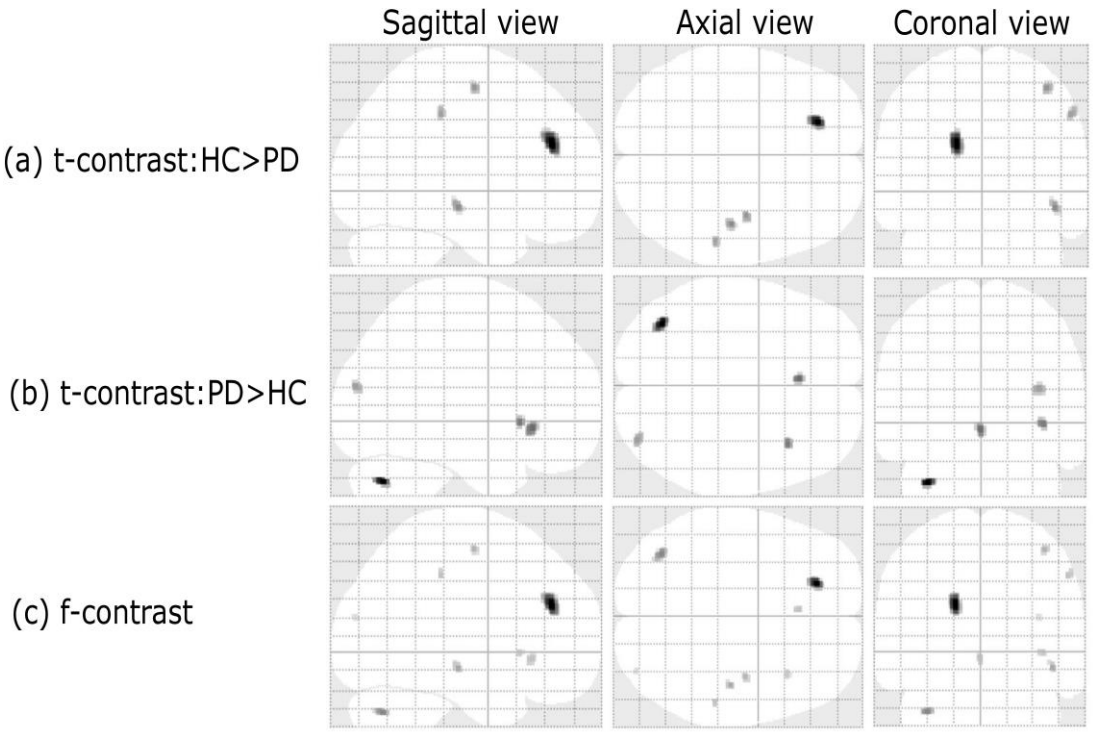

Figure 4.4: 3D masks for WM tissue with TIV as a covariate and different contrasts.

The number of features in the extracted 3D data and the number of obtained PCs explaining 95% of total variance after applying the PCA to the extracted 3D data are given in Table 4.1. Not only in GM, but also in WM the experiment results given in Table 4.2 show that the highest accuracy, 72.50% with f-contrast and 66.25% with t-

contrast, are obtained for using only TIV as a covariate. Furthermore, the experiments represent that using f-contrast outperforms using t-contrast remarkably. As a result, using TIV as a covariate and f-contrast gives the highest accuracy of 72.50%.

4.3.8 Analysis of Combining Extracted GM and WM VOIs

Through combining extracted GM and WM VOIs, new GM plus WM VOIs dataset is obtained. The experiments are done for the different combination of covariates and contrasts. The number of features in the extracted 3D data and the number of obtained PCs explaining 95% of total variance after applying the PCA to the extracted 3D data are given in Table 4.1.

Table 4.1: Number of features before PCA is applied and number of PCs for dimensionally reduced GM VOIs, WM VOIs, and combination of GM as well as WM VOIs

Cov^*-$Cont^*$	Number of Features[PC] *		
	GM	WM	GM+WM
AGE-F	67 [4]	61 [4]	128[6]
TIV-F	152[8]	198[10]	350[16]
TIV+AGE-F	217[9]	170[9]	387[16]
TIV+AGE+SEX-F	250[9]	86 [7]	336[16]
AGE-T	83 [2]	21 [2]	104[4]
TIV-T	301[9]	245[8]	546[14]
TIV+AGE-T	398[10]	234[8]	632[16]
TIV+AGE+SEX-T	419[10]	211[9]	630[16]

[*] [PC]:Number of principle components for 95% of total variance.
[*] Cov:Covariate, Cont:Contrast.

As in GM and also WM tissue analyses, for concatenated GM and WM VOIs dataset, the experiment results given in Table 4.2 show that the highest accuracies, 93.75% with f-contrast and 80.00% with t-contrast, are obtained for using only TIV as a covariate. What is more, the experiments represent that using f-contrast outperforms using t-contrast remarkably. As a result, using TIV as a covariate and f-contrast gives the highest accuracy of 93.75%.

Table 4.2: Performances of GM, WM, and fusion of GM as well as WM with different combinations of covariates and contrasts.

	GM			WM			GM+WM		
Cov^*-$Cont^*$	$\boldsymbol{ACC^*}$	$\boldsymbol{SEN^*}$	$\boldsymbol{SPE^*}$	$\boldsymbol{ACC^*}$	$\boldsymbol{SEN^*}$	$\boldsymbol{SPE^*}$	$\boldsymbol{ACC^*}$	$\boldsymbol{SEN^*}$	$\boldsymbol{SPE^*}$
AGE-F	71.25	70.00	72.50	71.25	55.00	87.50	90.00	87.50	92.50
TIV-F	**73.75**	80.00	67.50	**72.50**	67.50	77.50	**93.75**	95.00	92.50
TIV+AGE-F	66.25	92.50	40.00	62.50	45.00	80.00	92.50	92.50	92.50
TIV+AGE+SEX-F	66.25	97.50	35.00	58.75	65.00	52.50	81.25	87.50	75.00
AGE-T	61.25	85.00	37.50	60.00	50.00	70.00	67.50	87.50	47.50
TIV-T	**63.75**	82.50	45.00	**66.25**	72.50	60.00	**80.00**	92.50	67.50
TIV+AGE-T	61.25	90.00	27.50	58.75	52.50	65.00	75.00	92.50	57.50
TIV+AGE+SEX-T	58.75	90.00	27.50	61.25	72.50	50.00	71.00	87.50	55.00

[*] ACC:accuracy(%), SEN:sensitivity(%), SPE:specificity(%), Cov:Covariate, Cont:Contrast.

4.3.9 Identification of Affected Brain Regions in PD by Using VBM

The anatomical brain regions affected owing to PD are identified by using Neuromorphometrics labels [132]. The most important ROIs detected using different combinations of covariates and contrasts belong to the regions of left superior frontal gyrus, right middle temporal, left anterior cingulate gyrus, right anterior insula, right angular gyrus, left middle temporal gyrus, left inferior temporal, and right putamen. All of the obtained regions are reported in different studies in literature [32, 33, 71, 106, 118].

4.4 Discussion

In most studies, the classification of PD apart from HC is done by using GM or WM tissues exclusively [29, 32, 118]. In this study, not only the individual performances of extracted GM and WM VOIs but also the combination of them have been examined. Through concatenating the GM and WM modalities, the accuracy of PD classification has been increased remarkably as seen in Table 4.2. In order to detect tissue degenerations and improve the classification performances, using covariates in model building is required [108]. In this study, the role of using different covariates namely TIV, age, sex, and their different combinations have been analyzed [33, 108, 110].

Experimental results given in Table 4.2 show that using only TIV as a covariate achieves the highest level of accuracy compared to other covariates on GM, WM, and fusion of GM with WM datasets. Furthermore, among all different combinations of covariates, taking into account only TIV ensures the consistency of the highest classification accuracy for PD data. Unlike other studies which have focused only on decreases in PD brains, this study focuses on both increases and decreases in PD brains. In some studies, a 2-tailed t-contrast has been conducted in order to observe both the increased and decreased brain tissue parts in PD [32, 33, 117]. In this study, in addition to t-contrast, f-contrast has also been investigated. Because of the fact that t-contrast is directional (i.e. comparing the mean of two groups), classification accuracy performance of t-contrast with different covariates and tissue maps is lower than that of f-contrast which considers the effects of interest without any direction limitations (i.e. comparing variances of two groups). As it is seen from the Table 4.2, using f-contrast for model building outperforms using t-contrast remarkably. As a result, instead of focusing only on directional differences for PD data (i.e. comparing the means of two groups), searching for two directional differences (i.e. comparing the variances of two groups) improves the classification accuracy significantly.

Table 4.3: Classification performance comparison with the state-of-the-art.

Research Work	**HC/PD**	**Cov*-Cont***	**Volume(s)**	**ACC ***	**SEN ***	**SPE ***
Salvatore et al.[9]	28/28[a]	-	SSNMV*	85.80	86.00	86.00
Babu et al.[32]	112/117[b]	GM-T	GM	87.21	87.39	87.00
Rana et al.[27]	30/30[a]	TIV-T	GM	86.67	90.00	83.33
			WM	86.67	86.67	86.67
Rana et al.[33]	30/30[a]	TIV-T	GM	76.67	66.67	86.67
		TIV-T	WM	73.33	73.33	73.33
		TIV-T	GM+WM	88.33	90.00	86.67
Proposed Method	40/40[a]	TIV-F	GM+WM	**93.75**	**95.00**	**92.50**

[*] Cov:Covariate, Cont:Contrast, T:t-contrast, F:f-contrast.
[*] SSNMV: Skull-stripped & normalized MRI volumes.
[*] ACC:accuracy(%), SEN:sensitivity(%), SPE:specificity(%).
[a] Self-acquired dataset not available publically.
[b] Publically available PPMI dataset.

In Table 4.3, the results taken directly from the existing studies are compared with this study. However, in order to have a fair comparison, the algorithms in existing studies are applied to the 40 PD and 40 HC data from ppmi database which is used in this chapter.In order to compare the performance of [9] with this study, original 40 PD and 40 HC dataset are cropped, reoriented, and converted from DICOM format to 3D NifTi using MRICron software [133]. Skull stripping and normalization to MNI space are done by using FSL software [134]. PCA is used to reduce the dimension of the whole SSNMV data. After selecting the features, the SVM classification algorithm with LOOCV is used. However, in this study, the 3D masked data is obtained from GM and WM dataset and the PCA method is applied to the 3D masked data through the cross-validation in SVM algorithm. In other words, the PCA is applied to VOIs instead of the whole brain. Hence, irrelevant features are removed by masking the data before the dimensionality reduction. As a result, the classification performance improves significantly. While the classification accuracy in [9] reported as 85.80% for 28 PD and 28 HC self-acquired data, with the data used in this chapter, the obtained classification accuracy given in Table 4.4 is 68.75%.

In order to compare the performance of [32] with this study, the segmented, VBM normalized, and smoothed GM data are used for 3D mask generation. DARTEL normalization which improves inter-group registration and provides more precise as well as accurate localization of structural differences of sMRI data is not used in [32]. In statistical testing, only GM as a covariate and t-test for model building are used. The 480 features obtained from the 3D masked data are reduced to 10 features by using ICA. After selecting the features, SVM classification algorithm is used. In each trial, randomly 75% of total samples are selected for training and 25% for testing. The aim

of [32] is to show the superiority of a PBL-McRBFN-RFE) classification algorithm over SVM. While the obtained accuracy of SVM classification with 10 Independent Component Analysis (ICA) features reported in [32] is 67.44%, the one of PBL-McRBFN-RFE classification is 87.21%. Since the performance comparison of this study and [32] is examined, only SVM classification is applied to the dataset used in this chapter and the obtained classification accuracy given in Table 4.4 is 62.50%.

In [33], the VBM normalized, smoothed, and segmented 40 PD and 40 HC data studied in this chapter are used for 3D mask generation. DARTEL normalization is not used. In statistical testing, TIV as a covariate and t-test for model building are used. The mRMR is used for feature selection. After selecting the features, SVM classification algorithm with LOO validation is used. Feature addition is continued to achieve the maximum accuracy. The classification accuracy for the combination of GM and WM, given in Table 4.4, is obtained as 78.75%. In [33], with 30 PD and 30 HC self-acquired data, the obtained classification accuracy is 88.33%.

In [27], preselected regions, specifically SN, are drawn manually and classification is performed on these predefined ROIs. Therefore, it is limited to analyze predefined ROIs and affected by boundary selection that is operator dependent [33]. Since the experiments are performed with 30 PD and 30 HC self-acquired data, in this study, the same scheme might not be applicable to the data used in this chapter. It may not be a reasonable comparison owing to the fact that defining SN is not clearly given. Namely, the exact locations of the boundaries are not provided and the boundaries of regions depend on the operator. Having said this, the application of [27] to the 40 PD and 40 HC ppmi datasets used in this chapter is ignored.

As seen from the Table 4.3, using only TIV as a covariate, f-contrast, and combination of extracted GM and WM VOIs datasets outperforms similar studies reported in the state-of-the-art by means of classification accuracies.

Table 4.4: Comparing classification performances of existing methods with applying these methods to 40 PD and 40 HC ppmi datasets used in this chapter.

	Existing Methods			Applying Existing Methods		
Research Work	$\boldsymbol{ACC}^*$	$\boldsymbol{SEN}^*$	$\boldsymbol{SPE}^*$	$\boldsymbol{ACC}^*$	$\boldsymbol{SEN}^*$	$\boldsymbol{SPE}^*$
Salvatore et al.[9]	85.80	86.00	86.00	68.75	62.50	75.00
Babu et al.[32]	$67.44^{\dagger}$	$75.65^{\dagger}$	$58.00^{\dagger}$	62.50	57.00	68.00
Rana et al.[33]	88.33	90.00	86.67	78.75	85.00	72.50
Proposed Method	-	-	-	**93.75**	**95.00**	**92.50**

[*] ACC:accuracy(%), SEN:sensitivity(%), SPE:specificity(%).
[†] SVM classification performances reported in [32].

4.5 Conclusion

Detection and diagnosis of neurodegenerative diseases have become interesting research fields. In this chapter, the classification of PD apart from HC has been examined by using different combinations of covariates and two different contrasts for GM tissue, WM tissue, and fusion of them. Experimental results indicate that f-contrast gives higher classification accuracies for any combination of covariates and for all tissue types compared to t-contrast. Since f-contrast searches for regions of interest namely, any kind of differences and t-contrast tests null hypotheses against only 1-tailed alternative ones, the results for f-contrast are better than that of t-contrast. Among different covariate combinations for both f-contrast and t-contrast, using only TIV for GM tissue, WM tissue, and fusion of them give the highest classification accuracies. In this study, f-contrast with only TIV as a covariate for fusion of GM and WM tissues gives 93.75% classification accuracy. The experimental results provides that the use of f-contrast instead of t-contrast enhances the classification accuracies of

GM modality, WM modality, and the combination of GM and WM modalities, significantly.

Chapter 5

PERFORMANCE ANALYSIS OF DIFFERENT CLASSIFICATION ALGORITHMS USING DIFFERENT FEATURE SELECTION METHODS ON PARKINSON'S DISEASE DETECTION

5.1 Introduction

This chapter introduces an automated approach by comparing the performances of five classification algorithms for five feature ranking approaches by using an automatic FC stopping method on PD detection. The sMRI data have been used for classification of PD and HC. Differentiated voxels which are considered as VOIs are obtained by using the VBM method which discriminates the PD from the HC by analyzing group-wise comparisons of cross-sectional sMRI scans [32]. In order to improve inter-group registration and provide more precise and accurate localization of the structural differences of sMRI data, a DARTEL technique has been used with a combination of the VBM [106]. After preprocessing of the data, statistical analysis is studied for extracting VOIs through voxel-by-voxel comparison of the brains of PD and HC. In statistical analysis, in order to detect tissue abnormalities optimally, there are some parameters needed to be provided such as covariate and contrast [107]. In literature, for PD detection, in order to obtain the 3D masks of GM and WM tissue maps, mostly t-contrast is used [32, 33, 117]. While the t-contrast compares the mean of two groups, f-contrast compares variances of the two groups. Therefore, t-contrast is directional

(either mean of group one is larger than group two or vice versa), f-contrast checks for the differences between the groups. In this chapter, TIV is used as a covariate and f-contrast is used for model building. By doing statistical analysis, different number of clusters having different amounts of VOIs are obtained. The obtained clusters are used to generate a 3D mask [7]. Masking the segmented, warped, and smoothed data decreases the amount of raw data significantly. Hence, only the differentiated voxels clustered in VOIs are taken into account. In neurodegenerative disorder diagnosis, mostly the alterations in the GM tissue map using VBM have been considered [9, 32, 117]. However, in [16, 119], the alterations in the WM tissue map using VBM have been taken into account. In this chapter, the preprocessing of 40PD and 40HC data from PPMI datasets has been performed for both the GM and WM modalities and the 3D masks are obtained for GM and WM, separately. The obtained 3D masks are applied to the GM and WM tissues and only the differentiated voxels are extracted. In order to analyze the effects of GM and WM modalities together, a single vector is obtained by concatenating the 3D masked GM and WM datasets. By combining the GM and WM modalities, a feature-level fusion technique, the high dimensions of the combined GM and WM data are reduced to the lower VOIs level. Even though the feature extraction decreases the amount of raw data, since the number of samples is more than the number of observations, an automatic FS method might be used to prune off the irrelevant, redundant, and noisy information from the data.

Feature selection is an important task in machine learning, artificial intelligence, computer vision, and data mining. It aims to reduce the dimensionality of the data and noise in data sets. The purpose of FS is to remove redundant features from the feature sets and to keep the relevant ones. Therefore, a convenient FS method mostly provides

a better performance in learning and classification processes, since the chance of over-fitting increases with the number of features. Considering the evaluation strategy of the FS, there are two categories mostly used, filters [103] and wrappers [121]. In wrapper methods, to evaluate the possible feature subsets, a built-in classifier is required. However, in filter methods, correlation between features are evaluated using the criterion indicative of the ability to separate the classes [98]. In this chapter, five different filter based feature ranking approaches including both the supervised and the unsupervised learning have been investigated for the preprocessed, 3D masked, and the extracted GM as well as WM data sets. These approaches are Relief-F [122], LS [123], MCFS [100], UDFS [101], and CFS [103]. After ranking all the features, the optimum number of top-ranked features are determined by using an FC stopping method which maximizes the class separation between the PD and HC [124]. The purpose of this method is to select the optimal number of top-ranked features adaptively based on training data in each fold instead of identifying a fixed number of features. It also aims to determine a discriminative feature subset with high performance by using training data in each fold of classification algorithm [31]. In order to classify the data with the selected features, five different classification algorithms which are kNN [87], NB [104], ESD [105], EBT [105], and SVM [120] have been studied. The performances of all five feature ranking methods with all five classification algorithms are compared. Regarding to CFS method, a good feature subset has the features which are not correlated with each other, also they are correlated to the class labels. Hence, the performance of feature evaluation as subsets is better than that of evaluation individually [103]. The experimental results show that among all five feature ranking approaches with Fisher stopping criterion, the CFS one has the highest performances for all five classification algorithms, since it ranks feature

subsets rather than individual features. In order to improve the overall PD classification performance, a decision fusion technique among the outputs of all five classifiers for the CFS ranking with Fisher stopping criterion FS approach is proposed. The proposed decision fusion approach has been performed for GM, WM, and combination of GM as well as WM datasets. The obtained results indicate that fusion the outputs of five classifiers outperforms the performance of each classifier individually. Furthermore, the combination of the GM as well WM datasets, a feature-level fusion technique, significantly increase the classification performances of using the GM and the WM datasets, separately. By using the feature-level and decision fusion techniques, a higher classification result is achieved for PD detection.

The novelty of this chapter is to introduce an automatic approach by comparing the performances of five classification algorithms for five feature ranking approaches by using an automatic FC stopping method on PD detection. To the best of the authors' knowledge, it is the first study to compare the performances of several FS approaches with several classification algorithms and using both the feature-level and decision fusion techniques on PD detection. The experimental results indicate that for the combination of GM and WM datasets, using the CFS ranking with FC stopping method FS approach and fusion the outputs of the five classifiers gives the highest accuracy, sensitivity, and specificity scores of 95.00%, 90.00%, and 100%, respectively.

5.2 Materials

The materials and methodogy used in this chapter are same as the ones in chapter 4. The framework of this chapter is given in Fig. 5.1.

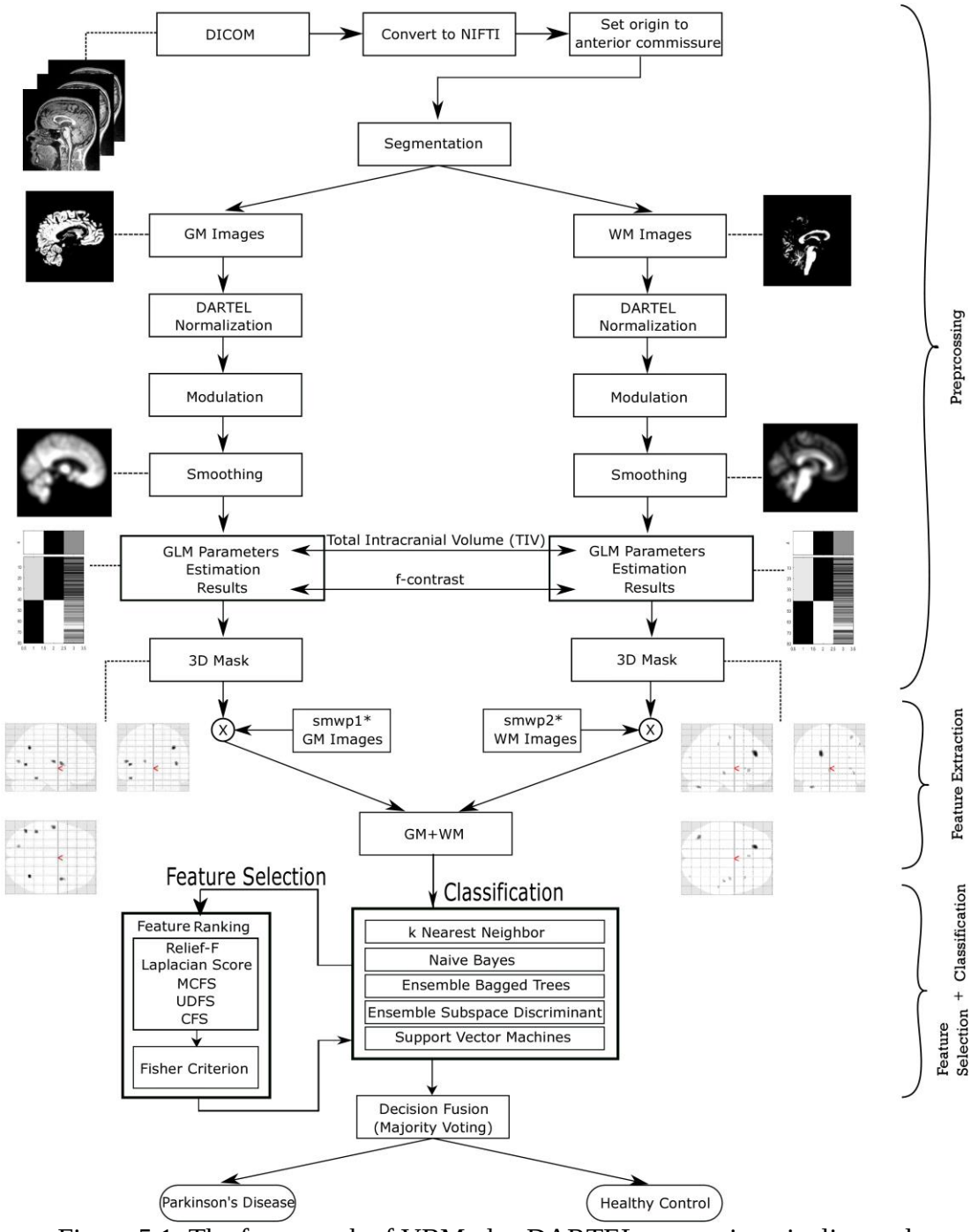

Figure 5.1: The framework of VBM plus DARTEL processing pipeline and classifying PD apart from HC.

smwp1*=DARTEL warped, modulated, smoothed GM tissue. smwp2*=DARTEL warped, modulated, smoothed WM tissue.

5.2.1 Feature Extraction

The generated 3D masks for GM and WM tissues are multiplied with the DARTEL normalized, modulated, and smoothed GM and WM tissues, respectively. Therefore, only the voxels clustered in the VOIs are extracted from the whole GM and WM tissue maps. The 3D masks of the GM modality is given in Fig. 5.2a and the WM modality is given in Fig. 5.2b. The dimensions of the extracted GM and WM data are reduced significantly through feature extraction. In this study, GM and WM VOIs are concatenated to investigate the effects of both modalities. The pipeline of the procedure is given in Fig. 5.1.

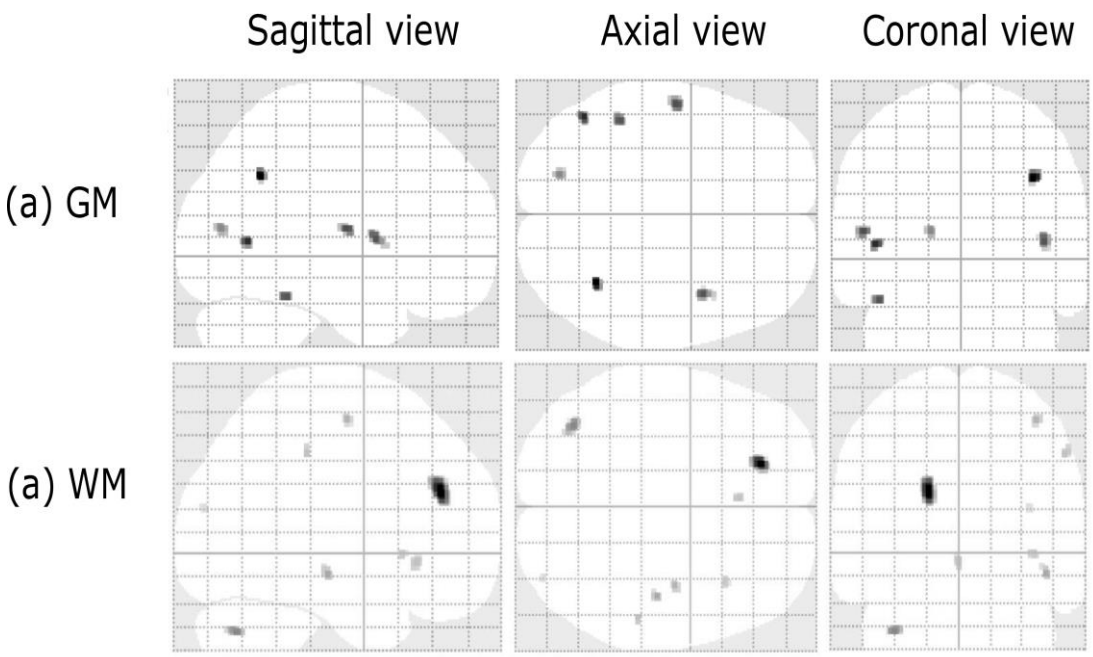

Figure 5.2: The 3D masks of GM and WM tissues with TIV as a covariate and f-contrast.

5.2.2 Feature Selection

In order to alleviate the effect of the curse of dimensionality, improve the performance of the designed model, reduce the learning process time, and enhance data understanding, an FS method is required. There are mainly three FS techniques, namely filters, wrappers, and embedded methods. In filter methods, irrelevant features

are filtered out independently from the learning algorithm. The classifier is ignored and only intrinsic properties of the data are examined. First, the features are taken into account exclusively and ranked according to their importance level. Then a feature subset is obtained from the ranked ones. In wrappers methods, classifiers are considered in order to score a given subset of features. Hence, in the wrapper approach, there is need to evaluate a specific learning algorithm and to determine which features are selected. In embedded ones, a selection process is injected into the learning of the classifier [98]. With wrappers and embedded methods, first, the feature subsets are sampled and evaluated. Then they are kept as the final outputs [98]. The wrappers and embedded methods tend to find the more optimal features for the specific learning algorithms, hence the computational cost is high. However, the filter methods are fast and easy to implement. In this chapter, filter feature ranking methods including both supervised and unsupervised ones are studied for detection of PD. A subset of relevant features from 3D masked data is ranked according to their degrees of relevance and importance. In this chapter, five different feature ranking methods, namely mRMR, LS, UDFS, MCFS, and CFS are used. After ranking the features, in order to select the most discriminative top-ranked ones, a FC stopping criteria is used. The details of each feature ranking method and feature selection procedure are provided in Chapter 3.

5.2.3 Classification Methods

In order to separate the PD patients apart from the HCs, classification algorithms need to be used. In literature, many different classifiers are used to train a model. In this chapter, five different classification methods are used for training the model, namely kNN, NB, ESD, EBT, and SVM. The procedure of each classification method is provided in Chapter 3. The pipeline of the 10-fold CV procedure is given in Fig. 5.3.

As shown in Fig. 5.3, the FS by using the ranked features is performed through the outer loop CV in the SVM classification algorithm.

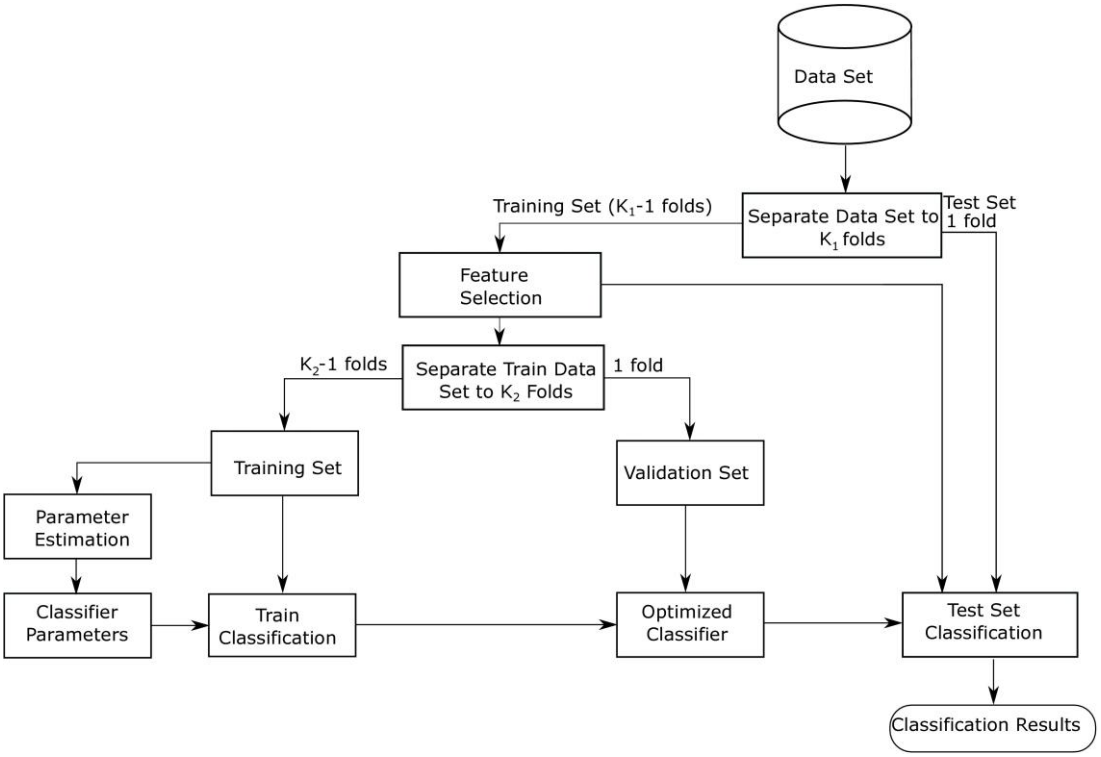

Figure 5.3: The pipeline of the 10-fold cross validation procedure [2].

5.2.4 Decision Fusion

In decision fusion, the prediction results of five classifiers are concatenated into a single 80x5 matrix. Assuming C_{KNN}, C_{NB}, C_{ESD}, C_{EBT}, C_{SVM} are the prediction binary results of the KNN, NB, ESD, EBT, and SVM, respectively. Hence, the decision fusion is designed as:

$$C_{DF} = [C_{KNN}, C_{NB}, C_{ESD}, C_{EBT}, C_{SVM}] \quad (5.1)$$

After generating the decision fusion matrix which has a size of 80x5, the MV method is employed as a final classification decision. Each sample having more than 3 votes are labeled as correctly classified. Since the MV is easy to implement, has high

performance on real data, and has a very simple scheme, it is one of the most used versatile combination methods in literature [31, 125].

5.3 Experimental Results

The experimental results obtained from 3D T1-weighted sMRI data by using SPM12 and CAT12 toolbox with DARTEL analysis have been studied. The segmented, normalized, modulated and smoothed data are used in order to generate a 3D mask by using TIV as a covariate, f-contrast, and the combination of GM as well as WM tissue maps. The 3D masks are the VOIs which encapsulate the most discriminative voxels between PD and HC. The 3D mask data of GM and WM are multiplied with the original GM and WM datasets, respectively and the 3D masked GM and WM data are obtained. In order to improve classification performance, the 3D masked GM and WM data are concatenated and a new fused GM+WM dataset is obtained. Since the number of the features in the combination of GM and WM datasets are more than that of samples, an FS approach needs to be applied to the data. In this chapter, filter based five feature ranking methods, including both the supervised and unsupervised learning, are applied to the combination of the GM and WM datasets. An adaptive FC algorithm is used in order to select the optimum number of top discriminative features among the ranked vector in each set of training fold. For classification of the data with selected features, five different methods are studied. Hence, the performances of all five different classification methods using the outputs of all five feature ranking approaches as an input are compared. The experimental results given in Table 5.1 indicate that among all five features, the CFS ranking with an FC stopping method has the highest accuracy, sensitivity, and specificity scores. FS, in a simple form, assess individual attributes and rank them regarding to their correlation with class labels. However, in [103] it is proved that a good feature subset has the features which are not correlated

with each other, besides they are correlated to the class labels. Therefore, the performance of feature evaluation as subset is better than that of evaluation individually. Since CFS assess and therefore ranks the feature subsets rather than the individual features, it mostly outperforms the Relief-f approach which does not identify redundant attributes [103]. The MCFS method which selects features in batch mode, performs especially well when the number of selected features is small [100]. However, in this chapter, for some training folds,the number of selected top-ranked features might be high.

Regarding the Table 5.1, among all five classification approaches by using five feature ranking method with FC stopping criteria as input, the CFS one indicate a better and robust performance for all the classifiers. Therefore, this approach has been evaluated for the GM, WM, and the combination of GM as well as WM datasets as inputs. The experiment results given in Table 5.2 show that using the combination of GM and WM datasets outperforms using GM and WM individually for all the classifiers. Since the number of features is increased by fusion of the GM and WM datasets, the variance is increased and the selected top-ranked features are also increased. Hence, the classification performances of all methods are improved. In Table 5.2, among all five classification methods, the SVM has the superiority classification performance for all the feature ranking approaches. As it is seen from Table 5.2, after applying the decision fusion technique by using MV technique, the specificities and accuracies of individually used GM and WM datasets are improved. Furthermore, applying this scheme to the combination of GM and WM datasets improve the whole classification performance.

Table 5.1: The five different classification performances by using five different feature ranking methods and FC feature selecting on each method for the combination of GM and WM datasets.

CM*	***FSM****	***ACC****	***SEN****	***SPE****
KNN	Relief-F	81.25	77.50	85.00
	LS	82.50	77.50	87.50
	MCFS	78.75	75.00	82.50
	UDFS	80.00	75.00	85.00
	CFS	**85.00**	**77.50**	**92.50**
NB	Relief-F	75.00	70.00	80.00
	LS	73.75	70.00	77.50
	MCFS	76.25	75.00	77.50
	UDFS	75.00	70.00	80.00
	CFS	**87.50**	**82.50**	**92.50**
ESD	Relief-F	76.25	80.00	72.50
	LS	73.75	75.00	72.50
	MCFS	66.25	72.50	60.00
	UDFS	70.00	72.50	67.50
	CFS	**78.75**	**85.00**	**72.50**
EBT	Relief-F	72.50	77.50	67.50
	LS	71.25	77.50	65.00
	MCFS	76.25	77.50	75.00
	UDFS	76.25	80.00	72.50
	CFS	**83.75**	**80.00**	**87.50**
SVM	Relief-F	86.25	82.50	90.00
	LS	83.75	77.50	90.00
	MCFS	83.75	80.00	87.50
	UDFS	86.25	82.50	90.00
	CFS	**90.00**	**87.50**	**92.50**

[*] CM:classification method, FSM:feature selection method, ACC:accuracy(%), SEN:sensitivity(%), SPE:specificity(%).

From the experiments, it is observed that the decision fusion technique improves the highest specificity scores from 75.00%, 87.50%, and 92.50% to 80.00%, 90.00%, and 100% for GM, WM, and combination of GM as well as WM, respectively. Additionally, the proposed technique improves the highest accuracy scores from 75.00%, 80.00%, and 90.00% to 77.50%, 81.25%, and 95.00% for GM, WM, and combination of GM as well as WM, respectively. Using the proposed methodology increases the sensitivity score from 87.50% to 90.00% for the combination of GM and WM datasets.

Table 5.2: The five different classification performances by using CFS feature ranking and FC feature selecting for GM, WM, and the combination of GM as well as WM datasets.

	GM			WM			GM+WM		
Classification	***ACC****	***SEN****	***SPE****	***ACC****	***SEN****	***SPE****	***ACC****	***SEN****	***SPE****
SVM	68.75	70.00	67.50	76.25	70.00	82.50	90.00	87.50	92.50
KNN	61.25	55.00	67.50	80.00	75.00	85.00	85.00	77.50	92.50
NB	72.50	70.00	75.00	68.75	57.50	80.00	87.50	82.50	92.50
ESD	68.75	75.00	62.50	61.25	52.50	70.00	78.75	85.00	72.50
EBT	75.00	77.50	72.50	70.00	70.00	70.00	83.75	80.00	87.50
Fusion-ALL(Proposed)	**77.50**	75.00	**80.00**	**81.25**	72.50	**90.00**	**95.00**	**90.00**	**100**

[*] ACC:accuracy(%), SEN:sensitivity(%), SPE:specificity(%).

5.4 Discussion

In literature, the performance of CFS technique is compared to the wrappers ones in many applications and in general, the CFS method outperforms the wrapper methods for the small datasets [103]. Additionally, the computational cost of the filter methods, including the CFS one, is less than the wrappers methods. In this study, five different FS methods following an adaptive FC stopping approach have been used for choosing the optimal number of top-ranked features. Additionally, five different classification methods are studied by using five different FS techniques. In order to improve the classification performance, a decision fusion scheme which combines the output of

Table 5.3: Classification performance comparison with the state-of-the-art.

Research Work	**HC/PD**	**Cov*-Cont***	**Volume(s)**	**ACC ***	**SEN ***	**SPE ***
Salvatore et al.[9]	28/28[a]	-	SSNMV*	85.80	86.00	86.00
Babu et al.[32]	112/117[b]	GM-T	GM	87.21	87.39	87.00
Rana et al.[27]	30/30[a]	TIV-T	GM	86.67	90.00	83.33
			WM	86.67	86.67	86.67
Rana et al.[33]	30/30[a]	TIV-T	GM	76.67	66.67	86.67
		TIV-T	WM	73.33	73.33	73.33
		TIV-T	GM+WM	88.33	90.00	86.67
Pahuja et al.[65]	30/30[b]	-T	GM	90.94	92.45	97.30
Proposed Method	40/40[a]	TIV-F	GM+WM	**95**	**90.00**	**100**

[*] Cov:Covariate, Cont:Contrast, T:t-contrast, F:f-contrast, SSNMV: Skull-stripped & normalized MRI volumes. [a] Self-acquired dataset not available publically. [b] Publically available PPMI dataset

each classifiers for the CFS method is used. In literature, the GM or WM modalities from sMRI datasets are used individually in order to detect the PD [27, 29, 32]. In this study, the performance of the proposed scheme is evaluated for the GM, WM, and the combination of the GM and WM modalities. In [33], a decision model is used for FS. In this model, the accuracy of an SVM classifier starting from the features of the first cluster is calculated and then the features are added incrementally into the classifier until the highest accuracy is obtained. The highest classification accuracy of 88.33% is obtained for combination of GM and WM datasets which are the modalities of self-acquired 30PD 30HC data. However, there is not an automatic FS scheme and the computational cost of this approach would be high for large datasets. In [32], a projection based learning and meta-cognitive radial basis function network with RFE as a classifier scheme is used for GM dataset. In 3D mask building, GM as a covariate and t-contrast for model building are used. The classification accuracy of 87.21% is obtained. However, the computational cost of the used wrapper FS method including RFE is expensive and the DARTEL normalization in VBM technique which improves inter-group registration and provides more precise and accurate localization of structural differences of sMRI data is not used. In [27], the preselected regions, especially SN, are drawn manually and classification is performed on these predefined ROIs. The t-test based ranking and mutual information based FS methods are taken into account. The GM, WM, and CSF modalities acquired from 30PD and 30HC self-acquired data are used separately and the classification accuracies of 86.67%, 86.67%, and 83.33% are obtained, respectively. However, using only manually drawn five ROIs which have the operator-dependent boundaries might skip the other brain regions affected by the PD. In [9], the PCA followed by Fisher's discriminant ratio is used to reduce the dimensionality of the features and SVM is used for classification. In [65],

the PD diagnosis based on extreme learning machine method along with GA feature subset selection is studied. The PD detection framework used in [65] is applied to the 40PD and 40HC datasets utilized in this thesis in order to compare its performance with the proposed method. The VBM normalized, modulated, 10mm FWHM smoothed, and segmented data are used for 3D mask generation. DARTEL normalization is not used. In statistical testing, two-sample t-test to generate the model and t-contrast to build the 3D GM mask are utilized. The GA is used for feature selection. After selecting the features, the performance of the ELM classifier is compared with SVM one. In each trial, 80% of total data is randomly selected for training and 20% of it is used for testing. Since the aim of this study is to compare the performance of the proposed method with the one provided in [65], only the SVM classifier is applied to our dataset and the classification accuracy of 75.00% is obtained whereas, it is reported as 90.94% with the given data in [65]. In Table 5.3, comparison of this study with the existing literature is given. As it is seen from the Table 5.3, the classification performance of the proposed PD detection methodology outperforms the similar studies reported in the state-of-the-art. As it is known, the GA has unguided mutation. The mutation operator in GA adds a randomly generated number to the parameter of an individual in the population. Therefore, it converges slowly which makes it computationally expensive.

5.5 Conclusion

Recently, the diagnosis of neurodegenerative diseases based on the neuroimaging data has been extensively studied. In this chapter, the classification of PD apart from HC have been investigated by using five different classification approaches and five different ranking methods followed with an adaptive Fisher criterion stopping method feature selection technique. A decision fusion method using MV technique is taken

into account in order to enhance the classification performance. The experiments are performed for the GM tissue, WM tissue, and the fusion of GM and WM ones. The experiments indicate that the CFS approach provides the best results with all five classification methods. Additionally, the SVM provides the highest classification performances for all five feature selection techniques. Therefore, comparing to the individual classification performances of the classifiers ,the fusion of them using the CFS approach improves the classification performance importantly. The obtained classification accuracies with the proposed methodology are 77.50% and 81.25% for the individual GM and WM tissues, respectively. However, the fusion of GM and WM modalities improves the classification accuracy of the proposed methodology up to 95.00%.

Chapter 6

AFFECTED BRAIN REGIONS DUE TO PARKINSON'S DISEASE

6.1 Introduction

In literature, there are many medical and engineering studies which focus on detection of deteriorated brain parts owing to PD [8, 39, 43, 59, 76]. The brain parts obtained by using MRI neuroimaging technique combined with CAD methods may provide the integrated and comprehensive interpretation of PD. In this chapter, the deteriorated brain regions owing to PD are obtained by using sMRI data and machine learning approaches. The captured regions are compared with the ones yielded in the state-of-the-art studies which analyzed the affected brain regions due to PD by using MRI data and CAD methods. The results indicate that the neurodegenerative PD may affect many neurotransmitters, brain regions, structural and functional connections, and neurocognitive systems [16, 35, 67].

6.2 Affected Brain Regions

The early diagnosis of PD is one of the most challenging tasks in neuroscience and the error rates in clinical PD detection may be extremely high, since there are not certain diagnosis methods. The detection accuracy has been enhanced by using MRI, both sMRI and fMRI, recently [32, 36, 37, 46]. In most of the studies using sMRI data with machine learning algorithms, only the atrophies that occurred in brain volumes due to PD have been considered. However, in this thesis, not only the atrophies but also the

increases in the brain parts of PD have been taken into account. In other words, while building both the 3D GM and WM masks between the brains of PD and HC, the directional t-contrast is used in many studies, yet in this thesis, any structural differences between the brains of PD and HC are utilized in the diagnosis of PD. The obtained brain regions are compared with the state-of-the-art studies and the results are provided in Table 6.1. The yielded affected brain regions may help the clinicians to diagnose the PD.

Table 6.1: Affected Brain Regions Owing to PD.

Author	Brain Region	Change	NI
Adeli et al.[59]	left red nucleus	WM+GM	MRI
	left middle cingulate gyrus	GM+CSF	
	right red nucleus	WM	
	left substantia nigra		
	right substantia nigra		
	pons		
	left superior temporal gyrus	GM	
	left middle frontal gyrus		
	right middle frontal gyrus		
Adeli et al.[60]	middle frontal gyrus	N/A	MRI
	pons		
	left substantia nigra		
	right substantia nigra		
	left red nucleus		
	left pallidum		
	left putamen		
	right caudate		
	left inferior temporal gyrus		
	right superior temporal gyrus		
Amoroso et [66]	middle temporal gyrus	N/A	MRI + CS
	superior temporal gyrus		
	sub-gyral		
	superior occipital gyrus		
	middle frontal gyrus		
	culmen		
	medial frontal gyrus		
	precentral gyrus		
	cingulate gyrus		
	precuneus		
Babu et al.[32]	superior temporal gyrus	GM Loss	MRI
Chen et al.[35]	limbic lobe	N/A	MRI
	frontal lobe		
	sub-lobar		
	midbrain		
	pons		
	posterior lobe		
	occipital lobe		
Chen et al.[74]	cingulo-opercular network	Connectivity Loss	MRI
	frontal-parietal network		
	cerebellum		

Table 6.1: Affected Brain Regions Owing to PD (continued).

Author	Brain Region	Change	NI
Cigdem et	**Left superior frontal** **right middle temporal** **left anterior cingulate** **right anterior insula** **right angular gyrus** **left middle temporal** **left inferior temporal** **right putamen**	**GM+WM**	**MRI**
Dunet et al.[75]	caudate nucleus	GM Loss	MRI
Focke et al.[37]	cerebellum brainstem basal ganglia	GM Loss WM Loss	MRI
Gao et al.[76]	cerebellum posterior lobe cingulate gyrus middle temporal gyrus lentiform nucleus putamen caudate nucleus lentiform nucleus	GM Loss WM Loss	MRI
Hikishima et	locus coerulus lateral hypothalamus	Volume Loss	MRI
Huppertz et	cerebrum frontal lobe temporal lobe parietal lobe occipital lobe cerebellum hippocampus amygdala caudate	GM Loss WM Increase Volume Loss	MRI
Kamagata et	right putamen right thalamus right postcentral gyrus left putamen right supramarginal gyrus right insula right postcentral gyrus right thalamus left pars triangularis	Local Local	MRI

Table 6.1: Affected Brain Regions Owing to PD (continued).

Author	Brain Region	Change	NI
Kazeminej al.[39]	cuneus	Metabolic Reduction	MRI
	left precuneus		
	left middle frontal gyrus	FA Reduction	
	right middle frontal gyrus		
	right superior frontal gyrus		
Lei et	left hippocampus	N/A	MRI + DTI
	right superior occipital		
	left amygdala		
	right calcarine		
	right lingual		
	left postcentral		
	left middle occipital gyrus		
	left thalamus		
	left putamen		
	middle frontal gyrus		
Lei et	left parahippocampal	N/A	MRI + DTI
	right parahippocampal		
	left rectus		
	right rectus		
	temporal superior		
	occipital superior		
	rolandic operculum		
	occipital middle		
	frontal middle orbital		
	temporal inferior		
Li et	thalamus	N/A	MRI
	temporal lobe		
	calcarine		
	cingulum		
Lin et	right superior temporal	N/A	MRI
	left amygdala		
	right amygdala		
	left fusiform gyrus		
	right fusiform gyrus		
	left anterior temporal lobe		
	right anterior temporal lobe		
	vermis cerebellum crus II		

Table 6.1: Affected Brain Regions Owing to PD (continued).

Author	Brain Region	Change	NI
Liu et al.[56]	precuneus	N/A	MRI
	thalamus		
Liu et al.[57]	postcentral gyrus	GM Loss	MRI
		GM Loss	
	right precuneus	GM Loss	
LLong et	right precentral gyrus	Volume	MRI
	right middle temporal	RFCS Loss	
	right posterior cingulate	GM Increase	
	middle frontal gyrus	ReHo	
	inferior frontal gyrus	WM Increase	
	rolandic operculum	ALFF	
Marquand et al.	cerebellum		MRI
	putamen		
Morales et al.[64]	left cerebral cortex		MRI
	left caudate		
Morisi et al.[43]	middle cerebellar peduncles	MD Loss	MRI
	right putamen		

Table 6.1: Affected Brain Regions Owing to PD (continued).

Author	Brain Region	Change	NI
Peng et al.[49]	orbitofrontal cortex postcentral gyrus middle frontal gyrus supramarginal gyrus	Connectivity Loss	MRI
Rana et al.[33]	middle gyrus frontal gyrus medial gyrus superior gyrus	GM+WM	MRI
Rana et al.[16]	right hippocampus left middle cingulum tri right precentral gyrus left precentral gyrus		MRI
Salvatore et al. [9]	midbrain pons	Volume	MRI

Table 6.1: Affected Brain Regions Owing to PD (continued).

Author	Brain Region	Change	NI
Singh et al.[55]	putamen	GM	MRI
	medial dorsal nucleus		
	pulvinar		
	posterior cingulate cortex		
	retrosplenial cortex		
	corpus callosum	WM	
	limbic cortex		
Tagare et al.[87]	left caudate	Dopamine	MRI
	right caudate		
	global pallidus		
	putamen		
Talai et al.[45]	left thalamus		MRI
	right thalamus		
	brainstem		
	left caudate		
	left putamen		
	inferior frontal gyrus		
	intra clacarine cortex		
	cingulate gyrus		
	frontal orbital cortex		
Zeng et al.[89]	cerebellar Crus I	GM Loss	MRI
	Vermis III		
	Vermis VIII		
Zeighami et	inferior frontal gyrus		MRI
	fusiform gyrus		
	putamen		
	cingulate gyrus		
	nucleus accumbens		
	cerebellum		
	substantia nigra		
	thalamus		
	caudet		

Chapter 7

CONCLUSIONS AND FUTURE WORK

7.1 Conclusion

This thesis analyses the impact of preprocessing of 3D-MRI scans on post-processing and to investigate the performances of different automated classification algorithms using different feature selection methods on PD detection. The classification of PD apart from HC has been examined by using different combinations of covariates and two different contrasts for GM tissue, WM tissue, and fusion of them. The experimental results represent that f-contrast yields higher classification accuracies for any combination of covariates and for all tissue types compared to t-contrast. Owing to the fact that f-contrast considers all differences between the brains of PD and HC by searching for regions of interest and t-contrast tests null hypotheses against only 1-tailed alternative ones, the results for f-contrast are better than that of t-contrast. Among different covariate combinations for both f-contrast and t-contrast, taking into account TIV alone for GM tissue, WM tissue, and fusion of them yield the highest classification accuracies. Having said these, using only TIV as a covariate in model building and f-contrast in 3D mask building give the highest accuracy at diagnosis of PD.

The discrimination of PD apart from HC has been investigated by using five different classification approaches and five different ranking methods followed with an adaptive Fisher criterion stopping method feature selection technique. A decision fusion method

using MV technique is used to improve the performance of accurate classification. The experiments are performed for the GM tissue, WM tissue, and the fusion of GM and WM tissues. The experiments indicate that among the five feature ranking approaches which are CFS, Relief-F, LS, MCFS, and UDFS, the CFS one provides the best results with all five classification methods. Additionally, among the five classification methods, kNN, NB, SVM, ESD, and EBT, the results of SVM provide the highest classification performances for all five feature selection techniques. Therefore, compared to the individual classification performances of the classifiers, the fusion of them using the CFS approach improves the classification performance importantly. The obtained classification accuracies with the proposed methodology are 77.50% and 81.25% for the individual GM and WM tissues, respectively. However, the fusion of GM and WM modalities improves the classification accuracy of the proposed methodology up to 95.00%.

7.2 Future work

In future studies on PD classification, the proposed preprocessing method can be applied to different combinations of GM, WM, and CSF tissues. In addition to considering only the sMRI data of the subjects, the fMRI, DTI, and SPECT data of them might be analyzed both individually and combined [22, 24, 61]. After obtaining enough datasets, the deep learning method may be applied directly to raw data and by doing so, the biases and other artifacts emerged during the preprocessing steps may be eliminated [126, 127, 128]. Furthermore, classification of three main neurodegenerative disorders, namely PD, AD, and HD may be studied by building a new convolutional neural network architecture in which two classification processes may be performed: first the classification of patients apart from HCs and then the classification of PD, AD, and HD patients.

REFERENCES

[1] W. Commons, File:blausen 0704 parkinsonsdisease.png — wikimedia commons, the free media repository, [Online; accessed 5-July-2018] (2018). https://commons.wikimedia.org/w/index.php?title=File:Blausen_0704_ParkinsonsDisease.png&oldid= 284404755.

[2] I. Beheshti, H. Demirel, Probability distribution function-based classification of structural mri for the detection of alzheimer's disease, Computers in Biology and Medicine 64 (2015) 208 – 216. https://doi.org/10.1016/j.compbiomed.2015.07.006.

[3] G. Bates, S. Tabrizi, L. Jones, Huntington's Disease, Oxford monographs on medical genetics, Oxford University Press, 2014. https://books.google.com.tr/books?id=i-4kAwAAQBAJ.

[4] I. McGrath, Spinocerebellar Ataxia, Foster Academics, 2015. https://books.google.com.tr/books?id=YEYwrgEACAAJ.

[5] K. Talbot, M. Turner, R. Marsden, R. Botell, Motor Neuron Disease: A Practical Manual, Oxford Care Manuals, OUP Oxford, 2010. https://books.google.com.tr/books?id=z8nXD17NdeoC

[6] C. Sumner, S. Paushkin, C. Ko, Spinal Muscular Atrophy: Disease Mechanisms and Therapy, Elsevier Science, 2016. https://books.google.com.tr/books?id=6_1PCwAAQBAJ.

[7] I. Beheshti, H. Demirel, F. Farokhian, C. Yang, H. Matsuda, Structural mri-based detection of alzheimer's disease using feature ranking and classification error, Computer Methods and Programs in Biomedicine 137 (2016) 177 – 193. doi:https://doi.org/10.1016/j.cmpb.2016.09.019. http://www.sciencedirect.com/science/article/pii/S0169260716301055.

[8] O. Cigdem, I. Beheshti, H. Demirel, Effects of different covariates and contrasts on classification of parkinson's disease using structural mri, Computers in Biology and Medicine 99 (2018) 173 – 181. doi:https://doi.org/10.1016/j.compbiomed.2018.05.006. http://www.sciencedirect.com/science/article/pii/ S0010482518301136.

[9] C. Salvatore, A. Cerasa, I. Castiglioni, F. Gallivanone, A. Augimeri, M. Lopez, G. Arabia, M. Morelli, M. Gilardi, A. Quattrone, Machine learning on brain mri data for differential diagnosis of parkinson's disease and progressive supranuclear palsy, Journal of Neuroscience Methods 222 (2014) 230 – 237. doi:https://doi.org/10.1016/j.jneumeth.2013.11.016. http://www.sciencedirect.com/science/article/pii/ S0165027013003993.

[10] R. Pizarro, X. Cheng, A. Barnett, H. Lemaitre, B. Verchinski, A. Goldman, E. Xiao, Q. Luo, K. Berman, J. Callicott, D. Weinberger, V. Mattay, Automated quality assessment of structural magnetic resonance brain images based on a supervised machine learning algorithm, Frontiers in Neuroinformatics 10. doi:10.3389/fninf.2016.00052. https://www.frontiersin.org/article/10.3389/fninf.2016.00052.

[11] B. Heim, F. Krismer, R. De Marzi, K. Seppi, Magnetic resonance imaging for the diagnosis of parkinson's disease, Journal of Neural Transmission 124 (8) (2017) 915–964. doi:10.1007/s00702-017-1717-8. https://doi.org/10.1007/s00702-017-1717-8.

[12] C. G Goetz, The history of parkinson's disease: Early clinical descriptions and neurological therapies, 1 (2011) a008862.

[13] J. Parkinson, An essay on the shaking palsy, The Journal of Neuropsychiatry and Clinical Neurosciences 14 (2) (2002) 223–236, pMID: 11983801. arXiv:https://neuro.psychiatryonline.org/doi/pdf/10.1176/jnp.14.2.223, https://neuro.psychiatryonline.org/doi/abs/10.1176/jnp.14.2.223.

[14] J. G. Goldman, C. G. Goetz, History of parkinson's disease, in: Parkinson's Disease and Related Disorders, Part I, Vol. 83 of Handbook of Clinical Neurology, Elsevier, 2007, pp. 107 – 128. doi:https://doi.org/10.1016/ S0072-9752(07)830053.
http://www.sciencedirect.com/science/article/pii/S0072975207830053.

[15] J.A. Obeso, M. Stamelou, C.G. Goetz, W. Poewe, A.E. Lang, D. Weintraub, D. Burn, G.M. Halliday, E. Bezard, S. Przedborski, S. Lehericy, D.J. Brooks, J.C. Rothwell, M. Hallett, M.R. Delong, C. Marras, C.M. Tanner, G.W. Ross, J.W. Langston, C. Klein, V. Bonifati, J. Jankovic, A.M. Lozano, G. Deuschl,H. Bergman, E. Tolosa, M. Rodriguez-Violante, S. Fahn, R.B. Postuma, D. Berg, K. Marek, D.G. Standaert, D.J. Surmeier, C.W. Olanow, J.H. Kordower, P. Calabresi, A.H.V. Schapira, A.J. Stoessl, Past, present, and future of parkinson's

disease: A special essay on the 200th anniversary of the shaking palsy, Movement Disorders 32 (9) (2017) 1264–1310. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/mds.27115, doi:10.1002/mds.27115. https://onlinelibrary.wiley.com/doi/abs/10.1002/mds.27115.

[16] B. Rana, A. Juneja, M. Saxena, S. Gudwani, S. Senthil Kumaranc, R. Agrawal, M. Behari, Relevant 3d local binary pattern based features from fused feature descriptor for differential diagnosis of parkinson's disease using structural mri 34 (2017) 134–143. doi:https://doi.org/10.1016/j.bspc.2017. 01.007 http://www.sciencedirect.com/science/article/pii/ S1746809417300058.

[17] H. Jones, T. Burns, M. Aminoff, S. Pomeroy, The Netter Collection of Medical Illustrations: Nervous System, Volume 7, Part 1 - Brain e-Book, Netter Green Book Collection, Elsevier Health Sciences, 2013. https://books.google.com.tr/books?id=kDpeeru0XD4C.

[18] S. Sigurlaug, The clinical symptoms of parkinson's disease, Journal of Neurochemistry 139 (S1) (2013) 318–324. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/jnc.13691, doi:10.1111/jnc.13691. https://onlinelibrary.wiley.com/doi/abs/10.1111/jnc.13691.

[19] L. V. Kalia, A. E. Lang, Parkinson's disease, The Lancet 386 (9996) (2015) 896 – 912.doi:https://doi.org/10.1016/S0140-6736(14)61393-3. http://www.sciencedirect.com/science/article/pii/ S0140673614613933.

[20] S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, C. Davatzikos, A review on neuroimaging-based classification studies and associated feature extraction methods for alzheimer's disease and its prodromal stages, NeuroImage 155 (2017) 530 –548.doi:https://doi.org/10.1016/j.neuroimage.2017.03.057. http://www.sciencedirect.com/science/article/pii/ S1053811917302823.

[21] M. R. Arbabshirani, S. Plis, J. Sui, V. D. Calhoun, Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls, NeuroImage 145 (2017) 137 – 165, individual Subject Prediction. doi:https://doi.org/10.1016/j.neuroimage.2016.02.079. http://www.sciencedirect.com/science/article/pii/ S105381191600210X.

[22] R. Schwartz, K. Rothermich, S. A. Kotz, M. D. Pell, Unaltered emotional experience in parkinson's disease: Pupillometry and behavioral evidence, Journal of Clinical and Experimental Neuropsychology 40 (3) (2018) 303–316, pMID: 28669253.doi:10.1080/13803395.2017.1343802. https://doi.org/10.1080/13803395.2017.1343802.

[23] A. Pilotto, E. Premi, S. Paola Caminiti, L. Presotto, R. Turrone, A. Alberici, B. Paghera, B. Borroni, A. Padovani, D. Perani, Single-subject spm fdg-pet patterns predict risk of dementia progression in parkinson disease, Neurology.

[24] J. Pasquini, R. Ceravolo, Z. Qamhawi, J.-Y. Lee, G. Deuschl, D. J. Brooks, Bonuccelli, N. Pavese, Progression of tremor in early stages of parkin- son's disease: a clinical and neuroimaging study, Brain 141 (3) (2018) 811–821. arXiv:/oup/backfile/content_public/journal/brain/141/3/10.1093_brain_awx376

/3/awx376.pdf, doi:10. 1093/brain/awx376. http://dx.doi.org/10.1093/brain/awx376.

[25] I. V. Zenyuk, D. Y. Parkinson, G. Hwang, A. Z. Weber, Probing water distribution in compressed fuel-cell gas-diffusion layers using x-ray computed tomography, Electrochemistry Communications 53 (2015) 24 – 28. doi:https://doi.org/10.1016/j.elecom.2015.02.005. http://www.sciencedirect.com/science/article/pii/ S1388248115000405.

[26] B. Ravina, D. Eidelberg, J. E. Ahlskog, R. L. Albin, D. J. Brooks, M. Carbon, Dhawan, A. Feigin, S. Fahn, M. Guttman, K. Gwinn-Hardy, H. McFarland, R. Innis, R. G. Katz, K. Kieburtz, S. J. Kish, N. Lange, J. W. Langston, K. Marek, L. Morin, C. Moy, D. Murphy, W. H. Oertel, G. Oliver, Y. Palesch, W. Powers, J. Seibyl, K. D. Sethi, C. W. Shults, P. Sheehy, A. J. Stoessl, R. Holloway, The role of radiotracer imaging in parkinson disease, Neurology 64 (2) (2005) 208–215.

[27] B. Rana, A. Juneja, M. Saxena, S. Gudwani, S. S. Kumaran, R. Agrawal, M. Behari, Regions-of-interest based automated diagnosis of parkinson's disease using t1-weighted mri, Expert Systems with Applications 42 (9) (2015) 4506 – 4516. doi:https://doi.org/10.1016/j.eswa.2015.01.062.http://www.sciencedirect.com/science/article/pii/S0957417415000858.

[28] D. Joshi, A. Khajuria, P. Joshi, An automatic non-invasive method for parkinson's disease classification, Computer Methods and Programs in Biomedicine 145 (2017) 135 – 145. doi:https://doi.org/10.1016/j.cmpb.2017. 04.007.

[29] E. Adeli-Mosabbeb, C.-Y. Wee, L. An, F. Shi, D. Shen, Joint feature-sample selection and robust classification for parkinson's disease diagnosis, in: Medical Computer Vision: Algorithms for Big Data, Springer International Publishing, 2016, pp. 127–136.

[30] C. Ornelas-Vences, L. P. Sanchez-Fernandez, L. A. Sanchez-Perez, A. Garza-Rodriguez, A. Villegas-Bastida, Fuzzy inference model evaluating turn for parkinson's disease patients, Computers in Biology and Medicine 89 (2017) 379 – 388. doi:https://doi.org/10.1016/j.compbiomed.2017.08.026.

[31] I. Beheshti, H. Demirel, Feature-ranking-based alzheimer's disease classification from structural mri, Magnetic Resonance Imaging 34 (3) (2016) 252 – 263. doi:https://doi.org/10.1016/j.MRI.2015.11.009. http://www.sciencedirect.com/science/article/pii/S0730725X15002945.

[32] G. S. Babu, S. Suresh, B. S. Mahanand, A novel pbl-mcrbfn-rfe approach for identification of critical brain regions responsible for parkinson's disease, Expert Systems with Applications 41 (2) (2014) 478 – 488. doi:https://doi.org/10.1016/j.eswa.2013.07.073. http://www.sciencedirect.com/science/article/pii/ S0957417413005605.

[33] B. Rana, A. Juneja, M. Saxena, S. Gudwani, S. S. Kumaran, R. K. Agrawal, 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, ICVGIP '16, 2016, pp. 22:1–22:6.

[34] J. J. Chen, Parkinson's disease: health-related quality of life, economic cost, and implications of early treatment, The American journal of managed care 16 Suppl Implications (2010) S87—93.

[35] Y. Chen, J. Storrs, L. Tan, L. J. Mazlack, J.-H. Lee, L. J. Lu, Detecting brain structural changes as biomarker from magnetic resonance images using a local feature based svm approach, Journal of Neuroscience Methods 221 (2014) 22 – 31.doi:https://doi.org/10.1016/j.jneumeth.2013.09.001. http://www.sciencedirect.com/science/article/pii/S0165027013003051.

[36] S. Duchesne, Y. Rolland, M. Vérin, Automated computer differential classification in parkinsonian syndromes via pattern analysis on mri, Academic Radiology 16 (1) (2009) 61 – 70. doi:https://doi.org/10.1016/j.acra. 2008.05.024. http://www.sciencedirect.com/science/article/pii/ S1076633208004042.

[37] F. N. K., H. Gunther, S. Sebstian, P. P. M., B. C. G., D. Peter, E. Jens, M. Alexander, P. Walter, T. Claudia, Individual voxel-based subtype prediction can differentiate progressive supranuclear palsy from idiopathic parkinson syndrome and healthy controls, Human Brain Mapping 32 (11) 1905–1915. doi:10.1002/hbm.21161. https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.21161.

[38] H. Hans-Jürgen, M. Leona, S. Martin, H. Rüdiger, H. Elke, E. Karl, A. Florian, R. Gesine, S. Maria, S. Alfons, P. E. H., O. W. H., K. Susanne, K. Jan, H. G. U., Differentiation of neurodegenerative parkinsonian syndromes by

volumetric magnetic resonance imaging analysis and support vector machine classification, Movement Disorders 31 (10) (2018) 1506–1517. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/mds.26715. doi:10.1002/mds.26715. https://onlinelibrary.wiley.com/doi/abs/10.1002/mds.26715.

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

[40] T. Li, W. Li, Y. Yang, W. Zhang, Classification of brain disease in magnetic resonance images using two-stage local feature fusion, PLOS ONE 12 (2) (2017) 1–19.doi:10.1371/journal.pone.0171749. https://doi.org/10.1371/journal.pone.0171749.

[41] D. Long, J. Wang, M. Xuan, Q. Gu, X. Xu, D. Kong, M. Zhang, Automatic classification of early parkinson's disease with multi-modal mr imaging, PLOS ONE 7 (11) (2012) 1–9. doi:10.1371/journal.pone.0047714.

[42] R. Morisi, G. Gnecco, N. Lanconelli, S. Zanigni, D. N. Manners, C. Testa, S. Evangelisti, L. L. Gramegna, C. Bianchini, P. Cortelli, C. Tonon, R. Lodi, Binary and multi-class parkinsonian disorders classification using support vector machines, in: Pattern Recognition and Image Analysis, Springer International Publishing, Cham, 2015, pp. 379–386.

[43] R. Morisi, D. Neil Manners, G. Gnecco, N. Lanconelli, C. Testa, S. Evange- listi, L. Talozzi, L. Gramegna, C. Bianchini, G. Calandra-Buonaura, L. Sambati, G. Giannini, P. Cortelli, C. Tonon, R. Lodi, Multi-class parkinsonian disor- ders classification with quantitative mr markers and graph-based features using support vector machines 47.

[44] R. Bharti, J. Akanksha, S. Mohit, G. Sunita, K. S. Senthil, B. Madhuri, R. K., 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 25 (3) 245–255. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/ima.22141. doi:10.1002/ima.22141. https://onlinelibrary.wiley.com/doi/abs/10.1002/ima.22141.

[45] J. S. N. D. F. Sahand Talai, Kai Boelmans, Automatic classification of patients with idiopathic parkinson's disease and progressive supranuclear palsy using diffusion mri datasets (2017). doi:10.1117/12.2254418. https://doi.org/10.1117/12.2254418.

[46] O. Cigdem, A. Yilmaz, I. Beheshti, H. Demirel, Comparing the performances of pdf and pca on parkinson's disease classification using structural mri images, 26th IEEE Signal Processing and Communication Applications (SIU) 2018, 2018.

[47] P. Justin Galvis, Adam F. Mezher, Effects of epi distortion correction pipelines on the connectome in parkinson's disease (2016). doi:10.1117/12.2217377.

https://doi.org/10.1117/12.2217377.

[48] K. Kamagata, A. Zalesky, T. Hatano, M. A. D. Biase, O. E. Samad, S. Saiki, K. Shimoji, K. K. Kumamaru, K. Kamiya, M. Hori, N. Hattori, S. Aoki, C. Pantelis, Connectome analysis with diffusion mri in idiopathic parkin- son's disease: Evaluation using multi-shell, multi-tissue, constrained spheri- cal deconvolution, NeuroImage: Clinical 17 (2018) 518 – 529. doi:https://doi.org/10.1016/j.nicl.2017.11.007.
http://www.sciencedirect.com/science/article/pii/ S2213158217302875.

[49] B. Peng, S. Wang, Z. Zhou, Y. Liu, B. Tong, T. Zhang, Y. Dai, A multilevel- roi-features-based machine learning method for detection of morphometric biomarkers in parkinson's disease, Neuroscience Letters 651 (2017) 88 – 94. doi:https://doi.org/10.1016/j.neulet.2017.04.034.
http://www.sciencedirect.com/science/article/pii/S030439401730335X.

[50] B. Peng, Z. Zhou, C. Geng, B. Tong, Z. Zhou, T. Zhang, Y. Dai, Computer aided analysis of cognitive disorder in patients with parkinsonism using machine learning method with multilevel roi-based features, in: 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Infor- matics (CISP-BMEI), 2016, pp. 1792–1796. doi:10.1109/CISP-BMEI.2016.7853008.

[51] G. Singh, M. Vadera, L. Samavedham, E. C.-H. Lim, Machine learning-based framework for multi-class diagnosis of neurodegenerative diseases: A study on parkinson's disease, IFAC-PapersOnLine 49 (7) (2016) 990 – 995, 11th IFAC

Symposium on Dynamics and Control of Process SystemsIncluding Biosystems DYCOPS-CAB 2016. doi:https://doi.org/10.1016/j.ifacol. 2016.07.331. http://www.sciencedirect.com/science/article/pii/ S2405896316305389.

[52] H. Zeng, Y. m. Cheung, Feature selection and kernel learning for local learning-based clustering, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (8) (2011) 1532–1547. doi:10.1109/TPAMI.2010.215.

[53] D. Gellerup, Discriminating parkinson's disease using functional connectivity and brain network analysis.

[54] G. Singh, L. Samavedham, Unsupervised learning based feature extraction for differential diagnosis of neurodegenerative diseases: A case study on early-stage diagnosis of parkinson disease, Journal of Neuroscience Methods 256 (2015) 30 – 40.doi:https://doi.org/10.1016/j.jneumeth.2015.08.011. http://www.sciencedirect.com/science/article/pii/ S0165027015003003.

[55] G. Singh, L. Samavedham, Algorithm for image-based biomarker detection for differential diagnosis of parkinson's disease, IFAC-PapersOnLine 48 (8) (2015) 918 – 923, 9th IFAC Symposium on Advanced Control of Chemical Processes ADCHEM2015.doi:https://doi.org/10.1016/j.ifacol.2015.09.087. http://www.sciencedirect.com/science/article/pii/ S2405896315011684.

[56] L. Liu, Q. Wang, E. Adeli, L. Zhang, H. Zhang, D. Shen, Feature selection based on iterative canonical correlation analysis for automatic diagnosis of parkinson's disease, in: Medical Image Computing and Computer-Assisted Intervention –

MICCAI 2016, Springer International Publishing, Cham, 2016, pp. 1–8.

[57] L. Liu, Q. Wang, E. Adeli, L. Zhang, H. Zhang, D. Shen, Exploring diagnosis and imaging biomarkers of parkinson's disease via iterative canonical correlation analysis based feature selection, Computerized Medical Imaging and Graphics 67 (2018) 21 – 29. doi:https://doi.org/10.1016/j.compmedimag.2018.04.002. http://www.sciencedirect.com/science/article/pii/ S0895611118301861.

[58] E. Adeli-Mosabbeb, C.-Y. Wee, L. An, F. Shi, D. Shen, Joint feature-sample selection and robust classification for parkinson's disease diagnosis, in: Medical Computer Vision: Algorithms for Big Data, Springer International Publishing, Cham, 2016, pp. 127–136.

[59] E. Adeli, F. Shi, L. An, C.-Y. Wee, G. Wu, T. Wang, D. Shen, Joint feature-sample selection and robust diagnosis of parkinson's disease from mri data, NeuroImage 141 (2016) 206 – 219. doi:https://doi.org/10.1016/ j.neuroimage.2016.05.054. http://www.sciencedirect.com/science/article/pii/ S1053811916301744.

[60] E. Adeli, K. H. Thung, L. An, G. Wu, F. Shi, T. Wang, D. Shen, Semi-supervised discriminative classification robust to sample-outliers and feature-noises, IEEE Transactions on Pattern Analysis and Machine Intelligence (2018) 1–1doi: 10.1109/TPAMI.2018.2794470.

[61] M. Banerjee, M. S. Okun, D. E. Vaillancourt, B. C. Vemuri, A method for automated classification of parkinson's disease diagnosis using an ensemble

average propagator template brain map estimated from diffusion mri, PLOS ONE 11 (6) (2016) 1–11. doi:10.1371/journal.pone.0155764. https://doi.org/10.1371/journal.pone.0155764.

[62] W.-C. Lin, K.-H. Chou, P.-L. Lee, N.-W. Tsai, H.-L. Chen, A.-L. Hsu, M.-H. Chen, , C. Huang, C.-P. Lin, C.-H. Lu, Parkinson's disease: diagnostic utility of volumetric imaging, Neuroradiology 59 (4) (2017) 367–377. doi: 10.1007/s00234-017-1808-0. https://doi.org/10.1007/s00234-017-1808-0.

[63] J. Wang, D. Long, Z. Chen, Early diagnosis of parkinson's disease patients using rvm-based classification with multi-characteristics, in: 2013 IEEE Third International Conference on Information Science and Technology (ICIST), 2013, pp. 54–58. doi:10.1109/ICIST.2013.6747499.

[64] D. A. Morales, Y. Vives-Gilabert, B. Gómez-Ansón, E. Bengoetxea, P. Larrañaga, C. Bielza, J. Pagonabarraga, J. Kulisevsky, I. Corcuera-Solano, M. Delfino, Predicting dementia development in parkinson's disease using bayesian network classifiers, Psychiatry Research: Neuroimaging 213 (2) (2013) 92 – 98. doi: https://doi.org/10.1016/j.pscychresns.2012.06.001. http://www.sciencedirect.com/science/article/pii/ S0925492712001254.

[65] G. Pahuja, T. N. Nagabhushan, A novel ga-elm approach for parkinson's disease detection using brain structural t1-weighted mri data, in: 2016 Second International Conference on Cognitive Computing and Information Processing (CCIP), 2016, pp. 1–6. doi:10.1109/CCIP.2016.7802848.

[66] N. Amoroso, M. L. Rocca, A. Monaco, R. Bellotti, S. Tangaro, Complex networks reveal early mri markers of parkinson's disease, Medical Image Analysis 48 (2018) 12 – 24. doi:https://doi.org/10.1016/j.media.2018. 05.004. http://www.sciencedirect.com/science/article/pii/ S1361841518302858.

[67] H. Lei, Z. Huang, J. Zhang, Z. Yang, E.-L. Tan, F. Zhou, B. Lei, Joint detection and clinical score prediction in parkinson's disease via multi-modal sparse learning, Expert Systems with Applications 80 (2017) 284 – 296. doi:https://doi.org/10.1016/j.eswa.2017.03.038. http://www.sciencedirect.com/science/article/pii/S0957417417301926.

[68] H. Lei, Z. Huang, T. Han, Q. Luo, Y. Cai, G. Liu, B. Lei, Joint regression and classification via relational regularization for parkinson's disease diagnosis, Technology and Health Care 26 (2018) 19–30. doi:10.3233/thc-174540.

[69] H. Lei, Y. Zhao, Y. Wen, B. Lei, Adaptive sparse learning for neurodegenerative disease classification, in: 2017 IEEE International Symposium on Multimedia (ISM), 2017, pp. 292–295. doi:10.1109/ISM.2017.51.

[70] A. Cherubini, R. Nisticò, F. Novellino, M. Salsone, S. Nigro, G. Donzuso, Quattrone, Magnetic resonance support vector machine discriminates essential tremor with rest tremor from tremor-dominant parkinson disease 29.

[71] E. Adeli, G. Wu, B. Saghafi, L. A. andFeng Shi, D. Shen, Kernel-based joint feature selection and max-margin classification for early diagnosis of parkinson's disease, Sci. Rep. 7, 41069, 2017. doi:doi:10.1038/srep41069.

https://www.nature.com/articles/srep41069.pdf.

[72] R. Prashanth, S. D. Roy, P. K. Mandal, S. Ghosh, High-accuracy detection of early parkinson's disease through multimodal features and machine learning, International Journal of Medical Informatics 90 (2016) 13 – 21. doi:https://doi.org/10.1016/j.ijmedinf.2016.03.001. http://www.sciencedirect.com/science/article/pii/ S1386505616300326.

[73] S. Haller, S. Badoud, D. Nguyen, I. Barnaure, M.-L. Montandon, K.-O. Lovblad, P. Burkhard, Differentiation between parkinson disease and other forms of parkinsonism using support vector machine analysis of susceptibility- weighted imaging (swi): initial results, European Radiology 23 (1) (2013) 12–19. doi:10.1007/s00330-012-2579-y. https://doi.org/10.1007/s00330-012-2579-y.

[74] Y. Chen, W. Yang, J. Long, Y. Zhang, J. Feng, Y. Li, B. Huang, Discrimina- tive analysis of parkinson's disease based on whole-brain functional connec- tivity, PLOS ONE 10 (4) (2015) 1–16. doi:10.1371/journal.pone. 0124153. https://doi.org/10.1371/journal.pone.0124153.

[75] V. Dunet, J. Deverdun, C. Charroud, E. Le Bars, F. Molino, S. Menjot de Champfleur, F. Maury, M. Charif, X. Ayrignac, P. Labauge, G. Castelnovo, F. Pinna, A. Bonafe, C. Geny, N. Menjot de Champfleur, Mri volumetric mor- phometry in vascular parkinsonism, Journal of Neurology 264 (7) (2017) 1511–1519. doi:10.1007/s00415-017-8561-5. https://doi.org/10.1007/s00415-017-8561-5.

[76] Y. Gao, K. Nie, B. Huang, M. Mei, M. Guo, S. Xie, Z. Huang, L. Wang, J. Zhao, Y. Zhang, L. Wang, Changes of brain structure in parkinson's disease patients with mild cognitive impairment analyzed via vbm technology, Neuroscience Letters 658 (2017) 121 – 132. doi:https://doi.org/10.1016/j.neulet.2017.08.028. http://www.sciencedirect.com/science/article/pii/ S0304394017306754.

[77] C. Gao, H. Sun, T. Wang, M. Tang, N. I. Bohnen, M. L. T. M. Müller, T. Herman, N. Giladi, A. A. Kalinin, C. Spino, W. T. Dauer, J. M. Hausdorff, I. D. Dinov, Model-based and model-free machine learning techniques for diagnostic prediction and classification of clinical outcomes in parkinson's disease, in: Scientific Reports, 2018.

[78] Q. Gu, H. Zhang, M. Xuan, W. Luo, P. Huang, S. Xia, M. Zhang, Automatic classification on multi-modal mri data for diagnosis of the postural instability and gait difficulty subtype of parkinson's disease 6.

[79] K. Hikishima, K. Ando, Y. Komaki, K. Kawai, R. Yano, T. Inoue, T. Itoh, M. Yamada, S. Momoshima, H. Okano, H. Okano, Voxel-based morphometry of the marmoset brain: In vivo detection of volume loss in the substantia nigra of the mptp-treated parkinson's disease model, Neuroscience 300 (2015) 585– 592. doi:https://doi.org/10.1016/j.neuroscience.2015.05.041. http://www.sciencedirect.com/science/article/pii/ S0306452215004832.

[80] S. Mangesius, A. Hussl, F. Krismer, P. Mahlknecht, E. Reiter, S. Tagwercher, Djamshidian, M. F. H. Schocke, R. Esterhammer, G. K. Wenning, C. E. Müller, C. Scherfler, E. R. Gizewski, W. Poewe, K. Seppi, Mr planimetry in neurodegen-

erative parkinsonism yields high diagnostic accuracy for psp., Parkinsonism & related disorders 46 (2018) 47–55.

[81] A. F. Marquand, M. Filippone, J. Ashburner, M. Girolami, J. Mourao-Miranda, G. J. Barker, S. C. R. Williams, P. N. Leigh, C. R. V. Blain, Automated, high accuracy classification of parkinsonian disorders: A pattern recognition approach, PLOS ONE8(7).doi:10.1371/journal.pone.0069237. https://doi.org/10.1371/journal.pone.0069237.

[82] Q. Noirhomme, D. Lesenfants, F. Gomez, A. Soddu, J. Schrouff, G. Garraux, Luxen, C. Phillips, S. Laureys, Biased binomial assessment of cross-validated estimation of classification accuracies illustrated in diagnosis predictions, NeuroImage: Clinical 4 (2014) 687 – 694. doi:https://doi.org/10.1016/j.nicl.2014.04.004.
http://www.sciencedirect.com/science/article/pii/S2213158214000485.

[83] S. Pazhanirajan, P. Dhanalakshmi, Mri classification of parkinson's disease using svm and texture features, in: Proceedings of the Second International Conference on Computer and Communication Technologies, Springer India, New Delhi, 2016, pp. 357–364.

[84] S. Pazhanirajan, P. Dhanalakshmi, Classification of parkinson's disease using mri images, Vol. 5, 2016, pp. 233 – 237. http://ijcsse.org/published/volume5/issue10/ p3-V5I10.pdf.

[85] P. Patrice, B. Gaetano, N. Federico, S. Maria, G. Monique, T. A. Pavy-Le,

P. Pierre, M. W. G., R. Olivier, Mri supervised and un- supervised classification of parkinson's disease and multiple system atrophy, Movement Disorders 33(4)600–08. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/mds.27307, doi:10.1002/mds.27307.https://onlinelibrary.wiley.com/doi/abs/10.1002/mds.2730.

[86] J. Schrouff, J. Crémers, Discriminant bold activation patterns during mental imagery in parkinson's disease.

[87] H. D. Tagare, C. DeLorenzo, S. Chelikani, L. Saperstein, R. K. Fulbright, Voxel-based logistic analysis of ppmi control and parkinson's disease datscans, NeuroImage 152 (2017) 299 – 311. doi:https://doi.org/10.1016/j.neuroimage.2017.02.067. http://www.sciencedirect.com/science/article/pii/S1053811917301787.

[88] Y. Zeighami, S.-M. Fereshtehnejad, M. Dadar, D. L. Collins, R. B. Postuma, B. Mišić, A. Dagher, A clinical-anatomical signature of parkinson's disease identi- fied with partial least squares and magnetic resonance imaging, NeuroImage. doi: https://doi.org/10.1016/j.neuroimage.2017.12.050. http://www.sciencedirect.com/science/article/pii/ S1053811917310741.

[89] L.-L. Zeng, L. Xie, H. Shen, Z. Luo, P. Fang, Y. Hou, B. Tang, T. Wu, D. Hu, Differentiating patients with parkinson's disease from normal controls using gray matter in the cerebellum, The Cerebellum 16 (1) (2017) 151–157. doi: 10.1007/s12311-016-0781-1. https://doi.org/10.1007/s12311-016-0781-1.

[90] Y.-W. Fu, H.-L. Chen, S.-J. Chen, L. Shen, Q. Li, A new evolutionary support vector machine with application to parkinson's disease diagnosis, in: Advances in Swarm Intelligence, Springer International Publishing, Cham, 2014, pp. 42–49.

[91] J. Ashburner, Spm: A history, NeuroImage 62 (2) (2012) 791 – 800, 20 YEARS OF fMRI. doi:https://doi.org/10.1016/j.neuroimage.2011. 10.025. http://www.sciencedirect.com/science/article/pii/ S1053811911011888.

[92] F. Farokhian, I. Beheshti, D. Sone, H. Matsuda, Comparing cat12 and vbm8 for detecting brain morphological abnormalities in temporal lobe epilepsy, in: Front. Neurol., 2017.

[93] J. Ashburner J. SPM: A history. Neuroimage. 2012;62-248(2):791-800. doi:10.1016/j.neuroimage.2011.10.025.

[94] W. P. John Ashburner, Karl J. Friston, Human Brain Function, 2003.

[95] N. K. Focke, M. Yogarajah, M. R. Symms, O. Gruber, W. Paulus, J. S. Duncan, Automated mr image classification in temporal lobe epilepsy, NeuroImage 59 (1) (2012) 356—362. doi:https://doi.org/10.1016/j.neuroimage.2011.07.068. http://www.sciencedirect.com/science/article/pii/ S1053811911008639.

[96] K. Friston, A. Holmes, C. Price, C. Büchel, K. Worsley, Multisubject fmri studies and conjunction analyses 10 (1999) 385–96.

[97] J. B. e. a. Rowe, Parkinson's disease and dopaminergic therapy—differential

effects on movement, reward and cognition, Brain (131.8). doi:10.1093/brain/awn112. http://dx.doi.org/10.1093/brain/awn112.

[98] G. Roffo, Ranking to learn and learning to rank: On the role of ranking in pattern recognition applications, CoRR abs/1706.05933. arXiv:1706.05933. http://arxiv.org/abs/1706.05933.

[99] R. J. Urbanowicz, M. Meeker, W. L. Cava, R. S. Olson, J. H. Moore, Relief-based feature selection: Introduction and review, CoRR abs/1711.08421. arXiv: 1711.08421. http://arxiv.org/abs/1711.08421.

[100] D. Cai, C. Zhang, X. He, Unsupervised feature selection for multi-cluster data, in: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '10, ACM, New York, NY, USA, 2010, pp.333–342.doi:10.1145/1835804.1835848. http://doi.acm.org/10.1145/1835804.1835848.

[101] Y. Yang, H. T. Shen, Z. Ma, Z. Huang, X. Zhou, L2,1-norm regularized discriminative feature selection for unsupervised learning, in: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelli- gence - Volume Volume Two, IJCAI'11, AAAI Press, 2011, pp. 1589–1594. doi:10.5591/978-1-57735-516-8/IJCAI11-267. http://dx.doi.org/10.5591/978-1-7735-516-8/IJCAI11-267.

[102] L. Du, Y.-D. Shen, Unsupervised feature selection with adaptive structure learning, in: KDD, 2015.

[103] M. A. Hall, Correlation-based feature selection for machine learning, Tech. rep. (1999).

[104] D. Jain, V. Singh, Feature selection and classification systems for chronic disease prediction: A review, Egyptian Informatics Journal. doi:https://doi.org/10.1016/j.eij.2018.03.002. http://www.sciencedirect.com/science/article/pii/ S1110866517300294.

[105] G. Roffo, Report: Feature selection techniques for classification, CoRR abs/1607.01327. arXiv:1607.01327. http://arxiv.org/abs/1607.01327.

[106] C.-H. Lin, C.-M. Chen, M.-K. Lu, C.-H. Tsai, J. C. Chiou, J.-R. Liao, J.-R. Duann, Vbm reveals brain volume differences between parkinson's disease and essential tremor patients 7 (2013) 247. doi:10.3389/fnhum.2013.00247. https://www.frontiersin.org/article/10.3389/fnhum.2013.00247.

[107] U. Ellfolk, J. Joutsa, J. O Rinne, R. Parkkola, P. Jokinen, M. Karrasch, Brain volumetric correlates of memory in early parkinson's disease 3.

[108] G. S. Pell, R. S. Briellmann, C. H. P. Chan, H. Pardoe, D. F. Abbott, G. D. Jackson, Selection of the control group for vbm analysis: influence of covariates, matching and sample size, NeuroImage 41 (4) (2008) 1324—1335. doi:https://doi.org/10.1016/j.neuroimage.2008.02.050. http://www.sciencedirect.com/science/article/pii/S1053811908001936.

[109] I. B. Malone, K. K. Leung, S. Clegg, J. Barnes, J. L. Whitwell, J. Ashburner,

N. C. Fox, G. R. Ridgway, Accurate automatic estimation of total intracranial volume: A nuisance variable with less nuisance, NeuroImage 104 (2015) 366-372.

[110] J. Barnes, G. Ridgway, J. Bartlett, S. Henley, M. Lehmann, N. Hobbs, M. Clarkson, D. MacManus, S. Ourselin, N. Fox, Head size, age and gender adjustment in mri studies: a necessary nuisance?, NeuroImage 53 (4) (2010) 1244–1255.

[111] C. D. Smith, H. Chebrolu, D. R. Wekstein, F. A. Schmitt, W. R. Markesbery, Age, and gender effects on human brain anatomy: A voxel-based morphometric study in healthy elderly, Neurobiology of Aging 28 (7) (2007) 1075 – 1087.

[112] Y. Taki, R. Goto, A. Evans, A. Zijdenbos, P. Neelin, J. Lerch, K. Sato, S. Ono, S. Kinomura, M. Nakagawa, M. Sugiura, J. Watanabe, R. Kawashima, H. Fukuda, Voxel-based morphometry of human brain with age and cerebrovascular risk factors, Neurobiology of Aging 25 (4) (2004) 455 – 463.

[113] E. Courchesne, H. J. Chisum, J. Townsend, A. Cowles, J. Covington, B. Egaas, M. Harwood, S. Hinds, G. A. Press, Normal brain development and aging: Quantitative analysis at in vivo mr imaging in healthy volunteers, Radiology 216 (3) (2000) 672–682.

[114] Y. Ge, R. I. Grossman, J. S. Babb, M. L. Rabin, L. J. Mannon, D. L. Kolson, Age-related total gray matter and white matter changes in normal adult brain. part i: Volumetric mr imaging analysis, American Journal of Neuroradiology 23 (8) (2002) 1327–1333.

[115] H. Lemaître, F. Crivello, B. Grassiot, A. Alpérovitch, C. Tzourio, B. Mazoyer, Age- and sex-related effects on the neuroanatomy of healthy elderly, NeuroImage 26 (3) (2005) 900 – 911.

[116] A. Besga, I. Gonzalez, E. Echeburúa, A. Savio, B. Ayerdi, D. Chyzhyk, J. L. M. Madrigal, J. C. Leza, M. Graña, A. Gonzalez-Pinto, Discrimination between alzheimer's disease and late onset bipolar disorder using multivariate analysis, in: Front. Aging Neurosci., 2015.

[117] S.-H. Lee, J. S. Lim, Parkinson's disease classification using gait characteristics and wavelet-based feature extraction, Expert Systems with Applications 39 (8) (2012) 7338 – 7344. doi:https://doi.org/10.1016/j.eswa.2012. 01.084. http://www.sciencedirect.com/science/article/pii/ S0957417412000978.

[118] X. Jia, P. Liang, Y. Li, L. Shi, D. Wang, K. Li, Longitudinal study of gray matter changes in parkinson disease., AJNR. American journal of neuroradiology 36 12 (2015) 2219–26.

[119] L. Christopher, Y. Koshimori, A. E. Lang, M. Criaud, A. P. Strafella, Uncovering the role of the insula in non-motor symptoms of parkinson's disease, Brain 137 (8) (2014) 2143–2154.

[120] C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines, ACM Transactions on Intelligent Systems and Technology 2 (2011) 27:1–27:27, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[121] L. Zhang, G. Sun, J. Guo, Feature selection for pattern classification problems, in: Proceedings of the The Fourth International Conference on Computer and Information Technology, CIT '04, IEEE Computer Society, Washington, DC, USA, 2004, pp. 233–237. http://dl.acm.org/citation.cfm?id=1025116.1025310.

[122] K. Kira, L. A. Rendell, The feature selection problem: Traditional methods and a new algorithm, in: Proceedings of the Tenth National Conference on Artificial Intelligence, AAAI'92, AAAI Press, 1992, pp. 129–134. http://dl.acm.org/citation.cfm?id=1867135.1867155.

[123] X. He, D. Cai, P. Niyogi, Laplacian score for feature selection, in: Proceedings of the 18th International Conference on Neural Information Processing Systems, NIPS'05, MIT Press, Cambridge, MA, USA, 2005, pp. 507–514. http://dl.acm.org/citation.cfm?id=2976248.2976312.

[124] C. Bouveyron, C. Brunet, Theoretical and practical considerations on the convergence properties of the fisher-em algorithm, Journal of Multivariate Analysis 109 (2012) 29 – 41. doi:https://doi.org/10.1016/j.jmva.2012. 02.012. http://www.sciencedirect.com/science/article/pii/ S0047259X1200053X.

[125] A. Narasimhamurthy, Theoretical bounds of majority voting performance for a binary classification problem, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (12) (2005) 1988–1995. doi:10.1109/TPAMI. 2005.249.

[126] S. Grover, S. Bhartia, Akshama, A. Yadav, S. K.R., Predicting severity of parkin-

son's disease using deep learning, Procedia Computer Science 132 (2018) 1788 – 1794, international Conference on Computational Intelligence and Data Science. doi:https://doi.org/10.1016/j.procs.2018.05.154. http://www.sciencedirect.com/science/article/pii/S1877050918308883.

[127] A. Tagaris, D. Kollias, A. Stafylopatis, Assessment of parkinson's disease based on deep neural networks, in: Engineering Applications of Neural Networks, Springer International Publishing, Cham, 2017, pp. 391–403.

[128] B. M. Eskofier, S. I. Lee, J. F. Daneault, F. N. Golabchi, G. Ferreira-Carvalho, G. Vergara-Diaz, S. Sapienza, G. Costante, J. Klucken, T. Kautz, P. Bonato, Recent machine learning advancements in sensor-based mobility analysis: Deep learning for parkinson's disease assessment, in: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016, pp. 655–658. doi:10.1109/EMBC.2016.7590787.

[129] Welcome Trust Centre for Neuroimaging, London, UK; available at: http:// www.fil.ion.ucl.ac.uk/spm. 04.09.2018.

[130] CAT Toolbox, (http://www. neuro.uni-jena.de/cat/) implemented in Matlab 2017b. 04.09.2018.

[131] Classification Learner App. https://www.mathworks.com/help/stats/classification-learner-app.html). 04.09.2018.

[132] Neuromorphometrics labels http://Neuromorphometrics.com/. 04.09.2018.

[133] MRICron software http://www.mccauslandcenter.sc.edu/crnl/mricron. 04.09.2018.

[134] FSL software https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/. 04.09.2018.