MAGNETIC RESONANCE IMAGING BASED DIFFERENTIAL DIAGNOSIS AND PROGNOSIS OF MILD COGNITIVE IMPAIRMENT IN PARKINSON'S DISEASE USING MACHINE LEARNING

by

Ozan Genç

B.Sc., Electrical & Electronics Engineering, Boğaziçi University, 2016

Submitted to the Institute of Biomedical Engineering
in partial fulfillment of the requirements
for the degree of
Master of Science
in
Biomedical Engineering

Boğaziçi University
2018

ACKNOWLEDGEMENTS

I would like to thank my advisor Asst. Prof. Dr. Esin Öztürk Işık for her support and advice. It has been a wonderful experience working in a large project with enthusiastic colleagues at Computational Imaging Laboratory at Boğaziçi University.

I am very glad to have worked with Gökçe Hale Hatay, Dilek Betül Arslan, Sevim Cengiz, Ani Kıçik, Özge Can Arslan and Ayhan Gürsan and appreciate their support and friendship. I am also grateful to my family for always supporting and encouraging me.

This study was supported by TÜBİTAK project #115S219 and the Ministry of Development project #2010K120330.

ACADEMIC ETHICS AND INTEGRITY STATEMENT

I, Ozan Genç, hereby certify that I am aware of the Academic Ethics and Integrity Policy issued by the Council of Higher Education (YÖK) and I fully acknowledge all the consequences due to its violation by plagiarism or any other way.

Name :

Signature:

Date:

ABSTRACT

MAGNETIC RESONANCE IMAGING BASED DIFFERENTIAL DIAGNOSIS AND PROGNOSIS OF MILD COGNITIVE IMPAIRMENT IN PARKINSON'S DISEASE USING MACHINE LEARNING

Parkinson's disease mild cognitive impairment (PD-MCI), which is one of the major risk factors for dementia, is present in 26.7% of PD patients. In this study, we classified PD-MCI, cognitively normal Parkinson's disease (PD-CN) and healthy control (HC) groups based on multimodal magnetic resonance imaging (MRI) using machine learning methods. We also investigated time dependent changes in PD-MCI patients through a longitudinal study. 33 PD-MCI, 27 PD-CN and 17 HC participated in this study. The participants were diagnosed by neurologists according to the neuropsychological test scores and physical examination results. MRI data was obtained at a 3T Philips clinical MR scanner using a 32-channel head coil. Mean cerebral blood flow (CBF), arterial blood volume (aBV) and bolus arrival time (BAT) maps obtained from arterial spin labeling MRI (ASL-MRI), fractional anisotropy (FA) and mean diffusivity (MD) maps obtained from diffusion tensor imaging (DTI), and metabolite peak ratios obtained from proton MR spectroscopic imaging (1H-MRSI) at various brain regions were used as features. Various machine learning methods were employed with appropriate hyperparameters. Random forest recursive feature elimination (RF-RFE) technique was used for feature selection. For longitudinal analysis, linear mixed effects model was utilized with age, education, gender, visuospatial disorder status, and genotype as covariants. The best classification accuracies were 77% for PD-MCI versus HC, 71% for PD-MCI versus PD-CN, and 86% for PD-CN versus HC. Machine learning based on multimodal MRI might be helpful in early diagnosis of PD-MCI. Reduced aBV and FA, and higher MD values were observed in time in PD-MCI. Future studies will aim to improve the classification of PD-MCI in a larger patient cohort.

Keywords: Parkinson's disease, mild cognitive impairment, multimodal MRI, machine learning, lineer mixed effects model.

ÖZET

PARKİNSON HASTALIĞI HAFİF KOGNİTİF BOZUKLUĞUNUN MANYETİK REZONANS GÖRÜNTÜLEME TEMELLİ MAKİNE ÖĞRENME YÖNTEMLERİYLE TANISI VE PROGNOZU

Parkinson hastalığı hafif kognitif bozukluğu (PH-HKB) demans için büyük bir risk faktörüdür ve PH'nin %26.7'sinde bulunur. Bu çalışmada, makine öğrenme yöntemleri kullanılarak Parkinson hastalığı hafif bilişsel bozukluğu (PH-HKB), bilişsel olarak normal Parkinson hastalığı (PH-KN) ve sağlıklı kontrol (SK) grupları multimodal manyetik rezonans görüntülemeye (MRG) dayalı sınıflandırılmıştır. Ek olarak, boylamsal çalışma ile PD-HKB hastalarında zaman içinde meydana gelen değişiklikler bulunmuştur. Çalışmaya 33 PH-HKB, 27 PH-KN ve 17 SK katılmıştır. Nöropsikolojik test ve muayene sonuçlarına göre katılımcılara nörologlar tarafından tanı konulmuştur. MRG verileri, 32 kanallı kafa bobini kullanılarak 3T Philips klinik MR sisteminde alınmıştır. Atardamar fırıl etiketleme (ASL) yönteminden elde edilen serebral kan akışı (SKA), atardamar kan hacmi (aKH) ve kan ulaşma zamanı (KUZ) verileri, difüzyon tensör görüntülemeden (DTG) elde edilen fraksiyonel anizotropi (FA) ve ortalama difüzivite (MD) verileri, proton MR spektroskopik görüntülemeden (1H-MRSG) elde edilen metabolit pik oranları makine öğrenme yöntemlerinde öznitelik olarak kullanılmıştır. Rassal ormanlar kullanarak özçağrılı öznitelik seçimi yapılmıştır. Boylamsal analiz için lineer karma model kullanılmıştır ve yaş, eğitim, cinsiyet, genetik bilgileri ve vizyospasyal bozukluk durumu kovaryant olarak kullanılmıştır. En iyi sınıflandırma doğrulukları, SK ve PH-HKB için % 77, PH-HKB ve PH-KN için % 71, SK ve PH-KN için % 86 olarak bulunmuştur. Multimodal MRG verilerine dayalı makine öğrenmesinin PH-HKB'nin erken tanısında yardımcı olabileceği düşünülmüştür. İlerideki çalışmalarda, daha büyük hasta popülasyonlarında multimodal MRG temelli PH-HKB sınıflandırmasının iyileştirilmesi hedeflenmektedir.

Anahtar Sözcükler: Parkinson hastalığı, hafif kognitif bozukluk, multimodal MRG, makine öğrenme, lineer karma modeller

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

β	Parameters of Fixed Effects
μ_X	Mean Value of Feature X
σ_X	Standard Deviation of Feature X
c	Actual Class Label
w	Weight Parameters of Features
y	Predicted Class Label

LIST OF ABBREVIATIONS

1H-MRSI	Proton Magnetic Resonance Imaging
aBV	Arterial Blood Volume
ACER	Addenbrooke's Cognitive Examination Revised
AD	Alzheimer's Disease
ASL	Arterial Spin Labeling
AdaBoost	Adaptive Boosting
bagging	bootstrap aggregating
BAT	Bolus Arrival Time
BJLOT	Benton Judgement of Line Orientation Test
CBF	Cerebral Blood Flow
COMT	Catechol-O-methyltransferase
CSF	Cerebrospinal Fluid
Cho	Choline
Cr	Creatine
DTI	Diffusion Tensor Imaging
FA	Fractional Anisotropy
FFE	Fast Field Echo
FLAIR	Fluid Attenuated Inversion Recovery
fMRI	Functional Magnetic Resonance Imaging
FOV	Field of View
GDS	Geriatric Depression Scale
Gln	Glutamine
Glu	Glutamate
HC	Healthy Control
Ins	myo-inositol
kNN	k Nearest Neighbor
LME	Linear Mixed Effects
LOOCV	Leave one out cross validation

MAPT	Microtubule Associated Protein Tau
MD	Mean Diffusivity
MMSE	Mini Mental State Examination
MNI	Montreal Neurological Institute
MRI	Magnetic Resonance Imaging
NAAG	N-acetylaspartylglutamate
NAA	N-acetyl aspartate
NPT	Neuropsychological test
PD-CN	Parkinson's Disease with Cognitively Normal
PD-MCI	Parkinson's Disease Mild Cognitive Impairment
PDD	Parkinson's Disease Dementia
PET	Positron Emission Tomography
PRESS	Point RESolved Spectroscopy
RF-RFE	Random Forest Recursive Feature Elimination
ROI	Region of Interest
RUSBoosted	Random Under Sampling Boosted
SBR	Striatal Binding Ratios
SFFS	Sequential Floating Forward Selection
SPECT	Single Photon Emission Computed Tomography
SVM	Support Vector Machine
TE	Echo Time
TR	Repetition Time
UPDRS	Unified Parkinson's Disease Rating Scale
WCST	Wisconsin Card Sorting Test

1. INTRODUCTION

Parkinson's disease (PD) is a progressive neurodegenerative disease and it has both motor and non-motor symptoms [1]. It is the second most common age related neurodegenerative disease after Alzheimer's disease (AD). It is estimated that seven to ten million people worldwide have PD [2]. The disease appears mostly over the age of 60 years and it is more frequent in men [3]. Motor symptoms of PD involve resting tremor, rigidity, bradykinesia and postural instability [4]. In U.S. average direct national cost per PD patient was $22,800 in 2010 and indirect costs (e.g., reduced employment) were estimated as $10,000 per patient [5]. Mild Cognitive Impairment (MCI) is an intermediate stage between cognitively normal and dementia patients [6]. It may include more severe problems with memory, language, judgment, and thinking than normal age related changes. 26.7% of PD patients meet the criteria for PD-MCI with a majority going on to develop Parkinson's disease dementia (PDD) [7]. Medications are available to treat the symptoms of the disease but unfortunately, a treatment to reverse the disease's effects is unknown. Early detection of the patients who will develop dementia is important in order to start suitable treatment to slow down the progression to dementia. Magnetic resonance imaging (MRI) is a non-invasive volumetric imaging method that doesn't use ionizing radiation. A single MR scanner can provide both anatomical and functional details of the tissue of interest by performing more than one MR modality, such as, functional MRI (fMRI), proton MR spectroscopic imaging (1H-MRSI), arterial spin labeling MRI (ASL-MRI), diffusion tensor MRI (DTI), T1-weighted and T2-weighted MRI. 1H-MRSI gives information about metabolites of the brain. In many brain diseases, differences in the levels of these metabolites and/or their ratios provide diagnostic information. ASL-MRI technique gives details of tissue perfusion in the brain by labeling the intrinsic diffusible tracer [8]. By the help of this technique, we get information about cerebral blood flow (CBF), bolus arrival time (BAT) and arterial blood volume (aBV). DTI gives us information about white matter connectivity patterns in the brain obtained using the diffusion anisotropy and the principal diffusion directions [9]. Degree of anisotropy of the diffusion process called

fractional anisotropy (FA) and measure of the total diffusion within a voxel called mean diffusivity (MD) maps are obtained from DTI. Machine learning (ML) is the scientific discipline focusing on how computers learn from data. Usually it is classified into two branches as supervised and unsupervised learning. ML is being used increasingly in biomedical science and medicine in classification of patients [10]. Some ML techniques that are used in many research are logistic regression, support vector machines (SVM), decision trees and k nearest neighbor classifiers (KNN). In this research, we employed processed multimodal MRI data as predictors to train many types of the classifiers. In biomedical research, repeated measurements of multiple outcomes are frequent. Repeated measurements of the same object over time (longitudinal analysis) are correlated to each other. Different types of measurements of the same data (multivariate analysis) are also correlated to each other. Therefore these correlations should be taken into account when conducting statistical analysis of multivariate longitudinal data [11]. In our study, we used multivariate longitudinal data coming from different MRI modalities measured at two time points that are 1.5 years apart. We conducted multivariate longitudinal data analysis by using linear mixed effects models and Wilcoxon signed rank test. In this project, we collected data of cognitively normal PD (PD-CN), PD-MCI and HC groups using multimodal MRI techniques. Our first purpose was the classification of PD-MCI, PD-CN and HC with high specificity and sensitivity based on multimodal MRI data (1H-MRSI, CBF, aBV, BAT, FA, MD). Our second purpose was the longitudinal analysis of MRI data and neuropsychological test results to detect changing patterns in PD-MCI.

2. BACKGROUND

2.1 Parkinson's Disease

2.1.1 Pathophysiology

Neurons in some regions of substantia nigra producing dopamine diminish in PD [12]. Dopamine is a neurotransmitter regulating motor functions, hence diminishing dopaminergic neurons leads to motor dysfunctions in patients. Deprivation of these dopaminergic neurons firstly occurs in ventrolateral substantia nigra, but as the disease progresses, the loss becomes widespread [13, 14]. Another change occurring in the brain is accumulation of an intracellular protein called α-synuclein inside the cytoplasm of neurons in different regions [15].

2.1.2 Potential Causes

Cause of PD is not fully understood but some patterns are prominent. For example, exposure to pesticides and brain injury increase the possibility of emergence of the disease, whereas consuming caffeine and smoking decrease it [16]. Moreover, the disease is relatively prevalent in some communities. For instance, PD is more common in Ashkenazi Jews having mutations in genes LRRK2 and GBA [17]. It is also relatively common in native American, Inuit, and Alaskan native communities. It is thought that genetic factors and organic pollutants may be risk factors in the prevalence of the disease in these communities [18].

2.1.3 Symptoms and Diagnosis

Both motor and non-motor symptoms exist in PD. Bradykinesia, rigidity, resting tremor are examples of motor symptoms [19] and incomplete list of non-motor ones

includes sleep disorders, cognitive impairment, dementia, hallucination, hyposmia and depression [20]. Physical examination and neuropsychological tests are performed to make diagnosis.

2.1.4 Parkinson's Disease Mild Cognitive Impairment and Dementia

PD-MCI is an intermediate stage between PD-CN and PDD [21]. This stage is a strong indicator for converting to dementia. Because dementia reduces the quality of life dramatically, PD-MCI stage should be examined carefully. Risk factors for evolving to dementia are older age, depression and non-tremor dominant phenotype [4, 22]. It has also been reported that individuals having visuopatial deficits convert to dementia in higher rates than the individuals having frontal executive phenotype [22, 23]. It is revealed that MAPT and COMT genotypes can give information about cognitive decline in PD patients [24]. Catechol-O-methyltransferase (COMT) gene produces a dopamine regulating enzyme [25] and microtubule associated protein tau (MAPT) gene encodes a phosphorylated protein expressed in the brain and it takes charge in stabilizing the cytoskeleton and axonal transport in neurons [26]. COMT gene has genotypes of Val/Met, Val/Val, Met/Met and MAPT gene has genotypes of H1/H2, H1/H1, and H2/H2. Single nucleotide polymorphism at 158. codon of COMT gene (Val158Met) causes decrease in activity of dopamine transporter enzymes [27]. Moreover, task based fMRI analyses revealed that individuals having Met/Met homozygote have less prefrontal activation than individuals having Val/Val homozygote [28]. On the other hand, subjects having H1/H1 genotype of MAPT gene showed more cognitive decline than the subjects carrying H2 haplotype [29].

2.1.5 Prognosis of Parkinson's Disease

Diagnosing PD correctly is a challenging issue. Misdiagnosis rate can be up to 24% in early stages [6]. Motor and non-motor symptoms in late stages result in severe burden for both patients and relatives giving care, hence early diagnosis is very crucial.

It is shown in different clinical studies that dopaminergic neurons demolishes severely before the motor symptoms show up [30, 31]. However, some non-motor symptoms may show up earlier than the onset of motor symptoms hence giving opportunity to diagnose PD earlier [6]. Information regarding the prognosis of PD might enable slowing down the progression of the disease by early intervention. In addition, it might allow patients for planning their future better and enable healthcare system to work better by budgeting necessities of the society [21].

2.1.6 Management of the Disease

PD is a non-reversible disease and no medication exists to stop the progression. However, some medicines are used to alleviate the effects of motor symptoms. Amino acid L-DOPA is a precursor of dopamine and used in treatment of the motor symptoms for a long time [6]. Non-motor symptoms don't respond to L-DOPA treatment [20] hence different medications are suggested for them. For example, cholinesterase inhibitors alleviate cognitive disturbances of PDD patients [32]. In addition, noradrenaline precursor droxidopa is used for orthostatic hypotension, anti-muscarinics are used for incontinence, and pro-kinetic drugs are used for constipation [32, 33].

2.2 Multimodal Magnetic Resonance Imaging

Clinical methods rather than neuroimaging methods are primarily used in diagnosing PD. Positron emission tomography (PET) and single photon emission computed tomography (SPECT) are some imaging techniques utilized conventionally in PD beside clinical methods [34]. However, multimodal MRI techniques also give valuable information about changes in the brains of PD patients. Structural changes and gray matter atrophy in brains of PD-CN and PD-MCI are unfortunately very subtle in early stages of the disease. These changes become prominent when motor symptoms have already been seen in late stages of the disease. Because axonal and synaptic changes occur in the early stages, examining these changes in white matter can allow

for diagnosing the disease earlier [21].

2.2.1 Diffusion Tensor Imaging

Diffusion rates of water in brain tissue are not same at every direction. Showing different diffusion rates for different directions is called anisotropy. Based on this anisotropic characteristic of the tissue, anatomical tracks can be viewed and white matter integrity can be assessed. DTI technique provides information about axonal changes in white matter by utilizing principles mentioned above, hence enabling possible detection of the disease in early stages. Fractional anisotropy map gives information about degree of anisotropy in each voxel, and mean diffusivity map tells about total diffusivity in each voxel. These maps calculated from tractography may give valuable information about white matter changes. FA and MD maps showed that alterations in white matter is positively correlated with increased cognitive impairment in PD patients [35, 36, 37].

2.2.2 Arterial Spin Labeling

ASL MRI technique gives opportunity to view perfusion patterns in the brain without using extrinsic contrast agent. In this technique, endogenous water in arterial blood is electromagnetically tagged with radiofrequency pulses and then tracked while it is perfused to different brain regions. We can obtain CBF, aBV and BAT maps that may give information about alterations of perfusion patterns in the brain. In a study, it was found that perfusion in cortex decreases in PD patients relative to the HC [38]. In another study, it was shown that a correlation exist between metabolic changes observed by fluorodeoxyglucose PET and perfusion patterns observed by ASL MRI [39].

2.2.3 Proton Magnetic Resonance Spectroscopic Imaging

1H-MRSI gives information about peak values of various metabolites inside the brain. Hydrogens spin in different frequencies in different metabolites and this technique utilizes this fact. We get different frequency patterns for different metabolites. Commonly assessed metabolites are N-acetyl aspartate (NAA), choline (Cho), creatine (Cr), myo-inositol (Ins), glutamine (Gln) and glutamate (Glu). By looking at the peak values of metabolites or their ratios in different brain regions we may get valuable information about the metabolic changes in PD. Unfortunately, getting clean information from the substantia nigra with MRS is compelling process because of small size, iron content and location of the region [34]. Despite the fact that some groups claimed that metabolic alterations exist in substantia nigra of PD patients by using 3T MR system [40, 41], these findings could not be verified by using neither 4T [42] nor 7T [43] MR system. In addition, some other studies investigating metabolic changes in substantia nigra of PD and HC, reported conflicting results about NAA/Cr [44, 45, 46]. For this reason, examining the metabolites in other regions such as putamen may be an option. Substantia nigra sends dense dopaminergic afferent projections to this structure thus investigating putamen may give clues about metabolic changes of PD [47]. However contradictions still exist among results of different studies investigating putamen. Decreased NAA/Cr ratio is observed in PD group relative to HC in one study [48], while other studies suggested that the ratio remains unchanged [49, 50].

2.3 Machine Learning Methods

Machine learning methods are powerful tools that helps us achieve tasks that cannot be handled by rule based methods [51]. We can feed machine learning algorithms with data and learn the patterns of this data. Using these patterns, we could classify new subjects based on their features.

2.3.1 Logistic Regression

Logistic regression is a linear classification method widely used in machine learning field. It is one of the simplest methods to understand and implement in machine learning and many practitioners try it first in classification tasks. In this method, response variable or class is tried to be linked by linear combination of input features. Weighted sum of input features is transformed by logistic or sigmoid function to obtain class prediction. A sigmoid function is shown in Fig. 2.1 [52]. During training phase, weights of the input features are "learned" from the data by cost function minimization procedure. During testing phase, weighted sum of attributes is calculated and sigmoid function is used to obtain final prediction output [53]. Equation 2.1 shows the prediction function for logistic regression. If the class density P(C1|x) is greater than 0.5 the instance is predicted as belonging to class C1, if P(C1|x) is smaller than 0.5 the instance is predicted as in class C2.

$$P(C1|x) = sigmoid(w^T x + w_0) = \frac{1}{1 + e^{-(w^T x + w_0)}} \tag{2.1}$$

Cross entropy function to minimize in logistic regression is,

$$E(w, w_0|X) = -\sum c^i \log(y^i) + (1 - c^i) \log(1 - y^i), \tag{2.2}$$

where c^i is actual class label and y^i is predicted class label. We can solve for feature weights minimizing the cost function by iterative numerical methods. Gradient descent algorithm is used commonly as an iterative optimization algorithm. In gradient descent algorithm, partial derivatives of cost function are taken with respect to each weight parameters. These partial derivatives are multiplied by a learning rate parameter and added to randomly initialized weight parameters in opposite direction.

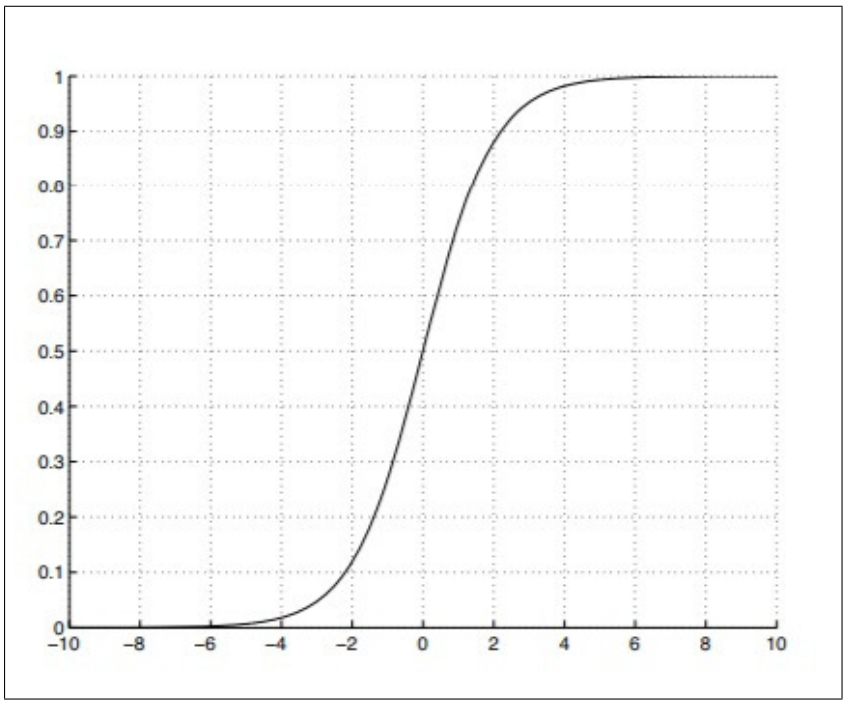

Figure 2.1 Sigmoid function [52]

When we reach to minimum point in cost function, all partial derivatives become zero and we obtain the solution for weights.

Figure 2.2 shows linear decision boundaries ($w_x + w_0$) and the logistic outputs for 10, 100 and 1,000 updates of weight parameters in gradient descent algorithm for an example univariate two class case [54].

This two class procedure could be extended to multiclass case by introducing softmax activation function shown in Eq. 2.3, where K is the number of classes. In this method, weighted sums of a subject's features for each class are calculated by using learned weight parameters specific to each class. Softmax activation function then takes these weighted sums and use them as exponents of Euler's number. After this exponentiation process, the function normalizes the results by dividing with the sum of all exponentials for that subject [55]. Softmax activation function ensures that the response for a class will be close to 1 if the weighted sum of the subject's features for that class is larger than the others, and close to zero for the other classes.

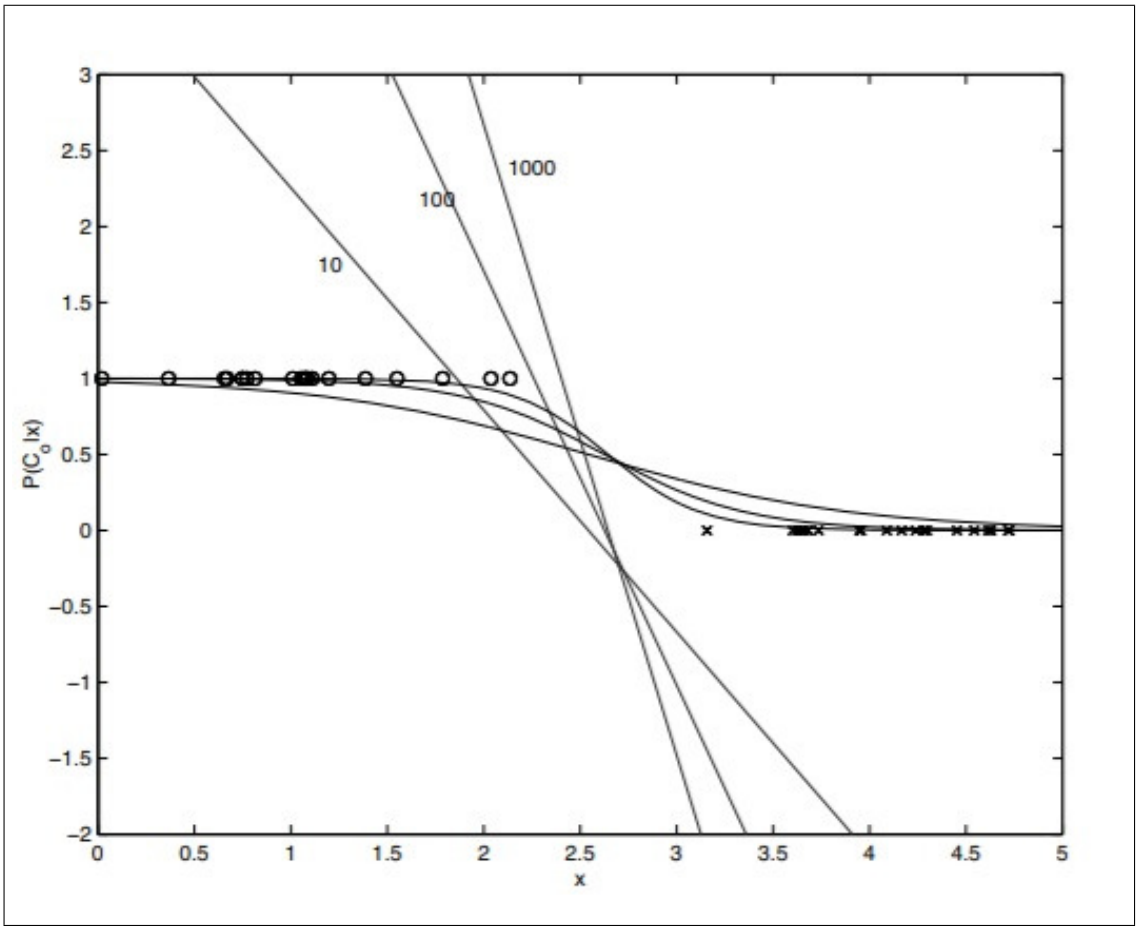

Figure 2.2 Logistic regression decision boundaries [54].

$$y_c = \frac{e^{w_c^T x + w_{c0}}}{\sum_{j=1}^{K} e^{w_j^T x + w_{j0}}}, \quad \text{c=1,...,K} \tag{2.3}$$

2.3.2 k-Nearest Neighbor Classifier

kNN is one of the simplest algorithms and most of the time it is the one tried first in classification tasks. Because it is a nonparametric method, we don't need strong assumptions about the data distribution. The algorithm is based on the assumption that similar inputs have similar outputs [56]. kNN is an instance based method, because it makes predictions for new subjects by looking at the similarities of it to the stored training subjects. In this method, the distances between test subject and all train-

ing subjects are calculated. Different types of distance measures such as Euclidean, Mahalanobis or Hamming distance could be used to evaluate similarity [57]. The test subject is classified according to majority votes of the closest training instances. The parameter k stands for the number of the closest training instances that vote for the classification of the new subject, and it should be an odd number to avoid ties. There are many advantages of this method. It can learn complex models with a simple logic and robust to noisy training data. Moreover, there is no training phase in this method, nothing is calculated until the prediction phase. But, the method has some disadvantages. It is computationally expensive and requires large amount of memory to store all the instances. Additionally, in high dimensional cases, efficiency of the method deteriorates dramatically.

Choosing the right distance metric is essential in correct implementation of the method. When correlations exist between input features using Mahalanobis distance over Euclidian distance is essential. Mahalanobis distance allows capturing these correlation and gives better results [58] as shown in Fig. 2.3. In a space that is spanned by two features x_1 and x_2, if Euclidian distance is used to determine the closest training subjects to the test subject shown with **x**, the test subject is misclassified as '$\circ$' for number of neighbors k=3. If Mahalanobis distance is used, it is correctly classified as 'x' by means of capturing correlation between features x_1 and x_2.

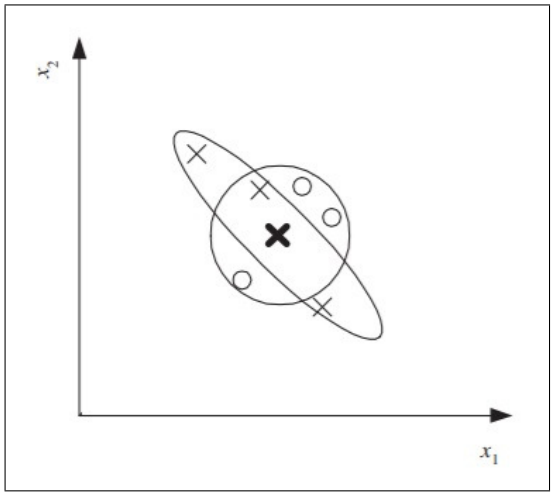

Figure 2.3 kNN classification example for different distance metrics [58].

2.3.3 Support Vector Machines

2.3.3.1 Linear Kernel Support Vector Machines

Support vector machines (SVMs) are useful methods that provide us to construct large margin classifiers. In these methods, finding weight parameters is a convex optimization problem. These problems could be solved analytically. Numerical methods such as iterative gradient descent algorithm are no longer needed in this method, hence we do not worry about learning rates, initialization procedures and convergence [59]. A hyperplane that separates the subjects to different classes in feature space is determined by the help of support vectors. Support vectors are some of the subjects that are in the vicinity of the hyperplane at least by a certain margin. In Figure 2.4, support vectors are the circled instances. Class that fall into left of the hyperplane has two support vectors whereas the other class has one support vector [60].

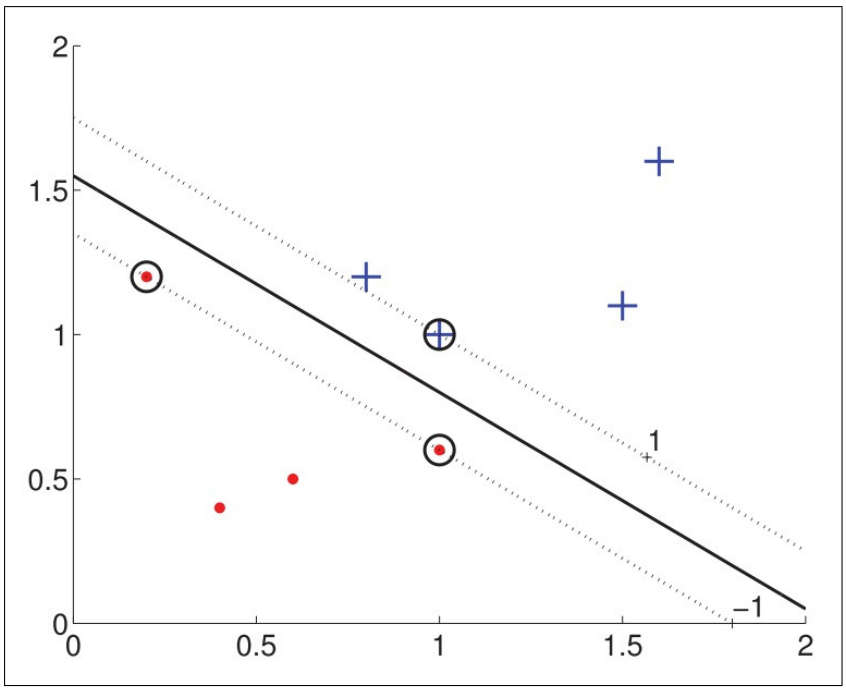

Figure 2.4 Separating hyperplane determined by support vectors of each class [60].

Instances other than support vectors have no contribution in determination of separating hyperplane hence they could be removed. Introducing maximum margins provides best hyperplane determination, and a more robust classification against noise.

In the case of that no linear hyperplane exist to separate the subjects perfectly, a hyperplane that results in the least error is determined by accounting for the deviations from the margin. We know that if $w_T x + w_0 > 0$, subject x belongs to class 1, if $w_T x + w_0 < 0$, subject x belongs to class 2, and if $w_T x + w_0 = 0$, subject x is on the hyperplane. If we have two arbitrary subjects x_1 and x_2 on the hyperplane we have,

$$\begin{aligned} w_T x_1 + w_0 &= w_T x_2 + w_0 \\ (w_T x_1 - w_T x_2) &= 0 \\ w_T (x_1 - x_2) &= 0, \end{aligned} \tag{2.4}$$

which would indicate that w vector is perpendicular to any vector that lies on the hyperplane. In SVM, margin is introduced by,

$$\begin{aligned} w_T x + w_0 &> 1 \quad \text{for class C1} \\ w_T x + w_0 &< -1 \quad \text{for class C2} \end{aligned} \tag{2.5}$$

and these two equations could be reduced to a single equation by adding the variable c^i that equals to 1 for subjects belonging to class C1 and equals to -1 otherwise. The new equation then becomes,

$$c^i (w_T x + w_0) > 1. \tag{2.6}$$

The equation for support vectors would then be,

$$\begin{aligned} c^i (w_T x_s + w_0) &= 1 \\ c^i (w_T x_s + w_0) - 1 &= 0 \end{aligned} \tag{2.7}$$

In Figure 2.5 we can see that the margin we want to maximize equals to the projection of difference vector of two support vectors from distinct classes to the unit vector that is perpendicular to the decision boundary. We see that w is perpendicular to separating hyperplane hence w/|w| is a unit vector that is perpendicular to the hyperplane.

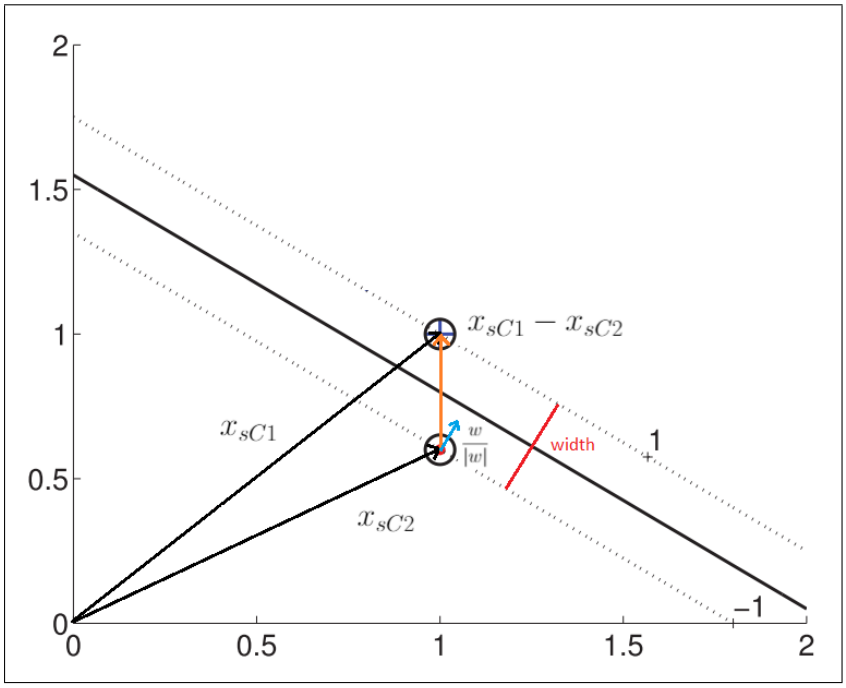

Figure 2.5 The margin and support vectors of SVM.

The margin of SVM could be defined by the following constraint equation,

$$\begin{aligned} &= (x_{sC1} - x_{sC2}) \cdot (w/|w|) \\ &= \frac{(x_{sC1} \cdot w) - (x_{sC2} \cdot w)}{|w|} \\ &= \frac{(1 - w_0) - (-1 + w_0)}{|w|} \\ &= 2/|w|, \end{aligned} \tag{2.8}$$

where $x_sC1 \cdot w$ equals to $w^T x_{sC1} = 1 - w_0$ and $x_C2 \cdot w$ equals to $w^T x_{sC2} = -1 + w_0$.

Maximizing 2/|w| is same as minimizing $|w|^2/2$. Because minimizing $|w|^2/2$ is mathematically more convenient than maximizing 2/|w|, we define margin maximiza-

tion problem as min $|w|^2/2$ subject to $c^i(w^T x + w_0) \geq 0$ for all x. By utilizing Lagrange multipliers we could solve the constrained minimization problem and find the support vectors. Then, we could find w parameters that are calculated as weighted sums of these support vectors.

2.3.3.2 Nonlinear Kernel Support Vector Machines

If the subjects we want to classify is not separable by a linear decision boundary, input feature space could be transformed by kernel functions. Linear separation could then be applied in transformed space. Polynomial function kernel, radial basis function kernel, and hyperbolic tangent function kernel are some examples of the transformation functions generally used. Kernel functions output the similarity degrees between the subjects. We use similarity values of an instance to other subjects as new features, instead of using initial input attributes [61]. Figure 2.6 shows an example transformation by using a radial basis kernel function [62].

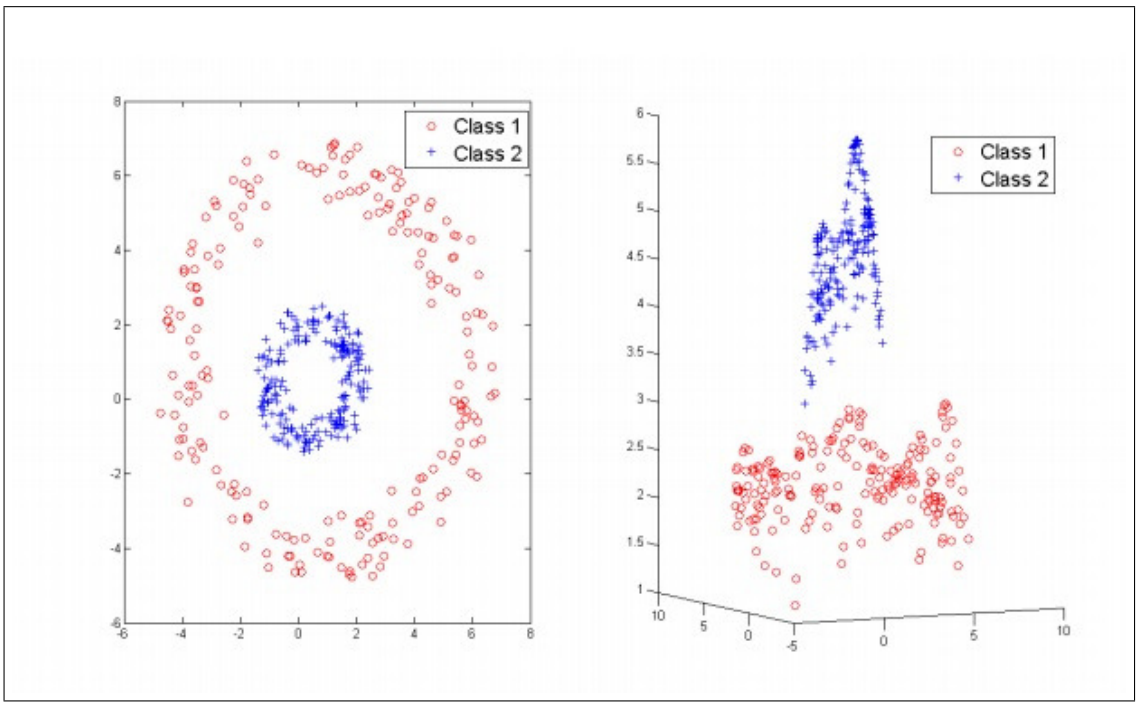

Figure 2.6 Input space transformation by applying a radial basis function kernel [62].

2.3.4 Decision Trees

Decision trees are non-parametric methods mainly used in classification tasks. A decision tree consists of decision nodes and terminal leaves that are indeed outputs. Terminal leaves are class information in the case of classification and numerical values in regression case. In decision nodes, discrimination is made according to feature values. In univariate decision trees, one feature is used at a decision node, whereas all features are used at each split in multivariate case. At each node, splits are made to decrease impurity. If all the subjects fall into one class after the split, it is said that split is pure and no further split is needed. There are different impurity measures such as entropy function, Gini index and misclassification error [63]. Figure 2.7 shows an example decision tree scheme (right) with the resultant decision boundaries that are orthogonal to each other (left) [64].

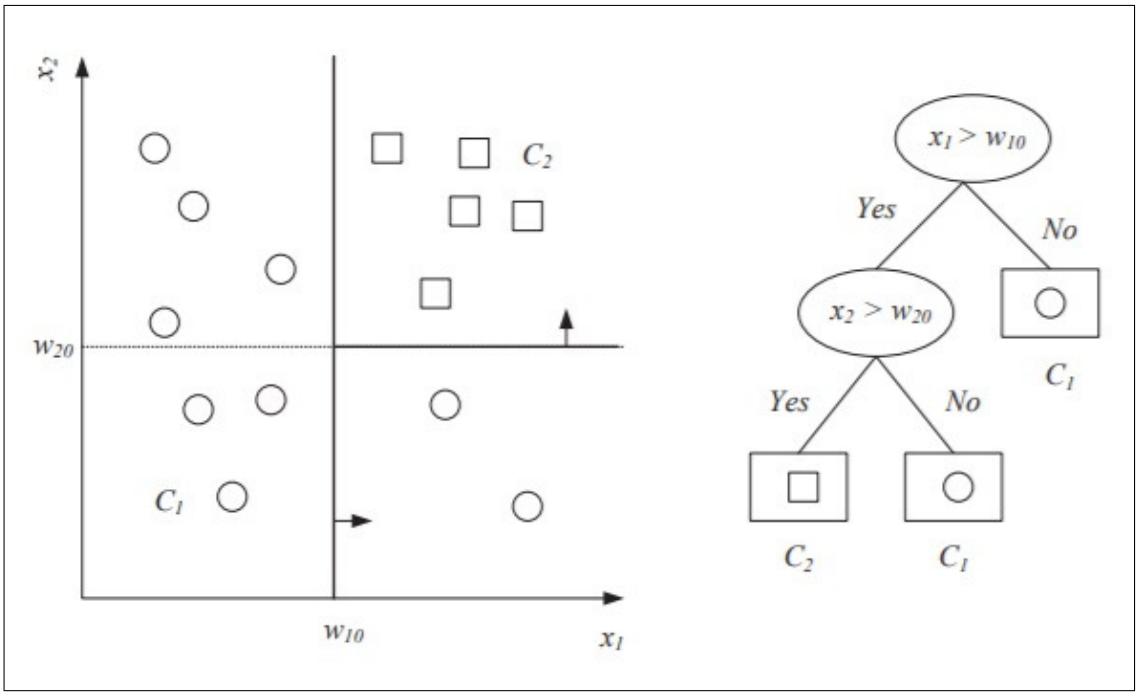

Figure 2.7 Decision boundaries and decision tree scheme [64].

In univariate case, features that are closer to the root are more important. Therefore, we can infer that decision trees can extract features inherently. Some features may not be used if they are not useful in decreasing impurity. One important issue to note is that features having many distinct values are preferred at decision

nodes, because more branches will be constructed and impurity will be lowered dramatically even if that feature doesn't have predictive power. This could lead to poor classification performance especially when there is noise. Prepruning and postpruning techniques are used to solve this problem. Prepruning is when the splitting stops early before impurity minimization. When a determined percentage of the training data cannot reach to a node, no further splitting is done in that node. Therefore, the constructed tree model generalizes better. On the other hand, postpruning is applied after the tree is fully constructed. Some portion of the training set (pruning set) is not used in splitting process to be used in postpruning process. Nodes decreasing the predictive power of the model are determined by the pruning set. Prepruning method is faster, but postpruning method usually provides better models. Univariate decision trees are helpful in interpreting the structure of the data. Rules could be found according to the splitting structure. In addition, when the splitting structure is found, subjects are no longer needed to be stored hence leading to less memory usage. It is also possible to use decision trees as feature extractors. The most informative features found in decision trees could be used in other models.

2.3.5 Ensemble Methods

Most widely used algorithms in machine learning are reviewed in previous sections. However, none of these algorithms may give the best accuracy by themselves. Using these algorithms together in various manners may lead to a higher accuracy. Ensemble models are constructed by combining different learning algorithms to increase the model performance. Models that constitute an ensemble model are called 'base learners'. Base learners should not have very high accuracies, in other words they should be weak. However, accuracy of each base learner should be greater than 0.5 to contribute to the overall accuracy [65]. It is advisable to include diverse base learners in the sense that having different algorithms, various hyperparameters, distinct feature sets or different training sets. Different algorithms have different assumptions about the distribution of data, hence including both parametric and nonparametric methods may be a reasonable option [66]. Using models that are constructed with the same al-

gorithm but having different hyperparameters may also increase overall accuracy. An example may be using different kernel functions in SVM models. In addition, when we have multiple sensors, we would have different feature representations. Additionally, training different models with different feature sets may be very beneficial. Both the input dimension, hence complexity per model, decreases by allocating feature sets and also diverse models could learn the training data from different perspectives. In addition to choosing base learners with different properties, how to combine these base learners is another issue. Two main ways of integration are multiexpert combination and multistage combination [67]. In multiexpert combination, base models work in parallel, whereas, in multistage combination they work in serial. Voting and stacking are two commonly used parallel combination schemes and cascading is an example of serial combination scheme. Voting is the most basic combination way. In this approach, final result is determined by a linear combination of outcomes of base learners. When all the weighs equal to 1, we have simple voting scheme. In stacked generalization [68] technique, weights of base models are learned by a combiner model. The combiner model should be trained on a different data set other than the training set that is used to construct base learners. Figure 2.8 shows an example multiexpert combination scheme (top) and a multistage combination (bottom) [69].

In cascading method, base learners are combined in a sequential way with ascending model complexity [70]. If learner in preceding stage is confident about its prediction the final result is its outcome. When the preceding classifier is not confident in some of its predictions, these unconfidently classified instances are forwarded to the next more complex model. In this way, the method tries to confidently classify all the instances with minimum complexity. Some other widely used ensemble methods are bootstrap aggregating (bagging) and boosting. In bagging method, different training subsets that have some common subjects are randomly allocated to train different base models. However in boosting, weak learners try to a better prediction on incorrectly classified instances by previous models [71]. Boosting method is effective, but it requires a large amount of training data that gets divided into distinct subsets. Adaptive boosting (AdaBoost) algorithm [72] that uses same training set repeatedly doesn't suffer from this problem. In this approach, selecting subjects for next classifier is not

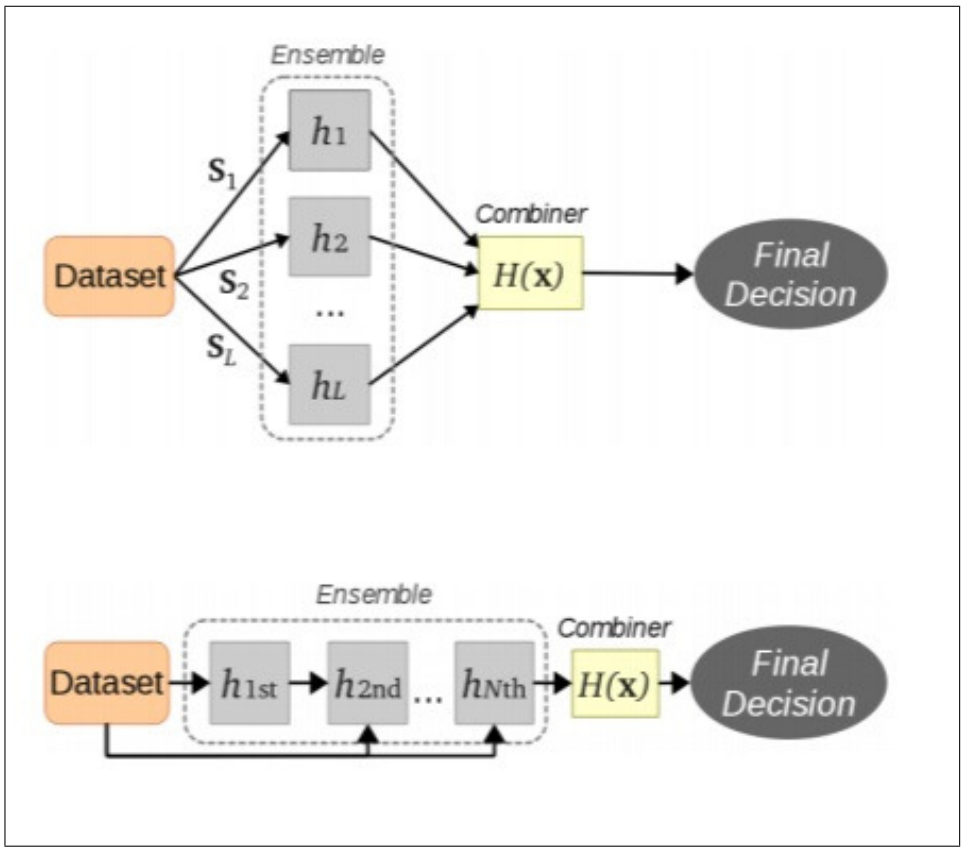

Figure 2.8 Two main combination schemes of ensemble learners [69]

completely random as in bagging, and the probability changes according to the error [73]. Ensemble classifiers are powerful models that could overcome complex tasks by dividing them into simpler ones. However, these models are not very easy to interpret.

2.3.6 Dimensionality Reduction

In machine learning, we seek the simplest model that could explain the structure. Simple models are more robust to noise and outliers, especially when the dataset is small. Higher dimensional data increases the model complexity. Therefore, decreasing the dimension of the data without decreasing the model performance is a useful approach to construct simpler models. Dimensionality reduction techniques could be used for this purpose. These techniques could be divided into two as feature selection methods and feature extraction methods. In feature selection methods, subset of initial feature set is selected that contains most informative features. In feature extraction

methods, new set of features are created using original feature set. In feature selection, wrapper methods and filter based methods are used. In wrapper methods, model's performance is evaluated for each feature subset. The feature subset leading to the best model performance is selected. There are many methods to create the feature subsets. Forward selection is one of the simpler algorithms of wrapper methods. In this method, feature subset is initially empty. Each feature's contribution to performance is calculated separately. The most informative feature is selected and kept in the subset. The procedure is repeated for other features and the feature subset grows until the model performance stops improving. In backward feature selection method, we start with all of the original features, and discard less informative features one by one while evaluating the model performance. Feature subset is shrunk until the model performance is not increasing anymore. We could improve these techniques by allowing multiple feature addition or removal at a time. This approach may produce a more optimal feature subset. Because, some features might not improve model performance individually, but together they might contribute dramatically. By changing number of added and removed features at each step, we could have more flexible methods. Examples of these methods are sequential floating forward selection (SFFS) and sequential floating backward selection [74]. When too many uninformative features are expected, performing forward selection is more advantageous to decrease computational expense and time. Recursive feature elimination is another wrapper based feature selection technique. In this technique, model is initially constructed using all the features and classifier scores the features according to their contribution to classification. The least informative feature is removed from the set and the same procedure is repeated until desired number of features are selected [75]. Feature subsets selected by wrapper methods extremely depend on the model used in calculation of feature contributions. In filter methods, feature selection procedure is applied without using any classifiers [76]. Filter methods are faster and computationally less expensive than the wrapper methods, but less effective in finding most suitable features for a specific model. Conducting statistical analysis for each individual feature is one of the filter methods. It is fast in finding the most discriminative individual features, but might miss the features that increase the model performance when they are used with other features. Correlation analysis is also another filter method that removes redundant features showing high

correlation with other features. Highly correlated features carry the same information, therefore, using only one of them is enough for classification purposes.

2.3.7 Evaluating Model Performance and Hyperparameter Tuning

Evaluating performance is another issue in machine learning. Accuracy is the most commonly used metric to assess performance. It is the proportion of correctly classified subjects to total number of subjects. However, we can't evaluate the model performance by looking at the accuracy value calculated for the training set. Parameters can be learned easily by a complex model and can be fit to the training data perfectly, but the model may fail to generalize explaining new data outside the training set. This problem is called 'overfitting' to the training set. On the contrary, if a model is too simple, it can't even learn the parameters that fit the training data, hence leading to 'underfitting' problem. We should check the accuracy of the model on separate datasets other than the training set. These datasets are validation set and test set. Validation set is used for hyperparameter tuning and test set is used for reporting final model performance. In any machine learning algorithm, some of the parameters are not learned from the data but determined by the practitioner. These parameters are called 'hyperparameters'. Most of these parameters affect the model complexity and should be chosen by maximizing generalizability power of the classifier. Some examples of hyperparameters are regularization parameter in logistic regression and svm, k in nearest neighbors, learning rate in any algorithm utilizing iterative optimization methods and gamma in gaussian kernel svm. Choosing these hyperparameters appropriately is essential part of the machine learning model design. Finding best parameters is nothing but a trial and error method. Model performance is evaluated for each trial on validation set and the parameter leading to the highest performance is selected. Usually, there are more than one hyperparameter to tune. For example in gaussian kernel svm, we need to tune both regularization parameter and gamma parameter which adjust the influences of training samples in decision boundary determination. We can try different combinations of these parameters and find the best combination. Grid search and randomized search are most popular hyperparameter searching techniques [77]. In

grid search, all combinations of desired values for different hyperparameters are tried, whereas in randomized search, randomly chosen values for each parameter are tried.

2.4 Longitudinal Analysis

Longitudinal analysis enables to study changing patterns in a group of subjects over time. There are two main problems about longitudinal data that complicates the analysis procedure. First the data is inherently correlated, because repeated measurements of the same subjects are analyzed. Second, variances are not homogenous. Many statistical methods assume that variance is the same along the regression line. If these issues are not taken into consideration, we make false conclusions. For example, if we do the analysis as if no correlation exist, we found too large p values for null hypothesis testing. If we make the analysis by ignoring heterogeneity in variances, we overestimate the goodness of the fit to the regression line.

2.4.1 Linear Mixed Effects Models

Linear Mixed Effects (LME) Models are powerful statistical methods used both in cross sectional and longitudinal analyses [78]. Correlations between repeated measurements and heterogeneity of variances are main problems in longitudinal data. Introducing random effects solves these problems [78]. In mixed effects model, we have both fixed effects and random effects. Definitions of fixed effects and random effects vary according to the purpose of the study. In our case, fixed effects were covariates and effects that we were interested about. Random effects were the effects we were not concerned about. Random effect in our case was different participants. If we include subjects as a variable in the model, we are able to capture the variances derived from the randomness of the individuals. Therefore, we are able to find variances that are accounted by fixed effects such as time. Fixed effects parameters are found by maximum likelihood estimation, whereas, random effects parameters are estimated using shrinkage [79].

A LME model can be shown as [80];

$$\begin{aligned} y_i &= X_i\beta + Z_i b_i + e_i \\ b_i &\sim N(0, D), e_i \sim N(0, Ri), \end{aligned} \tag{2.9}$$

where y_i is the response of subject i, β vector contains parameters of fixed effects, b_i vector contains parameters of random effects, X_i is a design matrix containing covariates of subject i, Z_i is a design matrix, e_i contains within subject random errors, R_i is a variance-covariance matrix of within subject measurements, and finally D_i is variance-covariance matrix of the random effects. In order to utilize LME models, our response variable should have normal distribution. Therefore, we have to perform normality check to our response variables. If some of them don't have normal distribution, we can apply data transformation and test for the normality again.

2.5 Literature Review

Some researchers tried to classify PD patients according to their cognitive status by the help of machine learning algorithms. In classification task, one challenge that many of the researchers encountered is choosing the most useful features from a vast number of available features. A previous study used only functional MRI (fMRI) data of the HC, cognitively normal PD (PD-CN) and PD-MCI [81]. Twenty-one edges (connectivity between each nodes) most frequently chosen across feature selection algorithm of randomized logistic regression (RLR) and leave one out cross validation (LOOCV) were selected as features. An accuracy of 80% was achieved in classifying PD-MCI and PD-CN. In another study, the low level ROI features (gray matter volume, cortical thickness, etc) and high level correlative features (connectivity between ROIs) were combined to create multilevel ROI features [82]. Multi kernel SVM algorithm was used to classify PD patients from HCs. Dimension reduction was performed by using both filter and wrapper based feature selection methods. 85.78% specificity and 87.79% sensitivity were attained as a result of using multilevel ROI features. Another study

employed features chosen from University of Pennsylvania smell identification test (UPSIT) scores, REM sleep behavior disorder screening questionnaire scores (RBDSQ), biomarkers from cerebrospinal fluid (CSF), and striatal binding ratios (SBR) from single photon emission computed tomography (SPECT) imaging [83]. After choosing most discriminatory features, SVM classifier successfully discriminated PD from HC with a specificity of 95.01% and sensitivity of 97.03%.

3. MATERIALS and METHODS

3.1 Data Acquisition

In this study, we examined MRI data of 27 PD-CN patients, 33 PD-MCI patients and 17 HCs collected at Hulusi Behcet Lifesciences Research Center of Istanbul University. Istanbul University Clinical Research Ethics Committee approved our study protocol and written informed consents were obtained from all the subjects before participating in this research. Neuropsychological tests (NPTs), genetic analysis for COMT and MAPT genotypes and MR examinations were performed at baseline for all the subjects. Longitudinal MRI data were collected only for the patients diagnosed as PD-MCI at first visit and for the patients converted from PD-CN to PD-MCI at second visit. For these patients, all NPTs and MRI data acquisition were repeated after 1.5 years of their first inspection. For the patients diagnosed as PD-CN at first visit and haven't converted to PD-MCI according to ACER test score at second visit, no further data were collected.

3.1.1 MR Data Acquisition

MR examinations of all the participants were carried out at a 3 Tesla clinical MR system (Philips Medical Systems, Best, Netherlands) using a 32 channel head coil. MR data acquisition included resting state fMRI (rs-fMRI), 1H-MRSI, T1-weighted, T2-weighted, ASL, DTI, and T2-weighted fluid attenuated inversion recovery (FLAIR) MRI, and total scan time was approximately 45 minutes. T1-weighted MRI were obtained by 3D Fast Field Echo (FFE) imaging (TR=9.1ms, TE=4.2ms, slice thickness 1mm, intersection space 0.3mm). T2-weighted FLAIR (TR=11000, TE=125ms, TI=2800ms, slice thickness 4mm, intersection space 1mm) were obtained with corresponding parameters. DTI images were taken with 32 distinct gradient direction (TR=5000ms, TE=105ms, matrix size 256x256x40, FOV=440x440x84mm, slice thick-

ness 2.1mm, b=1000 s/mm2). In addition to these images, 1H-MRSI data were taken with Point RESolved Spectroscopy (PRESS) sequence using short echo time (TR=1000ms, TE=35ms, 1000Hz, 1024 points, FOV=160x160x80mm, 10x10x10 voxel size, scan time 10min). We obtained ASL MRI data with echo planar imaging and signal targeting with alternating radio frequency pulses (EPISTAR) [84] technique using eight different inversion time in each slices for a total of six distinct MR slices (TR=250ms, TE=15.99ms, turning angle 40°). In Figure 3.1 some of the different MRI modalities and generated maps that we used in our study are presented.

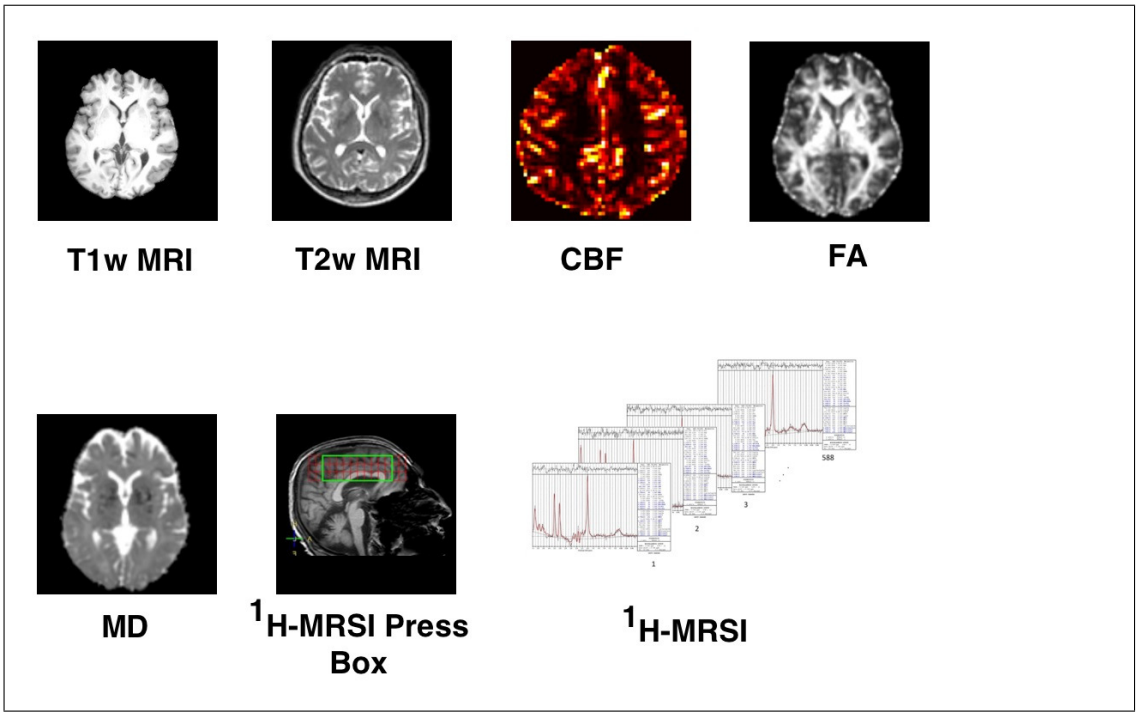

Figure 3.1 Magnetic resonance imaging modalities used in our study.

3.1.2 Neuropsychological and Genetic Tests

The cognitive status of the subjects were determined by expert neurologists by conducting neuropsychological tests such as Unified Parkinson's Disease Rating Scale (UPDRS), Addenbrooke's Cognitive Examination Revised (ACER), Mini Mental State Examination (MMSE), Wisconsin Card Sorting Test (WCST), Stroop test, phonemic fluency test, categorical fluency test, Benton Judgement of Line Orientation Test (BJLOT), and Geriatric Depression Scale (GDS). Additionally, the performance

of the subjects in drawing watch, pentagon and cube were assessed. Neurologists finally performed a physical examination as the last step of the patient visit. Additionally, blood samples were taken from all the subject to perform genetic analysis to detect COMT and MAPT genotypes.

3.2 MRI Data Preprocessing

After MRI data acquisition, data were preprocessed and analyzed. MR images taken from all the subjects were registered to their T2-weighted images, and head movement and eddy current corrections were made. Parameters generated from the registration of T2-weighted images to Montreal Neurological Institute (MNI) brain atlas were used to register ASL, DTI and MRSI images to MNI brain atlas. Carrying these images to MNI brain atlas is essential to be able to perform voxel based data analysis, because it aligns the brain regions of all subjects to the same locations in space. CBF, aBV and BAT maps were created from ASL data using an in-house software written in MATLAB (The Mathworks Inc., Natick, MA) [85]. The program fits the data to general kinetic model using a nonlinear algorithm [86]. The program estimates the main magnetization, M0, values for each pixel and maps are calculated by the help of estimated M0 values. Then the maps were masked by MNI structural and Harvard-Oxford cortical subcortical structural atlases [87]. 119 different brain regions were obtained after masking and average of pixel values at each region were calculated. DTI data were preprocessed in FMRIB Software Library (FSL, Analysis Group, FMRIB, Oxford, UK) to remove eddy current artefacts and to extract only brain tissue without eyes, skull and spine. Afterwards FA and MD maps were created and registered to MNI 152 brain atlas [88]. These maps were multiplied with 48 white matter (WM) masks generated from John Hopkins University (JHU)-81 WM atlas and 9 masks generated from MNI brain atlas. Pixel values of the remaining regions after masking were averaged. LCModel [89] was used to quantify 1H-MRSI metabolite concentrations. Amount of chemical shift of each metabolite was calculated and corrected. Chemical shift corrected metabolite concentration maps were created and were overlaid onto T2-weighted images. FSL was used to register these overlaid maps to MNI152

brain atlas. MNI structural and Harvard-Oxford cortical and subcortical structural atlases were used to select 119 different brain regions. Average metabolite concentration values were calculated in these regions for different metabolites. Metabolite values of NAA + N-Acetyl-aspartyl-glutamate (NAA+NAAG), Glu+Gln, Ins and Cho were divided by Cr values [90]. All these values coming from FA, MD, CBF, aBV, BAT, 1H-MRSI were used as features to train ML classifiers.

3.3 Classification with Machine Learning Methods

We had a feature set consisting of multimodal MRI data. Initial feature set contained 4 different metabolite ratios in 119 brain regions, FA and MD values in 57 brain regions and information of CBF, BAT and aBV values in 119 brain regions. Total number of features was 947, which was a big number for our case because we had relatively small number of subjects. We first tried to decrease the number of dimensions by selecting the most informative features. Second, we constructed the classifiers using scikit-learn v0.19.1 library in Python and Classification Learner Toolbox in MATLAB R2017b. We used different types of classifiers such as decision trees, SVM, logistic regression, nearest neighbor and ensemble classifiers. Many different techniques were tried and best feature selection and classification results were reported. Different feature selection followed by classifier construction pipelines were performed for classification of HC versus PD-CN, HC versus PD-MCI, PD-CN versus PD-MCI, and HC versus PD-CN versus PD-MCI.

3.3.1 Feature Selection Procedure

When there exist vast amount of features compared to number of subjects, classifiers are not able to detect useful features easily and give useless models. This phenomenon is called curse of dimensionality [91]. When an extra feature is added in model construction, we have to add extra subjects exponentially to ensure that feature space is dense enough to yield reasonable classification. When we used all

the features in the set, all the classifiers we tried gave poor results because of this phenomenon. Therefore, we tried many different feature selection methods, including SFFS with SVM, RF-RFE, filter based methods such as statistical methods. In order to find most informative features, we applied feature selection methods many times with different subject subsets. In this way, we could determine the features that are selected most of the time. We created different subsets by leaving one patient out, therefore we obtained number of subject times subsets. The subsets were created in leave one out manner, because we had small number of subjects, and it was reasonable to not to decrease subsets any further. Feature selection algorithms were tried in this manner and RF-RFE technique yielded a stable feature selection. If feature selection would be performed only in one subject set, selected features might be the ones fitting randomly to that particular subject set. This is a case of overfitting. The reason is that large amount of features are tried to be selected by using small number of subjects. Therefore, selected features might be best for the available dataset but might be uninformative for general population. If a feature subset is selected consistently at more than half of the trials from different subject subsets, the probability of overfitting is now much smaller. Therefore, we can say more confidently that we can select that particular feature subset. Feature selection procedure was applied to select feature subsets for four different classification tasks (HC vs PD-CN, HC vs PD-MCI, PD-CN vs PD-MCI, HC vs PD-CN vs PD-MCI). Within each classification task, the feature selection procedures were applied for six different feature types (FA, MD, CBF, aBV, BAT, 1H-MRSI). For each classification task-data type pairs (i.e. HC vs PD-CN - FA) feature subsets containing 2,3,4 and 5 features were found. Because only RF-RFE method yielded stable feature selection for some classification task-data type pairs, we chose this feature selection method. If a feature subset was not consistently selected at more than half of the trials for a pair, then we admitted that there weren't any informative features for that pair and didn't proceed to the classification step.

3.3.2 Classification Procedure

Classification procedure was performed both in Python and MATLAB environment, whereas feature selection procedure was performed only in Python in conjunction with the classification procedure.

3.3.2.1 Hyperparameter Tuning in Python Environment

As already mentioned before, hyperparameters can't be learned from the data and should be specified by the practitioner. Hyperparameter sets were formed for each classification method. For SVM and logistic regression, we chose candidate hyperparameter values in a way that the set included the default regularization parameter value C=1, and minimum and maximum values were very small and very big numbers respectively (10^{-24} and 10^{24}). Remaining 8 numbers were distributed evenly in logarithmic scale. We chose values with huge intervals, because we initially expected that selected C parameter value would converge to a value in the set. If the convergence would occur then parameter set would be formed again according to the converged value. However, the convergence didn't occur and the set was left as initial. The other hyperparameter that we formed a set was k parameter for kNN classifier. We chose k values mostly as odd numbers to break the tie. We only included 10 as an even number, because it was the default value for the kNN classifier in the MATLAB environment. kNN hyperparameter set contained values of 1, 3, 5, 7, 9 and 10.

After we created the hyperparameter sets we constructed the feature selection and classifier construction (main procedure) pipeline by using 'Pipeline' command. Pipeline function was essential in the sense that it allowed us to chain multiple procedures coherently. In the main procedure cross validation should be done to select the most suitable hyperparameter. We used 5 fold cross validation for hyperparameter selection. In general, feature selection should be done in the training set and then selected features are used in cross validation sets. In the pipeline these procedures were performed. First, a single subject was separated for final testing. Next, remaining

subjects were divided to 5 equal portions. Four portions were used for standardization, feature selection and training, and one portion was used for testing. Above step was repeated for each cross validation fold and for each hyperparameter. Mean accuracy values of all folds for each hyperparameter testing were calculated and hyperparameter yielding the highest cross validation accuracy was selected. Feature selection and training were performed with all subjects except one that was excluded at first step by using selected hyperparameter. Final classifier was tested with the single subject that was separated at the first step. All the procedure above was repeated for another single patient. Finally, accuracy, sensitivity and specificity values were calculated for all patients that were tested one by one.

In standardization procedure, all the features were scaled to have zero mean and unit standard deviation. It should be done in training set and then mean and standard deviation are found. These mean and standard deviation values should be used to transform testing data. We used Standardize function in Pipeline to perform standardization of all features as,

$$z = \frac{x - \mu_x}{\sigma_x} \tag{3.1}$$

In Equation 3.1 x is a feature, μ_x is mean of the feature and σ_x is the standard deviation of the feature.

Figure 3.2 shows the schematic of model construction and evaluation in Python environment.

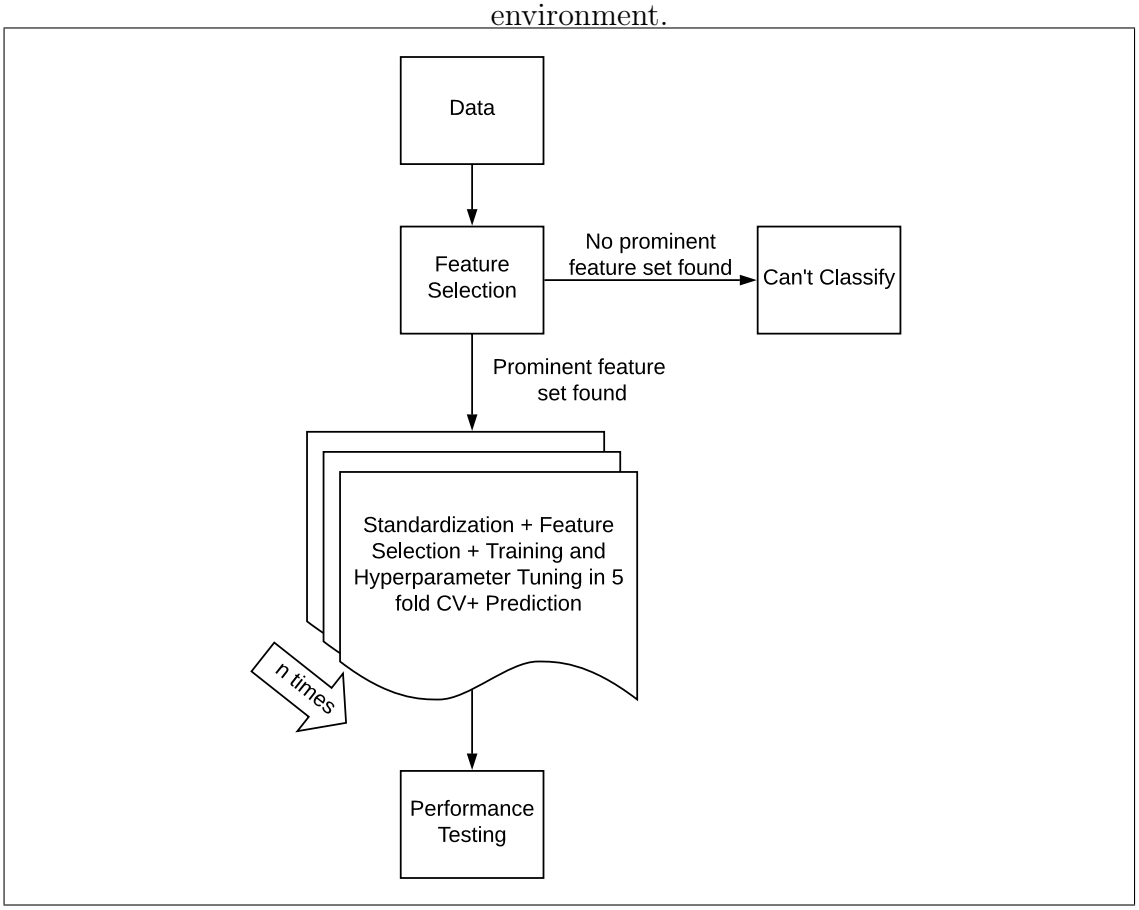

Figure 3.2 The schematic of model construction and evaluation in Python environment.

3.3.2.2 Classification in MATLAB with Default Parameters

After we performed hyperparameter tuning and evaluated the results in Python environment, we noticed that selected hyperparameters did not strongly converge to a single value. However, we observed some trends in selected hyperparameters. In MATLAB environment, classifiers were constructed with default hyperparameter settings by using most stably chosen features in Python. We used ClassificationLearner application in MATLAB to construct classifiers. The main advantage of this application is that it enables quick results by using only most popular hyperparameters. We can construct more stable classifiers by utilizing ClassificationLearner application. Two different kNN classifiers, one with Euclidian distance metric (kNN), and one with Mahalanobis distance metric (kNN*), were utilized along with other classifiers like SVM, bagged trees and logistic regression.

We used 10 fold cross validation, because it is the default value in the application. The smallest group in the classification was HCs with 17 subjects. We paid attention to include at least one subject from the smallest group in each fold. In this way accuracy, sensitivity and specificity values were more meaningful because at least one subject from HC could be tested in each fold. In 10 fold cross validation, data were split into 10 nearly equal parts. Proportion of subjects belonging to different groups was nearly preserved in each fold. One fold was separated for testing and remaining subjects were used for training. This procedure was repeated for each fold and thus each subject was predicted once. Accuracy, specificity and sensitivity values were calculated by using predictions and true response variables. MATLAB code was generated from the application to perform the above procedure 100 times utilizing a for loop. In this way, we obtained 100 different splitting for 10 fold cross validations. As a result, each subject was predicted 100 times in different splitting patterns. Finally, we obtained hundred accuracy, sensitivity and specificity values. Mean and standard deviations were calculated to get final accuracy, sensitivity and specificity values. Figure 3.3 shows the schematic of model construction and evaluation in MATLAB.

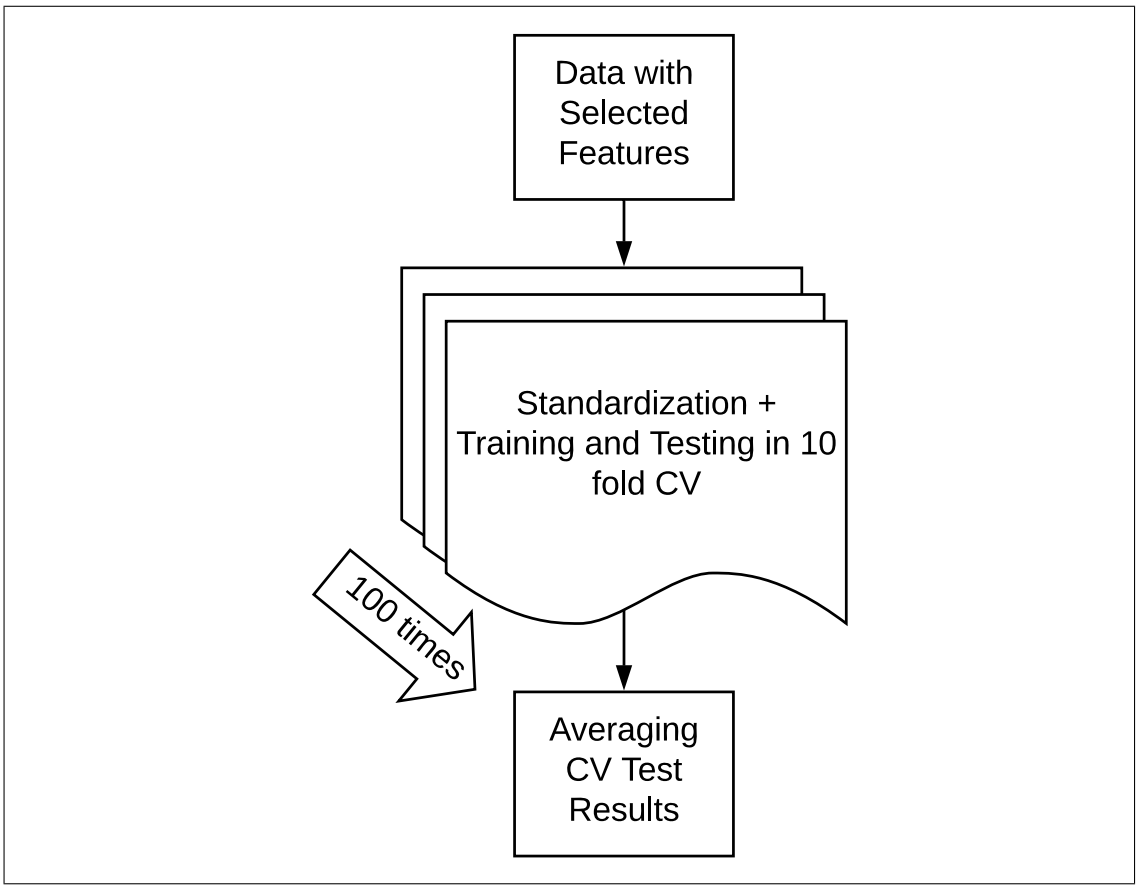

Figure 3.3 The schematic of model construction and evaluation in MATLAB environment.

3.4 Longitudinal Analysis

Multimodal MRI data acquisition, neuropsychological tests and physical examination were repeated for 24 PD patients. Only one patient was evolved to PD-MCI from PD-CN. That patient was excluded from the longitudinal analysis in order to examine the changing patterns over time of patients initially diagnosed as PD-MCI. Longitudinal analysis was conducted for multimodal MRI data using linear mixed effects model in MATLAB. Because linear mixed effects model requires data to be normally distributed, normality check was performed by using Shapiro-Wilk test. Random intercept fixed slope design was utilized and education, gender, age, visuospatial disorder status, and genetic information (MAPT and COMT) were added as covariates in the model. Subject variable containing patient identity was used as random effect and time variable that we want to analyze the effect of was used as fixed effect. Multiple

comparison corrections for P values were made using Holms method. Equation 3.2 depicts our longitudinal analysis model.

$$y \sim 1+VS+age+time+COMT+MAPT+gender+education+(1|subject) \quad (3.2)$$

In addition to multivariate MRI data analysis, neuropsychological test results were also analyzed. Normality check was performed for test results, and we noticed that some of the test results were not normally distributed. Therefore Wilcoxon signed rank test was utilized and multiple comparison correction was applied.

4. RESULTS

4.1 Machine Learning Results

4.1.1 Feature Selection Results

In feature selection procedure we found that some of the feature sets were selected more consistently than others (Table 4.1). Glu+Gln of caudate was selected both for three groups and HC vs PD-CN classifications. aBV of right planum temporale and right superior posterior temporal gyrus regions were selected both for all groups and PD-CN vs PD-MCI classifications. NAA+NAAG of left superior posterior temporal gyrus was selected for HC vs PD-CN classification. FA of body of corpus callosum, left cerebral peduncle and right tapetum were selected for HC vs PD-MCI classification. Lastly, GPC+PCh of left lateral superior occipital cortex was selected for all groups classification.

Table 4.1
Selected features for classification purposes.

Classification	Biomarker	Region
HC vs PD-CN vs PD-MCI	aBV	Right Planum Temporale
	aBV	Right Superior Temporal Gyrus, posterior division
	Gpc+Pch	Left Lateral Occipital Cortex, superior division
	Glu+Gln	Caudate
HC vs PD-CN	Glu+Gln	Caudate
	NAA+NAAG	Left Superior Temporal Gyrus, posterior division
HC vs PD-MCI	FA	Body of Corpus Callosum
	FA	Left Cerebral Penducle
	FA	Right Tapetum
PD-CN vs PD-MCI	aBV	Right Planum Temporale
	aBV	Right Superior Temporal Gyrus, posterior division

4.1.2 Classification Results Produced in Python Environment

We observed that some of the hyperparameters were selected more commonly. For example, in classification of PD-CN vs PD-MCI with kNN, hyperparameter N was tuned. 10 was selected as the best hyperparameter for this classification in 37.2% of the trials. In classification of the HC vs PD-MCI with logistic regression, regularization parameter C was tuned. 10^{-8} was selected as the best value for the hyperparameter C, but it was selected as the best only in 16% of the trials. Random forest yielded best classification performance for HC vs PD-CN (accuracy=74%) and PD-CN vs PD-MCI (accuracy=66%) classifications. Linear SVM gave the best result for HC vs PD-MCI classification (accuracy=71%).

Table 4.2
The classification results of HC vs PD-CN in Python.

HC vs PD-CN	Accuracy	Sensitivity	Specificity
kNN	0.55	0.66	0.33
Logistic Regression	0.64	0.70	0.53
Random Forest	0.74	0.85	0.53

Table 4.3
The classification results of HC vs PD-MCI in Python.

HC vs PD-MCI	Accuracy	Sensitivity	Specificity
kNN	0.60	0.69	0.44
Logistic Regression	0.71	0.78	0.56
Linear SVM	0.71	0.75	0.62

Table 4.4
The classification results of PD-CN vs PD-MCI in Python.

PD-CN vs PD-MCI	Accuracy	Sensitivity	Specificity
kNN	0.59	0.59	0.59
Random Forest	0.66	0.66	0.67

4.1.3 Classification Results Produced in MATLAB

As mentioned before, any classification was performed 100 times to obtain more reliable results, therefore we reported mean and standard deviation of these repetitive classification results. First we classified the three groups of HC, PD-CN and PD-MCI. The accuracy of three groups classifications were 0.91 ± 0.05 for HC, 0.65 ± 0.06 for PD-CN, and 0.64 ± 0.05 for PD-MCI groups. The best classifier leading to these accuracy values was constructed by Random Under Sampling Boosted (RUSBoosted) Trees using default values. These accuracy values were not high enough, therefore we performed pairwise two class classifications.

Table 4.5
The classification results of HC vs PD-CN in MATLAB.

HC vs PD-CN	Accuracy	Sensitivity	Specificity
Bagged Trees	0.86 ± 0.03	0.88 ± 0.04	0.82 ± 0.06
Logistic Regression	0.79 ± 0.01	0.85 ± 0	0.69 ± 0.03
Linear SVM	0.80 ± 0.01	0.85 ± 0	0.70 ± 0.03
kNN	0.78 ± 0.03	0.78 ± 0.03	0.79 ± 0.06
kNN*	0.78 ± 0.03	0.77 ± 0.04	0.79 ± 0.06

Table 4.6
The classification results of HC vs PD-MCI in MATLAB.

HC vs PD-MCI	Accuracy	Sensitivity	Specificity
Bagged Trees	0.77 ± 0.03	0.85 ± 0.03	0.62 ± 0.06
Logistic Regression	0.74 ± 0.02	0.85 ± 0.03	0.53 ± 0.03
Linear SVM	0.76 ± 0.02	0.89 ± 0.02	0.51 ± 0.02
kNN	0.79 ± 0.03	0.84 ± 0.03	0.69 ± 0.06
kNN*	0.77 ± 0.03	0.81 ± 0.04	0.70 ± 0.05

For the HC vs PD-CN and PD-CN vs PD-MCI classifications, bagged trees yielded the best accuracy (86% and 71%), sensitivity (88% and 73%) and specificity (82% and 69%) results (Tables 4.5 and 4.6). kNN method resulted in best performance for HC vs PD-MCI classification (accuracy=77%, sensitivity=85%, and specificity=62%) (Table 4.7). kNN and kNN* yielded similar performance in all cases.

Table 4.7
The classification results of PD-CN vs PD-MCI in MATLAB.

PD-CN vs PD-MCI	Accuracy	Sensitivity	Specificity
Bagged Trees	0.71 ± 0.03	0.73 ± 0.04	0.69 ± 0.04
Logistic Regression	0.51 ± 0.02	0.78 ± 0.02	0.18 ± 0.05
Linear SVM	0.50 ± 0.02	0.84 ± 0.04	0.09 ± 0.04
kNN	0.69 ± 0.01	0.69 ± 0.02	0.68 ± 0.03
kNN*	0.66 ± 0.02	0.67 ± 0.03	0.64 ± 0.03

4.2 Longitudinal Analysis Results

The variables showing statistically significant differences between the two time points are reported in Figures 4.1, 4.2, 4.3 and 4.4. Line orientation test result was analyzed using Wilcoxon signed rank test whereas other variables were analyzed with linear mixed effects model. Estimated beta parameters for second time points were $2.42\text{x}10^{-5}$ (P=$3.59\text{x}10^{-7}$) in temporal lobe, $3.85\text{x}10^{-5}$ (P=$4.69\text{x}10^{-5}$) in middle cerebellar peduncle, $2.66\text{x}10^{-5}$ (P=$1.18\text{x}10^{-4}$) in thalamus, and $2\text{x}10^{-5}$ (P=$1.39\text{x}10^{-4}$) in parietal lobe for MD. All these estimated beta parameters were positive, indicating an increase in MD values. It is also shown in Fig. 4.1 that MD values increased at the second time point. Estimated beta parameters for second time points were -0.03 (P=$6.49\text{x}10^{-7}$) for right inferior cerebellar peduncle and -0.02 (P=$1.60\text{x}10^{-5}$) for right medial lemniscus for FA. In this case estimated parameters were negative, indicating a decrease in FA values in time, which is also visible in Fig. 4.2. Lastly, estimated beta parameter was -0.04 (P=$5.46\text{x}10^{-5}$) for aBV at middle frontal gyrus at second time point, indicating a decrease of aBV in time (Fig. 4.3). According the results of the signrank test, line orientation test performance decreased over time for PD-MCI patients (P=$1.6\text{x}10^{-3}$) (Fig. 4.4).

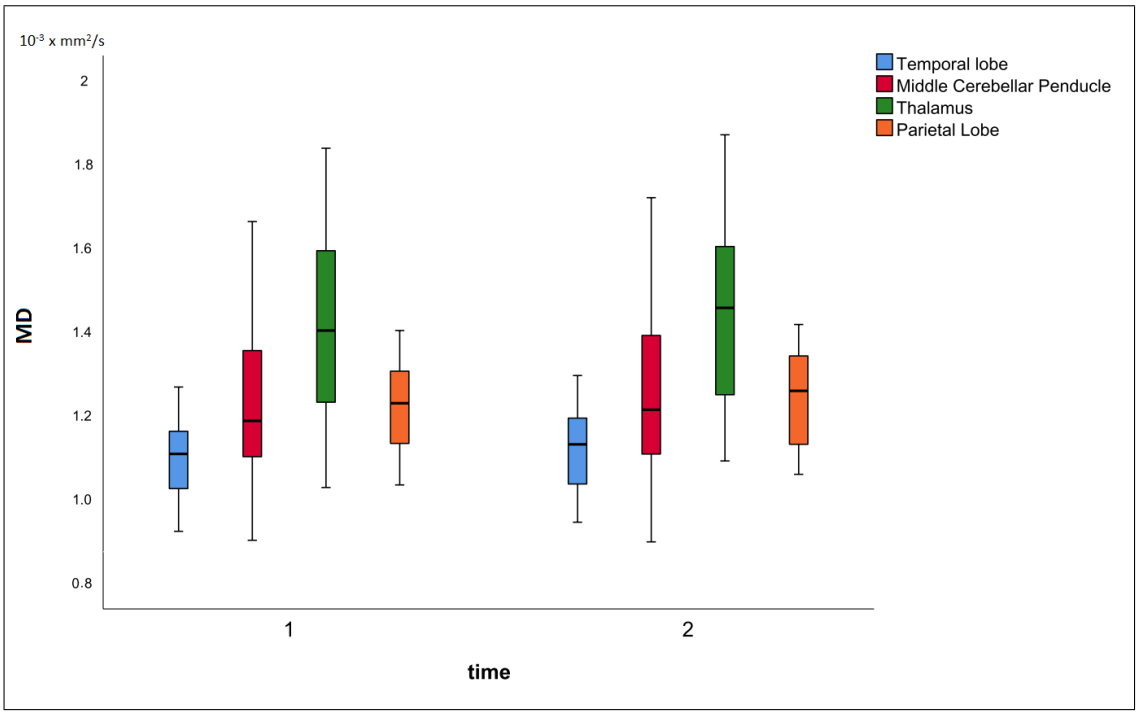

Figure 4.1 MD values increased over time for PD-MCI patients at temporal lobe (P=3.59x10^{-7}), at middle cerebellar penducle (P=4.69x10^{-5}), at thalamus (P=1.18x10^{-4}) and at parietal lobe (P=1.39x10^{-4}).

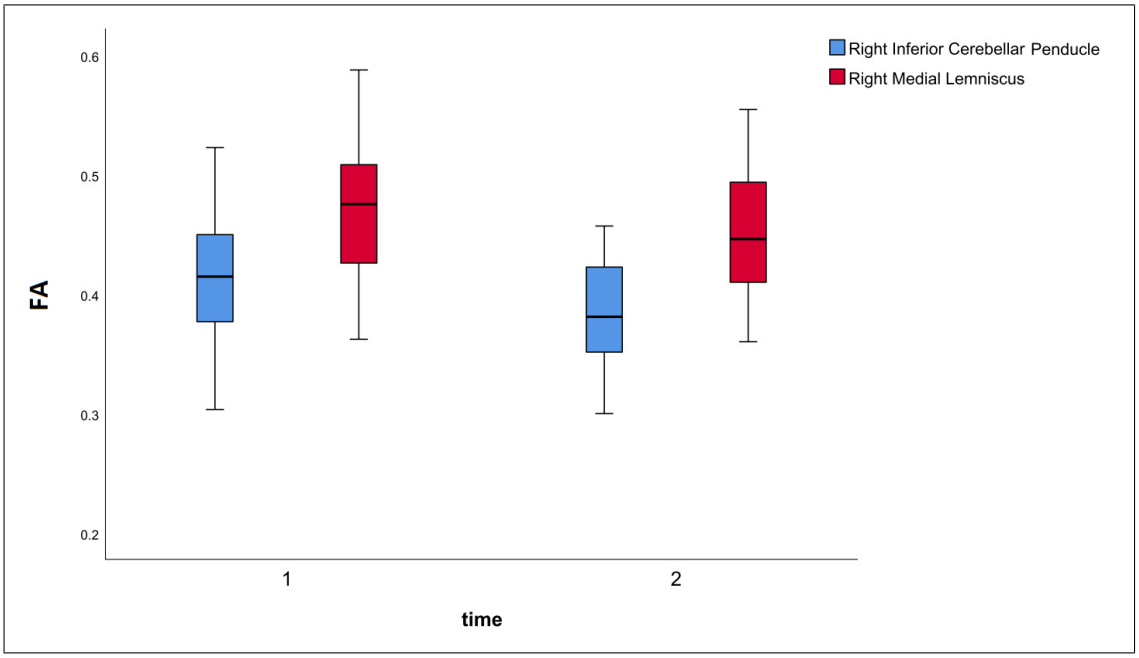

Figure 4.2 FA values decreased over time for PD-MCI patients at right inferior cerebellar peduncle (P=6.49x10^{-7}) and at right medial lemniscus (P=1.60x10^{-5}).

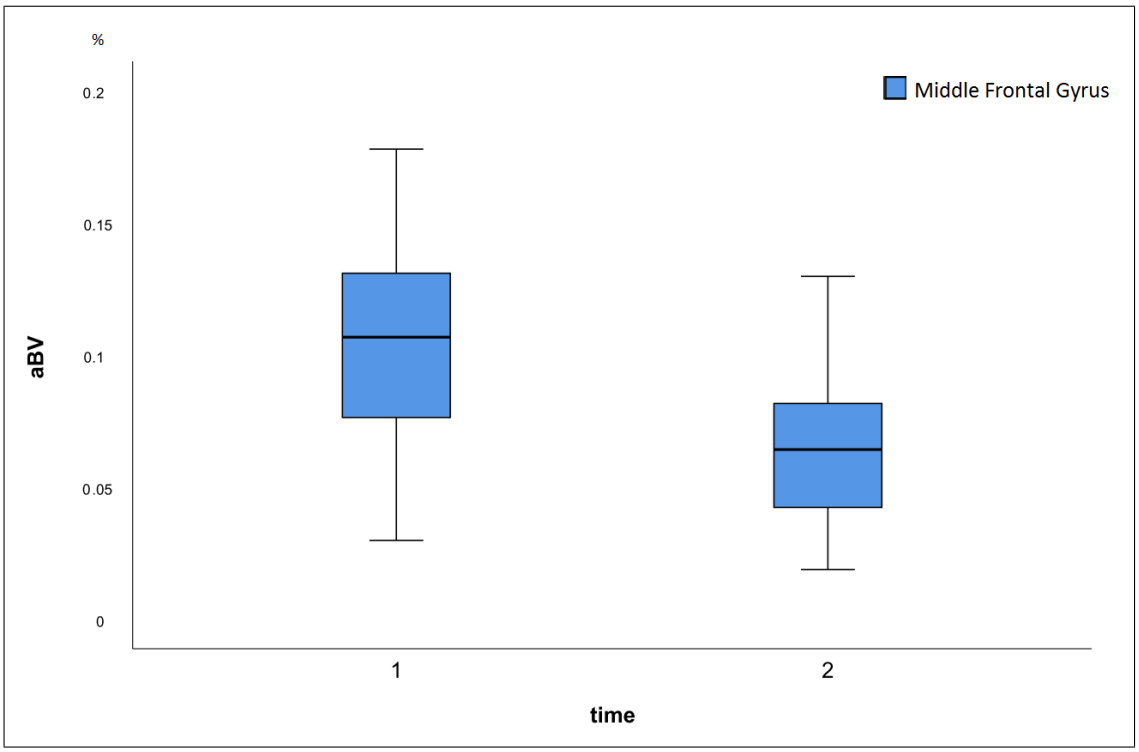

Figure 4.3 aBV values decreased over time for PD-MCI patients at middle frontal gyrus (P=5.46x10^{-5}).

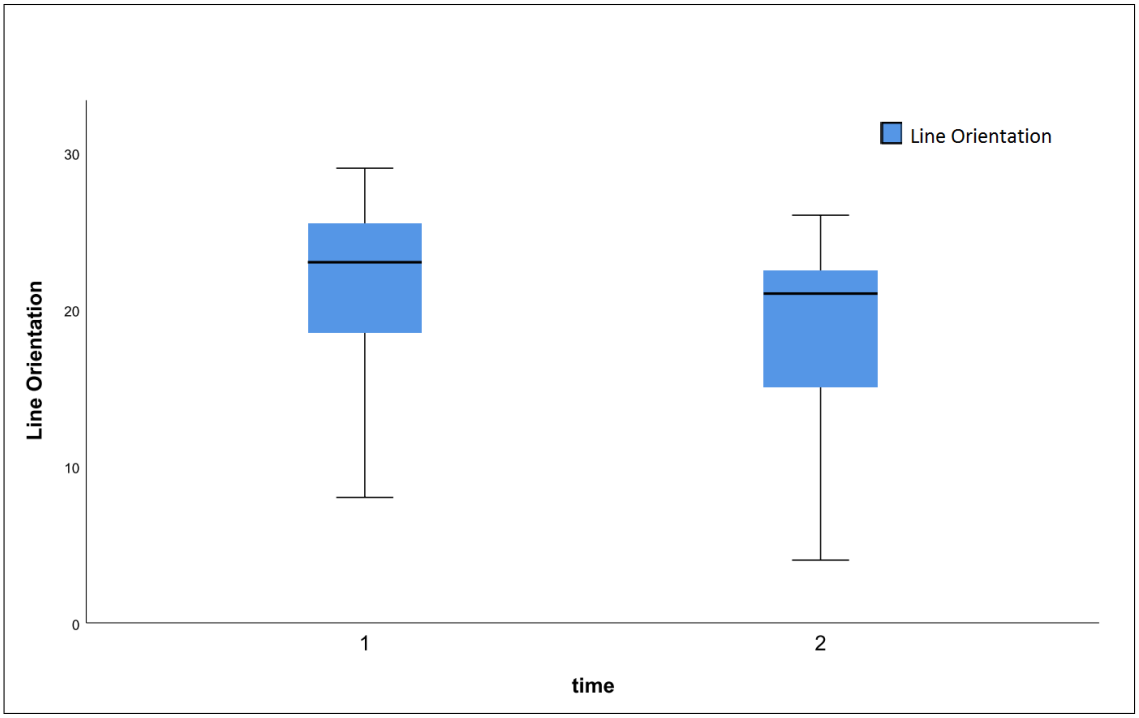

Figure 4.4 Line orientation test performance decreased over time for PD-MCI patients (P=1.6x10^{-3}).

5. DISCUSSION

In this study, first, we classified HC, PD-CN and PD-MCI with various machine learning techniques. Furthermore, we performed longitudinal analysis for some patients in PD-MCI group to see the effect of time on multimodal MRI data and neuropsychological test results. Our results indicated that PD-MCI, PD-CN and HC groups could be classified with reasonable accuracy. As line orientation test performance decreased over time, we observed a concurrent increase in MD and decrease in FA and aBV in some brain regions.

In classification part, we obtained some feature sets using the feature selection algorithms. Majority of these selected features were reasonable for our classification tasks. For example, in classification of HC vs PD-MCI, we selected FA values of body of corpus callosum, left cerebral peduncle and right tapetum. Body of corpus callosum allows communication of the two hemispheres. According to a study [92], PD patients had lower FA values and higher MD values than HC at the body of corpus callosum. They also concluded that as the disease progressed, these differences increased between PD and HC groups. In our data, we also observed this trend. Another study [93] suggested that compression of the cerebral peduncle by cerebral artery might result in PD. Cerebral peduncles refine motor movements, therefore changes in these tracks may give information in the classification. Tapetum may also give information about PD status of an individual. It enables hemispheric connection of prefrontal cortex. Prefrontal cortex is important in decision making and planning cognitive behavior [94]. Therefore, alterations in tapetum may affect prefrontal cortex connection and result in problems in cognitive behaviors which is common in PD-MCI. In classification of HC vs PD-CN, Glu+Gln of caudate and NAA+NAAG of left superior posterior temporal gyrus were selected as features. Caudate is well known for regulating motor movements [95]. Glu and Gln metabolites are neurotransmitters that take a role in signal transmission at synapses. However, high Glu may result in glutamate toxicity and possibly contribute to the development of PD [96]. In our data we saw that, Glu values were higher in

PD-CN than HC. Left superior posterior temporal gyrus is responsible for phonological processes [97]. An association with PD couldn't be found for this region at the literature. In classification of PD-CN vs PD-MCI, right planum temporale which is responsible for language functions [98] and right superior posterior temporal gyrus associated with insight based problem solving [99] were selected. However, association of these regions with PD couldn't also be found in the literature. In addition to caudate, right planum temporale and right superior posterior temporale gyrus; left lateral superior occipital cortex were selected. According to a study [100], decrease in cortical thickness of left lateral superior occipital cortex was observed in PD patients. An increase of Cho in this region might be due to cellular degeneration.

We examined the classification results and see a clear pattern of that tree based models and kNN models generally yielded better classification performance. Logistic regression and linear SVM have linear decision boundaries whereas with tree based and kNN methods nonlinear decision boundaries could be constructed. Groups may be classified better with models having nonlinear decision boundaries in feature space. Another reason might be utilization of random forest in feature selection. RF-RFE might have contributed to better performance of tree based methods.

In MATLAB environment, models that were constructed were more stable and resulted in better classification performance. This might be derived from the fact that we tried to perform feature selection and hyperparameter tuning jointly using large feature set and small number of subjects. Although we could select the same feature sets in most of the trials, selected feature sets varied with different subject subsets leading to unstable hyperparameter tuning. When we fixed the feature sets in MATLAB and used default hyperparameters we obtained better results. In three group classification, RUSBoosted tree method yielded the best accuracies. This might be expected, because this method is very suitable for data having class imbalance in multiclass classification.

We compared our results with the results of the studies in the literature. We examined three studies [81, 82, 83]. In all these studies, subject sets were larger than

our study. For example in the first study [81] 38 HC, 60 PD-CN and 35 PD-MCI were included. fMRI was used, and 21 brain connectomics out of 30,135 were selected and used as features. 80% accuracy was achieved in PD-CN vs PD-MCI classification with SVM. In the second study [82], 69 PD and 103 HC were included. 30 low level features, such as WM volume, and high level features, which was connectivity between regions, selected from 3393 possible ones were used as features. 85.78% accuracy was attained with multi-kernel SVM in classification of HC vs PD-CN. In the last study [83], 183 HC and 401 PD were included. The features were composed of SPECT measures and smell and sleep based test scores. 11 features selected from 13 with statistical tests were used in SVM and resulted in an accuracy of 96.40%. Our result for classification of HC vs PD-CN is very close to the results of the second study, but lower than the results of the last study. All these studies had an advantage of a larger patient population. However, first and second studies had a lot of features in the initial set. This makes it difficult to select stable and reliable feature subsets to be used in the classification. All these studies attained best results with SVM based models, whereas, we attained our best results with tree based and kNN based methods.

We found that line orientation test scores decreased over time. This test measures visuospatial performance of the participants. We also found statistically significant MRI based biomarker changes over time in some brain regions of PD-MCI patients. MD value increased in temporal lobe, middle cerebellar peduncle, thalamus and parietal lobe. Temporal lobe is associated with object perception and recognition [101]. In addition, according to a study [102] atrophy occurs in medial temporal lobe in PDD patients. Middle cerebral peduncle is responsible of refining movements and motor control [103]. Thalamus is associated with memory and visual consciousness [104], and parietal lobe is related with visuospatial abilities. Increase in MD value may indicate atrophy progression over time in these regions. In addition to changes in MD values, we observed decrease in FA values in right inferior cerebellar peduncle and right medial lemniscus. Inferior cerebellar peduncle is also associated with regulating motor control as middle cerebellar peduncle [103]. Right medial lemniscus is responsible for sense of proprioception. Decrease in FA values might be a sign of atrophy in these regions. Lastly, we observed a decrease in aBV for middle frontal gyrus. This region is

important in working memory [105]. Decrease in aBV may indicate possible blood flow changes in this region. As a result, these changes are mostly related to visuospatial tasks, motor control and memory tasks, moreover changes in biomarkers indicate possible atrophy and reduced blood flow. This might be expected for PD-MCI patients. Health status of these patients are expected to deteriorate. Although we only found a statistically significant change in line orientation test results, other neuropsychological test results may follow the decreasing pattern if we repeat the same tests after another 18 months.

6. CONCLUSION

In this project, we were able to combine information from multimodal MRI for classification of HC, PD-CN and PD-MCI. We also observed that multimodal MRI provides quantitative information about changes occurring in the brain during the course of PD. The number of participants was rather low, and the number of features were relatively high in our study. This situation was the main challenge in constructing stable classifiers. However, we still managed to define some prominent biomarkers. Findings of this study encouraged us to include more subjects and repeat the same procedure to increase the classification performance. Moreover, we expect that some of the patients will evolve to PDD. Therefore, if we repeat the study in the future, we could examine the MRI pattern changes in later stages of PD-MCI.

7. List of publications produced from the thesis

1. O. Genç, D. B. Arslan, S. Cengiz, G. H. Hatay, A. Kıçik, E. Erdoğdu, Ö. C. Kaplan, Z. Tüfekçioğlu, B. Bilgiç, H. Hanağası, H. Gürvit, T. Demiralp, A. M. Uluğ, E. Ö. Işık,Parkinson hastalığı hafif kognitif bozukluğunun MRG temelli makine öğrenme ile sınıflandırılması, *16.Ulusal Sinirbilim Kongresi*, pp. 101, Mayıs 20-23, 2018.

APPENDIX A. Software Packages

1. MATLAB (https://www.mathworks.com/downloads/)

2. Jupyter Notebook (https://www.anaconda.com/download/)

REFERENCES

1. Schneider, R. B., "Parkinson's disease psychosis: presentation, diagnosis and management," *Neurodegener Dis Manag*, Nov 2017.

2. Bhimani, R., "Understanding the burden on caregivers of people with parkinson's: A scoping review of the literature," *Rehabil Res Pract*, p. 718527, 2014.

3. Van Den Eeden, S.K., e. a., "Incidence of parkinson's disease: variation by age, gender, and race/ethnicity," *Am J Epidemiol*, Vol. 157, pp. 1015–22, Jun 2003.

4. Janvin, C. C., "Subtypes of mild cognitive impairment in parkinson's disease: progression to dementia," *Mov Disord*, Vol. 21, pp. 1343–9, Sep 2006.

5. Kowal, S. L., "The current and projected economic burden of parkinson's disease in the united states," *Mov Disord*, Vol. 28, pp. 311–8, Mar 2013.

6. Poewe, W., "Diagnosis and management of parkinson's disease dementia," *Int J Clin Pract*, Vol. 62, pp. 1581–7, Oct 2008.

7. Litvan, I., "Mds task force on mild cognitive impairment in parkinson's disease: critical review of pd-mci," *Mov Disord*, Vol. 26, pp. 1814–24, Aug 2011.

8. Petcharunpaisan, S., "Arterial spin labeling in neuroimaging," *World J Radiol*, Vol. 2, pp. 384–98, Oct 2010.

9. Conturo, T. E., "Tracking neuronal fiber pathways in the living human brain," *Proc Natl Acad Sci U S A*, Vol. 96, pp. 10422–7, Aug 1999.

10. Ramsay, J., "The elements of statistical learning: Data mining, inference, and prediction.," *Psychometrika*, Vol. 68, pp. 611–612, Dec 2003.

11. Bandyopadhyay, S., "A review of multivariate longitudinal data analysis," *Stat Methods Med Res*, Vol. 20, pp. 299–330, Aug 2011.

12. Poewe, W., "Parkinson disease," *Nat Rev Dis Primers*, Vol. 3, p. 17013, Mar 2017.

13. Fearnley, J. M., "Ageing and parkinson's disease: substantia nigra regional selectivity," *Brain*, Vol. 114 (Pt 5), pp. 2283–301, Oct 1991.

14. Damier, P., "The substantia nigra of the human brain. ii. patterns of loss of dopamine-containing neurons in parkinson's disease," *Brain*, Vol. 122 (Pt 8), pp. 1437–48, Aug 1999.

15. Braak, H., "Staging of brain pathology related to sporadic parkinson's disease," *Neurobiol Aging*, Vol. 24, pp. 197–211, Mar-Apr 2003.

16. Ascherio, A., "The epidemiology of parkinson's disease: risk factors and prevention," *Lancet Neurol*, Vol. 15, pp. 1257–1272, Nov 2016.

17. Chillag-Talmor, O., "Use of a refined drug tracer algorithm to estimate prevalence and incidence of parkinson's disease in a large israeli population," *Journal of Parkinsons Disease*, Vol. 1, no. 1, pp. 35–47, 2011.

18. Gordon, P. H., "Parkinson's disease among american indians and alaska natives: A nationwide prevalence study," *Movement Disorders*, Vol. 27, pp. 1456–1459, Sep 2012.

19. Kalia, L. V., L. A. E., "Parkinson's disease," *Lancet*, Vol. 386, pp. 896–912, Aug 2015.

20. Chaudhuri, K. R., S. A. H., "Non-motor symptoms of parkinson's disease: dopaminergic pathophysiology and treatment," *Lancet Neurol*, Vol. 8, pp. 464–74, May 2009.

21. Weil, R. S., "Mild cognitive impairment in parkinson's disease-what is it?," *Current Neurology and Neuroscience Reports*, Vol. 18, Apr 2018.

22. Williams-Gray, C. H., "The campaign study of parkinson's disease: 10-year outlook in an incident population-based cohort," *J Neurol Neurosurg Psychiatry*, Vol. 84, pp. 1258–64, Nov 2013.

23. Kehagia, A. A., "Cognitive impairment in parkinson's disease: the dual syndrome hypothesis," *Neurodegener Dis*, Vol. 11, pp. 79–92, Nov 2013.

24. Paul, K. C., "Apoe, mapt, and comt and parkinson's disease susceptibility and cognitive symptom progression," *J Parkinsons Dis*, Vol. 6, pp. 349–59, Apr 2016.

25. Bilder, R. M., "The catechol-o-methyltransferase polymorphism: relations to the tonic-phasic dopamine hypothesis and neuropsychiatric phenotypes," *Neuropsychopharmacology*, Vol. 29, pp. 1943–61, Nov 2004.

26. Pascale, E., "Genetic architecture of mapt gene region in parkinson disease subtypes," *Front Cell Neurosci*, Vol. 10, p. 96, 2016.

27. Chen, J., "Functional analysis of genetic variation in catechol-o-methyltransferase (comt): effects on mrna, protein, and enzyme activity in postmortem human brain," *Am J Hum Genet*, Vol. 75, pp. 807–21, Nov 2004.

28. Williams-Gray, C. H., "Catechol o-methyltransferase val158met genotype influences frontoparietal activity during planning in patients with parkinson's disease," *J Neurosci*, Vol. 27, pp. 4832–8, May 2007.

29. Nombela, C., "Genetic impact on cognition and brain function in newly diagnosed parkinson's disease: Icicle-pd study," *Brain*, Vol. 137, pp. 2743–58, Oct 2014.

30. Dijkstra, A. A., "Stage-dependent nigral neuronal loss in incidental lewy body and parkinson's disease," *Mov Disord*, Vol. 29, pp. 1244–51, Sep 2014.

31. Iacono, D., "Parkinson disease and incidental lewy body disease just a question of time?," *Neurology*, Vol. 85, pp. 1670–1679, Nov 2015.

32. Seppi, K., "The movement disorder society evidence-based medicine review update: Treatments for the non-motor symptoms of parkinson's disease," *Mov Disord*, Vol. 26 Suppl 3, pp. S42–80, Oct 2011.

33. Connolly, B. S., L. A. E., "Pharmacological treatment of parkinson disease: a review," *JAMA*, Vol. 311, pp. 1670–83, Apr 2014.

34. Pyatigorskaya, N., "A review of the use of magnetic resonance imaging in parkinson's disease," *Therapeutic Advances in Neurological Disorders*, Vol. 7, pp. 12–26, Jul 2014.

35. Melzer, T. R., "White matter microstructure deteriorates across cognitive stages in parkinson disease," *Neurology*, Vol. 80, pp. 1841–9, May 2013.

36. Agosta, F., "Mild cognitive impairment in parkinson's disease is associated with a distributed pattern of brain white matter damage," *Human Brain Mapping*, Vol. 35, pp. 1921–1929, May 2014.

37. Kamagata, K., "Relationship between cognitive impairment and white-matter alteration in parkinson's disease with dementia: tract-based spatial statistics and tract-specific analysis," *European Radiology*, Vol. 23, pp. 1946–1955, Jul 2013.

38. Fernandez-Seara, M. A., "Cortical hypoperfusion in parkinson's disease assessed using arterial spin labeled perfusion mri," *Neuroimage*, Vol. 59, pp. 2743–2750, Feb 2012.

39. Ma, Y. L., "Parkinson's disease spatial covariance pattern: noninvasive quantification with perfusion mri," *Journal of Cerebral Blood Flow and Metabolism*, Vol. 30, pp. 505–509, Mar 2010.

40. Groger, A., "Three-dimensional magnetic resonance spectroscopic imaging in the substantia nigra of healthy controls and patients with parkinson's disease," *European Radiology*, Vol. 21, pp. 1962–1969, Sep 2011.

41. Hattingen, E., "Phosphorus and proton magnetic resonance spectroscopy demonstrates mitochondrial dysfunction in early and advanced parkinson's disease," *Brain*, Vol. 132, pp. 3285–3297, Dec 2009.

42. Oz, G., "Proton mrs of the unilateral substantia nigra in the human brain at 4 tesla: Detection of high gaba concentrations," *Magnetic Resonance in Medicine*, Vol. 55, pp. 296–301, Feb 2006.

43. Emir, U. E., "Elevated pontine and putamenal gaba levels in mild-moderate parkinson disease detected by 7 tesla proton mrs," *Plos One*, Vol. 7, Jan 2012.

44. Nie, K., "Marked n-acetylaspartate and choline metabolite changes in parkinson's disease patients with mild cognitive impairment," *Parkinsonism & Related Disorders*, Vol. 19, pp. 329–334, Mar 2013.

45. Levin, B. E., "Whole-brain proton mr spectroscopic imaging in parkinson's disease," *Journal of Neuroimaging*, Vol. 24, pp. 39–44, Jan 2014.

46. Groger, A., "Differentiation between idiopathic and atypical parkinsonian syndromes using three-dimensional magnetic resonance spectroscopic imaging," *Journal of Neurology Neurosurgery and Psychiatry*, Vol. 84, pp. 644–649, Jun 2013.

47. Rudkin, T. M., A. D. L., "Proton magnetic resonance spectroscopy for the diagnosis and management of cerebral disorders," *Archives of Neurology*, Vol. 56, pp. 919–926, Aug 1999.

48. Abe, K., "Proton magnetic resonance spectroscopy of patients with parkinsonism," *Brain Research Bulletin*, Vol. 52, pp. 589–595, Aug 2000.

49. Watanabe, H., "Multiple regional 1h-mr spectroscopy in multiple system atrophy: Naa/cr reduction in pontine base as a valuable diagnostic marker," *J Neurol Neurosurg Psychiatry*, Vol. 75, pp. 103–9, Jan 2004.

50. Davie, C. A., "Differentiation of multiple system atrophy from idiopathic parkinsons-disease using proton magnetic-resonance spectroscopy," *Annals of Neurology*, Vol. 37, pp. 204–210, Feb 1995.

51. Alpaydın, E., *Introduction to Machine Learning*, pp. 1–3. Cambridge, Massachusetts: The MIT Press, third edition ed., 2014.

52. Alpaydın, E., *Introduction to Machine Learning*, p. 249. Cambridge, Massachusetts: The MIT Press, third edition ed., 2014.

53. Alpaydın, E., *Introduction to Machine Learning*, p. 248. Cambridge, Massachusetts: The MIT Press, third edition ed., 2014.

54. Alpaydın, E., *Introduction to Machine Learning*, p. 253. Cambridge, Massachusetts: The MIT Press, third edition ed., 2014.

55. Bishop, C. M., *Pattern recognition and machine learning*, Springer, 2006.

56. Alpaydın, E., *Introduction to Machine Learning*, pp. 190–191. Cambridge, Massachusetts: The MIT Press, third edition ed., 2014.

57. Alpaydın, E., *Introduction to Machine Learning*, pp. 196–198. Cambridge, Massachusetts: The MIT Press, third edition ed., 2014.

58. Alpaydın, E., *Introduction to Machine Learning*, p. 198. Cambridge, Massachusetts: The MIT Press, third edition ed., 2014.

59. Alpaydın, E., *Introduction to Machine Learning*, p. 350. Cambridge, Massachusetts: The MIT Press, third edition ed., 2014.

60. Alpaydın, E., *Introduction to Machine Learning*, p. 354. Cambridge, Massachusetts: The MIT Press, third edition ed., 2014.

61. Alpaydın, E., *Introduction to Machine Learning*, pp. 359–364. Cambridge, Massachusetts: The MIT Press, third edition ed., 2014.

62. Fletcher, T., "Support vector machines explained," Mar 2009.

63. Alpaydın, E., *Introduction to Machine Learning*, p. 217. Cambridge, Massachusetts: The MIT Press, third edition ed., 2014.

64. Alpaydın, E., *Introduction to Machine Learning*, p. 214. Cambridge, Massachusetts: The MIT Press, third edition ed., 2014.

65. Alpaydın, E., *Introduction to Machine Learning*, p. 499. Cambridge, Massachusetts: The MIT Press, third edition ed., 2014.

66. Alpaydın, E., *Introduction to Machine Learning*, p. 490. Cambridge, Massachusetts: The MIT Press, third edition ed., 2014.

67. Alpaydın, E., *Introduction to Machine Learning*, p. 491. Cambridge, Massachusetts: The MIT Press, third edition ed., 2014.

68. Wolpert, D., "Stacked generalizations," *Neural Networks*, no. 5, pp. 241–259, 1992.

69. Jr, M. P. P., "Combining classifiers: from the creation of ensembles to the decision fusion," *Neural Networks*, 2011.

70. Ulaş, A, O. Y., and E. Alpaydın, "Eigenclassifiers for combining correlated classifiers," *Information Sciences*, no. 187, pp. 109–120, 2012.

71. Schapire, R., "The strength of weak learnability," *Machine Learning*, no. 5, pp. 197–227, 1990.

72. Freund, Y., a. R. S., "Experiments with a new boosting algorithm," In Thirteenth International Conference on Machine Learning, 1996.

73. Alpaydın, E., *Introduction to Machine Learning*, p. 501. Cambridge, Massachusetts: The MIT Press, third edition ed., 2014.

74. Alpaydın, E., *Introduction to Machine Learning*, p. 118. Cambridge, Massachusetts: The MIT Press, third edition ed., 2014.

75. Samb, M. L., "A novel rfe-svm-based feature selection approach for classification," *International Journal of Advanced Science and Technology*, Vol. 43, Jun 2012.

76. Sanchez-Marono, "Filter methods for feature selection – a comparative study," in *Intelligent Data Engineering and Automated Learning - IDEAL*, pp. 178–187, Springer Berlin Heidelberg, 2007.

77. Bergstra, J., B. Y., "Random search for hyper-parameter optimization," *JMLR*, pp. 281–305, Feb 2012.

78. Wu, L., *Mixed Effects Models for Complex Data*, pp. 39–48. CRC Press, 2010.

79. Gelman, A., "Analysis of variance-why it is more important than ever," *The Annals of Statistics*, Vol. 33, no. 1, pp. 1–53, 2005.

80. Laird, N. M.; Ware, J. H., "Random-effects models for longitudinal data," *Biometrics*, Vol. 38, pp. 963–974, Dec 1982.

81. Abos, A., "Discriminating cognitive status in parkinson's disease through functional connectomics and machine learning," *Scientific Reports*, Vol. 7, Mar 2017.

82. Peng, B., "A multilevel-roi-features-based machine learning method for detection of morphometric biomarkers in parkinson's disease," *Neurosci Lett*, Vol. 651, pp. 88–94, Jun 2017.

83. Prashanth, R., "High-accuracy detection of early parkinson's disease through multimodal features and machine learning," *Neurosci Lett*, Vol. 90, pp. 13–21, Jun 2016.

84. Edelman, R., "Qualitative mapping of cerebral blood flow and functional localization with echo-planar mr imaging and signal targeting with alternating radio frequency.," *Radiology*, Vol. 192, pp. 513–520, Aug 1994.

85. Arslan, D. B., "Asl-mr ile parkinson hastalığı hafif kognitif bozukluğunda serebral perf"

86. Buxton, R., "A general kinetic model for quantitative perfusion imaging with arterial spin labeling," *Magn Reson Med.*, Vol. 40, pp. 383–96, Sep 1998.

87. Makris, N., "Decreased volume of left and total anterior insular lobule in schizophrenia," *Schizophrenia Research*, Vol. 83, pp. 155–171, Apr 2006.

88. Kaplan, O., O.-I. E., "Determination of diffusion weighted magnetic resonance imaging based biomarkers of mild cognitive impairment in parkinson's disease," (Istanbul), Biyomut, 2017.

89. Provencher, S. W., "Automatic quantitation of localized in vivo 1h spectra with lcmodel," *NMR Biomed*, Vol. 14, pp. 260–4, Jun 2001.

90. Cengiz, S., "Determination of biomarkers for mild cognitive impairment in parkinson's disease using magnetic resonance spectroscopic imaging," Master's thesis, Bogazici University, Istanbul, Turkey, 2016.

91. Keogh, E., and A. Mueen, *Encyclopedia of Machine Learning and Data Mining*, pp. 314–315. Boston: Springer, 2017.

92. Galantucci, S., and F. Agosta, "Corpus callosum damage and motor function in parkinson's disease," *Neurology*, Vol. 82, Apr 2015.

93. Jannetta, P. J., "Parkinson's disease: an inquiry into the etiology and treatment," *Neurology International*, Vol. 3, 2011.

94. Miller, E. K., "The prefrontal cortex categories, concepts and cognition.," *Philosophical Transactions of the Royal Society Biological Sciences.*, pp. 1123–1136., Aug 2002.

95. Malenka RC, Nestler EJ, H. S., *A Foundation for Clinical Neuroscience*, pp. 147–148. New York: McGraw-Hill Medical, second edition ed., 2009.

96. Pfeiffer, R. F., *Parkinson's Disease, Second Edition*, p. 596. CRC Press, Oct 2012.

97. Graves, W. W., "Left posterior superior temporal gyrus participates specifically in accessing lexical phonology," *J Cogn Neurosci*, Vol. 20, pp. 1698–1710, Sep 2008.

98. Shapleske, J., "The planum temporale: a systematic, quantitative review of its structural, functional and clinical significance," *Brain Research Reviews*, Vol. 29, pp. 26–49, Jan 1999.

99. Beeman, M. J., "Neural activity when people solve verbal problems with insight," *PLoS Biology*, Vol. 2, Apr 2004.

100. Pagonabarraga, J., "Pattern of regional cortical thinning associated with cognitive deterioration in parkinson's disease," *PLoS One*, Vol. 8, Jan 2013.

101. Schacter, D. L., *Psychology.* New York: Worth Publishers, second edition ed., 2010.

102. Silbert, L. C., "Neuroimaging and cognition in parkinson's disease dementia," *Brain Pathol.*, Vol. 20, pp. 646–653, May 2010.

103. J.., N., *The Human Brain: An Introduction to Its Functional Anatomy.* Mosby, 2002.

104. Jerath, R., "Functional representation of vision within the mind: A visual consciousness model based in 3d default space," *Journal of Medical Hypotheses and Ideas*, Vol. 9, pp. 45–56, Mar 2015.

105. Harms, M. P., "Structure-function relationship of working memory activity with hippocampal and prefrontal cortex volumes," *Brain Struct Funct.*, Vol. 218, Jan 2013.