COMPUTER-AIDED DIAGNOSIS OF ALZHEIMER'S DISEASE AND MILD
COGNITIVE IMPAIRMENT WITH MARS/CMARS CLASSIFICATION USING
STRUCTURAL MR IMAGES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ALPER ÇEVİK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
BIOMEDICAL ENGINEERING

SEPTEMBER 2017

Approval of the thesis:

COMPUTER-AIDED DIAGNOSIS OF ALZHEIMER'S DISEASE AND MILD COGNITIVE IMPAIRMENT WITH MARS/CMARS CLASSIFICATION USING STRUCTURAL MR IMAGES

submitted by **ALPER ÇEVİK** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Biomedical Engineering Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Hakan Işık Tarman
Head of Department, **Biomedical Engineering** _____

Prof. Dr. B. Murat Eyüboğlu
Supervisor, **Electrical and Electronics Engineering Dept.,** _____
METU

Prof. Dr. Gerhard-Wilhelm Weber
Co-supervisor, **Institute of Applied Mathematics, METU** _____

Examining Committee Members:

Prof. Dr. Gözde Bozdağı Akar
Electrical and Electronics Engineering Dept., METU _____

Prof. Dr. B. Murat Eyüboğlu
Electrical and Electronics Engineering Dept., METU _____

Prof. Dr. Kader Karlı Oğuz
Dept. of Radiology, Hacettepe University _____

Assoc. Prof. Dr. İlkay Ulusoy
Electrical and Electronics Engineering Dept., METU _____

Assist. Prof. Dr. Özlem Birgül
Biomedical Engineering Dept., Ankara University _____

Date: _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: ALPER ÇEVİK

Signature   :

ABSTRACT

COMPUTER-AIDED DIAGNOSIS OF ALZHEIMER'S DISEASE AND MILD COGNITIVE IMPAIRMENT WITH MARS/CMARS CLASSIFICATION USING STRUCTURAL MR IMAGES

Çevik, Alper

Ph.D., Department of Biomedical Engineering

Supervisor   : Prof. Dr. B. Murat Eyüboğlu

Co-Supervisor   : Prof. Dr. Gerhard-Wilhelm Weber

September 2017, 92 pages

Early detection of Alzheimer's disease (AD) and its prodromal stage, amnestic mild cognitive impairment (MCI), has drawn remarkable attention in recent years. Despite the impressive developments in fields of image analysis, pattern classification, and machine learning, no computer-aided diagnosis system has yet been a part of the clinical routine to diagnose the AD.

This thesis study aims to propose a thorough procedure which involves detecting the early signs of disease-originated deformations by fully-automated analysis of structural brain magnetic resonance images (MRI). A comprehensive review including the taxonomy of related biomarkers and state-of-the-art techniques is introduced.

Proposed methodology involves extraction of voxel intensity-based features (such as tissue probability maps) through segmenation and registration of brain MRI volumes. Voxel-based morphometry framework is employed to provide one-to-one correspondance between the images. Quality of the feature set is evaluated by an analysis including other approaches such as feature-based morphometry. A novel hybrid procedure involving both statistical analysis and utilization of domain knowledge is proposed for feature selection. Performance of the method is compared with these of well-known dimensionality reduction techniques. Multivariate adaptive regression

splines (MARS) and Conic MARS (CMARS) were utilized for construction of the class-separating hyperplanes through a parameter optimization procedure involving cross-validation. This study is the first-time engagement of both MARS and CMARS algorithms in field of medical image analysis. Qualitative and quantitative evaluations of classifier performances were presented including a comparison with benchmark studies in the field. Promising results are acquired through the tests performed on Alzheimer's Disease Neuroimaging Initiative (ADNI) data.

Keywords: Computer-Aided Diagnosis, Medical Image Analysis, Alzheimer's Disease, Pattern Classification, Machine Learning

ÖZ

ALZHEIMER HASTALIĞI VE HAFİF BİLİŞSEL BOZUKLUĞUN MARS/CMARS SINIFLANDIRMA İLE YAPISAL MR GÖRÜNTÜLERİ ÜZERİNDEN BİLGİSAYAR DESTEKLİ TANILANMASI

Çevik, Alper

Doktora, Biyomedikal Mühendisliği Bölümü

Tez Yöneticisi   : Prof. Dr. B. Murat Eyüboğlu

Ortak Tez Yöneticisi : Prof. Dr. Gerhard-Wilhelm Weber

Eylül 2017 , 92 sayfa

Alzheimer hastalığının (AH) ve prodromal evresi olan amnestik hafif bilişsel bozukluğun (HBB) erken tespiti, son yıllarda dikkat çeken bir konu haline gelmiştir. Görüntü analizi, örüntü sınıflandırma ve makine öğrenimi alanlarındaki etkileyici gelişmelere rağmen, henüz hiçbir bilgisayar destekli tanı sistemi, AH teşhisinin rutin klinik sürecinin bir parçası haline gelmemiştir.

Bu tez çalışması, yapısal beyin manyetik rezonans (MR) görüntülerinin analiziyle, hastalığa bağlı deformasyonların erken dönem belirtilerini saptamayı içeren kapsamlı bir prosedür önermeyi amaçlamaktadır. İlgili biyobelirteçlerin taksonomisi ve yaygın kabul görmüş tekniklerin incelemesini de barındıran, kapsamlı bir literatür taraması ortaya konmuştur.

Önerilen metodoloji, voksel yoğunluğuna dayanan (örneğin doku olasılık haritaları gibi) özniteliklerin, hacimsel beyin MR görüntülerinin bölütlenmesi ve hizalanması yoluyla çıkarılmasını içermektedir. Voksel tabanlı morfometri çerçevesi, imgeler arasında bire bir eşlemenin sağlanması için kullanılmaktadır. Öznitelik kümesinin kalitesi, öznitelik tabanlı morfometri gibi diğer yaklaşımları da içeren bir analizle değerlendirilmiştir. Öznitelik seçimi için hem istatistiksel analiz hem de alan bilgisi kullanımını içeren yeni bir hibrid prosedür önerilmiştir. Yöntemin performansı, bili-

nen boyut azaltma teknikleriyle karşılaştırılmıştır. Çapraz doğrulama içeren bir parametre optimizasyon prosedürü ile sınıf ayrımlı hiperdüzlemlerin oluşturulması için çok değişkenli adaptif regresyon eğrileri (MARS) ve konik MARS (CMARS) metotları kullanılmıştır. Bu çalışma kapsamında, tıbbi görüntü analizi alanında hem MARS hem de CMARS algoritmaları ilk kez kullanılmıştır. Sınıflandırıcı performanslarının nitel ve nicel değerlendirmeleri, alandaki önemli çalışmalarla karşılaştırmayı da içerir şekilde sunulmuştur. Alzheimer Hastalığı Nörogörüntüleme Girişimi (ADNI) verileri üzerinde yapılan testler yoluyla umut verici sonuçlar elde edilmiştir.

Anahtar Kelimeler: Bilgisayar Destekli Tanılama, Tıbbi Görüntü Analizi, Alzheimer Hastalığı, Örüntü Sınıflandırma, Makine Öğrenimi

To my family and friends...

ACKNOWLEDGMENTS

Finally, and most importantly, I would like to thank my beloved wife, Ajdan, who has made available her support in a number of ways throughout the last *chapter* of my life. This book you are holding in your hands, or scrolling down on the screen, would certainly have not existed if her *steel-strength* patience did not.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

AD	Alzheimer's disease
ADNI	Alzheimer's Disease Neuroimaging Initiative
ANN	Artificial neural networks
CAD	Computer-aided diagnosis
CART	Classification and regression trees
CDR	Clinical dementia rating
CMARS	Conic multivariate adaptive regression splines
CN	Cognitively normal
CQP	Conic quadratic programming
CSF	Cerebrospinal fluid
DARTEL	Diffeomorphic anatomical registration through exponentiated lie algebra
DoG	Difference of Gaussians
EER	Equal error rate
EM	Expectation-maximization
FBM	Feature-based morphometry
FDG PET	Fludeoxyglucose positron emission tomography
FDR	Fisher's discriminant ratio
GAM	Generalized additive models
GCV	Generalized cross-validation
GM	Grey (gray) matter
HOG	Histograms of oriented gradients
ICBM	International Consortium for Brain Mapping
LM	Linear models
LoG	Laplacian of Gaussian
MARS	Multivariate adaptive regression splines
MCI	Mild cognitive impairment
MCIc	Mild cognitive impairment (converting)
MCInc	Mild cognitive impairment (non-converting)
MKL	Multiple kernel learning

MNI	Montreal Neurological Institute
MoG	Mixture of Gaussians
MR	Magnetic resonance
MRI	Magnetic resonance imaging
MSE	Mental state examination
PRSS	Penalized residual sum of squares
RF	Random forest (classifiers)
ROC	Receiver operating characteristics
ROI	Region of interest
RSS	Residual sum of squares
SIFT	Scale-invariant feature transform
SPM	Statistical Parametric Mapping
TC	Tissue class
TPM	Tissue probability map
T1-W	T1-weighted
VAF	Voxel-as-feature
VBM	Voxel-based morphometry
WM	White matter

CHAPTER 1

INTRODUCTION

1.1 Motivation

Alzheimer's disease (AD) is the most common form of age-associated dementias which cause major problems with thinking, memory, and behavior. Symptoms usually develop slowly and get worse over time, and reach to a level which severely affects everyday life. It gradually interferes with all cognitive functions and may eventually lead to death. This neurodegenerative disease is known to be associated with "structural atrophy, pathological amyloid depositions, and metabolic alterations in the brain" [47]. The underlying mechanism responsible for these deformations has not yet been fully elucidated and remains a remarkable research subject in the areas of neurology and mathematical modeling [2].

According to the most recent annual report [10] of Alzheimer's Association[1] based on the U.S. data:

- Ten percent of the people which are 65 years old or older has Alzheimer's disease. This proportion increases with increasing age. (Only 3% of people are aged 65-74, whereas 32% of people are aged 85 and older have Alzheimer's disease.)

- In the USA, Alzheimer's disease is the 6th leading cause of death. Every 66 seconds, someone in the United States develops Alzheimer's dementia. By the year 2016, more than 5 million Americans suffer from Alzheimer's disease.

---

[1] 2017 Alzheimer's Disease Facts and Figures is a statistical resource for United States data related to Alzheimer's disease. `http://www.alz.org/documents_custom/2017-facts-and-figures.pdf`. [Accessed: 24- July- 2017] [98].

This figure is foreseen as 16 million by 2050.

- Since the year 2000, number of deaths from heart disease have decreased by 14% whereas the number of deaths related to Alzheimer's disease have increased by 89%. Death prevalence of AD is greater than those of breast cancer and prostate cancer combined.

- In 2017, Alzheimer's and other dementias will cost the nation $259 billion. By 2050, these costs are estimated to rise as high as $1.1 trillion.

- Thirty five percent of caregivers for AD (or another dementia) patients report that their health has deteriorated due to patient care responsibilities, compared to 19% of caregivers for older people without dementia.

Alzheimer's disease has no current cure. However, identification of AD in its preclinical stage has become extraordinarily important together with the attempts to postpone the onset or slow down progression of the disease and introduce an efficient symptomatic treatment. Thus, diagnosis of AD in its earliest stages - before irreversible brain damage or mental decline has occurred - is of vital importance. Quality of the treatment is strongly related to how early the disease is detected. Thus, early diagnosis improves quality of life for Alzheimer's patients and their caregivers.

Recently, no particular test that verifies a person suffers (or will suffer) from AD or *mild cognitive impairment (MCI)*[2] exists. Since the *amyloid plaques* and *neurofibrillary tangles* must be histopathologically confirmed for the exact diagnosis, it is only possible by an autopsy [94].

In the routine clinical practice, the disease is diagnosed through a full assessment involving physical examination, medical history, diagnostic tests (laboratory, clinical, and genetic - e.g., APOE-e4 gene), neurological evaluation (of coordination, reflexes, eye movement, muscle strength and tone, sensation, and speech), and surveys on the mental status (e.g., the mental state examination (MSE), mini-cog, and mood assessment). A medical workup for AD usually incorporates utilization of structural and

---

[2] MCI is a syndrome of brain functions which leads to a minor deterioration in cognitive capability and an elevated possibility of converting into AD. MCI can demonstrate various symptoms. In general, it is seen as a prodromal stage of Alzheimer's disease when impairments in the episodic memory is predominant over these symptoms.

functional imaging modalities like magnetic resonance imaging (MRI) and positron emission tomography (PET), relatively. Nevertheless, the aforementioned brain imaging methods are mostly elaborated to disqualify other reasonable situations such as stroke, tumor, fluid collection in intracranial compartments, and traumatic brain injury which may generate symptoms identical with AD but require different types of treatment.

Computer-aided diagnosis (CAD) of Alzheimer's disease has attracted considerable attention recently. Up to now, various biomarkers have been proved to be sensitive to the presence of AD and MCI. These biomarkers can be subsumed under three main groups: **(i)** MRI (for brain atrophy measurement), **(ii)** functional imaging (for hypometabolism quantification), and **(iii)** cerebrospinal fluid (CSF) (for quantification of specific proteins) [106]. A comprehensive review of the literature on these biomarkers is presented in Subsection 2.1.1.

Despite the growing interest in utilization of image analysis, statistical pattern classification and machine learning methods in field of medicine, no computer-aided diagnosis procedure has yet been a part of the clinical routine to help diagnosing the Alzheimer's disease (AD). Since the early diagnosis - or at least raised awareness of the early symptoms - would lead to a significant boost in quality of disease management, a computer-aided diagnosis tool which has the ability to discriminate between the effects of normal aging and AD or MCI on brain structure is of great value.

1.2 Objectives and Scope

Aim of this thesis study is to build up a procedure which improves the early detection of AD and MCI using structural brain MRI volumes, and builds a foundation for a fully-automated diagnostic software. The proposed approach is based on preparation of the MR images with appropriate image processing techniques, extraction and specification of the features with relatively higher strength to represent the distinguishing image properties, determination of the optimal model parameters, classification of image patterns utilizing the elements of statistical learning, and evaluation of the overall performance by qualitative and quantitative analysis involving an objective

comparison with other techniques examined in the field. The scope of this thesis study can be summarized in general terms of a machine learning framework as follows.

Data: Alzheimer's Disease Neuroimaging Initiative (ADNI) is a long-time study which involves observation of cognitively normal elders, and subjects with Alzheimer's disease and mild cognitive impairment in multiple centers. In scope of the thesis, magnetic resonance imaging (MRI), (18F)-Fludeoxyglucose positron emission tomography (FDG PET), urine serum, and cerebrospinal fluid (CSF) biomarkers are acquired together with the clinical and neuropsychological follow-up data at certain intervals from each participant [46]. In this thesis, datasets used for model training by cross-validation and performance evaluation were obtained from the ADNI database.

Feature Extraction: Image processing operations which are part of the *voxel-based morphometry* (VBM) [8] framework were employed in feature extraction procedure. Within this scope, 3D brain MR images acquired from the ADNI archive were segmented into the tissue probability maps belonging to the three main brain tissues; namely, gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). This process is called Unified Segmentation [9] and it also involves removal of the non-uniform bias field due to the imperfections in the acquisition process, and rigid-body registration of the images together with the tissue segmentation. Subsequent to the segmentation, images are non-linearly registered using the DARTEL algorithm [6]. Finally, all of the images were transformed into the MNI space by employing the DARTEL flow-fields, which were acquired in the preceding step. Performance outcome of this voxel-based approach is compared with those of widely-used feature description extraction algorithms such as *scale-invariant feature transform* (SIFT) [57] and *histograms of oriented gradients* (HOG) [24].

Dimensionality Reduction: Since the "*whole-brain analysis*" approach is embraced and the tissue probabilities derived from voxel intensities are treated as features in the beginning, total dimensionality is high enough to guarantee for the problem to suffer from the well-known "curse of dimensionality" phenomenon

[48]. This means introduction of a step for reducing the dimensionality towards a reasonable range is crucial. For this purpose, a novel 3-step procedure involving both statistical analysis and domain-specific knowledge is proposed. The reader may refer to our paper, "*Voxel-MARS*" [17], for mathematical definition of the methodology and justification of the preferences.

Classification: *Multivariate Adaptive Regression Splines* (MARS) [38] models were built and validated through a grid-search method combined with *N*-times replicated *k*-fold cross-validation procedure applied on the training data. Conic MARS (CMARS) [103] method is employed which replaces backward (model pruning) step of MARS. In this way, the coefficients for the basis-functions are updated rather than removing a number of them from the model, to avoid overfitting. Classification performance were compared to those of 28 other methods in terms of *sensitivity*, *specificity*, *positive predictive value* (PPV), and *negative predictive value* (NPV). Results (for MARS) were also reported in [17]. Additional results involving CMARS can be found in this thesis (see Chapter 6).

Main contributions of this study can be listed as follows: **(i)** A non-parametric, adaptive extension of *regression trees*, namely *Multivariate Adaptive Regression Splines* (MARS), was utilized for classification of high-resolution, volumetric MR images, for early detection of AD. MARS, unlike *Generalized Linear Models* and *Generalized Additive Models*, has the ability to model the interactions between variables through nonlinearities, and to automatically select the variables to be included in the model function. In this study, it was proved that these advantages maintained by the method, lead to a remarkable success in classification of high-dimensional data, even under the existence of the small sample size problem. **(ii)** This study also covers the application of a recently developed method, *Conic MARS* (CMARS) - a mathematically more integrated variant of MARS which involves concepts of *regularization* and *modern optimization*. CMARS replaces the pruning phase of MARS with a penalized cost function, which implies a trade-off between *accuracy* and *complexity*. In this study, it was shown that minimization of the cost function by constructing and solving a *Conic Quadratic Problem* enhances resulting models in terms of *robustness*, as well as increasing the amount of information preserved in the models, with respect to MARS. Utilization of both MARS and CMARS methods as pattern classifier is a

first-time study in the field of medical image analysis. **(iii)** A partially novel, 3-step, hybrid dimensionality reduction procedure involving *statistical analysis* and enabling inclusion of the *domain knowledge* was proposed for determination of significant tissue probabilities as features. Proposed approach was discussed qualitatively and quantitatively, in comparison with well-established techniques, and successful results were obtained. **(iv)** Features were represented by *voxel-based morphometry* through *whole-brain analysis* and this technique was compared with another widely-used approach, namely, *feature-based morphometry*. It was observed that the first approach, which is based on one-to-one correspondence between voxels, is more adequate than the second one, which involves representation of the images by descriptive features, for AD detection in the early stages of the disease.

As a result, a software routine which is capable of separating AD and MCI brain MRI from the cognitively healthy ones, and converting MCIs from non-converting ones, with high accuracy, sensitivity and specificity rates, was developed. It is important to mention that the proposed approach relies only on the assessment of structural MRI, which is already a part of standard clinical process for diagnosing AD.

1.3   Thesis Outline

The thesis begins with a brief introduction given in Chapter 1 (*Introduction*), where the motivation behind the work, main objectives and scope of the study are briefly introduced to the reader. The purpose and original contributions of the thesis are clearly stated in this chapter.

Chapter 2 (*Background*) is mainly composed of three parts. Firstly, a literature review is presented. This part (Section 2.1) includes summaries of research on both biomarkers of various formats and classification techniques utilized for AD diagnosis. In the following two sections (Section 2.2 and Section 2.3), the mathematical basis of the methods MARS and CMARS are explained, respectively. The discussions justifying our preferences in terms of biomarker and classification methodology can also be found in this chapter.

In the next chapter (Chapter 3), information on the data used in model training and

6

performance tests are given. Following section (Section 3.2) provides details of the feature extraction procedure which involves concepts of voxel-based morphometry. In the final section of the chapter, alternative methods in the field of feature-based morphometry are examined.

A novel, three-step, hybrid procedure for selection of significant features is put forward in Chapter 4. In addition to the description of method, this chapter contains a comparison of the proposed technique with other commonly-used ones in terms of overall performance.

Details regarding the implementation and performance evaluation are included in two consecutive chapters, namely, Chapter 5 and Chapter 6, relatively. The first chapter involves the details of model building, optimization and visualization of MARS and CMARS classifiers. Comparative results can be found in Chapter 6 (Performance Evaluation). This chapter also involves qualitative discussions that addresses the overall performance of the proposed methodology in several different aspects.

Finally, a brief conclusion, which is also including an outlook to the future scientific research plans, diagnostic tools, and possible applications is presented in the last chapter.

CHAPTER 2

BACKGROUND

2.1 Early Detection of AD and MCI

2.1.1 A Taxonomy of Neuroimaging-Based Biomarkers Employed in AD Diagnosis

It has been proven that, multiple neuroimaging-based biomarkers are sensitive to presence of AD and MCI; i.e., structural MRI for measuring brain atrophy, functional imaging with functional MRI (fMRI) and Positron Emission tomography (PET) for hypometabolism quantification, Single-Photon Emission Computed Tomography (SPECT) for perfusion, and Diffusion Tensor Imaging (DTI) for connectivity analysis (cf. Figure 2.1). Additionally, several studies demonstrated that multimodal feature sets obtained by combining biomarkers of different imaging techniques can also be efficiently used to diagnose AD. Throughout the next subsection, AD and MCI detection strategies based on the biomarkers derived from structural MR images are investigated in detail (see Subsection 2.1.2). A review regarding several of the most remarkable methods included in other groups is presented in this section.

Hinrichs et al. (2009) [44] presented a novel classification framework, which is based on FDG-PET imaging data. The proposed algorithm incorporates spatial priors by utilization of *boosting* of the *weak classifiers* corresponding to voxel intensity values. Similarly in [13], FDG-PET images were analyzed for prediction of conversion from MCI to AD at different stages of disease through a supervised learning algorithm based on *Support Vector Machines* (SVM). The study was completed with a prediction accuracy of 85%. Volume of interest-based segmentation of FDG-PET images was
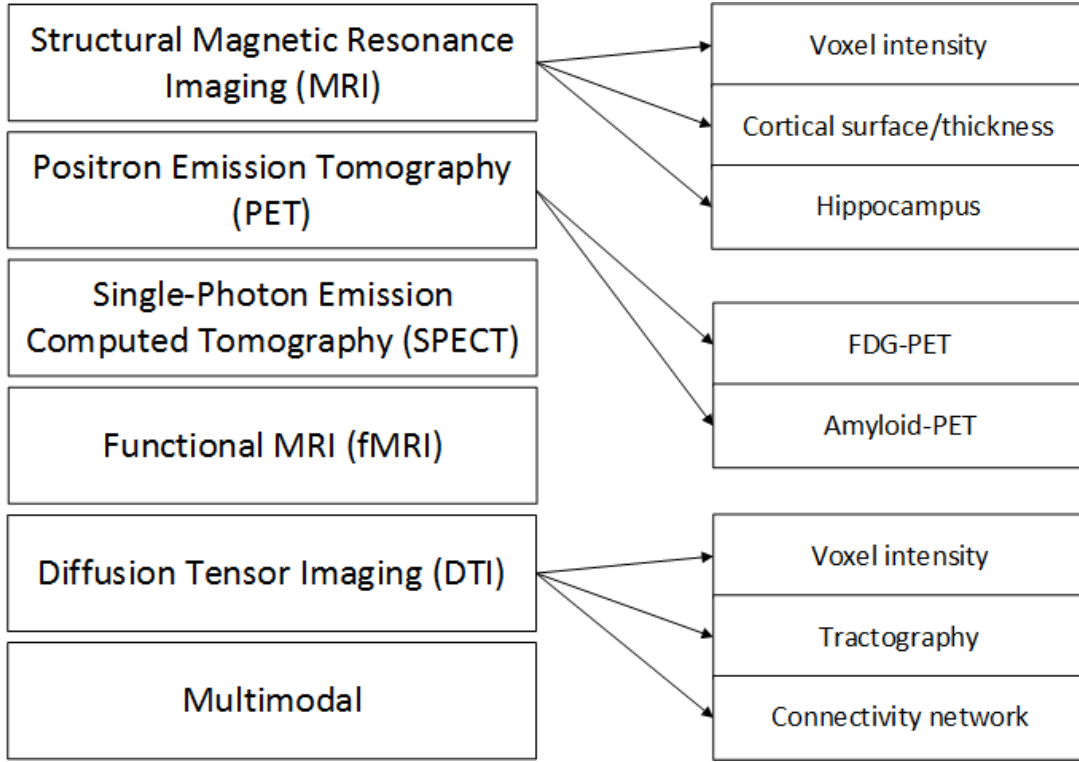
Figure 2.1: Block diagram: Neuroimaging-based biomarkers.

embraced for detection of converter MCIs by Pagani et al. (2015) and successful results - including the ability to predict conversion 2 years earlier than its occurrence - were presented in [75]. In [100], an algorithm which employs amyloid-PET images for quantification of *plaque*s at different stages of disease was introduced by Vandenberghe et al. (2013).

Studies presented in [41], [79], and [78] can be listed as examples that involve utilization of perfusion data acquired through SPECT. In [78], it was shown that the overall accuracy is inversely related to the number of input variables and the reported sensitivity was above 90%.

Chen et al. (2011) [19] conducted an analysis of large scale connectivity networks to classify AD, MCI, and healthy subjects by employing fMRI data and segmenting the brain into 116 separate regions of interest. The relations between network connectivity and behavior were analyzed through linear regression analysis. The research studies presented in [50] and [52] also involve successful approaches based on fMRI data.

The most important works involving AD diagnosis based on DTI data include tracto-graphy based method presented in [66], connectivity network analyses performed in [104] and [77], and the voxel intensity-based approach introduced in [31]. Highest accuracy (of 83%) in AD/CN classification among these was achieved by the latest one.

There also exist several studies that involve successful combinations of different bio-markers provided by different imaging modalities to form a classification framework. In [101], Vemuri et al. (2008) combined tissue density maps with demographics including age and gender along with the genetic profile information of the participants. A remarkable approach for dimensionality reduction, called STAND-score, was proposed by the authors. Davatzikos et al. (2011) [26] analyzed CSF proteins together with neuroimaging data to predict conversion from MCI to AD. Other attention-grabbing examples of multimodal studies that should be mentioned are [111] and [85]. A very up-to-date review on the neuroimaging biomarkers can be comprehended in [80].

2.1.2 AD and MCI Detection Based on Structural MRI

The most commonly used medical imaging modality for visualizing the anatomical structure of the brain is known as T1-W MRI (T1-weighted magnetic resonance imaging). T1-W MRI technology has the ability to provide volumetric images with respectively high resolution and contrast when compared with the alternatives. The underlying mechanism is based on a measure related with the recovery time of longitudinal magnetization established by an intense magnetic field, following the excitation pulse. This variable is called T1 relaxation time and it varies between different biological components forming the brain tissue. T1-W MRI technology simply utilizes this measure as the physical entity to convert into numerical distributions that form images. Generated digital data are presented in terms of voxels (volumetric pixels). Each voxel corresponds to an index by which the position is addressed, and an intensity value by which the gray level in this particular position is defined. Voxel intensities mostly take a decimal value between 0 and 1, where 0 represents black, 1 represents white, and any value lying between 0 and 1 is mapped into a different gray level. Bit-depth and resolution of the resulting images vary depending on the

hardware and the software qualifications of the MRI device.

Anatomical structures of the brain differ in their organic matter coverage. For example, the white matter (WM) which is composed of neuronal axons in large proportions has a high fat content, while the gray matter (GM) which is mostly formed from cell bodies is protein-rich, respectively. By this means, T1-W MRI modality, which is highly sensitive to the T1 relaxation times in different tissues, provide good contrast between WM and GM.

Voxel intensity-based biomarkers obtained through the analysis of brain as a whole were proved to be effective in early detection of AD, in many studies. This group of techniques is generally referred as "*whole-brain analysis*". In [1], the authors investigated MRI images of correctly and incorrectly diagnosed subjects with known current conditions in order to question how early can the conversion be detected. Davatzikos et al. (2008) [27] reported their findings on formation of subtle structural patterns of abnormality in structural brain MRI, in very early phases of MCI. In [53], linear suport vector machines used for discriminating effects of AD on brain from normal effects occurring in elderly subjects. The authors claimed that analyzing the brain as a whole was the essential approach at least in the beginning. The papers [107], [14], and [82] can also be noted as important studies falling into this group.

It has already been shown that biomarkers based on individual anatomical structures such as hippocampus, cortex, and amygdale may serve as biomarkers to detect AD. This group of feature extraction approaches is also called "*ROI-based methods*". ROI-based methods rely on *a priori* knowledge of particular shape deformations in specific brain structures. Enlargement of ventricles [4, 105], shrinkage of the hippocampus [54, 11, 102, 55], and local declines in the cortical thickness [76, 33, 32, 28] can be mentioned as widely-known anatomical deformations consequent to the existence of AD. Boutet et al. (2014) [11] demonstrated that even minor alterations in the hippocampal volume caused by AD can be detected using a 7T MRI. In papers [22] and [21], Chupin et al. proposed novel methods to perform early Alzheimer's detection through segmentation of amygdala and hippocampus.

In the work presented in [110], Zhang et al. (2015) utilized *displacement field* (DF), estimated by level-set method to track the morphometry from normal brains to the

brains of AD patients. For this purpose, the authors preferred performing "*whole-brain analysis*" - in which, all voxels in the brain are considered as a whole - rather than employing one of the ROI-based techniques. In the paper, rationale behind this preference is introduced through a list of drawbacks that the ROI-based methods are suffering from. These are: **(i)** requirement of the expert knowledge; **(ii)** dependency of the prediction accuracy to the experience level of the interpreters; **(iii)** difficulty in implementation of the mutual information among the pixels; **(iv)** potential existence of other regions connected with AD; and **(v)** examiners' tendency to segment the ROI manually rather than automatically.

In this study, *whole-brain analysis* is preferred rather than ROI-based approaches which suffer from several shortcomings listed above. In this context, commonly-used direct "*voxel-as-feature*" (VAF) methodology [41] was embraced for extraction of the baseline features. In addition to the aforementioned justification, our future intention to expand our work to handle computer-aided diagnosis of other neurodegenerative diseases can be mentioned as another motive to choose whole-brain approach and VAF methodology to obtain the largest set of initial baseline features for classification.

2.1.3 Classification Techniques Utilized for AD Detection

One of the most important components among the ones that form a framework allowing individual prediction of classes is the feature-based classification algorithm. Nowadays, there are numerous successful algorithms which are used for class prediction in many different areas. In particular, *Support Vector Machine* (SVM) classifiers are involved in a large portion of the research studies which have been carried out to distinguish AD patients from healthy subjects for more than a decade [80].

An SVM creates hypersurfaces that can function as both classifier and regression surface in high-dimensional spaces. In [53], Klöppel et al. (2008) employed a linear SVM classifier directly on the initial subset of extracted features, while in many other studies (such as [41, 18, 63, 58, 74, 101]), the researchers firstly apply an intermediate step on raw vectors of data matrix to reduce the number of dimensions in the feature space. This dimensionality is likely to be very high depending on the source

of information (e.g. MRI/fMRI/CAT/PET/SPECT images, numerical or categorical laboratory data, genetic data). Especially in case of the whole-brain MRI analysis, occurrence of the well-known phenomenon "curse of dimensionality" [48] is usually inevitable. Therefore, in many of the studies, including this one, investing an effort for dimensionality reduction has been of high importance.

There are many classification frameworks in which techniques other than SVM were embraced for model training and validation, in the field. For example, *Random Forest* (RF) classifiers were utilized in [42], a multi-class classification was performed based on an RF-based similarity measure. RF classifiers were also employed in [79] for grouping SPECT images. Liu et al. (2014) [56] analyzed multimodal data using *Multiple Kernel Learning* (MKL) approach. *Multiple Kernel*s were also utilized in [106] for diagnosing AD and MCI. In [83], the authors used *Probabilistic Neural Networks* (PNN) to diagnose AD over conventional and wavelet coherence. Additionally, four types of *Artificial Neural Networks* (ANNs), namely, *Backpropagation* (BP), *Radial Basis Networks* (RBF), *Learning Vector Quantization Networks* (LVQ) and PNN were employed in [84] together with VBM-based density features. Finally, there are several recent works which incorporated *Deep Learning* methods for class prediction. Promising applications can be listed as [92], [45], [93], and [88].

This thesis study involves the first time utilization of *Multivariate Adaptive Regression Splines* (MARS) and one of its recent variants *Conic MARS* (CMARS), in field of medical image analysis for computer-assisted diagnosis. Following sections include mathematical backgrounds of these methods, and the chapter ends with a brief discussion regarding justification of our preferences.

2.2 Multivariate Adaptive Regression Splines (MARS)

Multivariate Adaptive Regression Splines (MARS) [38] is a regression algorithm through which linear and nonlinear models can be built. MARS is *nonparametric* (which does not necessarily mean it is *none-parametric*), since the procedure involves no specific assumptions on the form of the distribution or the interdependence between the predictor variables. The algorithm is also *adaptive* since the model buil-

14

ding process which involves creation and pruning of the basis-functions (BFs) depends fully on the input data itself. MARS models are highly *flexible* since they are specified by superposition of hinge functions. The hinge functions can be stated as:

$$(x - \tau)_+ = \begin{cases} x - \tau, & \text{if } x > \tau, \\ 0, & \text{otherwise,} \end{cases} \tag{2.1}$$

and,

$$(\tau - x)_+ = \begin{cases} \tau - x, & \text{if } x < \tau, \\ 0, & \text{otherwise.} \end{cases} \tag{2.2}$$

The pair composed of the two similar looking functions introduced as Eqns. (2.1) and (2.2) is named a *reflected pair*. According to [43]: "The idea is to form *reflected pairs* for each input variable x_j with knots at each observed value x_{ij} of that variable." Consequently, the set of truncated linear functions from which the basis-functions are formed is:

$$\Gamma = \left\{ (x_j - \tau)_+, (\tau - x_j)_+ \mid \tau \in \{x_{1j}, x_{2j}, \ldots, x_{Nj}\}, j = 1, 2, \ldots, p \right\}. \tag{2.3}$$

Regression and classification problems can be handled by linear models. A linear model can be expressed in the form:

$$y = f(\boldsymbol{x}) + \epsilon, \tag{2.4}$$

where y is the response (which provides the class label information for the classification case), $\boldsymbol{x}$ is the vector of predictor variables, ϵ is a random error vector, and $f(\cdot)$ is the model function. MARS builds models of the form:

$$\hat{f}(\boldsymbol{x}) = \beta_0 + \sum_{m=1}^{M} \beta_m B_m(\boldsymbol{x}), \tag{2.5}$$

where M is the number of BFs forming the model function and each $B_m(\boldsymbol{x})$ is a BF formed by single element or by multiplication of two or more such elements that are available in Γ, multiplied by a coefficient β_m. The ith basis-function can be expressed in terms of hinge functions as follows:

$$B_i(\boldsymbol{x}) = \prod_{j=1}^{K_m} \left[s_{\kappa_j^m} \left(x_{\kappa_j^m} - \tau_{\kappa_j^m} \right) \right]_+ . \qquad (2.6)$$

Here, $s_{\kappa_j^m}$ is the sign (+ or −) of the jth truncated linear function forming the mth basis-function. Similarly, $\tau_{\kappa_j^m}$ gives the coordinate of the knot point on the jth predictor variable $x_{\kappa_j^m}$ corresponding to the jth truncated linear function in the mth BF.

A MARS model is built in two stages, namely, the *forward pass* and the *backward pass*. At the beginning, an initial model function with only one term (called the *intercept*) is generated through averaging the response values. After that, basis-functions are inserted in pairs, in an iterative scheme. At each individual iteration, basis-function pair providing the maximum decrease in a particular cost function - such as residual sum-of-squares (RSS) error - is found. Generally, the forward pass ends up with an overfit model. The model needs to be pruned in order to increase generalization ability. This is performed through the second step, the backward pass, by which least effective terms are removed from the model one by one until the remaining submodel is optimized. The contribution rates of basis-functions are computed by employing the *Generalized cross-validation (GCV)* criterion. Degree of any variable's significance is measured by the reduction in the recalculated GCV after it is eliminated from the model.

Equation (2.7) defines the GCV criterion as:

$$GCV = \frac{\sum_{i=1}^{N} (y_i - \hat{f}_\alpha(x_i))^2}{(1 - \tilde{C}(\alpha)/N)^2} . \qquad (2.7)$$

In Eqn. (2.7), N is the number of observations (data); $\hat{f}_\alpha$ is the estimation for the optimal submodel; $\tilde{C}(\alpha) = u_\alpha + dk$, where u_α is the number of linearly independent BFs that are kept in the pruned model, α; k is the number of knots at the end of the forward step; and d is a cost coefficient for optimization - also acting as a smoothing

parameter for the process.

2.2.1   MARS as a Classifier

The MARS method can be "extended to handle classification problems" [43]. When there are only two classes, the labels can be assigned as 0 and 1. The common practice is giving 0 to cognitively normal subjects and 1 to the AD patients. Consequently, the problem can be handled as a standard regression case. Primary drawback of a MARS classifier is that it is constructed by employing a least-squares loss function which does not enable direct interpretation of the class label as random variable with two realizations (0 and 1), at every point. Accordingly, the predictions are not assured to fall into the $[0, 1]$ interval. Therefore, it is not possible to anticipate these predictions as probabilities. Instead of this, they are evaluated as generalized probability scores.

To be able to decide for a label looking at the output function, determination of a threshold value is necessary. Since the boundaries of the output response are not definite and no parametric form for the output response is previously assumed, this threshold cannot be straightforwardly decided.

The widely-used method of [70] named "Automatic optimal threshold selection for image segmentation" was used for specification of that threshold for class separation.

2.3   Conic Multivariate Adaptive Regression Splines (CMARS)

As mentioned in the previous section (Section 2.2), MARS model building process consists of two consecutive steps: the *forward step* and the *backward step*. CMARS (Conic MARS) [103] replaces the backward step with a Tikhonov regularization problem by defining a new *penalized residual sum of squares (PRSS)* for the MARS model obtained at the end of the forward step. For solving this problem, a modern continuous optimization technique, *conic quadratic programming (CQP)* is applied.

Weber et al. [103] defined the PRSS function as:

$$PRSS := \sum_{i=1}^{N} (y_i - f(\tilde{\boldsymbol{x}}_i))^2 +$$

$$\sum_{m=1}^{M_{max}} \lambda_m \sum_{\substack{|\boldsymbol{\alpha}|=1 \\ \boldsymbol{\alpha}=(\alpha_1,\alpha_2)^T}}^{2} \sum_{\substack{r<s \\ r,s\in V_m}} \int \beta_m^2 \left[D_{r,s}^{\boldsymbol{\alpha}} B_m(\boldsymbol{t}^m) \right] d\boldsymbol{t}^m. \tag{2.8}$$

In Eqn. (2.8), V_m represents the variable set associated with the mth BF, $\boldsymbol{t}^m$ is the vector of variables contributing to the mth BF, and D is simply the derivative operator. The first term of the PRSS in Eqn. (2.8) gives the residual sum of squares for each observation, whereas the second term constitutes a measure of complexity. Thus, a trade-off between accuracy and model complexity is established by minimization of the PRSS function with penalty parameters λ_m.

In [103], the PRSS problem is approached as a Tikhonov regularization problem. After some rearrangements involving this approach and further simplifications (such as assuming that all multipliers λ_m are equal (to λ) and the model is differentiable everywhere), PRSS is approximated as stated in Eqn. (2.9):

$$PRSS \approx \left\| \boldsymbol{y} - \boldsymbol{B}(\tilde{\boldsymbol{d}})\boldsymbol{\beta} \right\|_2^2 + \lambda \left\| \boldsymbol{L}\boldsymbol{\beta} \right\|_2^2, \tag{2.9}$$

where $\boldsymbol{L}$ is an $M_{max} \times M_{max}$ diagonal matrix, which has values calculated by summation of L_{im} (see Equation ((2.10))) over each observation on its diagonal entries. Readers may refer to the paper [103] for further details involved in this simplification process. In the paper, L_{im} is expressed as:

$$L_{im} := \left[\left(\sum_{\substack{|\boldsymbol{\alpha}|=1 \\ \boldsymbol{\alpha}=(\alpha_1,\alpha_2)^T}}^{2} \sum_{\substack{r<s \\ r,s\in V_m}} \left[D_{r,s}^{\boldsymbol{\alpha}} B_m(\hat{\boldsymbol{x}}_i^m) \right] \right) \Delta \hat{\boldsymbol{x}}_i^m \right]^{\frac{1}{2}}. \tag{2.10}$$

The Tikhonov regularization problem given in Eqn. (2.9) is expressed through a conic quadratic programming (CQP) problem as introduced in Eqn. (2.11):

18

$$\min_{t,\boldsymbol{\beta}} \quad t$$

$$\text{subject to} \quad \left\| \boldsymbol{y} - \boldsymbol{B}(\tilde{\boldsymbol{d}})\boldsymbol{\beta} \right\|_2 < t, \tag{2.11}$$

$$\|\boldsymbol{L}\boldsymbol{\beta}\|_2 \le \sqrt{\tilde{M}}.$$

Here, the objective function is composed of a single artificial variable t which ranges between 0 and ∞. Two conic constraints associated with the two terms of the Tikhonov problem are defined. Furthermore, $\boldsymbol{B}(\tilde{\boldsymbol{d}})$ is the $(N \times (M_{max} + 1))$ matrix representing the input data operated with the basis-functions, and $\boldsymbol{B}(\tilde{\boldsymbol{d}})\boldsymbol{\beta}$ stands for the input data operated with the model (i.e., the output obtained by the model)[1].

To summarize, MARS is an adaptive and non-parametric extension of decision trees (especially, of the method CART), which is - in our case - elaborated for constructing nonlinear models functioning as class-separating hyperplanes. A MARS model is basically a superposition of truncated linear functions (or combinations of them) which are allowed to contribute the model only in certain intervals (defined by the knot points), with specific directions ($+$ or $-$), slopes (β_m), and the intercept term (β_0) determined by the forward algorithm.

CMARS replaces the backward step of classical MARS (in which the model is pruned by elimination of least effective basis-functions to decrease the overall complexity) with a process involving minimization of a PRSS function. It is accomplished by updating all of the unknown coefficients (β_m) instead of removing basis functions from the model, selectively. This is done by first expressing the PRSS as a Tikhonov regularization problem, and then transforming the domain into an optimization problem which is known to be solvable by the help of CQP. In this way, all of the relevant information learned from the observations is preserved, while overlearning from the data is avoided.

2.4 Why MARS?

In [43], a comparison study including 5 of the "off-the-shelf" data mining procedures (namely, Neural Networks, SVM, CART, k-Nearest Neighbor, and MARS) was pre-

---

[1] Here, $\tilde{M}$ should not be confused with number of basis functions, M.

sented. As a result of this study, MARS appeared to provide preferable support for **(i)** handling mixed-type data and missing data points, **(ii)** computational scalability, **(iii)** dealing with irrelevant inputs, and **(iv)** interpretability compared to support vector machines, which takes place as the most widely-used method for AD detection in the literature.

In [91], it was expressed that MARS has the ability to make quicker predictions than SVM, in which "every variable has to be multiplied by the corresponding element of every support vector". Moreover, a particular dominance of MARS lies in its capability to predict the contributions of some BFs that are allowed to influence the output value utilizing the additive and interactive effects of input variables [103]. Since conclusive anatomical information of very early effects of AD on brain structures is not present and, at the beginning, all voxel intensities in the brain MRI are assumed to be potential predictor variables, this facility of MARS - which is lacked by SVM - is of great importance for this thesis study.

MARS was noted to be computationally more efficient than the *Artificial Neural Networks* (ANN), which is another broadly employed technique in the field [108]. One disadvantage of ANN is that it generates models in structure of a "black box". This was explained in [37] as follows: "The functions fit by neural networks are difficult for the analyst to understand and difficult to explain to management. One of the very useful features of MARS is that it produces a regression like function which can be used to understand and explain the model." Both MARS and ANN are effective in modeling the nonlinearities introduced by interacting variables. Additionally, in [64], usage of MARS was addressed to be "tremendously advantageous over using *Linear Models* (LM), *Generalized Additive Models* (GAM), and CART for prediction".

Taking these aspects in consideration, we have decided on MARS - and CMARS, a mathematically more unified modification of the algorithmic section of MARS, including fewer heuristic components, and in accordance with modern optimization theory - as the method to be embraced for classification.

CHAPTER 3

DATA SPECIFICATION, PREPROCESSING AND FEATURE EXTRACTION FROM BRAIN MRI IMAGES

3.1 Study Data

3.1.1 ADNI Image Archive

Alzheimer's Disease Neuroimaging Initiative (ADNI) is a long-running study which involves observation of healthy elders, Alzheimer's disease and mild cognitive impairment (MCI) in multiple centers. In scope of the study, magnetic resonance imaging (MRI), (18F)-Fludeoxyglucose positron emission tomography (FDG PET), urine serum, and cerebrospinal fluid (CSF) biomarkers are acquired together with the clinical and neuropsychological follow-up data at certain intervals, from each participant [46].

ADNI was founded in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a 5-year public/private partnership with a budget of $60 million. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical/psychometric assessments can be put together to quantify the progression of MCI and early onset of AD. Specification of sensitive and specific biomarkers of AD existence in very early phases is intended to help researchers and physicians to improve treatments and follow-up their efficiency, as well as save time and money spent for clinical trials.

In this thesis, brain MRI datasets used for training, validation, and blind testing were

gathered from the ADNI image archive[1].

3.1.2 Image Subsets: Brain MRI of Normal, AD, and MCI Subjects for Training and Testing

In [23], Cuingnet et al. assessed the performances of 10 distinct procedures proposed by various researchers (5 methods based on voxel intensities, 3 techniques based on dilution in the cortex tissue, and 2 procedures grounded on the features derived from the hippocampus). The authors utilized MRI data belonging to 509 participants from the ADNI archive. Three 2-class classification cases were realized: **(i)** CN vs. AD, **(ii)** CN vs. MCIc (patients who converted from MCI to AD in 18 months; i.e., converting MCIs), and **(iii)** MCIc vs. MCInc (MCI patients who did not convert to AD in 18 months; i.e., non-converting MCIs). The researchers used T1-W (T1-weighted) MRI volumes as the data source for the experiments. MRI acquisition was performed in accordance with the ADNI protocol introduced in [46]. For each subject, when available, the baseline MR image was chosen. Otherwise, MRI scan was selected from the screening visit.

General criteria for inclusion and exclusion of the subjects were reported as follows: "cognitively normal" participants of the control group (CN) had mental state examination (MSE) scores in the interval of 24-30, and clinical dementia rate (CDR) of zero. They also did not have MCI, dementia, or depression. MCI patients again had MSE scores between 24 and 30, memory complaint along with a coexisting objective memory loss (specified by Wechsler Memory Scale), CDR of 0.5 and preserved activities of daily living without dementia. AD patients had MSE scores varying between 20 and 26, CDR of 0.5 or 1.0, and had probable AD according to NINCDS-ADRDA criteria (see [61]).

Cuingnet et al. [23] explained the preprocessing procedure of ADNI, more or less, as summarized in the following paragraph:

To support standardization among the centers and imaging platforms used in the ADNI study, image preprocessing operations for correction of several common ima-

---

[1] http://adni.loni.usc.edu. [Accessed: 19- Nov- 2016] [3].

Table 3.1: (Adapted from [23]) Demographic information regarding the subjects (CN: Cognitively normal, AD: Alzheimer's disease, MCIc: MCI converters, MCInc: MCI non-converters). Age and MSE score statistics are given in the format of "average $\pm$ st. dev. [min-max]". Number of participants in each group and the number of centers where the images were acquired are presented in columns indicated by "No." and "Cen.", respectively.

Diagnosis	No.	Age	Gender	MSE	Cen.
CN (Train)	81	76.1±5.6 [60-89]	38 M/43 F	29.2±1.0 [25-30]	35
AD (Train)	69	75.8±7.5 [55-89]	34 M/35 F	23.3±1.9 [18-26]	32
MCIc (Train)	39	74.7±7.8 [55-88]	22M/17F	26.0±1.8 [23-30]	21
MCInc (Train)	67	74.3±7.3 [58-87]	42M/25F	27.1±1.8 [24-30]	30
CN (Test)	81	76.5±5.2 [63-90]	38 M/43 F	29.2±0.9 [26-30]	35
AD (Test)*	67	76.0±7.1 [57-91]	32 M/35 F	23.2±2.1 [20-27]	33
MCIc (Test)	37	74.9±7.0 [57-87]	21 M/16 F	26.9±1.8 [24-30]	24
MCInc (Test)	67	74.7±7.3 [58-88]	42 M/25 F	27.3±1.7 [24-30]	31

ging artifacts were carried out. These preprocessing steps include geometrical corrections for nonlinearity in the image gradient and elimination of magnetic field intensity non-uniformities due to non-uniform receiver coil sensitivity. Both of these procedures for refinement can directly be managed from the MRI console. All participants were scanned two times at each visit. MR scans were scored qualitatively by the ADNI researchers for artifacts and overall acquisition quality. Each image was graded considering multiple criteria, including ghosting, homogeneity, blurring, intensity, flow artifact, etc. For every subject, the MRI volume which was recognized as the "best" quality scan by the ADNI researchers was used in the database. Here, "best" is described as the one which was chosen for the whole preprocessing procedure in ADNI methods web page[2].

In this thesis, it was decided to study on the exactly equivalent sets of images as used in [23] for model training and validation. This decision is made keeping the objectivity of the comparative analysis in mind. General information regarding the demographics of selected participants is introduced in Table 3.1. All of the experiment groups were formed by employing equivalent MRI volumes except for a single scan corresponding to an AD patient in the test group (which is pointed by "*" in Table 3.1). Since it was recognized that the current version (v12) of the SPM (Statistical

---

[2] `http://adni.loni.usc.edu/methods`. [Accessed: 19- Nov- 2016] [3].

23

Parametric Mapping) Toolbox lacks to provide valid results for this particular image at the end of the tissue segmentation phase (please refer to Subsection 3.2.1 for details of the segmentation procedure), the image is eliminated and the population statistics are recalculated.

Identifiers of the images which are used in this study are provided by the authors as supplementary to the original article (see Tables S1-S8 provided with the online copy of [23]).

3.2 Feature Extraction from Brain MRI Volumes by Voxel-Based Morphometry

In [8], Ashburner and Friston defined *voxel-based morphometry* (VBM) as a methodology for analyzing voxel-wise correspondence of the local concentrations of brain tissues between different groups of subjects. These groups can be composed of a sequence of images acquired from different subjects, or a time-series belonging to a particular subject. VBM involves brain tissue segmentation, spatial normalization and smoothing procedures applied on volumetric medical images with high-resolution.

The term *statistical parametric mapping* (SPM) was first used by Friston (1991) in [40] referring "construction of spatially extended statistical processes to test hypotheses about regionally specific effects". SPM software package[3] was designed for both standardizing the VBM procedures and handling the analysis of statistical parametric maps that are formed through VBM under the assumption of certain statistical properties.

Fig. 3.1 shows a block diagram representation of sub-procedures involved in VBM. In this section, a summary of the VBM methods that are employed in scope of this thesis study is introduced.

---

[3] `http://www.fil.ion.ucl.ac.uk/spm/`. [Accessed: 13- Aug- 2017] [35].
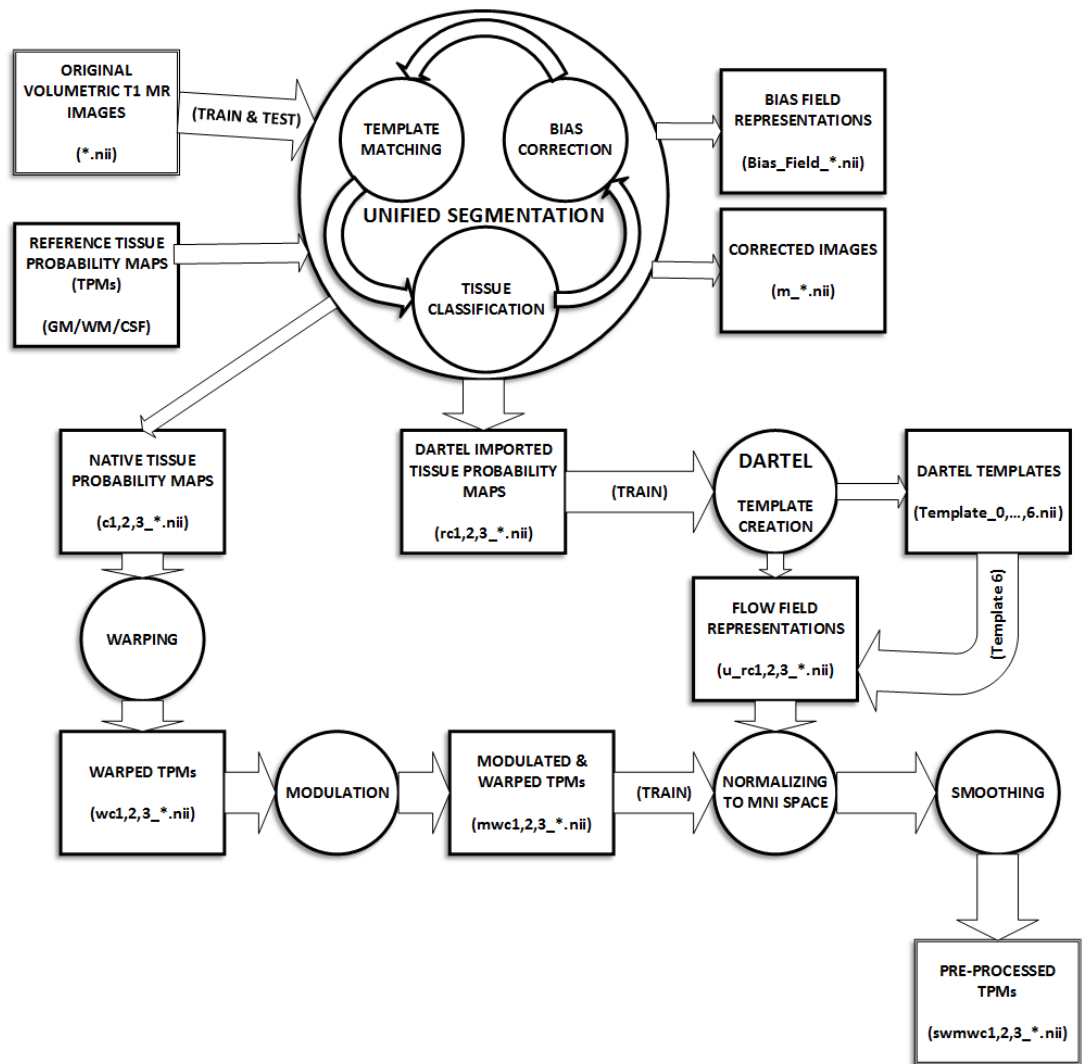
Figure 3.1: Flow diagram: Application of VBM procedures through SPM software pack.

3.2.1 Segmentation of Brain MRI into GM, WM, and CSF through the Method "Unified Segmentation"

The techniques used to segment brain images into integral parts (in our case, tissue classes/TCs) are based on two main approaches. The first is the classification of voxels by considering their probability of belonging to specific tissues, and the second is the direct alignment of images to predetermined templates. *Unified Segmentation* ([9]) is described as a probabilistic framework through which the tissue classification and image registration processes are put together within the same generative model. The algorithm is designed in an iterative scheme. Each iteration also covers bias removal phase, which is a crucial concern in analysis of MRI images. In [9], it is indicated that segmentation of brain tissues by estimating the input parameters of *Unified Segmentation* model results in more precise performance outcomes than applying each operation seperately and consecutively.

The objective function is derived from a *Mixture of Gaussians (MoG)* model and it is minimized through parameter optimization. Eqn. (3.1) represents the standard MoG model. Total probability distribution function can be modeled employing a superposition of K weighted Gaussians. The Gaussian with index k is modeled by its mean (μ_k), variance (σ_k^2), and mixing proportion (γ_k), where $\gamma_k \geq 0$ ($k = 1, 2, \ldots, K$) and $\sum_{k=1}^{K} \gamma_k = 1$:

$$P(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\gamma}) = \prod_{i=1}^{I} \left(\sum_{k=1}^{K} \frac{\gamma_k}{(2\pi\sigma_k^2)^{1/2}} \exp\left(-\frac{(y_i - \mu_k)^2}{2\sigma_k^2} \right) \right). \tag{3.1}$$

Following the inclusion of bias field (involving components corresponding to additive noise and scaling) and the information originating from the priors (TC templates), and taking the two-sided logarithm, the cost function (for minimization) becomes:

$$-\log P(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha}) = -\sum_{i=1}^{I} \log \left(\frac{\rho_i(\boldsymbol{\beta})}{\sum_{k=1}^{K} \gamma_k b_{i,k}(\boldsymbol{\alpha})} \sum_{k=1}^{K} \gamma_k b_{i,k}(\boldsymbol{\alpha}) (2\pi\sigma_k^2)^{-\frac{1}{2}} \right.$$
$$\left. \times \exp\left(-\frac{(\rho_i(\boldsymbol{\beta})y_i - \mu_k)^2}{2\sigma_k^2} \right) \right),$$
$$\tag{3.2}$$
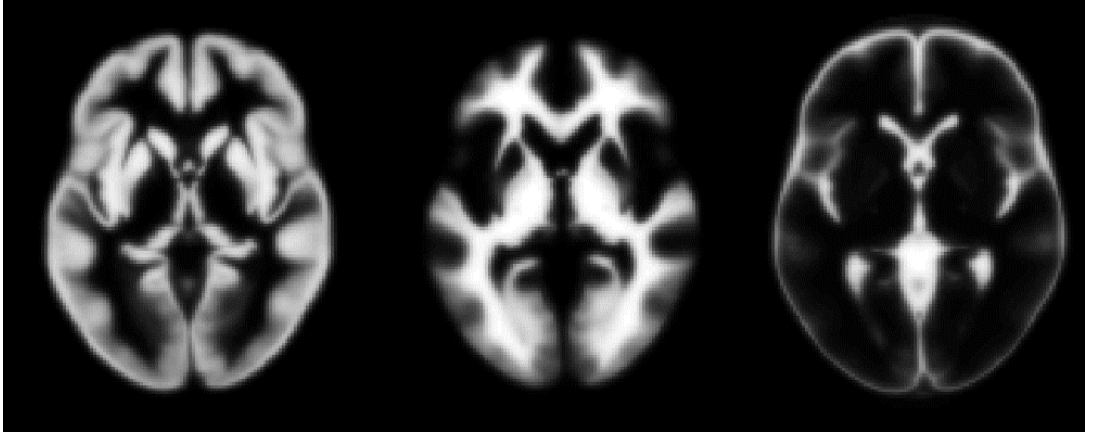
Figure 3.2: ICBM Tissue Probabilistic Atlases for GM (left), WM (middle), and CSF (right) [60].

where $\rho_i(\boldsymbol{\beta})$ states the bias field with parameter vector $\boldsymbol{\beta}$, and $b_{i,k}(\boldsymbol{\alpha})$ stands for the prior spatial information with parameter vector $\boldsymbol{\alpha}$ coming from TC templates $(i = 1, 2, ..., I; k = 1, 2, ..., K)$. Since there is a monotonical relation between the two functions, maximization of the probabilities included in the MoG (Eqn. (3.1)) is realized at the minimum value of right-hand side of Eqn. (3.2) (with respect to $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, and $\boldsymbol{\alpha}$). The reader is referred to the original work of [9] for further reading on details of the procedure and derivation of the objective function in Eqn. (3.2). An *Expectation-Maximization (EM)* based algorithm is utilized for solution of the optimization problem. The algorithm operates in a way that in the E-step the tissue probabilities are calculated, and in the M-step the cluster and the non-uniformity parameters are computed, iteratively [36].

Implementation of *Unified Segmentation* in SPM12 Toolbox [7] was utilized for obtaining native space aligned *tissue probability maps (TPMs)*. Spatial prior information related to the three main TCs was provided by ICBM (The International Consortium for Brain Mapping) Tissue Probabilistic Atlas[4]. 2D axial cross-sections of the ICBM prior templates are introduced in Figure 3.2.

The numbers of Gaussians for each TC were specified as 3 for GM, 2 for WM and

---

[4] All 452 ICBM subject T1-W scans were aligned with the atlas, corrected, and classified into GM, WM, and CSF. Each was averaged to create the TPs for each tissue type. These fields represent the likelihood of finding GM, WM, or CSF at a specified position for a subject that has been aligned to the space. (`http://www.loni.usc.edu/atlases/Atlas_Methods.php?atlas_id=7`. [Accessed: 19- Nov- 2016] [69].)

2 for CSF regarding the values reported in [9] as "typical". The procedure for optimizing the number of Gaussian terms per class was remarked to be an "issue of *model-order selection*" and was not descended to particulars in the original study. It is verified that typical K values that are used in this study are identical to the ones which are implemented as default model parameters presented by multiple versions of the SPM Toolbox, by cross checking.

An example volumetric brain MRI of a normal patient and the three tissue probability maps acquired at the end of *Unified Segmentation* procedure are presented in Figure 3.3. In Figure 3.3a a cross-sectional representation of the original volume is introduced; whereas, Figures 3.3b, 3.3c, and 3.3d show the probability distribution regarding GM, WM and CSF tissues, respectively.

3.2.2   Nonlinear Image Registration by DARTEL

Images acquired from distinct participants at different times windows should be registered and warped in order to make the data available for high quality analysis with sharpened localization, and, consequently, with greater sensitivity. For this reason, the *DARTEL (Diffeomorphic Anatomical Registration through Exponentiated Lie Algebra)* utility of SPM, which was built based on the proposal by [6], was employed.

Diffeomorphism is defined as a global, differentiable mapping between voxels, which is one-to-one, continuous, and invertible by means of having a nonzero Jacobian determinant. Therefore, modeling image transformations by considering diffeomorphism establishes preservation of desirable topological characteristics. DARTEL aims to incorporate diffeomorphism to keep the voxel-level, significant information, while removing larger-scale, anatomical differences between brain MR images of subjects.

First step of the process involves generating the DARTEL templates. This is accomplished through a procedure consisting of 6 main iterations with 3 inner loops per each. The main iteration cycle involves warping the images that form the training set to match the existing template, and taking the average of the warped images to update the template. The initial template is the mean of the starting data. Each inner loop
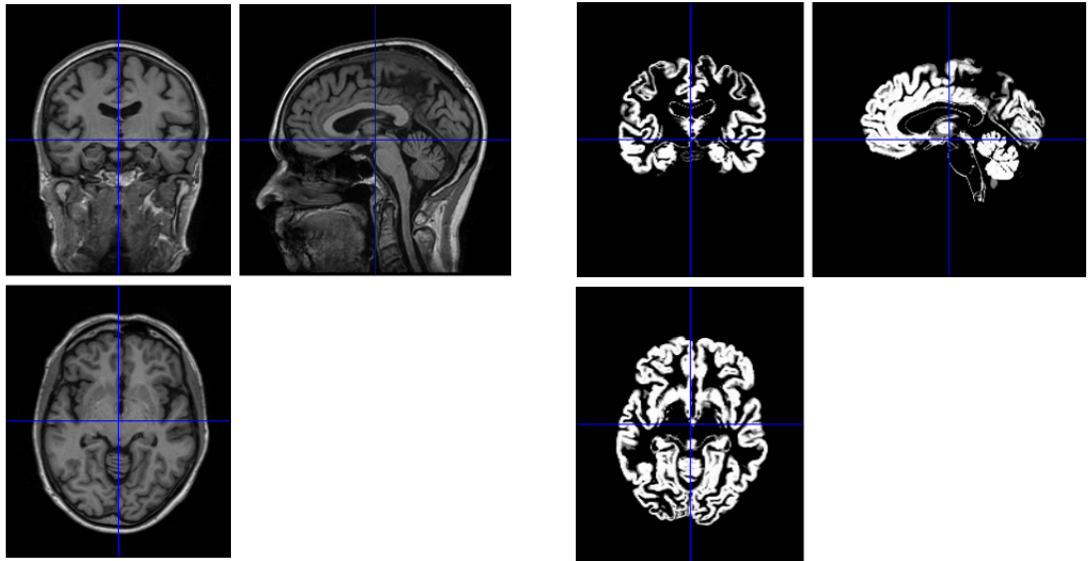
(a) Original MRI volume.

(b) Gray matter (GM).
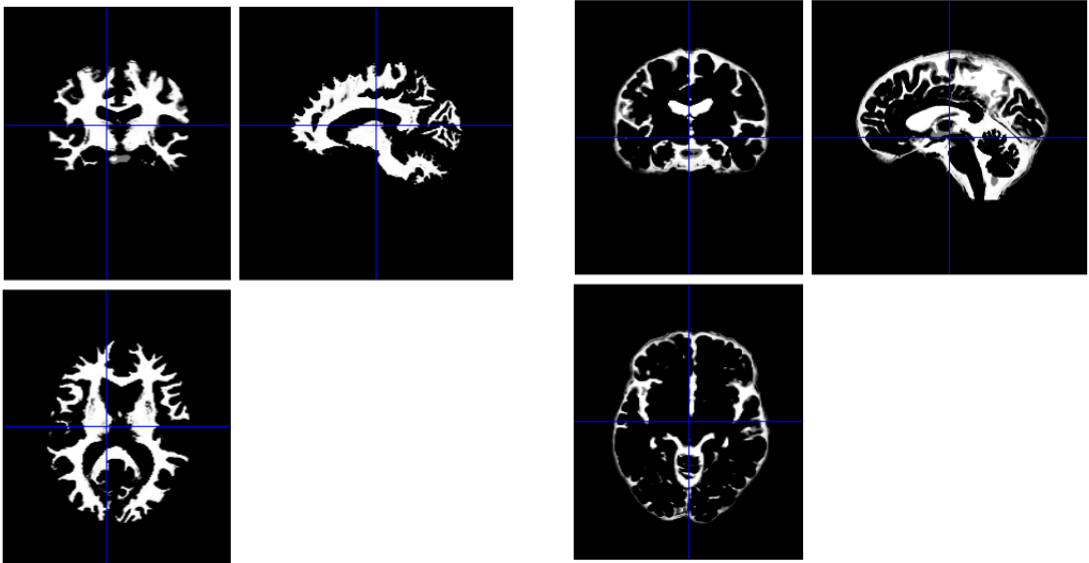
(c) White matter (WM).

(d) Cerebrospinal fluid (CSF).

Figure 3.3: Original MRI volume and the three tissue probability maps acquired at the end of *Unified Segmentation* procedure.
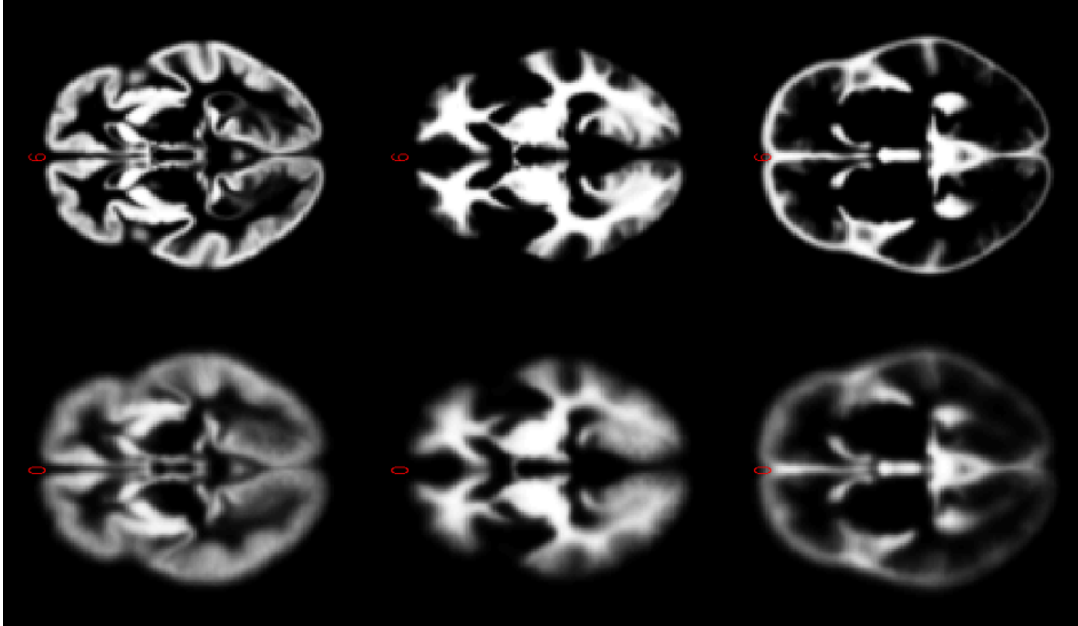
Figure 3.4: The axial cross-sections of volumetric DARTEL Templates at the beginning (lower row), and at the end of the 6th iteration (upper row). Columns represent different tissue classes, namely, GM, WM and, CSF from left to right.
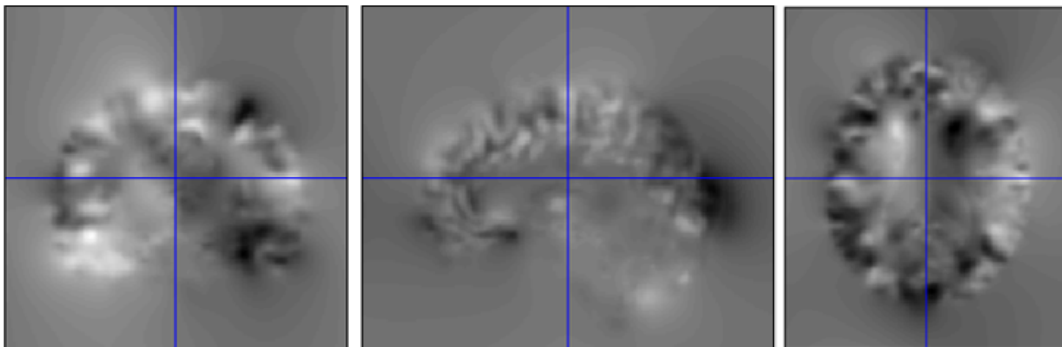
Figure 3.5: Cross-section of a gray matter DARTEL flow field.

corresponds to a Gauss-Newton iteration through which the deformations are updated within the main loop. The entire procedure begins with a coarse registration and expands towards a comparably more conclusive and stable state with finer structure and decreased regularization (see Figure 3.4), in iterations. The deformations were parametrized by and saved in the form of flow-fields.

Secondly, the images assigned for blind tests were nonlinearly registered, matching to the templates created in previous step, individually. In this step, it is of great significance to avoid the images of the test set from sharing information with each other for the sake of *objectivity* and the *extensibility* of the whole procedure. To be more specific, the flow-fields required for registering the test images were calculated employing the templates obtained by using the training set. If new templates derived from the blind test images were created, the warping operation would have resulted in errors within overall performance assessment. In addition, it would not be possible to achieve the same result for a given image when tackled with other test image groups or alone. Moreover, in order to add more images into the test set, one would have to repeat the process from start to end, causing changes to occur in other images.

Finally, the training and test images were normalized to the MNI-space[5], taking into account the issues raised in the previous paragraph (see Figure 3.6). Here, dimensions of all of the MRI volumes are standardized to $121 \times 145 \times 121$ with an isotropic resolution of $1.5mm \times 1.5mm \times 1.5mm$, as MNI suggests. Until the nonlinear warping is applied, the spatially normalized images are still group-wise unregistered. The final template in Figure 3.4, and the flow-fields (cf. Figure 3.5), created in the former step, were used for this purpose.

3.2.3   Final Steps: Intensity Modulation and Smoothing

Match Filter Theorem [96] states that the *signal-to-noise ratio* (SNR) is improved by convolving a smoothing filter of adequate size with the signal. The filter width is determined with the expected signal width, which is 5 to 10 times the smallest dimension of a voxel in case of volumetric images. The images were smoothed with a Gaussian smoothing kernel of size (full width at half maximum) 8 mm. Smoothing

---

[5] Standard brain space defined by *Montreal Neurological Institute* (MNI).
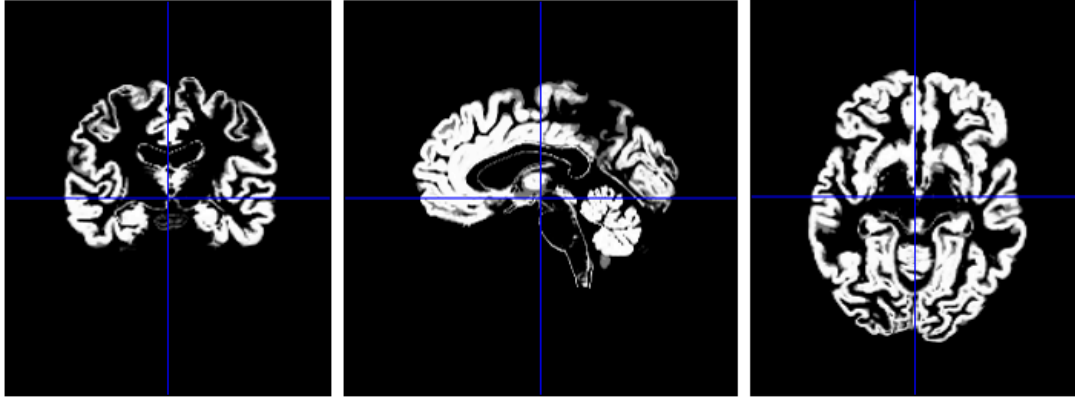
(a) Gray matter tissue probability map at the end of *Unified Segmentation* step.
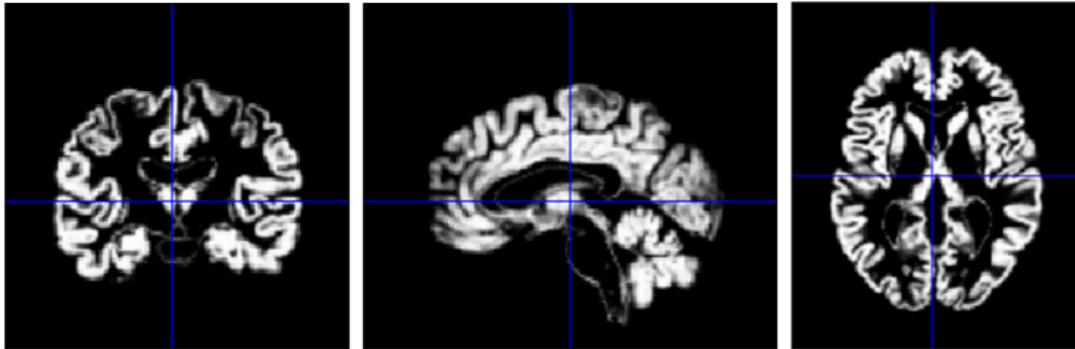
(b) Gray matter tissue probability map after MNI-space normalization applied.

Figure 3.6: Spatial normalization to the MNI-space.

was also performed in order to enhance the validity of statistical implications and to decrease the variation occurring both internally and between individual images. The reader is referred to [51] for a remarkably well workup on the effects of the filter size on VBM analyses.

Registering a set of images to match a fixed image results in alterations regarding the proportions of different tissue volumes. This adverse impact due to spatial normalization can be overcome by the application of intensity modulation through which the voxel intensity values are updated to compensate the aforementioned volume change. Therefore, as the final step, an intensity modulation operation regarding all three tissue probability maps was performed.

Entire image processing procedure was performed using a personal computer with 4-core, 2,40 GHz, Intel processing unit, 16 GB DDR3 RAM, SSD hard disk, and 64-bit Windows operating system. The highest portion of the total time-cost was consumed by the nonlinear registration of the training images. For example, for the AD/CN training dataset of 150 images, it took approximately 9.5 (of total processing time of 12) hours to create DARTEL templates and flow fields, using SPM12 package in MATLAB (r2013b). Additionally, the cross-validation procedure described in Section 6.1 took 3 hours 20 minutes of time for optimization of the MARS model parameters. However, only the training phase is subject to this kind of excessive time-costs. On the other hand, the time necessary for a single test image to be prepared for class prediction is never more than 20 minutes. This makes the proposed system highly practicable in the clinics for testing new images, in case of availability of an already trained model function and DARTEL templates derived from the training dataset. Nonetheless, it will require hours - again - to expand the training dataset with additional images, since it involves repetition of the warping process with inclusion of the whole training set.

3.2.4 Resulting Baseline Feature Space: Normalized Tissue Probability Maps

As a result of applying VBM procedures using the SPM software, a data matrix with a noteworthy large set of features, representing the images in the form of tissue probability maps, was obtained.
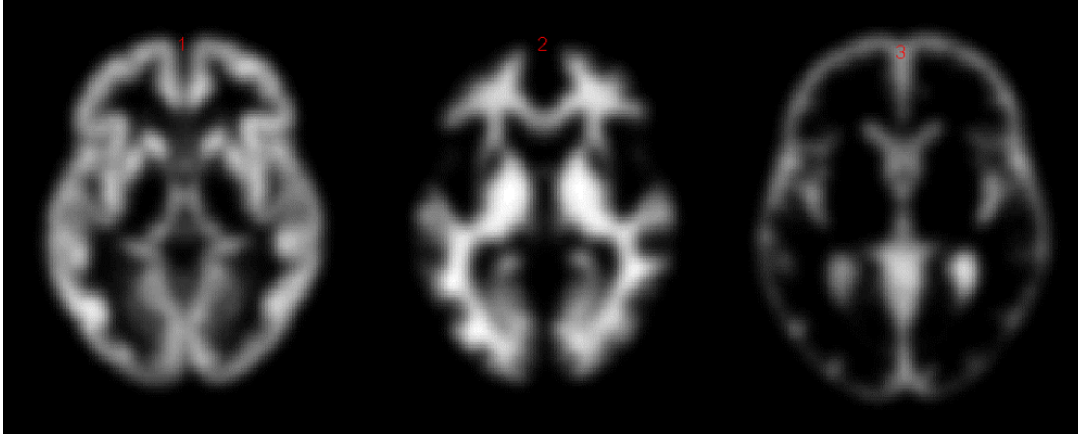
Figure 3.7: Final (segmented, spatially normalized, warped, modulated, and smoothed) version of tissue probability maps (axial view), GM (left), WM (middle), and CSF (right), available for classification.

With spatially normalizing to the MNI atlas, dimensions of all of the MRI volumes were standardized to $121 \times 145 \times 121 = 2,122,945$. Since the images were segmented into three brain tissues, this number is tripled in the data matrix. In other words, we end up with a data matrix with a size of 6,368,835 in columns. Considering the AD/CN classification case, the training dataset includes 150 images, each one of which is described by a vector of size 6,368,835, storing 8-byte double precision floating-point numbers in the entries. This means that the total size of the training data matrix is greater than 7 GB (not smaller than 5 GB even after the elimination of zero-voxels). These figures are emphasizing the necessity of utilization of an effective dimensionality reduction procedure.

3.2.5 Data Quality

At this point, the data quality is checked **(i)** by observing sample homogeneity with visual representation of sorted covariance matrices, **(ii)** through visual inspection. It is observed that no serious image artifacts occur (no outliers) in the training datasets, and only one problematic image in the test set containing subjects with AD exists. In order to eliminate the problem, this image -which failed to be segmented properly by SPM Unified Segmentation- is removed from the "AD (Test)" dataset (see Table 3.1).

3.3 Alternative Techniques for Extracting Feature Descriptors

In addition to the tissue probability maps acquired through VBM, several other feature extraction procedures were applied and the corresponding results were investigated. The first one of these methods is based on the image moments, and involves a scale-space analysis handled in different voxel sizes. The others analyzed in this context are two widely known techniques, used for image recognition and object detection, falling into the group of modern computer vision algorithms, namely, *scale-invariant feature transform* (SIFT) and *histograms of oriented gradients* (HOG).

In this section, the background information regarding those methods and their prior applications in the related literature are presented. Results acquired by different feature description approaches are introduced - with a qualitative discussion - in Chapter 6.

3.3.1 Image Moments by Block Processing

The feature space was enriched by deriving a new set of features, and the quality of the newly derived features was evaluated. For this purpose, the concept of *central moments* was employed. Normalized central moments up to the 4$^{\text{th}}$ degree were computed for blocks of varying dimensions (with different scales).

Normalized central moment (of degree k) of random variable x is expressed mathematically as:

$$\mu_k = \frac{E[(x - \mu)^k]}{\sigma^k}, \tag{3.3}$$

where, $E[\cdot]$ is the expected value operator, μ is the sample mean and σ is the standard deviation defined as:

$$\sigma = \sqrt{E[(x - \mu)^2]}. \tag{3.4}$$

The first and second normalized moments are equal to the *sample mean* and *vari-*

ance, respectively. The third moment is called *skewness* and it can be considered as a measure of the asymmetry characteristics of a distribution. The fourth moment is called *kurtosis*, which is a shape descriptor of a probability distribution similar to the *skewness*.

Updated dimensions were specified by downsampling the images in all 3 coordinates with increasing powers of 2, in an analogous fashion with scale-space analysis. More clearly, when the scale (downsampling factor) is equal to 0, each image dimension is divided by $2^0(=1)$, which means that the new blocks are identical with the original voxels. For scale 2, new dimensions are one-fourth of the originals (original image dimensions are 121-145-121, the MNI standard size), which means that each block is approximately composed of $4^3(=64)$ voxels. For scale 3, dimensions are divided by $2^3(=8)$, i.e., the new blocks are composed of $8^3(=512)$ voxels. Each central moment value (*mean*, *variance*, *skewness*, and *kurtosis*) corresponding to newly created blocks were computed regarding the original voxels composing the block for scale $s \in \{2, 3, 4\}$. Each computation was performed on probabilities belonging to all three tissue classes, namely, GM, WM, and CSF.

3.3.2 Scale-Invariant Feature Transform (SIFT)

In [57], Lowe (2004) proposed a method for extracting salient local features which are invariant to rotation and scale from images. These features can be employed for performing an association between image parts corresponding to alternate views of an object or scene. The method is called *scale-invariant feature transform* (SIFT), and it provides a set of highly distinctive features which are also invariant under different illumination levels and resistant to noise.

The algorithm consists of 4 steps (the reader may refer to [57] for further details):

Scale-space extrema detection: A *Difference-of-Gaussians* (DoG) function is utilized in order to identify potential points of interest in the image. These points are invariant under varying scales and orientations. They represent candidates for salient feature locations called *keypoints*.

In order to specify the scale-space minima and maxima points, the image is

convolved with Gaussians of different variances and grouped by octaves (an octave is formed by multiplying the standard deviation of a Gaussian function by two). Once the scale-space is obtained, the extremal points are specified as local minima/maxima of the images across varying scales. This is accomplished by comparing the value of every pixel to its neighbors within the same scale and neighboring pixels in each of the consecutive scales. If the intensity value is an extremum across these pixels, a candidate keypoint is identified.

Keypoint localization: The first step provides a high number of candidate keypoints, many of which are not stable enough to be specified as a keypoint. For this reason, a detailed fit to the nearby data is performed for location, scale, and ratio of principal curvatures. The algorithm calculates the interpolated location of the extrema by employing the quadratic Taylor expansion of the DoG function, and this approach improves matching and stability substantially.

Using the second order Taylor expansion and assigning a threshold to its value, some of the candidate keypoints are discarded. In this way, low-contrast keypoints were eliminated.

This procedure contains one more step to eliminate edge responses which are originated from non-robust candidates. For this purpose, principal curvature amounts are found through solving eigenvalues of the second-order Hessian matrix and defining a threshold for the ratio of two eigenvalues.

Orientation assignment: Based on the directions of local image gradients, single or multiple orientations are assigned to each keypoint location. A histogram of orientations is formed from the gradient orientations of points included in a region around the keypoint. A 360 degree range of orientations is covered by 36 bins of the histogram. Samples included in the histogram are multiplied by the magnitude of their gradient and by a circular Gaussian with a standard deviation which is 1.5 times the keypoint scale. Peaks that are within 80% of the highest peaks of the histogram are specified as dominant orientations.

Keypoint descriptor: For every keypoint, a descriptor vector is created. These descriptors are remarkably distinctive and invariant to dissimilarities in illumination, 3D viewpoint, scale, and noise level. For this purpose, first, a set of orien-

tation histograms is composed. Then, the magnitudes are further convolved by a Gaussian with σ equal to 50% of the descriptor window width. Finally, unit normalization is applied on the vector in order to lower its variance to affine illumination changes.

SIFT has been employed as a state-of-the art feature extraction technique in the areas of object recognition [90], robot localization and mapping [87], panorama stitching [12], and human action recognition [34, 86].

SIFT algorithm has also been utilized for the analysis of brain MRI with Alzheimer's disease in several studies. In [95], Toews et al. (2010) proposed *Feature-based morphometry* (FBM), specifically "to avoid the assumption of one-to-one inter-subject correspondence." The method consists of two major parts: extraction of scale-invariant features by SIFT and probabilistic modeling. SVM method is used for classification. Through the evaluation of proposed procedure on OASIS[6] database, the authors reached an equal error rate (EER) of 0.80 for patients of age between 60 and 80 with CDR[7]=1 (Group-A), and EER=0.71 for patients of age between 60 and 80 having CDR scores of 1 and 0.5 (Group-C).

In [25], Daliri (2012) presented an automated approach to be used for diagnosing Alzheimer disease from the whole brain MR images. Proposed method involves detecting scale and rotation invariant SIFT features and transforming them into 100-bin histogram vectors using *Bag of (Visual) Words* (BoW) approach. By reducing the number of features through *Fisher's discriminant ratio* (FDR), the algorithm improves the results obtained by Toews et al. (2010): EER = 0.86 for Group-A, and EER=0.75 for Group-C.

Another example to this group can be given as the work of Chen et al. (2014) [20], in which SVM models were successfully trained by SIFT features derived from 2D scans to detect AD in early phases.

---

[6] The Open Access Series of Imaging Studies (OASIS) is a project aimed at making MRI data sets of the brain freely available to the scientific community. (`http://www.oasis-brains.org/`. [Accessed: 16-July- 2017] [16].)

[7] The CDR is a 5-point scale used to characterize six domains of cognitive and functional performance applicable to Alzheimer's disease and dementias: Memory, Orientation, Judgment and Problem Solving, Community Affairs, Home and Hobbies, and Personal Care. According to this scale, a score of 0.5 represents "very mild dementia" and a score of 1 represents "mild dementia". (`http://alzheimer.wustl.edu/cdr/cdr.htm`. [Accessed: 16- July- 2017] [68].)

3.3.3 Histogram of Oriented Gradients (HOG)

Histogram of oriented gradients (HOG) is a feature description algorithm -like SIFT- which is applicable for object detection and image matching in the field of computer vision. Although the basic principals date back to a patent application in 1986, the algorithm became widespread when Dalal and Triggs presented their comprehensive work on pedestrian detection in 2005 [24] at the Conference on Computer Vision and Pattern Recognition (CVPR).

The algorithm is implemented in four steps as follows:

Computation of gradients: Gradients are computed by applying 1-D centered, point discrete derivative mask (or Sobel mask) in vertical and horizontal directions.

Orientation binning: Images are divided into smaller connected regions called *cells*. All cells are discretized into angular histogram channels according to the weighted vote that each voxel contributes.

Descriptor blocks: Regarding their sensitivity to illumination and contrast variations, the gradient magnitudes should be normalized. For this purpose, adjacent cells are combined to form spatial groups called *blocks*.

Block normalization: Gradient strengths are locally normalized to 1, with respect to L_2- or L_1-norm.

Besides the numerous applications of the method in detection of many kinds of objects, there exist several research works involving detection of Alzheimer's disease biomarkers in medical images. In [97], a method that uses HOG to search and retrieve MRI images with diagnosis of certain dementias is proposed. The method provides 70.7% sensitivity and 77.9% specificity in AD diagnosis for patients of age between 60 and 80 with CDR=1. Also, it is reported that the algorithm performance does not effectively depend on the database size.

In [15], status of β-amyloid burden in patients with probable AD is classified by extracting HOG descriptors from PET images instead of quantitative assessment, and classification accuracy beyond 90% is stated.

Unay and Ekin (2011) [97] used HOG descriptors to diagnose dementia in early phases of occurrence. Their algorithm provided an accuracy of 0.74.

The paper [67] presents a study on detecting morphological abnormalities caused by Alzheimer's disease using *speeded up robust features* (SURF) and HOG descriptors. Although good results are obtained in terms of specificity, the study suffers from very poor sensitivity outcome (best level reached among experiments is found to be 68%).

CHAPTER 4

A 3-STEP, HYBRID PROCEDURE TO INCREASE CLASSIFICATION ACCURACY BY REDUCING DIMENSIONALITY OF THE FEATURE SET

4.1 Dimensionality Reduction in Medical Image Analysis

In the previous chapter, data acquisition and image processing steps starting from the raw MR images and ending with the data matrix are explained in detail. The fact that the number of predictor variables is huge when compared to the number of samples directly suggests a well-recognized phenomenon called *curse of dimensionality*. In [48], Jain defined this phenomenon as follows: "a classifier's predictive accuracy does not have to increase with increasing number of predictor variables (or features) unless the class-conditional densities are completely known (i.e., all possible observations are made)."

In most of the neuroimaging studies, the number of samples is under 1,000, whereas the number of non-zero voxels (the ones that indicate brain tissue, therefore, having an intensity value other than zero after the skull stripping process) is in the order of hundred-thousands [65]. Either the situation in our study is not contradicting with this fact reported by Mwangi et al. (2014), where we ended up the feature extraction procedure with 100-150 training samples that are described by feature vectors containing more than 1,000,000 non-zero voxel intensities. As a result, the features with low relevance require to be removed before the model building process in order to have a model with higher generalization ability, and safe from the problem of over-fitting (i.e., over-learning or memorization). In addition to this, considering the high computational cost (time, memory) induced due to the aforementioned dimensionality, it is decided that utilization of a procedure for feature selection previous to the

model building phase is of great importance to us in this study.

One way to reduce the number of dimensionality of the feature space is to employ linear or nonlinear space-transforming reduction methods. However, these techniques do not provide a subset of the initial features, eventually. Instead of this, they create new -and completely original- feature spaces which are derived from the former ones. Thus, they should be applied only when there is no firm necessity to keep the reduced features as a subset of the initial set. Some of the commonly used methods falling into this group can be listed as *Principal Component Analysis* (PCA), *Kernel PCA*, *Multidimensional Scaling* (MDS), *Laplacian Eigenmaps*, *Generalized Discriminant Analysis* (GDA), and *Diffusion Maps*. A helpful review on the numerous methods involved in this concept can be inquired in [99].

Another way of reducing the dimensionality of the feature space is choosing the most significant predictor variables among the ones that form the initial feature set without exposing them to any kind of transformation. In this thesis, we embraced this approach regarding the following facts:

1. Marking some of the features as significant, rather than transforming all of them into another space, enables a deeper grasp of the main question of interest, e.g., by revealing the most relevant voxel locations in the brain volumes.

2. Space-transforming methods usually induce a higher computational cost involving operations in the frequency domain (e.g., solution of high-dimensional eigen-space problems). Additionally, these procedures are mostly limited with some inherent constraints. As an example, widely-used PCA does not apply to our domain, since it provides a maximum number of features which is smaller than number of samples by one, which is certainly not efficient for our problem domain. In [89], Shih et al. expressed this fact as follows: *"Conventional variable selection techniques are based on assumed linear model forms and cannot be applied in this 'large p and small N' problem."*

3. It is noticed that inclusion of domain knowledge into the feature selection phase through a relatively heuristic approach affects the performance positively (see Section 4.3).

Table 4.1: Number (and rate to the initial dimensions) of remaining features after zero-voxel elimination and feature selection procedures (AD/CN case).

	Initial	Non-zero voxels	Feature selection
No. of remaining features	6,368,835	2,112,054	3,320
Rate of remaining features	-	33.16%	0.05%

This approach is explained in the following section.

4.2 A 3-Step, Hybrid Procedure

In this investigation, a partially novel procedure to specify the optimal subsets (constrained to presumed parameters) of the initial set of predictor variables is developed, rather than directly applying space-transforming dimensionality reduction methods. Our methodology involves three steps, namely, *Statistical Analysis*, *Tissue Probability Criteria*, and *Within-class Norm Thresholding* [17].

Our proposed procedure is characterized as *hybrid*, since it involves both an unsupervised part (in which the algorithm makes decisions by statistically analyzing the data without any prior assumptions) and components which includes logical rules relying on domain-specific knowledge. Step I, namely, *Statistical Analysis* is involved in the first group, whereas, the following two steps (*Tissue Probability Criteria* and *Within-class Norm Thresholding*) are involved in the second.

The proposed method is mentioned to be *partially novel*, because we do not claim novelty for each individual step. For example, Step I (*Statistical Analysis*), is based on the well-known GLM-based mathematical approach. Novelty is claimed for the modification of the STAND-score in Step II, the heuristic approach in Step III, and use of all three steps in conjunction.

It is also important to emphasize that each step is applied on the initial feature set separately, without imposing a specific order. After application of all three steps, the voxels to be eliminated are determined by the union set of marked voxels at each particular step.

Table 4.1 demonstrates the decrease in the number of features by elimination of
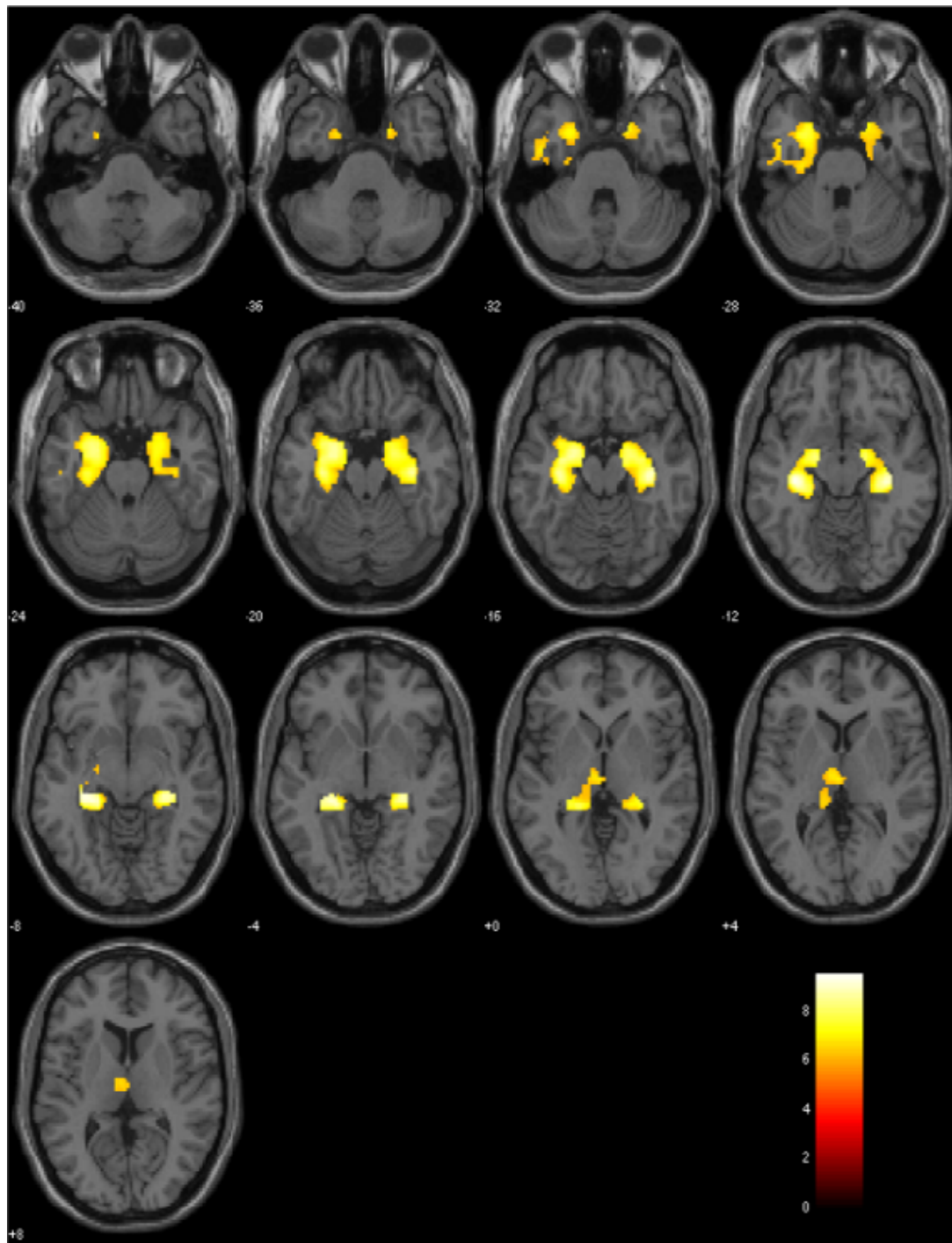
Figure 4.1: Obtained binary mask involving "significant" voxels to distinguish AD and CN subjects.

the zero-voxels and application of the proposed dimensionality reduction method. It is seen that by utilization of the proposed technique, only 0.05% (3,320 out of 6,368,835) of the voxels composing the raw data were left as significant features.

In Figure 4.1, the obtained mask involving the voxels which were marked as "significant" by the application of our procedure is highlighted. One interesting observation can be mentioned here, as the resulting region of interest showed up to be asymmetric between the two lobes of the brain. However, this fact does not involve any contradictions with the expected anatomical effects due to AD.

A positive impact of dimensionality reduction on the overall prediction performance was observed. For this, a demonstration can be found in Section 4.3. Steps of the methodology are presented in the following subsections in detail.

4.2.1 *Step I:* Statistical Analysis

In order to investigate the voxel-wise gray matter tissue probability statistics, a multivariate *General Linear Model* (GLM) (cf. Eqn. (4.1)) is utilized. A GLM is stated - in mathematical terms - as:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{u}. \tag{4.1}$$

In Eqn. (4.1), $\boldsymbol{y}$ represents an *N*-dimensional output vector (where *N* is the number of observations), $\boldsymbol{X}$ is an $N \times (p/3)$ matrix by which the GM tissue probabilities of each image are encapsulated, and $\boldsymbol{b}$ is a vector of $(p/3)$ unknown parameters to be estimated. The total number of the initial predictor variables by which the three tissue probability maps are expressed is p. Thus, considering only the GM tissue probabilities, final dimensionality becomes equal to the number of voxels in an image, which is $(p/3)$. Elements of the error vector $\boldsymbol{u}$ are independent and identically distributed (i.i.d.) variables with a mean value of zero. A GLM expresses the output response as described by a linear combination of the explanatory variables. In frame of this study, the output response is the label which indicates whether the subject is an AD patient or he/she is cognitively healthy, whereas, the explanatory variables are the indices of
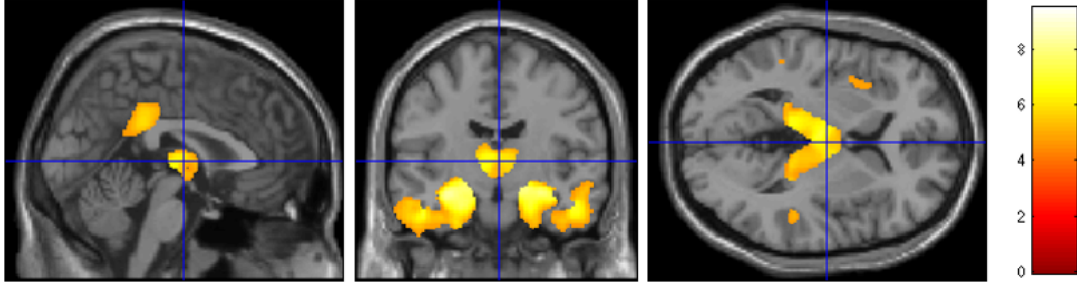
Figure 4.2: Significant voxels acquired through the Statistical Analysis step of the algorithm, alone. The input parameters are specified as; Height threshold: $p-value = 0.05$ (FWE), $T = 4.53$; extent threshold: $k = 4 \times 4 \times 4 = 64$ voxels, $p-value = 0.015$.

the GM tissue probability map, or simply, voxels.

Means of the two populations, i.e., brain MRI groups formed by AD patients and cognitively healthy participants are voxel-wisely compared by using two-sample *t*-test. In this way, the voxels which have relatively higher statistical significance are determined and the *design matrix* (the matrix X in Eqn. (4.1) is built. Various masks can be obtained depending on two input parameters. The first one is called *height threshold* and it limits similarities at voxel-level, whereas, the second one is referred as the *extent threshold* and it limits the similarities at cluster-level. More clearly, by changing the values given to these two thresholds, different combinations formed by columns of the design matrix can be designated as significant feature subsets. An example of this process involving analysis of fMRI (functional MRI) data can be found in [39].

In Figure 4.2, the outcome acquired through this single step of the algorithm, in case of extent and height threshold settings specified in the figure caption, is shown.

4.2.2 *Step II:* Tissue Probability Criteria

Authors of the paper [101] proposed a novel procedure to assign weights to the voxel indices of a segmented structural MRI volume, indicating their distinctiveness. The measure is referred as *STAND-score*, or *Structural Abnormality Index Score*.

The *STAND-score*s are computed through several steps. First of all, the TPMs are downsampled to have a voxel size of 8 mm in each direction. Secondly, the voxels having the sum of GM and WM tissue densities as less than 10% in all images are removed. Next, the volumes having CSF in half or more of the voxels are canceled out. Finally, weighting factors for each remaining voxel are computed by applying a linear SVM. These factors are employed for further removal of the insignificant features from the model. The logic behind the procedure is based on the assumption that relatively high and low tissue probabilities in particular locations of the brain imply anatomical relevance for early onset of Alzheimer's. Thus, the approach involves a direct use of field knowledge.

Step II of our procedure was proposed under the inspiration of the 2$^{\text{nd}}$ step of the *STAND-score* methodology with a slight modification and an addition. Firstly, the lower boundary for the sum of GM and WM contents to decide for erasing an index was not straightforwardly assigned as 10%. This threshold was tested for various values and it was observed that better performance outcomes were attained with higher percentages. Secondly, a new constraint for the sample mean of the tissue content was introduced. The motivation behind this approach was to prevent removal of numerous feature candidates from the model in an instant, on the account of a very low threshold value. Meanwhile, the effect of this choice was compensated for by employing a second operator which removes additional features by evaluating sample means.

In the first pass, the voxel indices satisfying the condition that sum of GM and WM tissue probabilities being above a certain threshold τ_1 are marked to be kept, and the remaining variables are indicated as insignificant. Next, the voxel indices with sum of sample means of GM and WM tissue probabilities below the second threshold τ_2 are marked to be removed. Both operations were applied on the initial data. Therefore, the order is interchangeable. The union set of features, which has been indicated as insignificant through consecutive steps, are deleted from the data. To be more precise:

$$P_{GM}(i,j) + P_{WM}(i,j) < \tau_1, \quad \forall i \in \{1, 2, ..., n\}, \tag{4.2}$$

or,

$$\overline{P_{GM}(j)} + \overline{P_{WM}(j)} < \tau_2. \tag{4.3}$$

By the expressions given in Eqns. (4.2) and (4.3), the elimination criteria for the jth variable is pointed out, mathematically. Here, i and j are sample index and variable index, the distributions P_{GM} and P_{WM} stand for GM and WM probabilities, and, $\overline{P_{GM}}$ and $\overline{P_{WM}}$ are sample means of the tissue probabilities, respectively.

In this thesis, the best values for the thresholds τ_1 and τ_2 were determined through manual observations among successive trials. The input parameters of MARS were kept constant (see Section 6.1) during this experiments. Meanwhile, τ_1 and τ_2 values were varied to form different combinations. The MARS model was trained using the remaining data at each trial, and the best combination of τ_1 and τ_2 were determined as the ones providing the highest *AUC* (*Area Under the ROC Curve*). In the AD/CN classification case, the best values that the two thresholds take became: $\tau_1 = 0.5$ and $\tau_2 = 0.7$.

4.2.3 *Step III:* Within-Class Norm Thresholding

Proposed method is finalized with *Step III*, which can be simply described as a comparison between **(i)** Euclidean norm of a vector defined for a variable through observations, and **(ii)** average of these norms computed for variables of the same tissue class (TC). The fundamental logic behind this approach can be explained as: for a "significant" variable, intensity variation among observations having different class labels must induce a relatively greater norm. Firstly, the L_2-norm of samples of each predictor variable is calculated by:

$$\|\boldsymbol{P}(j)\|_2 = \sqrt{\sum_{i=1}^{n} \left(\boldsymbol{P}(i,j)\right)^2}. \tag{4.4}$$

Subsequently, means of the previously calculated norms are computed for the 3 TCs, separately. These are named as within-class norm means. Namely, $\hat{\mu}_{GM}$, $\hat{\mu}_{WM}$, and $\hat{\mu}_{CSF}$ are computed by:

Table 4.2: Comparison of the proposed approach with other commonly known methods in terms of sensitivity (SEN), specificity (SPE), positive predictive value (PPV), and negative predictive value (NPV) [17].

	SEN (%)	SPE (%)	PPV (%)	NPV (%)
None	67.16	81.48	75.00	75.00
Voxel-MARS	**83.58**	**86.42**	**83.58**	**86.42**
PCA	82.09	71.60	70.51	82.86
MDS	82.09	71.60	70.51	82.86
Laplacian Eigenmaps	79.10	74.07	71.62	81.08
Kernel PCA	2.99	95.06	33.33	54.23
Diffusion Maps	91.04	1.24	43.26	14.29
GDA	0	100	-	54.73

$$\hat{\mu}_c = \frac{\sum_{j \in c} \|\boldsymbol{P}(j)\|_2}{j_c}, \tag{4.5}$$

where c states the TC ($c \in \{GM, WM, CSF\}$), and j_c is the number of predictor variables from the corresponding TC. The elimination rule for a variable is described as its norm $\|\boldsymbol{P}(j)\|_2$ being smaller than a ratio of the within-class norm mean ($\hat{\mu}_c$) of the corresponding class. That ratio can be referred as ϵ, where the value of ϵ must lie between 0 and 1. For determination of the ϵ value, the same methodology as clarified in Subsection 4.2.2 was used. In the AD/CN case, the highest classifier accuracy among several trials was obtained by setting $\epsilon = 0.9$.

Considering the tissue probabilities are always greater than zero, a basic summation opeator or the L_1-norm could also have been used instead of L_2-norm. This would have affected the value of ϵ, however, final set of significant variables would not have changed since the relation between vector magnitude and the Euclidean norm is monotonic. Nevertheless, L_2-is preferred since it is a more common and general notation of representing vectorial distances.

4.3 A Comparison of Our Approach with Commonly Used Dimensionality Reduction Techniques

Performance of the proposed 3-step, hybrid methodology for feature selection was quantitatively assessed by comparing with six commonly known techniques which are

mentioned above, in terms of *sensitivity* (SEN), *specificity* (SPE), *positive predictive value* (PPV), and *negative predictive value* (NPV) outcomes (see Table 4.2).

Previous to the application of each method, the first step (*Statistical Analysis* via *General Linear Model*) of our procedure was applied on the baseline gray matter tissue probability maps prepared for training and testing. Hence, dimensionality of the feature space was decreased to the value of 26,448, which is suitable for the restrictions enforced by any of the methods. Following to that, the training data was assessed for intrinsic dimensionality through *Maximum Likelihood Estimation* (MSE), which became 11 for the AD/CN group. The techniques included in Table 4.2 (from PCA to GDA) were executed on training and test samples to scale down the dimensionality of both to the recently computed intrinsic dimension [1]. Input parameters were assigned as $M_{max} = 11$ and $K_{max} = 1$, where M_{max} refers to "the maximum allowed number of basis-functions (BFs) to have in the model at the end of the forward step", and K_{max}, stands for "the maximum degree of interaction between variables to form a BF". Keeping them constant, MARS models were composed and classification was performed. The models were applied on the blind test data. In the case labeled as "None", model was built immediately after the *Statistical Analysis* step, without further reduction in the dimensionality. In the case demonstrated as "Voxel-MARS", our proposed method was executed and the number of resulting dimensions for the model building process was observed to be 3,320.

Conforming to the results shown in Table 4.2, just 3 in 6 techniques (i.e., *PCA*, *MDS*, and *Laplacian Eigenmaps*) became able to produce variable vectors with admissible discriminative strength for classification. It is observed that they provided higher sensitivity than the "None" case did, in which the dimensionality was not reduced at all. However, the proposed feature selection methodology outperformed all of the other methods presented on Table 4.2 in terms of *sensitivity*, *specificity*, *positive predictive value*, and *negative predictive value*.

---

[1] *The MATLAB Toolbox for Dimensionality Reduction* [99] was used for computations.

CHAPTER 5

UTILIZATION OF MARS AND CMARS FOR EARLY DETECTION OF ALZHEIMER'S DISEASE

Up to this point, the brain MRI volumes have already been processed in view of the *voxel-based morphometry* framework to extract features for cognitively normal, Alzheimer's disease and MCI (mild cognitive impairment) classification. A 3-step feature selection routine has been applied to determine an optimum subset of the significant predictor variables. Also, the mathematical basis on which MARS and CMARS methods were founded has been detailed in previous chapters.

This chapter focuses on information regarding details in the application of our classification procedure. Model building phase of MARS and optimization phase of CMARS (by which the classical backward process of MARS is replaced) are introduced in detail. The aim of this chapter is to make available, the information essential to reproduce the results presented in this thesis, for the reader.

5.1 Model Building

Multivariate Adaptive Regression Splines is the name of a brand trademarked as MARS® by the company *Salford Systems*, as well as it is the name of the regression method. Therefore, researchers are not allowed to implement and use the algorithm by its original name for commercial purposes. For this reason, the method is alternatively named as "Earth" by majority of the academic community. Although a free version of the software has been made available for academic research studies by the firm, in this thesis, using an open source implementation of the method is preferred.

Two of those implementations, namely, the *ARESLab*[1] and the *Earth*[2] packages were investigated. Since it was implemented in C programming language, the second one (*Earth*) outperforms *AresLAB* (which was programmed entirely as a native MATLAB script) in terms of processing time. Additionally, the package Earth includes a MATLAB *mex* interface, which means the included functions can be called over MATLAB easily. Based upon these, the *Earth* package is chosen.

Although both of the software products allow their users to control several input parameters, generally, only two of these inputs are utilized. First one of these parameters is M_{max}, which refers to the maximum allowed number of basis-functions (BFs) to have in the model at the end of the forward step, and the second one is K_{max}, standing for the maximum degree of interaction between variables to form a BF. If another parameter named *prune* is set to "true", immediately after the forward pass, the backward process starts, and approximately half of the BFs (with least relevance) are removed from the model. Otherwise, the model does not be pruned at all.

At the end of the model building (training) phase, *Earth* produces three matrices, namely, "*Dirs*" (carrying the $+/-$ sign information of the hinge functions contributing to each one of the basis-functions), "*Cuts*" (carrying the information regarding the knot points at each predictor variable contributing to each one of the BFs), and "*Betas*" (carrying the coefficients associated with each particular BF). Information encapsulated by these 3 matrices is sufficient to express the resulting MARS hyperplane in the feature space.

5.2 Optimization

CMARS replaces the backward step with an algorithm involving *Tikhonov regularization* and *conic quadratic optimization*; thus, the widest MARS model is required for this part of the algorithm. In order to produce the widest model (the model with M_{max} basis-functions), the "prune" property of the *Earth* object is set to "false", in other words, the backward step of the MARS algorithm is canceled out. CMARS

---

[1] Jekabsons G., ARESLab: Adaptive Regression Splines toolbox for MATLAB/Octave, 2011, available at http://www.cs.rtu.lv/jekabsons/ [49].

[2] S. Milborrow. Derived from MDA:MARS by T. Hastie and R. Tibshirani., "Earth: Multivariate Adaptive Regression Splines," 2011 [62].

optimizes coefficients of the linearly independent terms in this widest model obtained without pruning.

The first step of the procedure is the computation of the L_{im} values which form the $\boldsymbol{L}$ matrix appearing in Equations (2.9) and (2.10), respectively. In order to accomplish that calculation, the data matrix is required to be updated to guarantee differentiability everywhere. For this purpose, at each predictor variable (column), the data vector is sorted, and two additional rows before the minimum and after the maximum row are introduced. Values for these rows are specified as 0.5 smaller than the minimum, and 0.5 greater than the maximum value in the original vector, respectively. The value 0.5 is selected by evaluating the results of a number of trials, however, it has been observed that the results are not considerably sensitive to this value. In order to find L_{im} values at the diagonals of $\boldsymbol{L}$, the D and the $\Delta \hat{\boldsymbol{x}}_i^m$ values are computed for each sample, in an iterative loop. Details regarding the discretization scheme and calculation of D values are presented in [103], including a numerical example.

Having the data matrix, the $\boldsymbol{L}$ matrix, *Dirs*, *Cuts*, and *Betas*, now, the problem can be expressed as a CQP problem as shown in Eqn. (2.11). For this purpose, a commonly used, modern optimization toolbox, *MOSEK* [5], is employed. In order to be able to construct the *MOSEK* optimization model matrix, the constraints given in Eqn. (2.11) must be restated as linear constraints. Thus, N new variables (also constraints), named β_i, and associated with the N samples (indexed by $i \in \{M+1, M+2, ..., M+N\}$), are defined as:

$$
\begin{aligned}
y_1 - \boldsymbol{B}(\boldsymbol{x_1})\beta &= \boldsymbol{\beta}_{M+1}, \\
y_1 - \boldsymbol{B}(\boldsymbol{x_2})\beta &= \boldsymbol{\beta}_{M+2}, \\
&\vdots \\
y_N - \boldsymbol{B}(\boldsymbol{x_N})\beta &= \boldsymbol{\beta}_{M+N}.
\end{aligned}
\tag{5.1}
$$

Correspondingly, the first conic constraint of Eqn. (2.11) can be rewritten in terms of these new variables as:

$$\left(\sum_{i=M+1}^{M+N} \beta_i^2 \right)^{\frac{1}{2}} < t. \tag{5.2}$$

Similarly, the second conic constraint is expressed in terms of M new variables (linear constraints):

$$
\begin{aligned}
L_1 \beta_1 &= \boldsymbol{\beta}_{M+N+1}, \\
L_2 \beta_2 &= \boldsymbol{\beta}_{M+N+2}, \\
&\vdots \\
L_M \beta_M &= \boldsymbol{\beta}_{M+N+M},
\end{aligned} \tag{5.3}
$$

and the corresponding single conic constraint with $\sqrt{\tilde{M}}$,

$$\left(\sum_{i=M+N+1}^{M+N+M} \beta_i^2 \right)^{\frac{1}{2}} \leq \sqrt{\tilde{M}}. \tag{5.4}$$

By these conversions, we end up with a set of constraints which are available now for being modeled and solved in *MOSEK* toolbox.

5.3 Visualizing the Model Function

In Figure 5.1a, a MARS model function built with input parameters $M_{max} = 14$, $K_{max} = 2$, and *prune* property set to "false" is shown. Thus, it is an example to the widest model acquired at the end of the *forward pass*. Each line represents a linearly independent term which is associated to a BF multiplied by a weighting factor. Each BF is expressed as a product of *hinge functions*, described by *predictor variables* and corresponding *knot points*.

As it is described before, the *backward step* of the MARS algorithm prunes the model, in other words, removes the least significant terms in the model function in order to enhance generalization ability of the classifier (the number of BFs decreases to 12 from 14, and the accuracy rises to 0.6757). Fig. 5.1b illustrates how this procedure is

```
0.541                                                             // 0
-0.317 * max(0, x[2561] - 0.497)                                  // 1
-1.29 * max(0, 0.497 - x[2561])                                   // 2
+0.247 * max(0, x[994] - 0.0838) * max(0, 0.497 - x[2561])        // 3
+0.691 * max(0, 0.0838 - x[994]) * max(0, 0.497 - x[2561])        // 4
-4.72 * max(0, 0.497 - x[2561]) * max(0, x[2967] - 0.0715)        // 5
+0.117 * max(0, 0.497 - x[2561]) * max(0, 0.0715 - x[2967])       // 6
+1.88 * max(0, x[2542] - -0.226) * max(0, 0.497 - x[2561])        // 7
+0.138 * max(0, -0.226 - x[2542]) * max(0, 0.497 - x[2561])       // 8
-0.626 * max(0, x[912] -   1.03)                                  // 9
+0.0952    * max(0, 1.03 - x[912])                                // 10
+0.201 * max(0, x[2155] - -0.574)                                 // 11
+0.251 * max(0, -0.574 - x[2155])                                 // 12
+0.269 * max(0, x[2840] - 0.165)                                  // 13
+0.604 * max(0, 0.165 - x[2840])                                  // 14
```

(a)
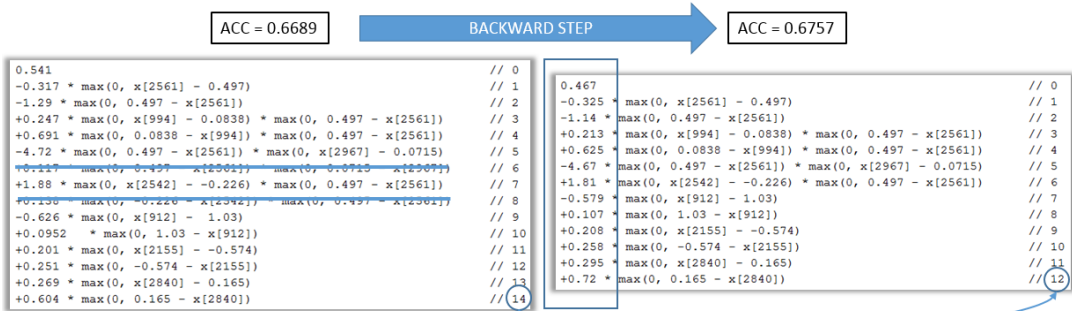
(b)
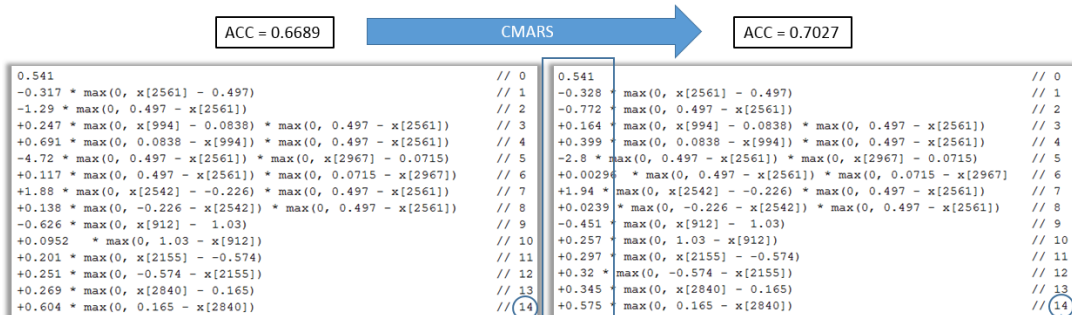
(c)

Figure 5.1: (a) MARS model without pruning, (b) Evolution of the model through backward step of MARS, (c) Evolution of the model through CMARS.

55

(a)

(b)

(c)

(d)
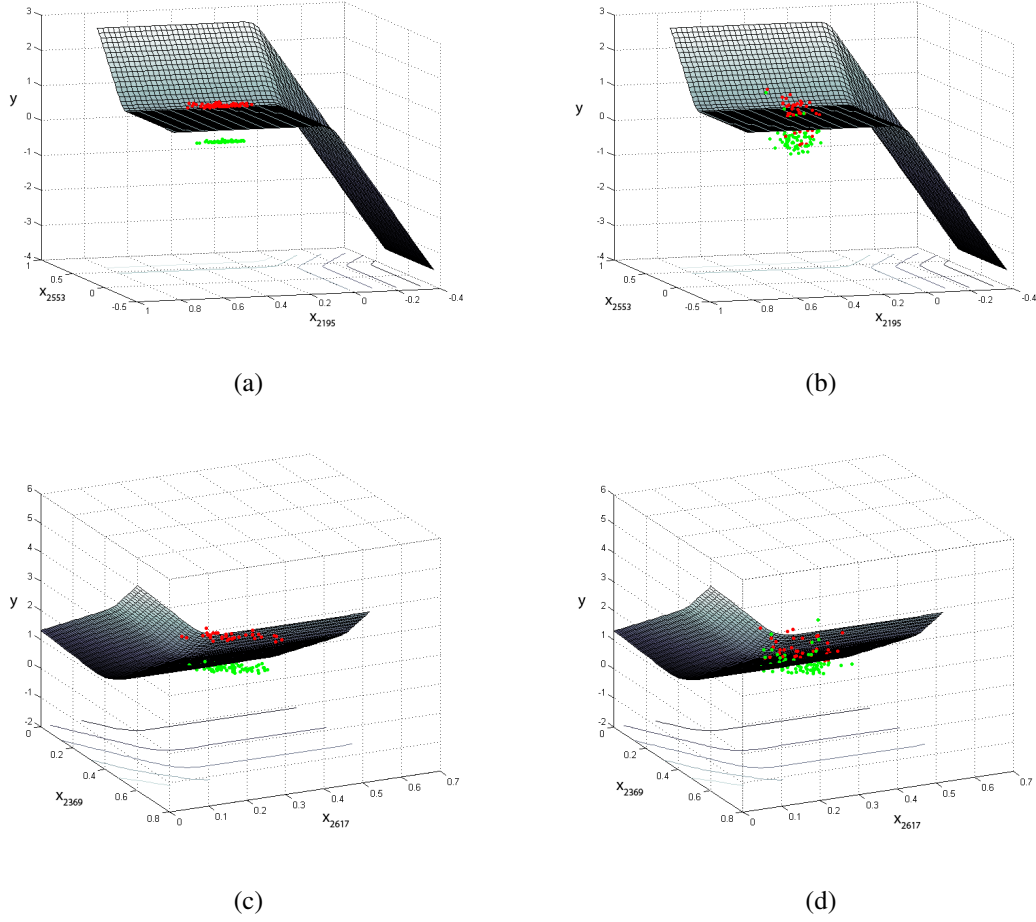
Figure 5.2: Projections of the two hypersurfaces obtained in AD/CN and MCI/CN classification cases onto the axes defined by the most contributory variables [17].

resulted. Finally, in Fig. 5.1c, it is observable that CMARS increases accuracy of the classifier by optimizing the coefficients without canceling out any of the BFs from the model (the number of BFs remains 14, and the accuracy rises to 0.7027, which is greater than the accuracy obtained through the *backward step* of MARS).

Visualizing the problem spaces in these degrees of dimensionality has always been a problematic issue. To provide a visual sense of the separating surface which divides the problem space into two subspaces, projections were employed. Fig. 5.2 shows two examples to these kind of projections, obtained by projecting the model functions onto two axes described by predictor variables. Figures 5.2a and 5.2b show the output responses acquired from the training and test samples, and the surface obtained in AC/CN classification case projected onto the axes of predictor variables

indicated by indices 2553 and 2195 (which correspond to BF1 and BF4), respectively. Similarly, Figures 5.2c and 5.2d show projections of a sample model constructed in MCI/CN case onto the axes of predictor variables defined by indices 2369 and 2617, respectively. Both of the MARS model functions were built with input parameters $M_{max} = 11$, $K_{max} = 2$, and "prune" property set to "true". Aforesaid voxel indices were selected in order to visualize nonlinearity through corresponding hinge functions occurring in a nonlinear term in multiplication form. In both cases, a good separation of the *healthy* and the *diseased* samples was seen.

In both of the cases shown in Fig. 5.2, we decided on the variable axes on which the surfaces were projected, considering the contribution of the basis-functions (coinciding with these variables) to the overall discriminative strength of the trained model function. *Residual Sum of Squares* (RSS) was used as the measure to choose the binary combination of these basis-functions among all of them. Thus, the two variables which produce the minimum RSS on the blind test data through corresponding hinge functions were selected for projection.

CHAPTER 6

PERFORMANCE EVALUATION

6.1 Parameter Optimization with Grid Search and Cross-Validation

Definitive characteristics of MARS models are determined by assigning values to the two primary input parameters. The first one of these parameters is M_{max}, which represents maximum number of basis-functions (BFs) that is possible to have in the model at the end of the forward step. The second parameter is K_{max}, which refers maximum degree of interactions between variables that form a basis-function. An increment in M_{max} increases the flexibility of the final model. Unfortunately, high flexibility means high complexity at the same time. Nonlinearities and variable interdependencies can be modeled by choosing a K_{max} value greater than 1.

A coarse-to-fine grid search procedure was conducted for optimization of these two major MARS parameters. Classifiers were administered on AD/CN, MCI/CN, and MCIc/MCInc subject groups in two iterations. Firstly, M_{max} was varied with step size of 10, taking the values from the set $\{11, 21, ..., 101\}$. For each M_{max}, K_{max} was varied taking each value from the set $\{1, 2, 3\}$. Classification was performed and the *Area Under the ROC Curve* (AUC) is calculated for each input combination. Next, M_{max} was varied with a smaller step size (of 2), near the parameter value providing the maximum AUC in the former step, and, K_{max} was again varied between the elements of $\{1, 2, 3\}$.

$\tilde{N}$-times replicated $\tilde{k}$-fold cross-validation ([59]) method was adapted to validate the model parameters[1]. In this design, the initial training dataset was randomly parti-

---

[1] The letters N and k are included with a tilde ($\sim$) sign on top of them to avoid with the N used for stating the sample size and the k appearing in Eqn. (2.7) to express number of knots, respectively.

tioned into $\tilde{k}$ subsets (folds). While each fold was employed as the test data, the remaining $\tilde{k} - 1$ subsets were assembled to train the classifiers. This process was rei-terated $\tilde{N} - 1$ more times with randomly updated partitions and the $\tilde{k} \times \tilde{N}$ results - in this case, AUC (area under the ROC curve), sensitivity, and specificity - were avera-ged. The parameter pair maintaining the best AUC value was assigned as the eventual input parameter set. Ultimate model functions were built using all of the training data with these parameters, and classification was performed on the - previously reserved - blind test data, utilizing resulting models to attain the final results.

On the other hand, the inequality between the numbers of scans included by the 3 dif-ferent classification groups caused a problem in specification of the cross-validation parameters $\tilde{N}$ and $\tilde{k}$. (The AD/CN cohort includes 150 scans, the MCI/CN cohort in-cludes 120 scans, and the MCIc/MCInc cohort includes 104 scans in the training sets, see Table 3.1.) Besides, it was seen that when the training sample size was reduced below 100, the classification performance remarkably decreased. Taking this situa-tion into consideration, the value of $\tilde{k}$ was chosen as 3 for the AD/CN group, 6 for the MCI/CN group, and 18 for the MCIc/MCIcn group. This decision was made to keep the number of training samples close ($N \times (\tilde{k} - 1)/\tilde{k} \approx 100$, where N is the total num-ber of samples) in each classification case, throughout the cross-validation process. The $\tilde{N}$ values are assigned considering compensation of the differences in number of total repetitions introduced by using different $\tilde{k}$ values, accordingly ($\tilde{N} = 18$ for AD/CN case, $\tilde{N} = 9$ for MCI/CN case, and $\tilde{N} = 3$ for MCIc/MCInc case).

Classifiers were formed by training MARS models fed by the parameters obtained through the parameter optimization approach based on grid search and cross-validation as clarified in this subsection. The results presented in the form of confusion matrix in Table 6.2 were attained by application of the acquired models on the blind test data (see Section 6.2). The individual demonstrations of parameter optimization procedure for AD/CN, MCI/CN, and MCIc/MCInc classification groups are shown in Figures 6.1, 6.2, and 6.3, respectively. Table 6.1 introduces the maximum average AUC value observed during the search process, matching parameter pair, and the averages of the performance metrics for each case. In the figures, the coarse search progress is dis-played by the graphs in the left-hand side column, and the fine tuning of the model training inputs is shown by the ones in the right-hand side column. Each row corre-

Table 6.1: Optimal parameters with corresponding averaged performance outcomes obtained through the grid search procedure (see Section 6.1 for the definitions of M_{max} and K_{max}).

	AD/CN	MCI/CN	MCIc/MCInc
M_{max}	11	11	71
K_{max}	1	1	1
AUC	0.8656	0.7025	0.5477
SEN	0.7826	0.6328	0.5404
SPE	0.8025	0.6607	0.4973

Table 6.2: Confusion matrix regarding all 3 experiment results. "H" stands for "Healthy" and "D" stands for "Diseased". Vertical division shows the ground truth data, where the horizontal division shows the prediction results.

Confusion Matrix		True Classes					
		AD/CN		MCI/CN		MCIc/MCInc	
		H	D	H	D	H	D
Predictions	Healthy	70	11	72	8	40	14
	Diseased	11	56	9	29	27	23
	TOTALS	81	67	81	37	67	37

sponds to a different value (1, 2, or 3) of K_{max}. The dashed black line represents the maximum *AUC* observed in each run. The ultimate maximum *AUC* is pointed out by the yellow rectangles in each figure.

6.2 Classification of Normal, AD, and MCI Brain MRI by MARS

Classification experimentations were executed by training MARS models with the input parameters attained through the parameter optimization process as clarified in the former subsection. Constructed models were employed for making class predictions on the blind test data. Class predictions at the end of the classification routine are presented in the format of confusion matrix provided in Table 6.2.

6.2.1 Description of the Performance Metrics

ACC (accuracy), SEN (sensitivity), SPE (specificity), PPV (positive predictive value), NPV (negative predictive value), and AUC (area under the ROC curve) me-
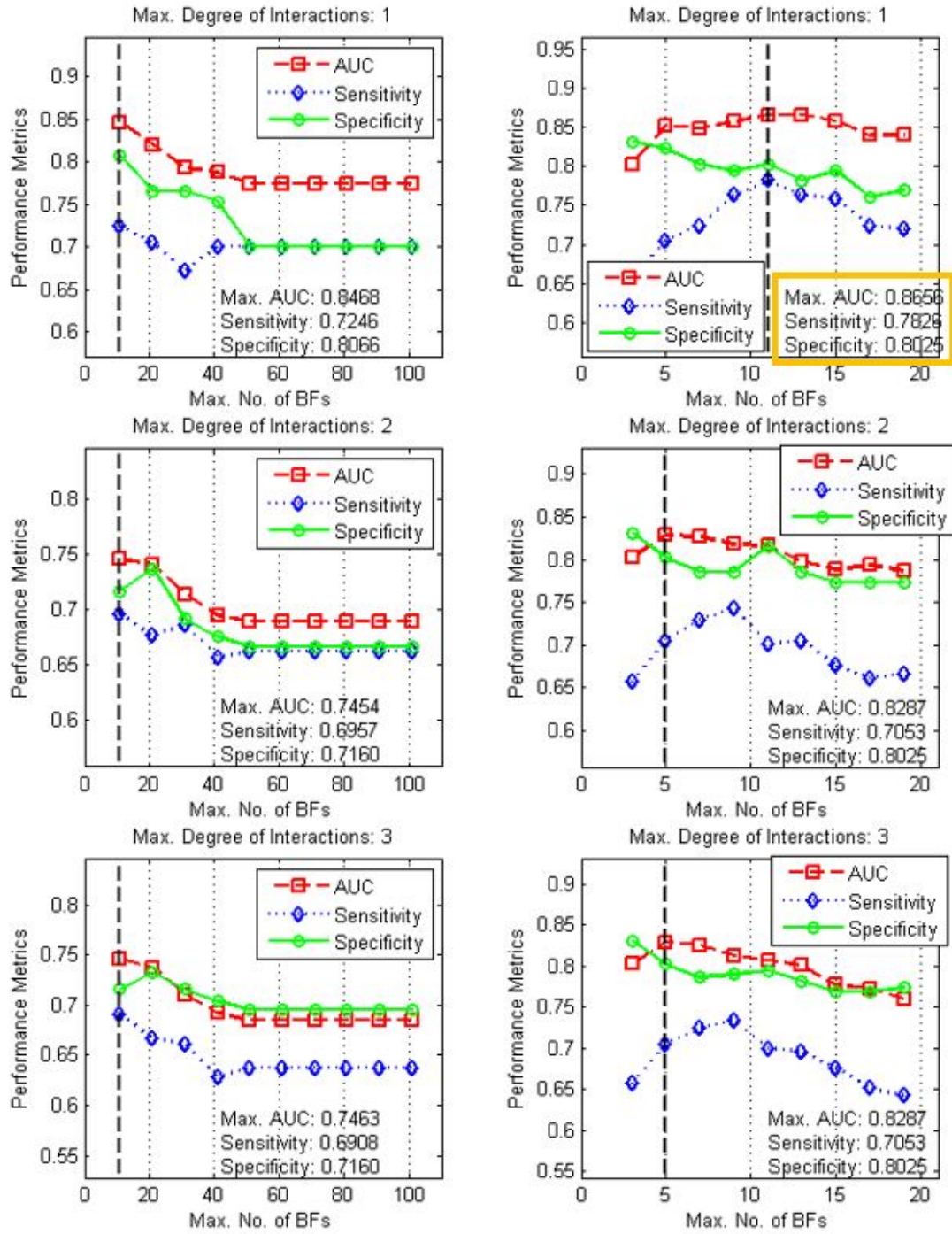
Figure 6.1: (Adapted from [17]) Optimization of the MARS model parameters with grid search and cross validation for the AD/CN group. $AUC = 0.8656$ with $M_{max} = 11$ and $K_{max} = 1$.
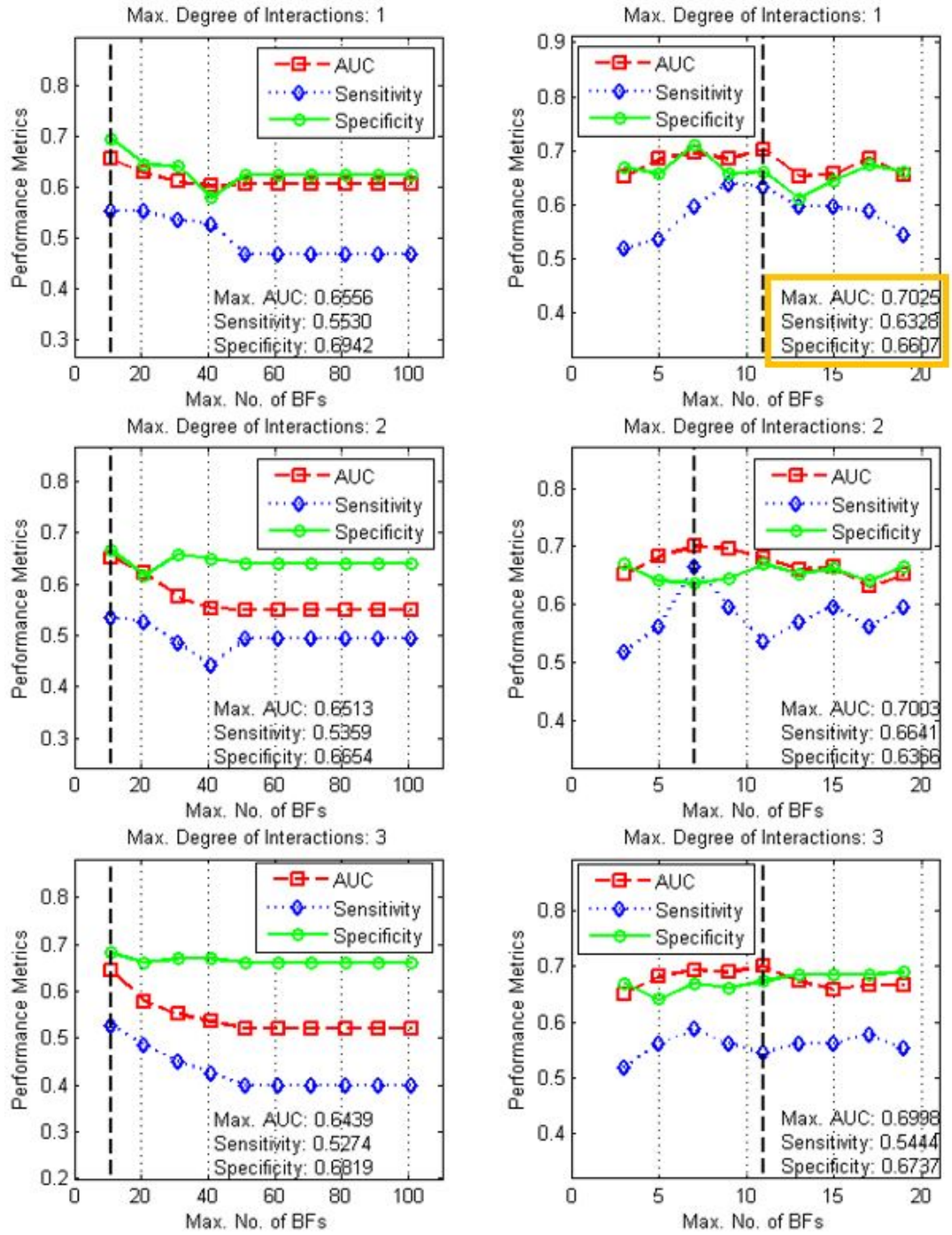
Figure 6.2: (Adapted from [17]) Optimization of the MARS model parameters with grid search and cross validation for the MCI/CN group. $AUC = 0.7025$ with M_{max} = 11 and $K_{max} = 1$.
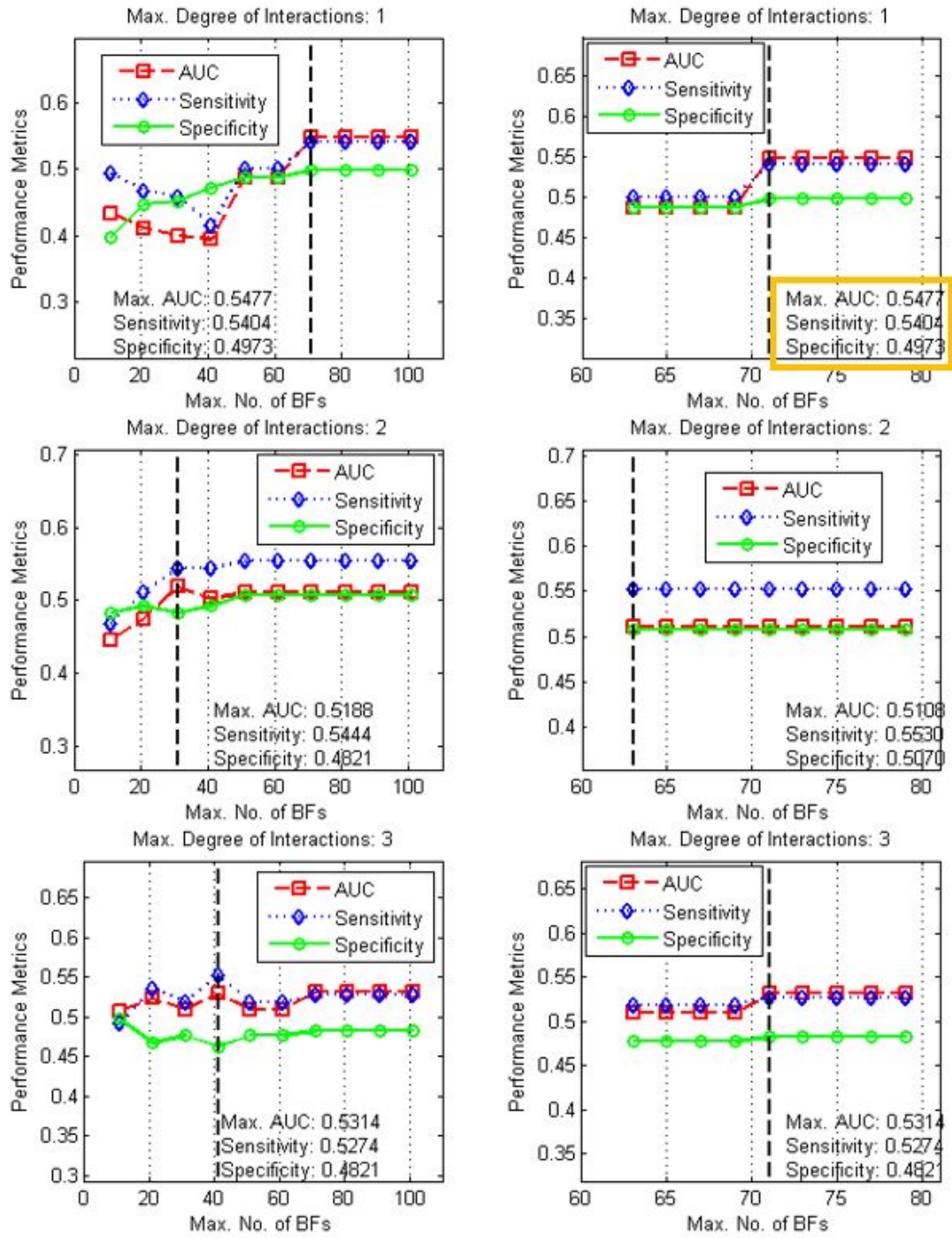
Figure 6.3: (Adapted from [17]) Optimization of the MARS model parameters with grid search and cross validation for the MCIc/MCInc group. $AUC = 0.5477$ with $M_{max} = 71$ and $K_{max} = 1$.

Table 6.3: Performance of *Voxel-MARS* is compared with the average outcomes of others. Average performance rates of other methods are indicated as "mean ± standard deviation [range]".

Case	Metric	Others (%)	VM (%)	Diff. (%)	Rank
AD/CN	**SEN**	71.46±5.65 [59-82]	83.58	+12.12	1
	SPE	89.39±5.03 [77-98]	86.42	-2.97	22
	PPV	85.18±6.23 [72-96]	83.58	-1.60	21
	NPV	78.93±3.88 [70-86]	86.42	+7.49	1
MCIc/CN	**SEN**	54.39±12.39 [22-73]	78.38	+23.99	1
	SPE	88.82±7.27 [73-99]	88.89	+0.07	17
	PPV	71.50±11.97 [50-89]	76.32	+4.82	12
	NPV	81.21±3.60 [73-87]	90.00	+8.79	1
MCIc/MCInc	**SEN**	44.20±15.22 [22-70]	62.16	+17.96	2
	SPE	76.47±8.77 [61-91]	59.70	-16.77	16
	PPV	51.33±7.25 [39-67]	46.00	-5.33	12
	NPV	68.18±5.02 [66-79]	74.07	+5.89	7

trics were calculated assigning positive to the *diseased* label and negative to the *healthy* label. Therefore, true positives (TP) were described as the *diseased* samples which are predicted to be *diseased*, true negatives (TN) were described as *healthy* ones which are predicted to be *healthy*, false positives (FP) were described as *healthy* ones which are predicted to be *diseased*, and false negatives (FN) were defined as *diseased* ones which are predicted to be *healthy*. In AD/CN classification group, ADs were assumed as *diseased* and CNs were assumed as *healthy*. Similarly, in MCI/CN case, the *diseased* set included subjects with MCI. In the third case, *diseased* samples were presumed to be the converter MCIs (MCIc), and the (non-converting MCI) MCInc samples were assigned as *healthy*.

6.2.2 AD vs. Normal

Table 6.4 shows the results acquired through classification by performing different methods on the same dataset formed for AD/CN classification. The proposed methodology was designated as *Voxel-MARS* (ID: 0) being inspired by the naming convention embraced by Cuingnet et al. (2011) in [23]. All 28 versions of the 10 main methods evaluated in [23] were sorted by resulting sensitivity, in descending order. The most successful 5 of them were chosen for comparison. The columns in the ta-

Table 6.4: Performance of MARS compared with five methods (top five in sensitivity) in AD/CN case.

ID	Method Name	SEN	SPE	PPV	NPV
0	**Voxel-MARS**	**83.58%**	**86.42%**	**83.58%**	**86.42%**
1.5.1 a	Voxel-COMPARE-D-gm	82%	89%	86%	86%
1.1.1 a	Voxel-Direct-D-gm	81%	95%	93%	86%
1.4.1 b	Voxel-Atlas-D-all	81%	90%	87%	85%
2.2	Thickness-Atlas	79%	90%	87%	84%
1.4.1 a	Voxel-Atlas-D-gm	78%	93%	90%	83%

Table 6.5: Performance of MARS compared with five methods (top five in sensitivity) in MCI/CN case.

ID	Method Name	SEN	SPE	PPV	NPV
0	**Voxel-MARS**	**78.38%**	**88.89%**	**76.32%**	**90.00%**
1.3.1 a	Voxel-STAND-D-gm	73%	85%	69%	87%
3.1.1	Hippo-Volume-F	73%	74%	56%	86%
3.1.2	Hippo-Volume-S	70%	73%	54%	84%
1.4.2 a	Voxel-Atlas-S-gm	68%	95%	86%	87%
2.3	Thickness-ROI	65%	94%	83%	85%

ble involve unique hierarchical IDs given in the original paper, abbreviations given to the methods, sensitivity, specificity, positive predictive value, and negative predictive value outcomes, from left to right.

Results on Table 6.4 suggest that the proposed technique (*Voxel-MARS*) provided the best sensitivity (SEN) and NPV among all of the other 28 procedures, with an adequate specificity of 86.42%. Besides, *Voxel-MARS* ranks 12th at specificity (SPE) and 21th at PPV among the procedures. However, difference of sensitivity in the positive direction (+12.12) is much greater than the loss in specificity (-2.97) when compared with the average performance levels of all of the other methods (see Table 6.3).

6.2.3 MCI vs. Normal

MCI/CN classification steps were executed on previously assigned training and test datasets. Results were demonstrated in Table 6.5. Similar to the AD/CN case, other methods assessed in [23] were rearranged in the order of attained sensitivity and the

Table 6.6: Performance of MARS compared with five methods (top five in sensitivity) in MCIc/MCInc case.

ID	Method Name	SEN	SPE	PPV	NPV
3.1.1	Hippo-Volume-F	70%	61%	50%	79%
0	**Voxel-MARS**	**62.16%**	**59.70%**	**46.00%**	**74.07%**
3.1.2	Hippo-Volume-S	62%	69%	52%	77%
1.5.1 a	Voxel-COMPARE-D-gm	62%	67%	51%	76%
1.3.1 a	Voxel-STAND-D-gm	57%	78%	58%	76%
1.5.1 b	Voxel-COMPARE-D-all	54%	78%	57%	75%

uppermost 5 were included in Table 6.5. Again, it was observed that the proposed method provides better sensitivity and NPV than all of the other techniques included in the research. In specificity, *Voxel-MARS* ranked 17th, and in PPV it ranked 12th, whereas, both of the observed results were above the average rates attained through the other techniques (see Table 6.3).

In both of the AD/CN and MCI/CN cases, optimal values for M_{max} and K_{max} became 11 and 1, respectively. Similarly, eventual number of BFs at the end of pruning appeared as 7 in both cases.

6.2.4 Converting MCI vs. Non-converting MCI

The authors in [23] stated that, only 15 out of 28 methods managed to produce meaningful results (with SEN and SPE values different from 100% or 0) in MCIc/MCInc classification. As with the former cases, these 15 techniques were ordered by SEN and the highest 5 were included in Table 6.6. *Voxel-MARS* ranked 2nd in terms of sensitivity. Proposed procedure ranked 16th in specificity (SPE), 12th in PPV, and 7th in NPV among other approaches. Like in the AD/CN group, the advance in SEN is greater than the downturn in terms of SPE (see Table 6.3).

A comprehensive quantitative assessment involving the performances of all methods including our approach is presented in Table 6.3. Success statistics of the other 28 techniques are introduced in the format of "average (%) ± standard deviation (%) [range (%)]". In both of the AD/CN and AD/MCI cases, all 28 methods managed to provide meaningful outcomes, whereas in prediction of conversion, only 15 of

Table 6.7: Confusion matrix regarding results acquired by CMARS in all three classification groups. These are the scores observed through cross-validation where the best AUC condition is met.

Confusion Matrix		True Classes					
		AD/CN		MCI/CN		MCIc/MCInc	
		H	D	H	D	H	D
Predictions	Healthy	71	20	67	13	42	16
	Diseased	10	47	14	24	25	21
	TOTALS	81	67	81	37	67	37

Table 6.8: SEN and SPE outcomes acquired by MARS and CMARS with best AUC values are compared.

		AD/CN	MCI/CN	MCIc/MCInc
MARS	SEN (%)	83.58	78.38	62.16
	SPE (%)	86.42	88.89	59.70
CMARS	SEN (%)	70.15	64.86	56.76
	SPE (%)	87.65	82.72	62.69

them did. Thus, in the third group, 13 methods producing null sensitivity were not included in calculation of performance statistics. The column *Diff.* shows the difference between the results acquired by our method and the averages achieved by other techniques.

6.3 Classification of Normal, AD, and MCI Brain MRI by CMARS

CMARS validation procedure was handled in a similar way to the MARS experiments, demonstrated in Section 6.2. Firstly, MARS models were built with the input parameters obtained through the parameter optimization procedure. *Prune* property of *Earth* package was set to "false" in order to acquire the model in its widest form in the beginning. The regulatory parameter ($\tilde{M}$) of CMARS was varied between 10^{-3} and 10^3 in logarithmic steps, and $\tilde{M} = 1$ was observed to be the value giving the maximum AUC among the trials.

In Table 6.7, prediction results, which were gathered through CMARS classifier applied on AD/CN, MCI/CN, and MCIc/MCInc groups, were shown. Additionally, a comparison between MARS and CMARS techniques in terms of the two major per-

Table 6.9: SEN and SPE scores acquired by utilization of central moments computed at different scales. Experiments were performed on AD/CN group and MARS method were employed for classification. (*) Scale 0 means the original image. (**) Earth package cannot construct MARS models at these dimensions.

Scale	0* (original)		2		3		4	
Dimensions	121-145-121		30-36-30		15-18-15		8-9-8	
Block sizes	1		64		512		4096	
(%)	SEN	SPE	SEN	SPE	SEN	SPE	SEN	SPE
Mean (0th)	83.58	78.38	79.01	73.13	87.65	52.24	80.25	71.64
Variance (1st)	-	-	80.25	65.67	79.01	76.12	83.95	70.15
Skewness (2nd)	-	-	79.01	56.72	79.01	73.13	67.90	68.66
Kurtosis (3rd)	-	-	67.90	73.13	69.14	70.15	59.26	64.18
0th +1st	-	-	-**	-**	**88.89**	**71.64**	88.89	68.66
All 4 combined	-	-	-**	-**	74.07	71.64	80.25	47.76

formance statistics, SEN and SPE, can be found in Table 6.8. It was observed that the maximum performance scores acquired by MARS cannot be reached, in terms of sensitivity. However, in the AD/CN classification case, a higher specificity (87.65%) compared to that of MARS (86.42%) was obtained.

In Figure 6.4, overall ACC scores provided by MARS and CMARS under varying M_{max} input (for $K_{max} \in 1, 2$) are introduced. As seen on the first graph, MARS outperformed CMARS at point where the maximum accuracy was obtained, however, CMARS reached up to a better accuracy score when the number of basis functions was increased. Similarly, in the nonlinearity case demonstrated by the second graph, higher accuracy was observed by utilization of CMARS. These observations verify the theoretical proposition in [103] that CMARS performs better when the complexity of the model increases, thus, enhances *robustness*.

6.4 Classification Based on Feature-Based Morphometry

6.4.1 Image Moments

The feature sets were prepared by downsampling and computation of image moments at different scales, as it is described in Subsection 3.3.1. MARS classifier was applied on the AD/CN dataset. Optimal parameters acquired by the cross-validation proce-
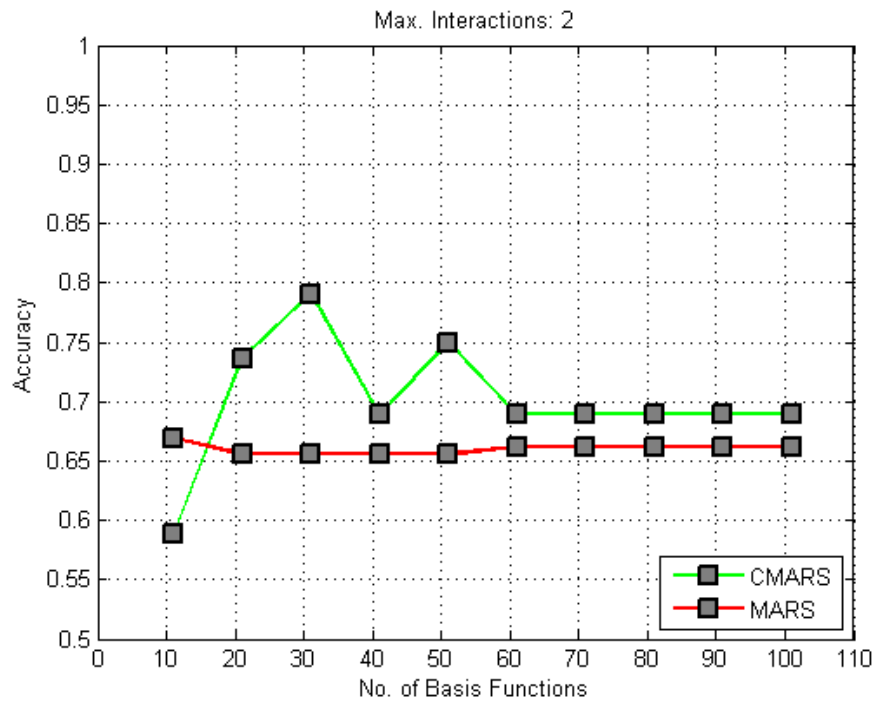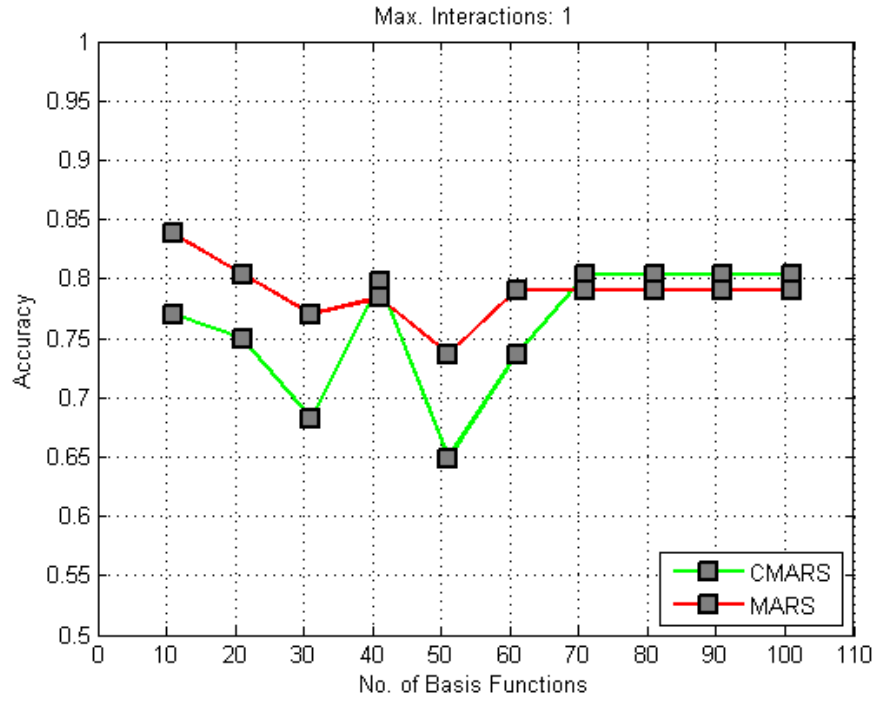
Figure 6.4: Overall accuracy (ACC) of classical MARS and CMARS under varying number of basis-functions (M_{max}) with (above) $K_{max} = 1$ (below) $K_{max} = 2$ (AD/CN group).

Table 6.10: Classification accuracy acquired by SIFT and HOG feature descriptors. Results are compared with our methodology. (*) The study does not involve diagnosis in early phase, it implicates quantification of amyloid status.

	SIFT	HOG	Voxel-MARS
Our implementation	0.71	0.74	**0.85**
(Toews et al., 2010) [95]	0.71	-	-
(Daliri, 2012) [25]	0.75	-	-
(Chen et al., 2014) [20]	0.74	-	-
(Cattell et al., 2016) [15]	-	0.90 *	-
(Ameer et al., 2017) [67]	-	0.68	-
(Unay and Ekin, 2011) [97]	-	0.74	-

dure explained in Section 6.1 were used in model training. Previously reserved blind data were employed for testing the models.

As it is presented in Table 6.9, a successful outcome was obtained by combining mean and variance (0th and 1st moments, respectively) features calculated in the 3rd scale. In this case, number of features became 24,300, with $AUC = 0.8778$. Both AUC and SPE values were remarkably high with this feature setting, however, a low SPE value was obtained.

6.4.2 Feature Descriptors

SIFT and HOG descriptors were derived from the original MRI volumes belonging to AD/CN group. For the extraction of SIFT and HOG features from 3D image data, recently implemented *bbrister/SIFT3D*[2] and *3D Voxel HOG*[3] [30] software packages were used in MATLAB environment, respectively.

The number of descriptor vectors which are provided by SIFT, varies between the keypoints. Hence, these descriptor vectors cannot be directly employed to form a feature space for classification model training. In order to solve this problem, a *Bag of (Visual) Words* (BoW) [109] model was constructed to compose a feature space by using the SIFT descriptors for classification. The co-dimensional vectors - of size 768 - were clustered into 100 groups using *k*-means clustering approach. Represen-

---

[2] MATLAB *bbrister/SIFT3D* package. [81].
[3] MATLAB package for Histogram of Orientated Gradients 3D (*3D Voxel HOG*) [29].

ting each keypoint, a 100-bin histogram of descriptors was formed by considering vectorial distances of the descriptors to the computed cluster centroids. In the second case, HOG descriptors were directly used for training the classifier. In both cases, class predictions were made by MARS method.

Table 6.10 shows the performance outcomes reported in the papers included in Subsections 3.3.2 and 3.3.3. Among the multiple results presented by the authors, the ones that were gathered by the datasets having the most analogous characteristics with ours were selected. In the table, if it is available, the overall accuracy values were used. Otherwise, the equal error rates (EER) were included as the performance measure. The first row shows the results that we acquired by the AD/CN dataset through our implementation. It is observed that, our proposed methodology outperformed FBM-based approach in early detection of AD. A wider discussion on this issue can be found in the following section.

6.5 Discussion

6.5.1 Challenges

ADNI image archive contains thousands of high resolution, volumetric brain MR images belonging to the subjects who consulted to the health centers with symptoms of dementia. Usually, a participant's brain was scanned multiple times, including the baseline examination and the regular visits lasting at least for 18 months. In the beginning, it seemed very difficult to compose a data subset with appropriate heterogeneous characteristics to work on. This problem was tackled with adopting the dataset that was formed by Cuignet et al. (2011) [23], as a part of their comprehensive study involving comparison of major structural MRI-based methods in the field. By embracing the equivalent image set, having heterogeneous and balanced data classes was ensured. Undoubtedly, this preference affected the objectivity of the validation part of our study, in the positive direction.

The derivation of tissue probability maps from the 509 high resolution images of our dataset involved a process which induces a respectively high memory and time cost, along with the programming efforts given. It was an error-prone process and the repe-

titions were coming along with a high return. Learning the script-based programming interface of the *SPM Toolbox* helped in dealing with such problems.

In consequence, working with the intensity-based features yielded a small sample size problem, which is very familiar to the researchers in field of medical image analysis. Nonetheless, throughout the literature investigation, it was observed that larger datasets does not necessarily imply more successful results. The issue was solved with the help of a properly designed dimensionality reduction procedure, which became one of the core components of this thesis.

MARS has the ability to produce fairly high-dimensional, possibly nonlinear and complex models. This, together with the high dimensional data vectors, caused a problem in data visualization. Although it was defeated by the approach explained in Chapter 5, it is an open issue to find better ways of demonstrating high-dimensional problem spaces.

Finally, confronting with the complexities behind the optimization part included in CMARS algorithm enriched this thesis, substantially. Applying the discretization procedure as it was introduced in [103], and transforming the highly-coupled constraints into a form that they can be modeled by the *MOSEK Optimization Toolbox* were the most challenging programming tasks involved in our study.

6.5.2 Voxel-Based Morphometry vs. Feature Descriptors

At the first glance, performing the class prediction over the feature descriptors like SIFT and HOG may seem advantageous in several aspects. The most important superiority of these methods are their independence from certain image properties such as scale, illumination and orientation. As a consequence, they remove the obligation to ensure voxel-wise one-to-one correspondence between the images. Therefore, the information loss due to image manipulations are minimized.

However, it was observed that, neither in the previously published studies, nor by our implementation, the desired performance level cannot be reached by employing the Feature-based Morphometry approach (see Subsections 3.3.2, 3.3.3, and 6.4.2). Indeed, this is understandable by virtue of a clear reasoning.

Obviously, the addressed problem is to identify signs of the effect of AD in "very early" phases - sometimes, even before the existence - of the disease. This time period coincides with the interval that the disease related microscopic changes are just beginning to sum up to the voxel-level, macroscopic structural alterations. In this phase, these changes are barely visible - in most of the cases, invisible - to the naked eye of an expert. The aforementioned feature descriptors, on the other hand, are known to work well with certain structures, such as edges, corners and blobs. However, it is noteworthy that, utilization of HOG in [15] resulted in a remarkably well accuracy in prediction of the amyloid status (see Table 6.10). This inference suggests that, even they are not sufficiently successful in detecting the existence of AD prior to the onset, these feature descriptors are promising in quantification of the amyloid plaques, which is directly associated with the stage of disease progression in the oncoming periods.

6.5.3 Feature Selection Methodology

As explained in Chapter 4, a hybrid, 3-step feature selection procedure incorporating both statistical analysis and the domain knowledge was developed to reduce the dimensions. For inclusion of the domain knowledge, a novel, heuristic approach was established, validated and formulated in mathematical terms.

Employing a procedure that is not transforming the feature space into another coordinate system, made it possible to locate the actual points of interest. This makes our system a potential tool for detailed anatomical assessment, beyond being a *black-box*, producing plain class predictions. As with being more applicable in related problem domains, the generated algorithm was proved to be more effective in determination of significant features, when compared with several of the widely-employed techniques (see Table 4.2).

6.5.4 MARS as a Classifier

Many benefits are provided with choosing MARS as the method for classification. First of all, the model functions, which are gathered at the end of the training phase,

has the ability to reveal points of interest, as well as interconnections between these points, by its nature. The model functions are easy to interpret, since they exist in form of linear combination of linearly independent basis-functions, which carry the information corresponding to one or more voxels. As with the feature selection methodology, this property of MARS classifiers provided an insight into the anatomical alterations in the early phases of AD. Also, the global, adaptive behavior of MARS - and CMARS - is in an analogous manner with our *brain-as-a-whole* approach.

In addition to these, MARS facilitates highly flexible models which are able to fit remarkably complex data points. High flexibility implies higher ability to learn, however, overuse of this capability may cause overfitting problem. Examples of this condition were presented throughout this chapter.

6.5.5 MARS vs. CMARS

MARS resolves the overfitting problem with a process termed as the *backward-step*, which has been gone over multiple times throughout this thesis. The process involves simplification of the model with the help of a technique, named as *generalized cross-validation*. Model simplification is performed by elimination of the least relevant basis-functions from the model function. This can be reasonably interpreted as throwing off a certain amount of - possibly valuable - information.

CMARS replaces this part of the algorithm in order to avoid the information loss. It incorporates the famous Tikhonov regularization and Conic Quadratic Optimization techniques to enhance generalization ability of the model, instead. The amount of regularization can be controlled over an additional parameter, as explained in detail in Chapter 2.

Along the experiments, CMARS was observed to succeed performing well. In AD/CN and MCIc/MCInc classification cases, CMARS outperformed MARS in terms of specificity with scores of 87.65% and 62.69%, respectively (see Table 6.8). In general, it managed to produce accuracy and sensitivity values in acceptable ranges. It is seen that in higher dimensional models, the accuracy of CMARS outperformed that of MARS (cf. Figure 6.4). However, it did not achieve to exceed the maximum scores

that were acquired by MARS in terms of these two measures.

As a result, MARS and CMARS were utilized for the first time in field of medical image analysis. It was seen that, MARS performed remarkably well when compared with well-known techniques which are previously tested and validated. Even though the peak performance of MARS cannot be reached by CMARS, the algorithm was observed to be promising in terms of robustness and stability.

CHAPTER 7

CONCLUSIONS AND OUTLOOK

7.1 Conclusions

In this study, a complete procedure which improves the early detection of AD and MCI by analyzing structural brain MRI volumes, and builds a foundation for a fully-automated computer-aided diagnosis system was introduced. An elaborate image processing scheme involving derivation of voxel intensity-based GM, WM, and CSF tissue probabilities as feature vectors was applied to 3D T1-W MRI volumes gathered from the ADNI database. Results obtained through this approach is quantitatively examined and compared with several previously developed methods (based on searching for more "descriptive" features) falling into the discipline of computer vision, such as SIFT and HOG. It was observed that, when considered within the limitations of our work, the intensity-based approach has demonstrated higher success rates in performance than the other. In the *Discussion* section (Section 6.5), a comprehensive discussion on these consequences is presented.

The decision of treating tissue probability map entries assembling to "whole brain" as potential features (i.e., *voxel-as-feature* approach) effectively yielded a small sample size problem, which necessitates a procedure capable to deal with the problems caused by the existence of high dimensionality. A partially new, 3-step, hybrid feature selection methodology has been developed to identify features with a fairly high discriminatory strength. The procedure begins with the creation of an initial mask using a *general linear model*, continues with the insertion of domain rules based on expert knowledge into the model to remove some features by looking at the tissue probability distributions, and ends with the *within-class norm thresholding* stage. This final step

involves a method which removes the dimensions from the data matrix, when corresponding Euclidean sample norms are not large enough compared to the average norms of feature vectors within their tissue class. In addition, an evaluation of this procedure involving a comparison with commonly used space-transforming dimensionality reduction techniques was performed. It was seen that the proposed method outperformed all of those techniques and, what is more, it performed better than the "no reduction" case.

A non-parametric, adaptive extension of decision trees (particularly, of CART), namely MARS, was used to generate linear and nonlinear models which function as separating surfaces. CMARS is a recent technique which replaces the backward step of classical MARS with a process involving minimization of a PRSS function. It is accomplished by updating all of the coefficients instead of removing basis functions from the model. In the frame of this study, CMARS was also implemented and performance of the method as a classifier was evaluated. It was observed that, in this way, all of the relevant information coming from the observations are preserved, while overfitting problem is avoided. Our CMARS implementation produced acceptable results which are also comparable with the ones obtained by MARS in previous phases of the study.

Since it resulted in the first time utilization of MARS in classification of data with such a high degree of dimensionality, use of MARS in the field of medical image analysis provided one of the most important contributions of this study. It ended up with a notable success among the techniques that employ voxel intensity-based features derived from different parts (or whole) of the brain MRI as the predictor variables. It was proved that MARS can perform outstandingly as a nonlinear classifier with the ability to generate complex and stable models with very high-dimensional training data and no fixed parametric form presumption. Use of the methods made it possible to detect the pixel-level macroscopic effects induced by accumulation of microscopic changes occurring in the early stages of AD and converting MCI.

The latest technological advances that apply to any kind of measuring device (an MRI scanner in our case) create a massive and ever-increasing source of information (i.e., *big data*) to analyze. This phenomenon encourages us, engineers, medical scientists,

and mathematicians to incorporate data analysis into our focus of interest. When considered in this context, concepts of statistical learning and applied mathematics find a widespread use. Computer-aided diagnosis of AD (and MCI) at early onset through the analysis of medical images involves utilization of techniques from engineering sciences such as *statistical learning*, *mathematical modeling*, and *optimization*. For this reason, the development of computer-assisted AD and MCI diagnostic tools can still be seen as an important research topic in the fields of Biomedical Engineering and Neuroscience.

This comparative study showed that our approach is one of the most promising advances in the domain.

7.2 A Future Outlook

The research on MARS and CMARS will be extended to include very new mathematics-supported methods of *Statistical Learning*, *Machine Learning* and *Data Mining*, such as RMARS [72], RCMARS [73], RCGPLM [71], and upcoming methods, which can be used comfortably, elegantly and successfully, for further advances in detection of AD. Our global approach of analyzing the brain as a whole, and treating all voxels as candidate features will also be enlarged to question if the proposed approach may be effectively used for early diagnosis of neurodegenerative diseases other than AD and MCI.

With this thesis, a solid basis for an efficient and stable software system is founded. In the near future, essential parts of our implementation will be refactored and packed in order to form a fully-automated software product with the support of graphical user interfaces (GUI) and powerful visualization tools. A complete system that can effectively be employed in clinics for both screening and patient follow-up will be purposed.

Another possible future direction could be to extend the procedure in the way that it is able to incorporate outputs in terms of probabilities of having AD, MCI, and converting MCI, rather than producing plain class labels as predictions. This could be accomplished through reassessment of the model output response function as a

generalized probability distribution. Also, a study on "to which extent the proposed approach could be modified to handle early detection of other neurodegenerative diseases" will be conducted, in the near future. Finally, it would be interesting to enrich the datasets using images from multiple sources to investigate how multi-modality in the feature sets effects the overall prediction performance.

REFERENCES

[1] S. Adaszewski, J. Dukart, F. Kherif, R. Frackowiak, B. Draganski, A. D. N. Initiative, et al. How early can we predict Alzheimer's disease using computational anatomy? *Neurobiology of Aging*, 34(12):2815–2826, 2013.

[2] E. Álvarez-Miranda, H. Farhan, M. Luipersbeck, and M. Sinnl. A bi-objective network design approach for discovering functional modules linking Golgi apparatus fragmentation and neuronal death. *Annals of Operations Research*, pages 1–26, 2016.

[3] Alzheimer's Disease Neuroimaging Initiative. Alzheimer's Disease Neuroimaging Initiative. `http://adni.loni.usc.edu`, 2003. [Online; accessed 10-Aug-2017].

[4] L. G. Apostolova, A. E. Green, S. Babakchanian, K. S. Hwang, Y.-Y. Chou, A. W. Toga, and P. M. Thompson. Hippocampal atrophy and ventricular enlargement in normal aging, mild cognitive impairment and Alzheimer's disease. *Alzheimer disease and associated disorders*, 26(1):17, 2012.

[5] M. ApS. The MOSEK optimization toolbox for MATLAB manual. version 7.1 (revision 28). `http://docs.mosek.com/7.1/toolbox/index.html`, 2015. [Online; accessed 10-Aug-2017].

[6] J. Ashburner. A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1):95 – 113, 2007.

[7] J. Ashburner. Computational anatomy with the SPM software. *Magnetic Resonance Imaging*, 27(8):1163 – 1174, 2009. Proceedings of the International School on Magnetic Resonance and Brain Function.

[8] J. Ashburner and K. J. Friston. Voxel-based morphometry—the methods. *Neuroimage*, 11(6):805–821, 2000.

[9] J. Ashburner and K. J. Friston. Unified segmentation. *NeuroImage*, 26(3):839 – 851, 2005.

[10] A. Association et al. 2017 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 13(4):325–373, 2017.

[11] C. Boutet, M. Chupin, S. Lehéricy, L. Marrakchi-Kacem, S. Epelbaum, C. Poupon, C. Wiggins, A. Vignaud, D. Hasboun, B. Defontaines, et al. Detection of volume loss in hippocampal layers in Alzheimer's disease using 7 t mri: A feasibility study. *Neuroimage: Clinical*, 5:341–348, 2014.

[12] M. Brown and D. G. Lowe. Recognising panoramas. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 1218–1225 vol.2, 2003.

[13] C. Cabral, P. M. Morgado, D. C. Costa, M. Silveira, A. s Disease Neuroimaging Initiative, et al. Predicting conversion from mci to ad with fdg-pet brain images at different prodromal stages. *Computers in biology and medicine*, 58:101–109, 2015.

[14] R. Casanova, C. T. Whitlow, B. Wagner, J. Williamson, S. A. Shumaker, J. A. Maldjian, and M. A. Espeland. High dimensional classification of structural MRI Alzheimer's disease data based on large scale regularization. *Frontiers in neuroinformatics*, 5, 2011.

[15] L. Cattell, G. Platsch, R. Pfeiffer, J. Declerck, J. A. Schnabel, C. Hutton, A. D. N. Initiative, et al. Classification of amyloid status using machine learning with histograms of oriented 3d gradients. *NeuroImage: Clinical*, 12:990–1003, 2016.

[16] W. U. A. D. R. Center. The open access series of imaging studies (OASIS). `http://www.oasis-brains.org/`. [Online; accessed 10-Aug-2017].

[17] A. Çevik, G.-W. Weber, B. M. Eyüboğlu, and K. K. Oğuz. Voxel-MARS: a method for early detection of Alzheimer's disease by classification of structural brain MRI. *Annals of Operations Research*, Feb 2017.

[18] R. Chaves, J. Ramìrez, J. Górriz, M. López, D. Salas-Gonzalez, I. Álvarez, and F. Segovia. SVM-based computer-aided diagnosis of the Alzheimer's disease using t-test NMSE feature selection with feature correlation weighting. *Neuroscience Letters*, 461(3):293 – 297, 2009.

[19] G. Chen, B. D. Ward, C. Xie, W. Li, Z. Wu, J. L. Jones, M. Franczak, P. Antuono, and S.-J. Li. Classification of alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional mr imaging. *Radiology*, 259(1):213–221, 2011.

[20] Y. Chen, J. Storrs, L. Tan, L. J. Mazlack, J.-H. Lee, and L. J. Lu. Detecting brain structural changes as biomarker from magnetic resonance images using a local feature based svm approach. *Journal of neuroscience methods*, 221:22–31, 2014.

[21] M. Chupin, E. Gérardin, R. Cuingnet, C. Boutet, L. Lemieux, S. Lehéricy, H. Benali, L. Garnero, and O. Colliot. Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. *Hippocampus*, 19(6):579–587, 2009.

[22] M. Chupin, A. Hammers, R. Liu, O. Colliot, J. Burdett, E. Bardinet, J. Duncan, L. Garnero, and L. Lemieux. Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: Method and validation. *NeuroImage*, 46(3):749 – 761, 2009.

[23] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M.-O. Habert, M. Chupin, H. Benali, and O. Colliot. Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. *NeuroImage*, 56(2):766 – 781, 2011.

[24] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[25] M. R. Daliri. Automated diagnosis of alzheimer disease using the scale-invariant feature transforms in magnetic resonance images. *Journal of medical systems*, 36(2):995–1000, 2012.

[26] C. Davatzikos, P. Bhatt, L. M. Shaw, K. N. Batmanghelich, and J. Q. Trojanowski. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiology of aging*, 32(12):2322–e19, 2011.

[27] C. Davatzikos, Y. Fan, X. Wu, D. Shen, and S. M. Resnick. Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. *Neurobiology of Aging*, 29(4):514 – 523, 2008.

[28] R. S. Desikan, H. J. Cabral, C. P. Hess, W. P. Dillon, C. M. Glastonbury, M. W. Weiner, N. J. Schmansky, D. N. Greve, D. H. Salat, R. L. Buckner, et al. Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. *Brain*, 132(8):2048–2057, 2009.

[29] R. Dupre. Histogram of orientated gradients 3D (3D Voxel HOG). `https://www.mathworks.com/matlabcentral/fileexchange/55978-histogram-of-orientated-gradients-3d--3d-voxel-hog-`, 2016. [Online; accessed 10-Aug-2017].

[30] R. Dupre and V. Argyriou. 3d voxel hog and risk estimation. In *Digital Signal Processing (DSP), 2015 IEEE International Conference on*, pages 482–486. IEEE, 2015.

[31] M. Dyrba, M. Ewers, M. Wegrzyn, I. Kilimann, C. Plant, A. Oswald, T. Meindl, M. Pievani, A. L. Bokde, A. Fellgiebel, et al. Robust automated detection of microstructural white matter degeneration in alzheimer's disease using machine learning classification of multicenter dti data. *PloS one*, 8(5):e64925, 2013.

[32] S. F. Eskildsen, P. Coupé, V. S. Fonov, J. C. Pruessner, D. L. Collins, A. D. N. Initiative, et al. Structural imaging biomarkers of Alzheimer's disease: predicting disease progression. *Neurobiology of aging*, 36:S23–S31, 2015.

[33] S. F. Eskildsen, P. Coupé, D. García-Lorenzo, V. Fonov, J. C. Pruessner, D. L. Collins, A. D. N. Initiative, et al. Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the adni cohort using patterns of cortical thinning. *Neuroimage*, 65:511–521, 2013.

[34] G. Flitton, T. Breckon, and N. Megherbi Bouallagu. Object recognition using 3D SIFT in complex CT volumes. In *Proceedings of the British Machine Vision Conference*, pages 11.1–11.12. BMVA Press, 2010. doi:10.5244/C.24.11.

[35] W. T. C. for Neuroimaging. Statistical Parametric Mapping. `http://www.fil.ion.ucl.ac.uk/spm/`, 2011. [Online; accessed 10-Aug-2017].

[36] R. Frackowiak, K. Friston, C. Frith, R. Dolan, C. Price, S. Zeki, J. Ashburner, and W. Penny. *Human Brain Function*. Academic Press, 2nd edition, 2003.

[37] L. A. Francis, C. S. J. Taylor, F. Morrisette, and C. Yaure. 269 martian chronicles: Is MARS better than neural networks? `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.375.9523`, 2003. [Online; accessed 10-Aug-2017].

[38] J. H. Friedman. Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1):pp. 1–67, 1991.

[39] K. Friston, A. Holmes, J.-B. Poline, C. Price, and C. Frith. Detecting activations in PET and fMRI: Levels of inference and power. *NeuroImage*, 4(3):223 – 235, 1996.

[40] K. J. Friston, C. Frith, P. Liddle, and R. Frackowiak. Comparing functional PET images: the assessment of significant change. *Journal of Cerebral Blood Flow & Metabolism*, 11(4):690–699, 1991.

[41] G. Fung and J. Stoeckel. SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information. *Knowledge and Information Systems*, 11(2):243–258, 2007.

[42] K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, D. Rueckert, A. D. N. Initiative, et al. Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *NeuroImage*, 65:167–175, 2013.

[43] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, 2 edition, 2009.

[44] C. Hinrichs, V. Singh, L. Mukherjee, G. Xu, M. K. Chung, S. C. Johnson, A. D. N. Initiative, et al. Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. *Neuroimage*, 48(1):138–149, 2009.

[45] V. K. Ithapu, V. Singh, O. C. Okonkwo, R. J. Chappell, N. M. Dowling, S. C. Johnson, A. D. N. Initiative, et al. Imaging-based enrichment criteria using deep learning algorithms for efficient clinical trials in mild cognitive impairment. *Alzheimer's & Dementia*, 11(12):1489–1499, 2015.

[46] C. R. Jack, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward, A. M. Dale, J. P. Felmlee, J. L. Gunter, D. L. Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C. S. DeCarli, G. Krueger, H. A. Ward, G. J. Metzger, K. T. Scott, R. Mallozzi, D. Blezek, J. Levy, J. P. Debbins, A. S. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler, and M. W. Weiner. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*, 27(4):685–691, 2008.

[47] C. R. Jack, D. S. Knopman, W. J. Jagust, L. M. Shaw, P. S. Aisen, M. W. Weiner, R. C. Petersen, and J. Q. Trojanowski. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology*, 9(1):119–128, 2010.

[48] A. Jain, R. P. W. Duin, and J. Mao. Statistical pattern recognition: a review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(1):4–37, 2000.

[49] G. Jekabsons. ARESLab: Adaptive Regression Splines toolbox for MATLAB/Octave. `http://www.cs.rtu.lv/jekabsons/`, 2011. [Online; accessed 10-Aug-2017].

[50] B. Jie, D. Zhang, W. Gao, Q. Wang, C.-Y. Wee, and D. Shen. Integration of network topological and connectivity properties for neuroimaging classification. *IEEE transactions on biomedical engineering*, 61(2):576–589, 2014.

[51] D. K. Jones, M. R. Symms, M. Cercignani, and R. J. Howard. The effect of filter size on VBM analyses of DT-MRI data. *Neuroimage*, 26(2):546–554, 2005.

[52] A. Khazaee, A. Ebrahimzadeh, and A. Babajani-Feremi. Identifying patients with Alzheimer's disease using resting-state fmri and graph theory. *Clinical Neurophysiology*, 126(11):2132–2141, 2015.

[53] S. Klöppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, C. R. Jack, J. Ashburner, and R. S. J. Frackowiak. Automatic classification of MR scans in Alzheimer's disease. *Brain*, 131(3):681–689, 2008.

[54] R. La Joie, A. Perrotin, V. De La Sayette, S. Egret, L. Doeuvre, S. Belliard, F. Eustache, B. Desgranges, and G. Chételat. Hippocampal subfield volumetry in mild cognitive impairment, Alzheimer's disease and semantic dementia. *NeuroImage: Clinical*, 3:155–162, 2013.

[55] S. Li, F. Shi, F. Pu, X. Li, T. Jiang, S. Xie, and Y. Wang. Hippocampal shape analysis of Alzheimer disease based on machine learning methods. *American Journal of Neuroradiology*, 28(7):1339–1345, 2007.

[56] F. Liu, L. Zhou, C. Shen, and J. Yin. Multiple kernel learning in the primal for multimodal alzheimer's disease classification. *IEEE journal of biomedical and health informatics*, 18(3):984–990, 2014.

[57] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[58] B. Magnin, L. Mesrob, S. Kinkingnéhun, M. Pélégrini-Issac, O. Colliot, M. Sarazin, B. Dubois, S. Lehéricy, and H. Benali. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. *Neuroradiology*, 51(2):73–83, 2009.

[59] W. L. Martinez and A. R. Martinez. *Computational Statistics Handbook with MATLAB*. London: Chapman and Hall, CRC, 2002.

[60] J. Mazziotta, A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, T. Paus, G. Simpson, B. Pike, et al. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 356(1412):1293–1322, 2001.

[61] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 34(7):939, 1984.

[62] S. Milborrow. earth: Multivariate Adaptive Regression Splines. `https://cran.r-project.org/package=earth`, 2011. [Online; accessed 10-Aug-2017].

[63] C. Misra, Y. Fan, and C. Davatzikos. Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to ad: Results from ADNI. *NeuroImage*, 44(4):1415 – 1422, 2009.

[64] G. G. Moisen and T. S. Frescino. Comparing five modelling techniques for predicting forest characteristics. *Ecological modelling*, 157(2):209–225, 2002.

[65] B. Mwangi, T. Tian, and J. Soares. A review of feature reduction techniques in neuroimaging. *Neuroinformatics*, 12(2):229–244, 2014.

[66] T. M. Nir, J. E. Villalon-Reina, G. Prasad, N. Jahanshad, S. H. Joshi, A. W. Toga, M. A. Bernstein, C. R. Jack, M. W. Weiner, P. M. Thompson, et al.

Diffusion weighted imaging-based maximum density path analysis and classification of Alzheimer's disease. *Neurobiology of aging*, 36:S132–S140, 2015.

[67] A. Nisha.S, S. Nisha.S, and M. Sathik. A study on SURF and HOG descriptors for Alzheimer's disease. *International Research Journal of Engineering and Technology (IRJET)*, 4:626–632, 2017.

[68] W. U. S. of Medicine. The clinical dementia rating. `http://alzheimer.wustl.edu/cdr/cdr.htm`. [Online; accessed 10-Aug-2017].

[69] L. L. of Neuroimaging. ICBM Probabilistic Atlases Methods. `http://www.loni.usc.edu/atlases/Atlas_Methods.php?atlas_id=7`. [Online; accessed 10-Aug-2017].

[70] N. Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.

[71] A. Özmen and G. W. Weber. Robust conic generalized partial linear models using rcmars method-a robustification of cgplm. In *AIP Conference Proceedings*, volume 1499, pages 337–343. AIP, 2012.

[72] A. Özmen and G. W. Weber. Rmars: robustification of multivariate adaptive regression spline under polyhedral uncertainty. *Journal of Computational and Applied Mathematics*, 259:914–924, 2014.

[73] A. Özmen, G. W. Weber, İ. Batmaz, and E. Kropat. Rcmars: Robustification of cmars with different scenarios under polyhedral uncertainty set. *Communications in Nonlinear Science and Numerical Simulation*, 16(12):4780–4787, 2011.

[74] P. Padilla, M. Lopez, J. Gorriz, J. Ramirez, D. Salas-Gonzalez, and I. Alvarez. NMF-SVM based cad tool applied to functional brain images for the diagnosis of Alzheimer's disease. *Medical Imaging, IEEE Transactions on*, 31(2):207–216, 2012.

[75] M. Pagani, F. De Carli, S. Morbelli, J. Öberg, A. Chincarini, G. Frisoni, S. Galluzzi, R. Perneczky, A. Drzezga, B. van Berckel, et al. Volume of interest-based [18 F] fluorodeoxyglucose PET discriminates MCI converting to alzheimer's disease from healthy controls. A European Alzheimer's Disease Consortium (EADC) study. *NeuroImage: Clinical*, 7:34–42, 2015.

[76] H. Park, J.-j. Yang, J. Seo, J.-m. Lee, et al. Dimensionality reduced cortical features and their use in predicting longitudinal changes in Alzheimer's disease. *Neuroscience letters*, 550:17–22, 2013.

[77] G. Prasad, S. H. Joshi, T. M. Nir, A. W. Toga, P. M. Thompson, A. D. N. I. (ADNI, et al. Brain connectivity and novel network measures for Alzheimer's disease classification. *Neurobiology of aging*, 36:S121–S131, 2015.

[78] J. Ramìrez, J. Górriz, D. Salas-Gonzalez, A. Romero, M. López, I. Álvarez, and M. Gómez-Rìo. Computer-aided diagnosis of Alzheimer's type dementia combining support vector machines and discriminant set of features. *Information Sciences*, 237(0):59 – 72, 2013.

[79] J. Ramírez, J. Górriz, F. Segovia, R. Chaves, D. Salas-Gonzalez, M. López, I. Álvarez, and P. Padilla. Computer aided diagnosis system for the Alzheimer's disease based on partial least squares and random forest spect image classification. *Neuroscience letters*, 472(2):99–103, 2010.

[80] S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, and C. Davatzikos. A review on neuroimaging-based classification studies and associated feature extraction methods for alzheimer's disease and its prodromal stages. *NeuroImage*, 2017.

[81] B. Rister. bbrister/SIFT3D. `https://www.mathworks.com/matlabcentral/fileexchange/52800-bbrister-sift3d`, 2015. [Online; accessed 10-Aug-2017].

[82] C. Salvatore, A. Cerasa, P. Battista, M. C. Gilardi, A. Quattrone, I. Castiglioni, A. D. N. Initiative, et al. Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach. *Frontiers in neuroscience*, 9, 2015.

[83] Z. Sankari and H. Adeli. Probabilistic neural networks for diagnosis of alzheimer's disease using conventional and wavelet coherence. *Journal of neuroscience methods*, 197(1):165–170, 2011.

[84] A. Savio, M. García-Sebastián, C. Hernández, M. Graña, and J. Villanúa. Classification results of artificial neural networks for alzheimer's disease detection. In *IDEAL*, volume 5788, pages 641–648. Springer, 2009.

[85] T. M. Schouten, M. Koini, F. de Vos, S. Seiler, J. van der Grond, A. Lechner, A. Hafkemeijer, C. Möller, R. Schmidt, M. de Rooij, et al. Combining anatomical, diffusion, and resting state functional magnetic resonance imaging for individual classification of mild and moderate Alzheimer's disease. *NeuroImage: Clinical*, 11:46–51, 2016.

[86] P. Scovanner, S. Ali, and M. Shah. A 3-dimensional SIFT descriptor and its application to action recognition. In *Proceedings of the 15th ACM International Conference on Multimedia*, MM '07, pages 357–360, New York, NY, USA, 2007. ACM.

[87] S. Se, D. Lowe, and J. Little. Vision-based mobile robot localization and mapping using scale-invariant features. In *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164)*, volume 2, pages 2051–2058 vol.2, 2001.

[88] B. Shi, Y. Chen, P. Zhang, C. D. Smith, J. Liu, A. D. N. Initiative, et al. Non-linear feature transformation and deep fusion for alzheimer's disease staging analysis. *Pattern Recognition*, 63:487–498, 2017.

[89] D. T. Shih, S. B. Kim, V. C. P. Chen, J. M. Rosenberger, and V. L. Pilla. Efficient computer experiment-based optimization through variable selection. *Annals of Operations Research*, 216(1):287–305, 2014.

[90] B. Sirmacek and C. Unsalan. Urban-area and building detection using SIFT keypoints and graph theory. *IEEE Transactions on Geoscience and Remote Sensing*, 47(4):1156–1167, April 2009.

[91] J. Strickland. *Predictive modeling and analytics*. Lulu. com, 2014.

[92] H.-I. Suk, S.-W. Lee, D. Shen, A. D. N. Initiative, et al. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage*, 101:569–582, 2014.

[93] H.-I. Suk, S.-W. Lee, D. Shen, A. D. N. Initiative, et al. Deep ensemble learning of sparse regression models for brain disease diagnosis. *Medical image analysis*, 37:101–113, 2017.

[94] P. Tiraboschi, L. A. Hansen, L. J. Thal, and J. Corey-Bloom. The importance of neuritic plaques and tangles to the development and evolution of AD. *Neurology*, 62(11):1984–1989, 2004.

[95] M. Toews, W. Wells, D. L. Collins, and T. Arbel. Feature-based morphometry: Discovering group-related anatomical patterns. *NeuroImage*, 49(3):2318–2327, 2010.

[96] A. Tomescu and F. Tomescu. Matched filters for system identification. *IEEE Transactions on Automatic Control*, 15(3):398–399, 1970.

[97] D. Unay and A. Ekin. Dementia diagnosis using similar and dissimilar retrieval items. In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pages 1889–1892. IEEE, 2011.

[98] A. A. USA. 2017 Alzheimer's disease facts and figures. `http://www.alz.org/documents_custom/2017-facts-and-figures.pdf`, 2017. [Online; accessed 10-Aug-2017].

[99] L. J. van der Maaten, E. O. Postma, and H. J. van den Herik. Dimensionality reduction: A comparative review. *Journal of Machine Learning Research*, 10(1-41):66–71, 2009.

[100] R. Vandenberghe, N. Nelissen, E. Salmon, A. Ivanoiu, S. Hasselbalch, A. Andersen, A. Korner, L. Minthon, D. J. Brooks, K. Van Laere, et al. Binary classification of 18 F-flutemetamol PET using machine learning: Comparison with visual reads and structural MRI. *NeuroImage*, 64:517–525, 2013.

[101] P. Vemuri, J. L. Gunter, M. L. Senjem, J. L. Whitwell, K. Kantarci, D. S. Knopman, B. F. Boeve, R. C. Petersen, and C. R. J. Jr. Alzheimer's disease diagnosis in individual subjects using structural MR images: Validation studies. *NeuroImage*, 39(3):1186 – 1197, 2008.

[102] L. Wang, F. Beg, T. Ratnanather, C. Ceritoglu, L. Younes, J. C. Morris, J. G. Csernansky, and M. I. Miller. Large deformation diffeomorphism and momentum based hippocampal shape discrimination in dementia of the Alzheimer type. *IEEE transactions on medical imaging*, 26(4):462–470, 2007.

[103] G.-W. Weber, I. Batmaz, G. Köksal, P. Taylan, and F. Yerlikaya-Özkurt. CMARS: a new contribution to nonparametric regression with multivariate adaptive regression splines supported by continuous optimization. *Inverse Problems in Science and Engineering*, 20(3):371–400, 2012.

[104] C.-Y. Wee, P.-T. Yap, W. Li, K. Denny, J. N. Browndyke, G. G. Potter, K. A. Welsh-Bohmer, L. Wang, and D. Shen. Enriched white matter connectivity networks for accurate identification of MCI patients. *Neuroimage*, 54(3):1812–1822, 2011.

[105] B. S. Ye, Y. Lee, K. Kwak, Y.-H. Park, J. H. Ham, J. J. Lee, N.-Y. Shin, J.-M. Lee, Y. H. Sohn, and P. H. Lee. Posterior ventricular enlargement to differentiate dementia with Lewy bodies from Alzheimer's disease. *Journal of Alzheimer's Disease*, 52(4):1237–1243, 2016.

[106] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen. Multimodal classification of Alzheimer's disease and mild cognitive impairment. *NeuroImage*, 55(3):856 – 867, 2011.

[107] T. Zhang and C. Davatzikos. ODVBA: Optimally-discriminative voxel-based analysis. *Medical Imaging, IEEE Transactions on*, 30(8):1441–1454, 2011.

[108] W. Zhang and A. T. Goh. Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geoscience Frontiers*, 7(1):45–52, 2016.

[109] Y. Zhang, R. Jin, and Z.-H. Zhou. Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics*, 1(1-4):43–52, 2010.

[110] Y. Zhang and S. Wang. Detection of Alzheimer's disease by displacement field and machine learning. *PeerJ*, 3:e1251, 2015.

[111] Y. Zhang, S. Wang, and Z. Dong. Classification of Alzheimer disease based on structural magnetic resonance imaging by kernel support vector machine decision tree. *Progress In Electromagnetics Research*, 144:171–184, 2014.

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Çevik, Alper
Nationality: Turkish (TC)
Date and Place of Birth: 10.02.1983, Malkara/TEKİRDAĞ
Marital Status: Married
E-Mail: alper.cevik@metu.edu.tr

EDUCATION

Degree	Institution	Year of Graduation
M.S.	Biomedical Engineering, METU	2011
B.S.	Electrical and Electronics Engineering Department, METU	2007
High School	Bursa AOS Science High School	2000

PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
2007-2017	Hemosoft IT & Training Services, Inc.	Software Engineer
2006	Hemosoft IT & Training Services, Inc.	Internship (Software Development Department)
2004	TAI (Turkish Aerospace Industries, Inc.)	Internship (R&D Department)

PUBLICATIONS

International Journal Publications

Çevik, A., Weber, G. W., Eyüboğlu, B. M., Oğuz, K. K., & Alzheimer's Disease Neuroimaging Initiative. (2017). Voxel-MARS: a method for early detection of Alzheimer's disease by classification of structural brain MRI. Annals of Operations Research, SI: OR in Neuroscience.

International Conference Publications

Çevik A., Eyuboglu B. M. (2014) Mumford-Shah Based Unsupervised Segmentation of Brain Tissue on MR Images. In: Roa Romero L. (eds) XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013. IFMBE Proceedings, vol 41. Springer, Cham.

National Journal Publications

Çevik, A., Eyüboğlu, B. M. (2011). Doku Anomalisi İçeren Beyin MR İmgeleri Üzerinde Mumford-Shah Tabanlı Bölütleme/Mumford-Shah Based Segmentation of Brain MR Images With Tissue Abnormalities. EMO Bilimsel Dergi, 1(2), 103-107.

Thesis Studies

Çevik, A. (2011). A Medical Image Processing and Analysis Framework (MS Thesis, Biomedical Engineering, Middle East Technical University, Ankara, Turkey).