

# **Impact of MRI technology on Alzheimer's disease detection**

by

Saruar Alam

A Thesis Presented in Partial Fulfillment

Of the Requirement for the Degree of Master of Research

Department of Computing

Macquarie University

22 April 2018



**MACQUARIE**  
University  
SYDNEY • AUSTRALIA

## **Statement of Originality**

This work has not previously been submitted for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself

*Saruar Alam*

Date: 22 April 2018

Saruar Alam

# Abstract

Alzheimer's disease (AD) can be detected using magnetic resonance imaging (MRI) based features and supervised classifiers. The subcortical and ventricular volumes change for AD patients. These volumes can be extracted from MRI by tools such as FreeSurfer and multi-atlas-based likelihood fusion (MALF) algorithm. Medical imaging centers typically use MRI protocols for brain scanning. These protocol differences include different scanner models with various operating parameters. The scanner models can have the same or different field strengths. A key factor in classifying multicentric MR subject images having different protocols is how different scanner models affect the extraction of features, and subsequent classification performance of a supervised classifier. We have investigated the classification performance of FreeSurfer and MALF based volume features together with Radial Basis Function Support Vector Machine and Extreme Learning Machine across different imaging protocols. We have also investigated both FreeSurfer and MALF, whose defined regions of the brain are most effective for the detection of the disease over different protocols. Our study result indicates marginal differences in classification performance across scanner models with the same or different field strengths when differentiating AD, Mild Cognitive Impairment, and Normal Controls. We have also observed differences in ranking order of the most effective regions.

## Table of Contents

|  |      |
|--|------|
| Statement of Originality .....   | i    |
| Abstract .....   | ii   |
| Table of Contents .....  | iii  |
| List of Figures .....  | v    |
| List of Tables .....   | vi   |
| Acknowledgement .....  | vii  |
| Alzheimer's Disease Neuroimaging Initiative (ADNI) Acknowledgement .....   | viii |
| 1.....   | 1    |
| Introduction.....  | 1    |
| 1.1    Background .....  | 1    |
| 1.2    Contribution of the study.....                                      | 4    |
| 1.3    Thesis structure.....   | 5    |
| 2.....   | 7    |
| Background and Related works .....   | 7    |
| 2.1    Background .....  | 7    |
| 2.2    MRI as a biomarker.....   | 8    |
| 2.3    MRI based features.....   | 9    |
| 2.3.1    Voxel-based features or Tissue diffusion Map based features ..... | 9    |
| 2.3.2    Vertex-based /Cortical Surface-based features .....               | 10   |
| 2.3.3    Pre-determined ROI-based features .....                           | 11   |
| 2.4    MRI based tools.....  | 12   |
| 2.4.1    FreeSurfer: .....   | 12   |
| 2.4.2    Statistical Parametric Mapping (SPM) .....                        | 12   |
| 2.4.3    FMRIB Library(FSL).....   | 12   |
| 2.5    Analysis of using different field strengths/scanners in CAD .....   | 12   |
| 2.6    Conclusion.....   | 16   |
| 3.....   | 17   |
| AD Diagnostic Models.....  | 17   |
| 3.1    Background .....  | 18   |
| 3.2    Multi-atlas based pipeline .....                                    | 18   |
| 3.3    FreeSurfer features based pipeline .....                            | 22   |
| 3.4    Supervised learning algorithms .....                                | 23   |
| 3.4.1    Support Vector Machine .....                                      | 24   |
| 3.4.2    Extreme Learning Machine.....                                     | 26   |

|                                  |  |    |
|----------------------------------|--|----|
| 3.5                              | Conclusion.....  | 27 |
| 4.....                           |  | 28 |
| Data and experimental work ..... |  | 28 |
| 4.1                              | Dataset .....  | 28 |
| 4.1.1                            | Data for the Multi-Atlas Likelihood fusion (MALF) based algorithms ..... | 29 |
| 4.1.2                            | Data for the FreeSurfer based Method .....                               | 29 |
| 4.2                              | Methods to avoid double dipping.....                                     | 31 |
| 4.3                              | Feature Ranking by the two-sample $t$ test .....                         | 32 |
| 4.4                              | Support Vector Machine Recursive Feature Elimination (SVM-RFE).....      | 33 |
| 4.4.1                            | SVM-RFE algorithm:.....  | 33 |
| 4.5                              | Evaluation matrices .....  | 34 |
| 4.6                              | Cross-Validation.....  | 35 |
| 4.7                              | Used Tools/Packages.....   | 36 |
| 5.....                           |  | 37 |
| Result and Discussion .....      |  | 37 |
| 5.1                              | Feature setup for classification.....                                    | 37 |
| 5.2                              | Classification performance of Multi-atlas-based pipeline.....            | 38 |
| 5.2.1                            | Performance of RBF kernel SVM classifier .....                           | 38 |
| 5.2.2                            | Performance of ELM classifier .....                                      | 40 |
| 5.3                              | Classification performance of FreeSurfer based methods .....             | 42 |
| 5.3.1                            | Performance using RBF SVM classifier .....                               | 43 |
| 5.3.2                            | Performance using ELM classifier.....                                    | 44 |
| 5.4                              | Ranking MALF and FreeSurfer based features.....                          | 44 |
| 5.5                              | Conclusion.....  | 48 |
| 6.....                           |  | 50 |
| Conclusion .....                 |  | 50 |
| 6.1                              | Research Summary.....  | 50 |
| 6.2                              | The limitations of our study .....                                       | 51 |
| 6.3                              | Future Work .....  | 52 |

## List of Figures

|  |    |
|--|----|
| Figure 3-1: MALF based segmentation of four different subjects over four scanner models .....                                  | 20 |
| Figure 3-2: MALF based segmentation of a subject image at two granularity levels .....   | 20 |
| Figure 3-3: Schematic diagram of MALF based segmentation .....   | 22 |
| Figure 3-4: FreeSurfer based subcortical segmentation of four different subjects over four scanner models .....                | 23 |
| Figure 3-5: Schematic diagram of the proposed approach .....   | 24 |
| Figure 3-6: Linear Support Vector Machine .....  | 25 |
| Figure 3-7: ELM SLFN approach .....  | 26 |
| Figure 4-1: Illustration of 5-fold cross-validation .....  | 36 |
| Figure 5-1: Performance accuracy of MALF based method using RBF-SVM over four different protocols .....                        | 39 |
| Figure 5-2: Performance result of MALF based features at granularity level 5 using RBF-SVM over four different protocols ..... | 40 |
| Figure 5-3: Performance accuracy of MALF based method using ELM over four different protocols .....                            | 41 |
| Figure 5-4: Performance result of MALF based features at granularity level 5 using ELM over four different protocols .....     | 41 |
| Figure 5-5: Comparison of ELM and SVM at granularity level 5 of MALF based method .....  | 42 |
| Figure 5-6: Performance result of FreeSurfer based method using RBF-SVM over four difference protocols .....                   | 43 |
| Figure 5-7: Performance result of FreeSurfer based method using ELM over four difference protocols .....                       | 43 |
| Figure 5-8: Ranking of MALF based features at granularity level 3 while classifying AD/CN .....                                | 45 |
| Figure 5-9: Ranking of MALF based features at granularity level 3 while classifying AD/MCI ....                                | 46 |
| Figure 5-10: Ranking of MALF based features at granularity level 3 while classifying CN/MCI ...                                | 46 |
| Figure 5-11: Ranking of FreeSurfer based features while classifying AD/CN .....  | 47 |
| Figure 5-12: Ranking of FreeSurfer based features while classifying AD/MCI .....   | 47 |
| Figure 5-13: Ranking of FreeSurfer based features while classifying CN/MCI .....   | 48 |

## List of Tables

|   |    |
|---|----|
| Table 2-1: Importance of T1-weighted, T2-weighted, and PD-weighted MR image.....                                | 8  |
| Table 2-2: A summary of classification accuracy of different classifiers using different feature type.<br>..... | 13 |
| Table 3-1: MALF based segmented structures (54 Nos) at 3rd granularity level.....                               | 20 |
| Table 3-2: FreeSurfer based subcortical segmented structures.....   | 23 |
| Table 4-1: Different phases of ADNI project .....   | 30 |
| Table 4-2: Imaging parameters used in MALF based approach .....   | 30 |
| Table 4-3: Imaging parameters used in FreeSurfer.....   | 32 |
| Table 4-4: Selected data for MALF and FreeSurfer based approach .....   | 32 |
| Table 4-5: Confusion matrix for a binary classifier .....   | 35 |

## **Acknowledgement**

I would like to express my sincere appreciation to all those who played a part in the accomplishment of my dissertation, supported and guided me throughout the study of my MRes program. Although, I have covered this work within 9 months available time, the journey was invaluable regarding learning various techniques, and algorithms and their practical implementation.

First of all, I would like to express my sincere gratitude and special appreciation to my advisor, Dr. Len Hamey and Dr. Kevin Ho-Shon for their guidance, encouragement, continuous support, and all the invaluable lessons learned. Under their supervision, I have learned to identify and approach research questions. Under their thoughtful guidance, the exposure I have achieved during my research would be a real treasure for my life.

I express my profound gratitude to my parents and family for their love, support, and encouragement throughout of my entire life. Their moral and emotional support has made the way apparent to pursue my MRes degree.

Lastly, I would like to acknowledge the financial assistance provided by Macquarie Research Excellence Scholarship program (iMQRES) which helped me to carry out this research. Also, I would like to acknowledge the facilities provided by CBRAIN, and MRI cloud.



## **Alzheimer's Disease Neuroimaging Initiative (ADNI) Acknowledgement**

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant no. U01 AG024904) and DOD ADNI (Department of Defense Award no. W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging and the National Institute of Biomedical Imaging and Bioengineering and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development LLC; Johnson & Johnson Pharmaceutical Research & Development LLC; Lumosity; Lundbeck; Merck & Co. Inc.; Meso Scale Diagnostics LLC; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research provide funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (<http://www.fnih.org/>). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California.

Data used in preparation of this thesis were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](http://adni.loni.usc.edu)). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: [http://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf)

# Introduction

## 1.1 Background

Dementia is a chronic and prolonged decline in cognitive performance which progressively develops, and depending on the type of dementia, can be associated with the formation of the  $\beta$ -amyloid plaques, synaptic dysfunction, damage to brain cells, and brain shrinkage. It subsequently causes the deterioration of memory, cognitive abilities, reasoning, and language [1]. According to a statistical report, over 135 million people worldwide will suffer from dementia by 2050 [2]., which is triple the current number of affected patients. Dementia is now a primary global health and social threat which will bring an enormous financial burden on families, and national healthcare system.

Alzheimer's disease (AD) and Frontotemporal dementia (FTD) are the most common forms of dementia. AD accounts for 60-80% of dementia cases [3], and FTD accounts for 4-20% [4]. The majority of dementia cases affect individuals above the age of 65 [5]. Other forms of dementia such as Parkinson's Disease, Huntington's disease, and vascular dementia are also prevalent [6]. The support cost of all AD patients alone is estimated to be \$220 billion in the USA and \$605 billion per year globally. The developed countries are also supporting research activities related to AD [2].

Alzheimer's disease has primarily three stages: mild AD, moderate AD, and severe AD. It also includes preclinical AD, Mild Cognitive Impairment (MCI), and other dementias associated with AD [7]–[9]. Clinical studies typically identify three groups, including normal controls (CN), MCI, and AD patients. MCI cohorts have some symptoms which are common with AD cohorts, but MCI subjects can do daily activities normally as healthy people do. An MCI subject stays cognitively stable for several years, then progresses to some type of dementia, especially to AD. The percentage of the MCI cohorts likely to convert to AD every year is between 10% and 30%, whereas the conversion rate from normal controls (CN) to MCI or AD is just 1-2% [10], [11]. Recent research suggests that early treatment of MCI may slow down the progression to AD [12], so it is highly valuable to study and predict the early stages of MCI.

AD is diagnosed using both invasive and non-invasive methods. Non-invasive diagnosis methods don't involve surgery, which reduces the risk of infection, the duration of hospital stay, and the possibility of trauma [13]. Further, non-invasive methods are quicker, cheaper, and don't require as much expertise to implement [14], [15]. For these reasons, it is preferable to use non-invasive or minimally invasive diagnostic methods provided they have sufficient accuracy. Our study is based on non-invasive medical imaging technology, specifically MRI, as a biomarker for AD/MCI detection. The original non-invasive diagnostic methods are based on clinical observation, patient history, and cognitive testing. Some clinical tests can reasonably assess the severity of AD such as MMSE (Mini-Mental State Examination) [16], AMTS (Abbreviated Mental Test Score) [17], modified MMSE (3MS) [18], CASI (Cognitive Abilities Screening Instrument) [19], and CDT (Clock Drawing Test) [20]. Despite their ability to recognize the cognitive status and various types of dementia, they are inadequate since the clinical test score does not necessarily correlate with memory and thinking disorders [21]. Also, some clinical tests tend to have educational, social, and cultural biases [21]. Other effective non-invasive diagnostic methods are based on medical imaging. Medical imaging assists to visualize degenerative histological changes, which includes amyloid plaques, hypo-metabolism, and the structural changes caused by neurological disorders. These histological can become apparent long before the disorder is clinically detectable [22]. The use of medical imaging revolutionizes the noninvasive diagnosis of dementia. The commonly used imaging methods are MRI (Magnetic Resonance Imaging), fMRI (functional MRI), SPECT (Single Photon Emission Computed Tomography), and PET (Positron Emission Tomography). MRI is also called structural MRI (sMRI) to differentiate it from functional MRI (fMRI). Throughout this thesis, we will use the terms MRI and MR imaging to refer to structural MRI. MRI applies radio waves in a strong magnetic field to construct a pictorial presentation of organs and tissue of the human body. Hence MRI can potentially be used as a biomarker for MCI and to assess the progression to AD.

The current approach to dementia diagnosis based on medical imaging needs human expertise which is labor and time intensive and prone to operator bias [23]. Therefore, the Computer Aided Diagnosis (CAD) is emerging so that the drawbacks of manual detection can be overcome. The goals of CAD-based methods for dementia are 1. Detecting AD from normal cohorts [24] 2. Differentiating AD from other types of dementia [25] 3. Separating several stages of dementia such as AD and MCI [26] 4. Finding the importance of various Regions of Interest (ROIs) of the human brain which are sensitive to the progression of AD [27]. Although a cure for the disease is not available, the CAD method assists in analyzing subtle change which may be a biomarker for progression of the disease. Specifically, CAD could be used to monitor the progression of brain atrophy to show the effectiveness of medications [12]. In addition, CAD methods may also help to predict the severity of the disease in the long run [28].

MRI has the potential to detect biomarkers for AD/MCI. Three types of feature are usually extracted from MRI: Voxel-based, Vertex-based, and predefined ROI based. The voxel-based methods measure volumes of brain matter, specifically Grey Matter (GM), White Matter (WM), and Cerebral-Spinal Fluid (CSF) [29], [30]. The Vertex-based methods measure primarily the cortical thickness based on structural features [31]. The predefined ROI based methods measure the volume of specific regions of the brain such as hippocampus, amygdala, and the entorhinal cortex [32]–[34]. For all three feature extraction methods, an efficient segmentation or parcellation method is required to obtain an accurate measurement of a region of interest (ROI). These methods use template or atlas-based parcellation algorithms to obtain improved segmentation accuracy [35], [36]. The template or atlas provides prior knowledge of the structure of the required brain regions. However, a standard template or atlas may not always be applicable. Moreover, a single atlas may have some biases such as age, gender, and protocol which influences the segmentation. Therefore, multi-atlas or multi-template based methods are used which diversify the collection of atlases regarding age, gender or protocol. The use of multi-atlas methods achieves better segmentation accuracy [37].

CAD-based methods differentiate the severity of dementia with the aid of a supervised classification algorithm [38]. The supervised classification algorithm uses features extracted from the MRI to train the classifier to differentiate dementia stages. Corresponding features from independent images are used in the testing phase to assess the performance of the classifier. A classifier extracts generalization knowledge during training from the labeled subjects, and it predicts the label of an unknown subject based on that generalization knowledge. To achieve efficiency, the classifier requires input of an appropriate feature representation. However, the extraction methods for rich representative features should be computationally feasible. Also, the features should be non-redundant and relevant to avoid poor training of the classifier [39]. The relevant features are either

combined or reduced dimensionally if the feature dimension is high before using them as input to the classifier [26]. We require an experimental design to assess the performance of the classifiers where each subject (or sample) is used for either training or testing, but not for both in a single classifier instance. Using the same subject for both training and testing introduces bias, and the performance measures would be unreliable [40].

MRI data originates in medical diagnostic centers, each of which employs different MRI acquisition protocols [41]. The protocols may differ in field strength (1.5 T or 3 T), scanner model or operating parameters of the scanner. Further, protocols used in a single center may change due to the replacement or upgrade of a scanner. Over the last two decades, the studies which are based on quantitative neuroimaging data have been primarily conducted on 1.5 T MRI images. But imaging equipment with a higher field strength has become available more recently, and gradually it is replacing the lower field strength image-based clinical assessment. A brain imaging center archives data from both low and high field strength scanners, and from different brands and models. Typical centers use different scanner models from different manufacturers such as Siemens, and GE Healthcare [41]. For example, two available models of Siemens are Symphony and TrioTim, and two available models of GE Healthcare are Signa HDX and Signa Excite. The protocol differences creates a dilemma for longitudinal studies because the study needs to collect data acquired in different time phases of an individual patient, but the imaging may be performed using different protocols [41].

The protocol differences may affect the segmentation or parcellation of the entire brain whether the method is single or multi atlas/template based [42]. The pooling of the MRI data from multiple imaging centers is required to collect large-scale data to reliably analyze the progression stages of AD. However, the MRI acquisition methods followed by individual centers use different imaging field strengths or scanners with different data processing parameters, and several post-processing smoothing parameters. This difference may result in inaccurate measurement of volume or thickness of different regions, especially those regions where image quality is sensitive to differences in magnetic field strength. As a result, MRI based diagnostic methods may be affected. In other words, these differences may also affect the performance of a supervised classifier that is used to detect AD and MCI by processing features extracted from MRI data. The primary purpose of this thesis is to investigate the impact of protocol differences on AD and MCI detection.

## 1.2 Contribution of the study

To date, a few comparative studies have explored the acquisition differences across scanner models with the same or different field strengths, and methods of measurement. These studies will be

reviewed in chapter 2. The analysis of the acquisition differences across different methods related to diagnosis of AD using a supervised classifier falls outside of the scope of those studies. In our study, we have addressed this research question.

We have used two popular methods for structural volumetric measurements of various ROIs. These methods are FreeSurfer and Multi-atlas likelihood fusion (MALF) algorithm. We have considered two types of features from MRI, multi-atlas likelihood fusion-based structural volume, and FreeSurfer based volume. We have used two supervised classifiers, the frequently used Radial Basis Function SVM (RBF-SVM) [43], [44], and the more recently proposed Extremely Learning Machine (ELM) [45].

In our thesis we have addressed the following research questions:

- We have investigated whether classification performance using MALF based structural volume features varies across different protocols.
- We have investigated whether classification performance using FreeSurfer based volumetric features varies across different protocols.
- We have compared performance result of two efficient supervised classifiers using both MALF and FreeSurfer based volumetric features across different protocols
- We have studied both the MALF and FreeSurfer defined ROIs which are most significant when detecting AD or MCI across different protocols.

We have employed a data selection method described in chapter 4 to avoid double dipping and calculated six terms (i.e., accuracy, sensitivity, specificity, precision, F1 score, and gmean) to measure the performance of two classifiers (i.e., RBF-SVM, and ELM)

### 1.3 Thesis structure

The remainder of this thesis is organized as follows.

Chapter 2 briefly describes various types of features extracted from MRI used for the classification, the effect of protocol difference in extracting those features and some prior works related to this.

Chapter 3 describes the multi-atlas likelihood fusion method (MALF), and the reason for choosing that method. It also includes a brief description of FreeSurfer based features, and two supervised classifiers that are used in our study-Radial Basis Function Support Vector Machine (RBF-SVM), and Extreme Learning Machine (ELM).

Chapter 4 provides details of the experimental approach. This chapter includes a brief description of data, the different preprocessing steps it undergoes, feature selection method, feature ranking method, parameters to measure the performance of a classifier, and the cross-validation technique.

Chapter 5 presents the results we have obtained, including the comparison of the two supervised classifiers, and the two feature extraction methods. The chapter also includes the analysis of affected brain ROIs which are extracted by the feature extraction method.

Finally, chapter 6 summarises the findings of our study mentioning some limitations of our work, and also raises some research questions which we plan to address in future.

## Background and Related works

The chapter briefly introduces the brain tissues relating to Alzheimer's disease, and introduces different types of features extracted from Magnetic Resonance Imaging (MRI) which include Voxel-based, Vertex based, and Pre-defined ROI based. This chapter also discusses articles which use those features as a biomarker, and supervised classifiers such as Support Vector Machine to analyze the severity of the brain atrophy related to AD. Finally, this chapter also discusses the protocol differences of MRI in relation to various feature extraction methods.

### 2.1 Background

Alzheimer's disease severely affects the memory, thinking capability, and intelligence of a patient [1]. These factors are measured by cognitive tests such as MMSE (Mini-Mental State Examination) [16] and AMTS (Abbreviated Mental Test Score) [17]. The severity of AD patient's AD is diagnosed with by these cognitive scores. However cognitive reserve influences these scores [46]. Cognitive reserve or brain resilience depends on race, gender, class, education, age, etc. For example, people from different education backgrounds have different cognitive scores despite having a similar severity of brain damage [47]. Neuroimaging measurements are less affected by the biases mentioned above [48]. Hence neuroimaging analysis has a higher potential to predict disease severity as compared to neuropsychological measurement. In other words, the neuroimaging measurement is more reliable or robust.



The brain undergoes several structural changes when a subject progresses to AD/MCI. The White Matter (WM), Gray Matter (GM), and Cerebrospinal fluid (CSF) abnormalities are a diagnostic pattern of Alzheimer's disease. WM is nerve tissue existing in the central nervous system containing primarily myelinated fibers, and it is almost white colored. GM is a reddish-gray colored nerve tissue primarily composed of nerve cell bodies and dendrites. It also exists in the central nervous system like WM. CSF is a watery fluid which is produced, absorbed, and flows in the ventricles of the brain and around the surface of the brain and spinal cord. Atrophy of these tissues is observed when the disease progresses. The regions of the brain where atrophy starts are the hippocampus, entorhinal cortex, and amygdala [49], [50]. The major cause of cerebral atrophy in the hippocampus and entorhinal cortex is the formation of senile plaques consisting of amyloid beta-42 protein [51]. Another reason for atrophy in the hippocampus is neurofibrillary tangles (NFT) which consist of tau proteins [51]. When the disease progresses, hippocampus affects at the beginning [32]. But Xu et al. [33] suggests entorhinal cortex affects at the beginning.

## 2.2 MRI as a biomarker

Magnetic resonance imaging (MRI) can discriminate between different types of tissue, and is a noninvasive imaging technology. Therefore, the imaging marker, MRI, continues to be considered a strong tool for the analysis of disorders in the central nervous system, particularly the brain. MRI facilitates to quantify the density, shape, and volume of Gray Matter, White Matter, and CSF structures in the cerebral brain. Various techniques are employed for morphometric analysis which measures shape or volume of gray matter structures [52], [53]. Similarly, the shape or volume of white matter and CSF structures are also measured. The tissue can be differentiated by different MR imaging schedules. Three different types of MRI sequences are primarily available such as T1-weighted, T2-weighted, and PD(Photon density)-weighted. Their significance regarding characterizing of different tissues are shown in Table 2-1 below. The most commonly used schedule for AD diagnosis is T1 weighted [54].

Table 2-1: Importance of T1-weighted, T2-weighted, and PD-weighted MR image

|             | Importance  |
|-------------|---|
| T1-weighted | Good contrast between GM and WM.<br>GM-Dark gray, WM-lighter gray, and CSF- black |
| T2-weighted | Good contrast between CSF and brain tissues<br>CSF- Bright                        |
| PD-weighted | Good contrast between GM and WM<br>Little contrast between CSF and brain          |

## 2.3 MRI based features

The anatomical imaging such as MRI is frequently utilized in the clinical dementia diagnosis [33]. In CAD, different types of features from MRI is used which corresponds to structural abnormalities of the brain. The MRI-based features can be roughly categorized into three groups: voxel-based [55], vertex-based [39], and pre-determined ROI-based [32], [50].

### 2.3.1 Voxel-based features or Tissue diffusion Map based features

A voxel is a volume element which represents a crucial region of the three dimensional (3D) space. The voxel-based feature is obtained from a 3D volumetric space image of human the cerebral brain. Voxel-based methods originated into the mid-1990s when Wright at al. [56] studied the gray matter and white matter voxel values to diagnose schizophrenia. In the voxel-based method, the voxels of the whole brain are partitioned or segmented into three different tissues (GM, WM, and CSF). The GM voxel tends to provide more discriminatory information for AD or MCI detection as compared to WM or CSF. Hence the focus of various studies is on GM voxels [29], [30]. The GM volume based feature of entire brain with linear support vector machine stratifies AD from CN with promising accuracy [57]. The study claims that the GM volume of hippocampus shrinks, subsequently other neighboring cortex areas are affected. These conclusions are supported by other studies [32], [58], [59]. Also, GM volume feature can be used to predict the conversion of MCI patients to AD using supervised classifiers such as SVM, RVM (Relevance Vector Machine), and nearest neighbors [60], [61]. SVM performs well as compared to the other classifiers applied in these studies. Another recent study uses the GM volume patterns and SVM to stratify AD from normal controls [62].

The dimensionality of volume features becomes large when the voxels of the whole brain are considered. Therefore, the region of Interest (ROI) based methods are implemented to overcome the problem. In ROI methods, the voxels of brain regions are registered to a template by nonlinear registration [63], [64]. As a result, the brain regions are deformed to match the template. The methods such as Voxel-based morphometry (VBM) [65], deformation-based morphometry (DBM) [66], and Tensor-based morphometry (TBM) [67] are most frequently used. In these methods, the density of tissues (GM, WM, and CSF) is calculated from the regions which are defined by the atlas. VBM-type methods quantify the regional tissue density of the original brain volume without considering the deformation of it. However, two types approaches, DBM-type approaches and TBM-type procedures measure the deformation field and the Jacobian of deformation, respectively. In MRI based AD diagnosis, these measurements can be considered as input patterns for multivariate methods such as Support Vector Machine or Extreme Learning Machine. Several studies report a promising accuracy using these methods [68]–[70].

The parcellation of the whole brain or automatic warping of a brain to an atlas is not a trivial task. The parcellation may not be adaptive to anatomical regions. That's why adaptive parcellation is applied where the whole brain image is grouped into the most discriminative areas and adaptive features of those areas are extracted based on the correlations between the tissue volumes and other classification parameters [71]. The predefined region based atlas may not always be available. Also, the parcellation based on a single atlas is prone to be affected by registration noise. Therefore, the parcellation may not yield suitable discriminatory regions or features. The multi-atlas based warping is used to address the problem [37], [72]. Several studies show that multi-atlas based method outperforms single atlas-based method regarding the performance of AD detection [37], [72]. However, these methods are computationally intensive. The tissue maps from multiple atlases are conventionally concatenated in those methods. The tissue information from all the atlases usually doesn't contribute uniformly to detect the severity of the disease. Although these atlases may contribute equally, there will be redundant features. To eliminate the unnecessary features, the weighted concatenation methods are applied using an ensemble SVM classifier [73]. To overcome the potential bias relating to single template based studies, multi-template based feature extraction method in conjunction with relationship induced sparse selection and ensemble SVM is proposed to classify AD/MCI [74]. They have trained the feature selection method based on the knowledge of the relationship among different templates and also similarity among different subjects

### 2.3.2 Vertex-based /Cortical Surface-based features

A vertex is an identifiable structural point in the brain. According to clinical studies, not only voxel analysis helps to categorize the severity of AD but also vertex atrophy provides information to distinguish AD, MCI from CN. Analysis of the cortical surface is a particular form of vertex-based analysis. The cortical thickness is defined as thickness of different regions of cerebral cortex measured in the cortical surface. The cortical thickness is correlated with the atrophy and the histopathological changes caused by dementia [75]. Hence it can be used as an essential surrogate marker to diagnose the disease. As a volumetric feature, cortical thickness gives promising classification performance [31]. In addition to it, the cortical surface area is also an important feature even though it is biologically uncorrelated with cortical thickness measurement [76]. For example, Li et al. [77] have used cortical thickness, surface area, and other volumetric and geometric measures for SVM based MCI classification. However, the features extracted from all vertices of the cortical surface have the problem of large dimensionality. Park et al. [78] have used cortical thickness features which have been derived from parameterized 3D meshes. They distinguished AD/MCI from CN using principal component analysis (PCA) as a dimensionality reduction technique and support vector machine (SVM).

Manual parcellation is time intensive and prone to operator bias depending on an individual expert. Researchers and the neuroanatomists propose various automated methods to parcellate the cortical surface to overcome the drawback of manual parcellation. The automated methods are primarily based on one of a specific template [79], watershed segmentation [80], and graph algorithm [81]. In a template-driven warping method, local correspondence is formed between the applied template and an individual subject. In a watershed-based approach, the cortical sulci are segmented by the watershed transformation and manually labeled the detected regions by a trained neuroanatomist. The cortical sulci are represented by vertices of a graph in graph-based methods. The arcs which connect them establish their relationships. An automated parcellation method may incorporate prior statistical information and cortical geometry [82]. Using the template driven method, Desikan et al. [83] have stated that cortical thickness of entorhinal cortex and supramarginal gyrus is most affected in AD. A similar claim has been made by another study where they observed 88.2% accuracy in SVM based AD/CN classification [84]. Wee et al. [39] have used regional mean cortical thickness based features and multi-kernel Support Vector Machine. They have improved the performance by forming additional correlative features where feature set is formed based on the similarity of cortical thickness between a pair of brain ROIs. They achieved an accuracy of 92.35% for AD/CN, 83.75% for AD/MCI, and 75.05% for MCI-C/MCI-NC classification.

### 2.3.3 Pre-determined ROI-based features

Apart from the above voxel- and vertex-based methods, the predefined ROI-based feature discriminates the severity of the disease. Specific regions of the brain are affected when the disease progresses. The neurodegeneration happens progressively, starting at the medial temporal lobe, successively affecting the entorhinal cortex, hippocampus, limbic system, and neocortical areas [85]. Several studies found the difference of hippocampal atrophy between AD/MCI and normal controls [39], [86]. Hence it is used as an essential biomarker in numerous studies [64], [87].

In the ROI-based method, segmentation is done before feature extraction. Manual segmentation is time-consuming and operator dependent. So several semi-automated methods of segmenting hippocampi and amygdala are studied [88]. Also, several fully automated methods are proposed which use probabilistic priors derived from an atlas and anatomical information from local anatomical patterns for segmentation [88]. This automated approach outperforms several semi-automatic methods regarding segmentation accuracy. The volume or shape of hippocampi or a weighted combination of both is used as a discriminatory feature to SVM which predicts MCI conversions from AD [89].

## 2.4 MRI based tools

Researchers frequently use different automated tools such as FreeSurfer, Statistical Parametric Mapping (SPM), and FMRIB library (FSL) to segment brain tissues and measure volume from different regions of interest (ROIs). Among them, FreeSurfer and SPM are most commonly used.

### 2.4.1 FreeSurfer:

FreeSurfer is an automated, efficient brain imaging tool to analyze structural MRI and functional MRI image. We have provided a brief detail of FreeSurfer in chapter 3.

### 2.4.2 Statistical Parametric Mapping (SPM)

SPM (accessible at <http://www.fil.ion.ucl.ac.uk/spm/>) utilizes the unified segmentation [90]. This unified segmentation is a probabilistic model which integrates tissue degradation, bias correction and image registration within the same generative model [90]. It generates partial volume segmentation outcomes for every type of tissue. To get the segmentation output, it uses tissue prior probability maps based on intensity values. From the segmented output, it calculates gray matter (GM), white matter (WM) and Cerebrospinal Fluid (CSF). Also, later version such as SPM8 includes the segmentation of some soft tissue maps, bone, and air/background. As a result, the probability of misclassification of non-brain tissue is reduced.

### 2.4.3 FMRIB Library(FSL)

FSL (accessible at <http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>) utilizes the SIENAX package for estimating brain tissue volumes one by one serially [91]. SIENAX starts by extracting brain and skull from the single full head MR image [92]. The brain image is then affine-registered to MNI152 space [93]. Next, tissue-type segmentation with partial volume estimation is performed [94], finally measuring GM, WM, and ventricular CSF.

## 2.5 Related MRI based AD diagnosis studies

There is a great variety of specific MRI-based image features that have been used for AD diagnosis. Here, we review frequently used and recently proposed features.

Cuingnet et al. [64] contrast ten different high-dimensional feature sets, classified using an SVM classifier. They have employed 509 baseline ADNI 1.5T MR images for classification of AD/MCI vs CN. Two approaches use only the hippocampal shape or volume, while the rest are whole-brain based approaches. These approaches use either cortical thickness steps or voxel-wise tissue density maps

for gray matter, white matter, and CSF. High accuracies in identifying AD patients from CN (around 81% sensitivity and 95% specificity) are reported while using whole-brain approaches. Four of the ten approaches were capable of differentiating MCI patients who later

Table 2-2: A summary of classification accuracy of different classifiers using different feature type.

| Study                    | Type of Features                    | Dataset  | Category of features         | Classifier       | Classification accuracy |        |        |           |
|--------------------------|-------------------------------------|--|------------------------------|------------------|-------------------------|--------|--------|-----------|
|                          |                                     |  |                              |                  | AD/CN                   | AD/MCI | CN/MCI | sMCI/pMCI |
| (Liu et al. [73])        | M-ROI                               | AD-97<br>pMCI-117<br>sMCI-117<br>CN-128            | Tissue density map based     | SVM              | 92.51                   | -      | -      | 78.88     |
| (Misra et al. [95])      | S-ROI                               | AD-56<br>pMCI-27<br>sMCI-76<br>CN-66               | Tissue density map based     | SVM              | -                       | -      | -      | 81.50     |
| (Salvatore et al. [96])  | SUFR                                | AD-137<br>pMCI-76<br>sMCI-134<br>CN-162            | Tissue density map based     | SVM              | 76.00                   | -      | 72.00  | 66.00     |
| (Li et al. [77])         | All vertices                        | MCI-24<br>CN-26                                    | Cortical Surface based       | SVM              | -                       | -      | 80     | -         |
| (Wee et al., [39])       | Atlas                               | AD-198<br>pMCI-89<br>sMCI-111<br>CN-200            | Cortical Surface based       | Multi-kernel SVM | 92.35                   | 79.24  | 83.75  | 75.05     |
| Lama et al. [97]         | Cortical thickness and surface area | AD-70<br>MCI-74<br>CN-70                           | Cortical Surface-based       | Regularized ELM  | 76.61                   | -      | -      | -         |
| (Sorensen et al. [98])   | Hippocampus                         | AD-101<br>MCI-233<br>pMCI-93<br>sMCI-140<br>CN-169 | Pre-defined ROI based        | SVM              | 91.20                   | -      | 76.40  | 74.20     |
| (Chincarini et al. [99]) | Biologically selected regions       | AD-144<br>pMCI-136<br>sMCI-166<br>CN-189           | Pre-defined ROI based        | SVM              | 97.00                   |        | 92.00  | 74.00     |
| Zu et al. [100]          | Feature concatenation               | AD-51<br>pMCI-43<br>sMCI-56<br>CN-52               | Multimodal (MRI and FDG-PET) | Multi-kernel SVM | 95.95                   | 80.26  | -      | 69.78     |
| Alam et al. [101]        | MRI bases texture                   | AD-86<br>CN-86                                     | Texture                      | Twin SVM         | 92.65                   | -      | -      | -         |

\*SUFR-Supervised/unsupervised feature reduction, S-ROI-Single set adaptive ROIs, M-ROI-Multiple set adaptive ROIs, sMCI-stable mild cognitive impairment, pMCI-progressive mild cognitive impairment, and ELM-Extreme Learning Machine

progressed to AD (progressive MCI) from people who stayed stable for 18 weeks (stable MCI) marginally more correctly than the usual random classifier.

The classification accuracy of different recent approaches such as SVM, ELM, and TSVM using different types of features and are summarised in Table 2-2. This table can't be seen as the comparison of those methods because the different methods are tested on different data sets, some of which were captured using different imaging protocols. It provides a glance at contemporary approaches, their classification accuracy on different types of features. In these studies, MRI is acquired from either a single scanner model or multiple scanner models with the same or different field strengths. However, the limitation of these approaches is that they didn't consider these protocol differences while using MRI data in their studies. They merged MR images ignoring these protocol differences.

## **2.6 Effects of MRI acquisition protocol difference**

The classification of AD is performed using features from MRI which may be acquired by different protocols. The protocol differences include different field strengths, different scanner models and different operating parameters used to acquire the MR images. These differences may impact the classification performance. Here, we review research on the effects of protocol differences in general and also as related specifically to AD diagnosis.

The use of different field strength has pros and cons. For example, lower field strengths are less susceptible to chemical artifacts, and higher field strength images provide higher depth information for shape analysis and between class comparison [102], [103]. However, higher field strength has some limitations such as it is more vulnerable to chemical shift artifacts [102], [103]. Several articles regarding atrophy measurement distinguish 1.5 T and 3T field strength scans [104]–[106]. Some studies have encouraged to combine to use both field strengths [34], [107], some have posed concern regarding it [108], and some have proposed techniques to combine them [109].

There is the difference in the tissue signal when comparing 1.5T scan with its 3T counterpart [110]. These concerns related to topographic localization of regions or tissues across field strengths images. Because 3T images, using their higher contrast and greater signal-to-noise ratio (SNR), may improve the topographic localization of atrophy [104]. For instance, the number of lesion detection and observer agreement varies while rating 1.5 T and 3 T images by experts in this field [108]. 3T MRI had the greater inter-observer agreement. 3T shows more dysplasias, while 1.5T demonstrated more tissue reduction and mesial temporal sclerosis (MTS). Therefore, the diagnosis of patients may benefit from 3T scans if the patient suffers from dysplasia. The article also claims that the identification of lesion depends on scanner models.

In CAD regarding AD, the cortical thickness and volume features are extracted after segmenting brain tissues, and estimating Grey Matter (GM), WM (White Matter) and CSF (Cerebrospinal Fluid) volume. The segmentation and volume estimation of these tissues could be affected by field strengths, and scanner models. For example, a study revealed that gray matter and white matter volumes of a subject calculated by a tissue classification algorithm could vary depending on field strengths [111]. The estimated intracranial volume (eTIV) also differs systematically between 1.5 T and 3 T images of healthy subjects [105]. Additionally, the preliminary outcome from another study implies that 3T images may have the ability to detect volume differences that are not evident in 1.5T [104]. Also, a multi-atlas-based method claims that higher granularity level structural information of an image varies across different imaging protocols [112]. The level of granularity level is defined by the number of structural regions. Higher granularity level signifies higher number of structural regions. When granularity level increases, the protocol effect increase. The protocol difference in this particular paper includes three scanner manufactures and two magnetic field strength scans (1.5 T and 3T). Such disparity discourages to merge data across different imaging field strengths to measure subcortical brain volume measurement unless the difference is systematic, regular, detectable and suitable for correction [113].

In contrast, Scorzin et al. [34] have claimed no apparent differences of estimated volumes of hippocampus and amygdala based on 1.5 T and 3 T images irrespective of whether the segmentation method is manual or VBM based. Ho et al. [114] showed that sample size estimates for finding decaying rate atrophy using Tensor-based Morphometry (TBM) did not vary between 1.5T and 3T images. Goodro et al. [107] showed that for both the strengths, the correlation between subcortical volume of healthy cohorts and their age is strong. The correlation is similar for both the automated tools, Free Surfer and FSL, which are also used for subcortical segmentation. However, they have found that the degree of correlation varies based on age group.

Several studies have attempted to find a way to combine data from 1.5 T and 3.0 T scanners. Most of these studies are template or atlas driven. For example, Keihaninejad et al. [105] have proposed an approach to measure intracranial volume (ICV) across the two field strengths. This approach is based on tissue probability mapping in MNI template space and reverse brain mask (RBM) for reverse normalization to native space. By using this approach, the intracranial volume becomes more similar across field strengths as compared with two other automated methods, FSL and SPM, which are also used to measure ICV. Also, a multi-atlas based automated hippocampal segmentation has been found to be accurate at both imaging field strengths [106]. By using this method, the estimated atrophy rate is similar for both 1.5 T and 3 T images of a longitudinal study of a subject. Also, another multi-atlas-based study has proposed that lower granularity level features are ideal to use to achieve less impact



on different protocols [112]. Pfefferbaum et al.[109] have successfully integrated T1-weighted MR imaging data obtained with different field strengths. They have used a suitable regression-based correction method to boost the correspondence between two different field strengths with reference to estimating regional volume.

## **2.7 Conclusion**

We have discussed several MRI based features relating to Alzheimer's disease, various tools such as FreeSurfer, SPM, and FSL which are used to extract features from MRI. We have presented various classification approaches which use MRI based features and several classifiers such as SVM, ELM, and TWSVM. Among them, we have found that SVM is more frequently used in AD diagnosis. We have also presented various challenges because of having different protocols of MRI and reviewed several methods which attempted to solve the problem.

## AD Diagnostic Models

This chapter describes the AD detection algorithms that we have used to investigate the impact of protocol differences on the diagnosis of AD/MCI. The algorithms primarily comprise first extracting feature representation from a subject MR image and then training a supervised classifier using those features. The features have been extracted using either MALF based subcortical and ventricular structures, or FreeSurfer based subcortical volumes. We have chosen these two methods as widely used representatives of the current state-of-the-art in brain feature extraction for AD diagnosis. Our primary focus is to address the research question of whether the protocol differences affect the performance and whether these two mainstream methods are robust to those differences. These two methods, MALF based and FreeSurfer based, are briefly described below. We have also chosen supervised classifiers SVM and ELM to analyze the classification performance across protocol differences. These two algorithms are also briefly described. This chapter also discusses the reasons for choosing these algorithms for the investigation of the impact of protocol differences.

### 3.1 Background

It is important to evaluate the way the progression of AD affects specific anatomical structures and to know the correlation between the various cognitive scores and anatomical phenotypes of the brain structure. To achieve this, it is necessary to first segment the brain structures or regions of interest. Manual segmentation done by a neuroanatomist is not feasible as it is time intensive, and operator dependent. Therefore, the fully automated and accurate segmentation techniques are required to enable large-scale neuroimaging research. Atlas-based segmentation techniques are widespread in the research area of brain image analysis.

A brain atlas comprises the information of several brain structures/regions in order to supervise the delineation of the structures while performing segmentation of a brain image. The easiest approach is to provide a visual representation of certain brain structures, and then let the anatomists specify the borders between structures that are nearby. An atlas might be generated from a subject or an average of various subjects or the data based on a particular study/project by the trained neuroanatomist in this field [115]. To achieve anatomical variability, it is recommended to employ multiple atlases, so as to diversify the information that delineates of a structure. This diversification encourages researchers to use various multi-atlas-based segmentation procedures. The use of atlas-driven features from MRI is widespread in automated AD diagnosis. The multi-atlas-based study is a trend which addresses several drawbacks of protocol difference in several studies as we have mentioned in chapter 2. For these reasons, we have chosen a multi-atlas based approach as one of the feature extraction methods to study the impact of protocol differences.

### 3.2 Multi-atlas based pipeline

The two-level hierarchical pipeline algorithm, Multi-Atlas Likelihood Fusion (MALF) has been discussed in this section [116]. Most automated multi-atlas-based methods don't include skull stripping. However, the MALF method has integrated two stages sequentially: skull stripping and brain tissue segmentation.

The brain regions need to be separated from non-brain regions to achieve accurate registration and segmentation of a subject image. An essential step in multi-atlas-based segmentation would be to register the MR image of a subject to an atlas template by nonlinear registration. To guarantee accurate registration between a subject and an atlas, it normally requires an initial step to separate the main brain regions from additional "non-brain" regions. This method is known as "skull-stripping." The method largely relies on the intensity and geometric features of the input MR image. The noise originating from the fluctuation of intensity and geometric features of the image poses a huge

challenge to achieve robust skull stripped image output across different field strengths and scanner models. Other factors such as age of the subject and the resolution of the MR image may also influence the robustness and segmentation accuracy of a method.

Accurate skull stripping is necessary for the three types of MR based feature that we have described in chapter 2 which include atlas-driven volume features, cortical thickness/surface area-based features, and pre-defined ROI based features. Unintentional incorporation of brain tissue with non-brain tissue may lead to inexact feature measurement of various regions/structures of the brain which are located near the skull boundary. This may affect the classification performance. Several automated skull-stripping approaches are commonly used [117], [118].

MALF based skull tripping relies on the downsampling the subject image and applying small deformations. Therefore, this process is much faster. The process creates a preliminary brain mask of four groups: GM, WM, CSF, and lateral ventricles (LV) where the background pixels and skull information are filtered out. Then the brain mask is post-processed such as filling the holes, and smoothing the broken boundary and eliminate small regions by applying morphological operations.

This automated pipeline comprises two hierarchical stages as shown in Figure 3-3. The first level follows three preprocessing steps as explained below.

- Firstly, transform the MR images of a subject and 19 atlases it has employed linearly to Montreal Neurological Institute (MNI) standard template space. In the standard space, the N4 algorithm is used to sharpen the image histogram peak and correct the intensity homogeneity.
- Secondly, perform the nonlinear transformation between the MR image of a subject image and 19 atlases employing Large Deformation Diffeomorphic Metric Mapping (LDDMM). This subject image is aligned with all defined structures from those atlases.
- From the transformed space, a bias field is created which is measured from the dissimilarity in intensity of defined structures between the subject image and the atlases.

The second level again uses LDDMM with three steps varying the degree of elasticity to three different values. Thus, the subject image gets all defined structure levels from those atlases for second time. This level segments the subcortical and ventricular structures of the subject image. This method provides a rich representation of features from a subject image at five different ontology levels which are called granularity levels. The number of brain structures is 8, 19, 54, 136, and 282 for each of the five granularity levels, respectively. All the structures defined at granularity level 3 are mentioned in Table 3-1. One slice of MALF based structures of each protocol is shown in Figure 3-1 where each ROI colors are encoded by  $z$  score. At higher granularity levels, the structures of the lower granularity

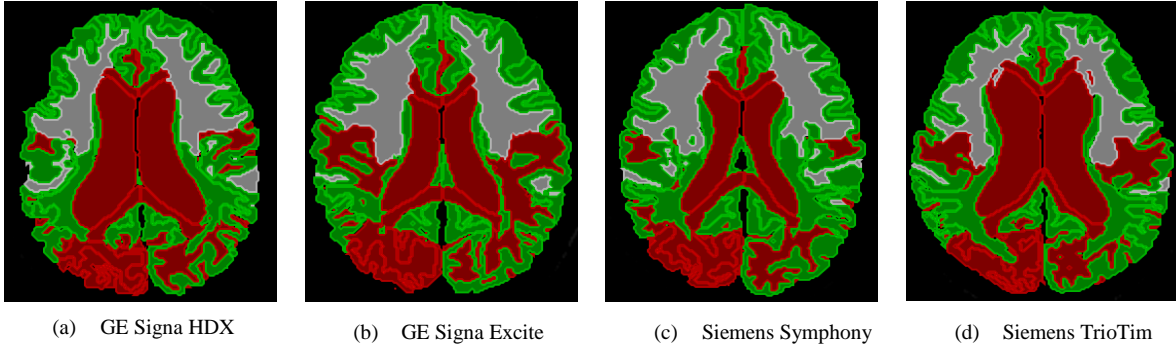


Figure 3-1: MALF based segmentation of four different subjects over four scanner models

Table 3-1: MALF based segmented structures (54 Nos) at 3rd granularity level

| Structure Name   | Structure Name   | Structure Name     | Structure Name  | Structure Name      |
|------------------|------------------|--------------------|-----------------|---------------------|
| AnteriorWM_L     | CinguSul_R       | Insula_R           | midbrain_R      | PosteriorWM_L       |
| AnteriorWM_R     | CorpusCallosum_L | IV_ventricle       | Occipital_L     | PosteriorWM_R       |
| BasalForebrain_L | CorpusCallosum_R | LateralVentricle_L | Occipital_R     | SylvianFissureExt_L |
| BasalForebrain_R | Frontal_L        | LateralVentricle_R | Occipital_Sul_L | SylvianFissureExt_R |
| BasalGang_L      | Frontal_R        | Limbic_L           | Occipital_Sul_R | Temporal_L          |
| BasalGang_R      | FrontSul_L       | Limbic_R           | Parietal_L      | Temporal_R          |
| CentralSul_L     | FrontSul_R       | LimbicWM_L         | Parietal_R      | TempSul_L           |
| CentralSul_R     | III_ventricle    | LimbicWM_R         | Parietal_Sul_L  | TempSul_R           |
| Cerebellum_L     | InferiorWM_L     | Medulla_L          | Parietal_Sul_R  | Thalamus_L          |
| Cerebellum_R     | InferiorWM_R     | Medulla_R          | Pons_L          | Thalamus_R          |
| CinguSul_L       | Insula_L         | midbrain_L         | Pons_R          |                     |

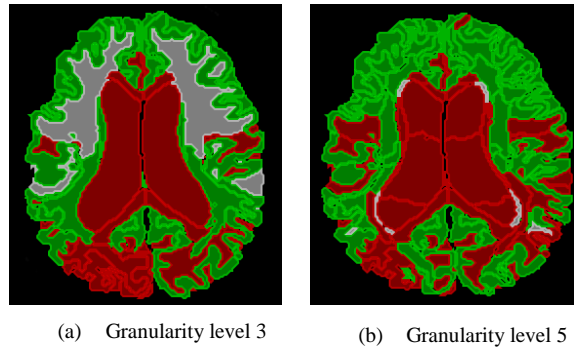


Figure 3-2: MALF based segmentation of a subject image at two granularity levels

are further subdivided to extract more microstructure. For example, all the structures of level 3 and level 4 are subdivided further to get the structures at level 5 as shown in Figure 3-2. Therefore, when the granularity level increases, the number of structures increases, and the size/volume of each structure reduces.

Several articles investigate the segmentation accuracy and robustness of the MALF based approach [112], [116]. For example, Tang et al. [116] have analyzed the segmentation accuracy of MALF algorithm at both the first and second levels, and compared with other conventional skull stripping and fusion based methods. They have considered two kinds of dataset such as pediatric data (3T) and elderly cohorts suffering from dementia (1.5 T) for their experiment. At the first level, the

method overcomes the drawbacks such as segmented distortion and inaccuracy manifested by both the hybrid watershed algorithm (HWA, adopted by FreeSurfer) and by the brain extraction tool (BET, implemented by FSL). At the second level, the segmentation of subcortical and ventricular structures is compared with FreeSurfer and FSL. MALF outperforms both other methods [116]. They have also compared the MALF based method with other level fusion-based segmentation approaches (STAPLE, Spatial STAPLE, and ANTS+PICS) [119]–[121]. Regarding segmentation accuracy, MALF is superior to two of them (STAPLE, and Spatial STAPLE), and similar to the remaining approach (ANTS+PICS). However, they have also revealed that MALF achieves higher segmentation accuracy for the pediatric data (3T) as compared to the elderly data (1.5 T).

The MALF method provides better segmentation accuracy compared to the other methods discussed above, both in skull stripping and segmentation of subcortical and ventricular structures. For these reasons, we have chosen the MALF pipeline to represent the multi-atlas based approaches in our study.

As mentioned in chapter 2, Liang et al. [112] studied the robustness of MALF using AD/MCI data over six different protocols. In particular, they considered the impact of granularity levels over protocol differences. These protocols include three scanner manufactures with two magnetic field strengths. The whole brain was segmented into subcortical and ventricular structures at five granularity levels. Total five types of granularity levels have been defined and the total number of brain regions range from 6 to 286 throughout the phases. The variability of brain volumes is studied corresponding to age, the protocol, and diagnosis of AD from NC and MCI. They considered 120 cohorts from ADNI with six different protocols comprised of 72 normal subjects and 48 AD cohorts. They found no significant protocol difference with images having granularity levels 1-4, but the effect had been observed at granularity level 5. In other words, when granularity increases, the measurement precision decreases and the protocol effect increases. It is harder to define reproducibility of smaller regions, and more contrast difference is apparent when the granularity of image is higher. The age effect is observed throughout granularity levels. Hence, lower granularity level study is ideal to achieve less impact of different protocols. The degree of effect of protocol difference is minor while categorizing the well described (conventional) anatomical features of AD patients.

However, we have not come across any study to investigate how the protocol difference varies the classification performance of AD/MCI from CN, and the ranking of different ROIs which contributes to detecting AD/MCI.

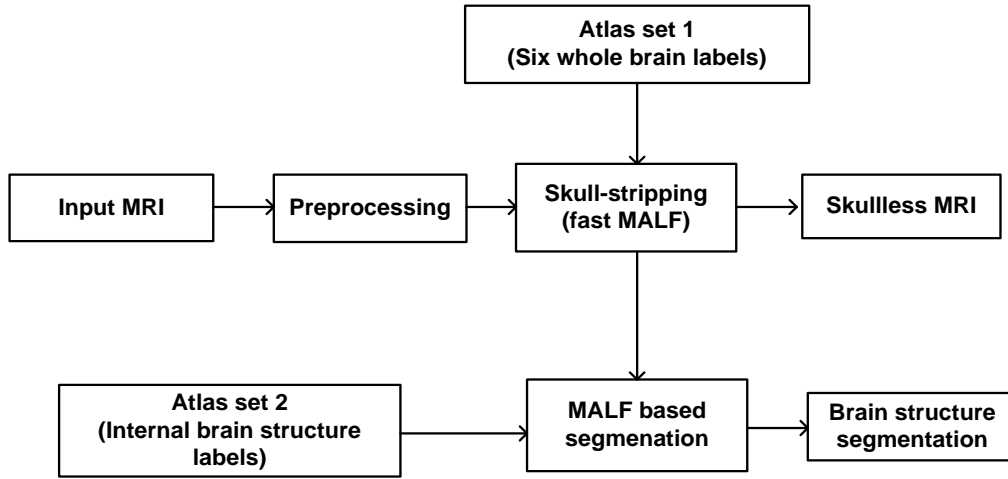


Figure 3-3: Schematic diagram of MALF based segmentation

Therefore, we have investigated the classification performance of the MALF based method over different imaging protocols. The protocol differences include four different scanner models, three with 1.5T field strength and one with 3T field strength. .

The MALF based approach is fully automated and is implemented in MRICloud with multiple sets of atlases. We have selected 19 adult atlases with age range from 50-90 which is similar to our dataset.

### 3.3 FreeSurfer features based pipeline

FreeSurfer is a frequently used tool that provides a fully automated processing stream to extract and analyse features from structural MRI brain images. FreeSurfer is open source and free. Considering its widespread use in MRI based CAD studies, we have chosen this tool to study the impact of protocol differences. While feeding a subject MR image to FreeSurfer pipeline, the image undergoes various stages such as strength non-uniformity correction [122], affine transformation to a common template, intensity normalization, elimination of non-brain tissue [123], linear and non-linear registration to a probabilistic brain atlas and labeling of cortical and subcortical structures/regions [124]. FreeSurfer extracts volumetric features, and thickness and surface area features of cortical parcellation [125]. The volume of subcortical structures is extracted by using FreeSurfer. FreeSurfer also measures estimated total intracranial volume (eTIV), the volumes of the ventricular and corpus callosum regions after aggregating each volume from its corresponding sub-regions. The FreeSurfer package also provides visualization tools to examine various functional/anatomical regions of the brain. We have run the FreeSurfer tool with default settings with the *recon-all* command. We have extracted all volumes the defined regions from the *stat* documents that FreeSurfer produces using the '*asegstats2table*' command. The subcortical volume regions are mentioned in Table 3-2. One slice of subcortically segmented images of four protocols are shown in Figures 3-4(a) through 3-4(d). In these

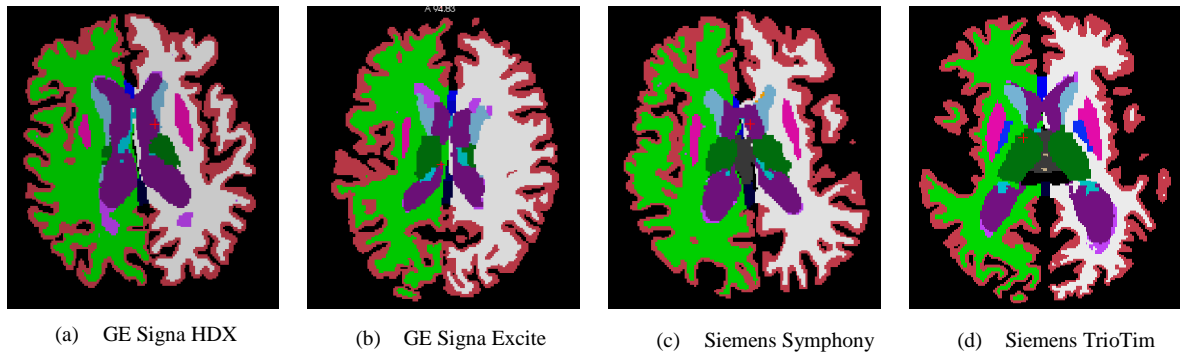


Figure 3-4: FreeSurfer based subcortical segmentation of four different subjects over four scanner models

Table 3-2: FreeSurfer based subcortical segmented structures

| Structure Name         | Structure Name        | Structure Name         | Structure Name       | Structure Name        |
|------------------------|-----------------------|------------------------|----------------------|-----------------------|
| 3rd-Ventricle          | CC_Posterior          | Left-Hippocampus       | rhSurfaceHoles       | Right-Pallidum        |
| 4th-Ventricle          | CortexVol             | Left-Inf-Lat-Vent      | Right-Accumbens-area | Right-Putamen         |
| 5th-Ventricle          | CSF                   | Left-Lateral-Ventricle | Right-Amygdala       | Right-Thalamus-Proper |
| BrainSegVol            | EstimatedTotalIntraCr | Left-Pallidum          | Right-Caudate        | Right-VentralDC       |
| BrainSegVolNotVent     | anialVol              | Left-Putamen           | Right-Cerebellum-    | Right-vessel          |
| BrainSegVolNotVentSurf | Left-Accumbens-area   | Left-Thalamus-Proper   | Cortex               | SubCortGrayVol        |
| BrainSegVol-to-eTIV    | Left-Amygdala         | Left-VentralDC         | Right-Cerebellum-    | SupraTentorialVol     |
| Brain-Stem             | Left-Caudate          | Left-vessel            | White-Matter         | SupraTentorialVolNotV |
| CC_Anterior            | Left-Cerebellum-      | lhCortexVol            | Right-choroid-plexus | ent                   |
| CC_Central             | Cortex                | lhSurfaceHoles         | Right-Hippocampus    | SupraTentorialVolNotV |
| CC_Mid_Anterior        | Left-Cerebellum-      | MaskVol                | Right-Inf-Lat-Vent   | entVox                |
| CC_Mid_Posterior       | White-Matter          | MaskVol-to-eTIV        | Right-Lateral-       | SurfaceHoles          |
|                        | Left-choroid-plexus   | Optic-Chiasm           | Ventricle            | TotalGrayVol          |
|                        |                       | rhCortexVol            |                      | WM-hypointensities    |

images, each ROI has a specific RGB color code. We have also investigated the classification performance using those FreeSurfer based features over different imaging protocols.

### 3.4 Supervised learning algorithms

We have briefly discussed the algorithms we have applied in our study. As we have seen in chapter 2, several studies based on SVM and K-SVM using ADNI dataset have shown that these have good potential to distinguish AD from NC and MCI. SVM performs better as compared to other classifiers such as Bayesian, and random forest. The referenced studies either used exclusively 1.5 T images or used both field strengths (1.5 T and 3T) without considering the impact of the protocol differences due to different scanner models with possibly different field strengths. Therefore, in our study, KSVM (RBF Kernel) has been trained and tested to obtain the performance result for detection of AD and MCI while considering the protocol differences. The performance is compared with another efficient classifier, Extreme Learning machine (ELM). The proposed overall classification approach is shown in Figure 3.5.



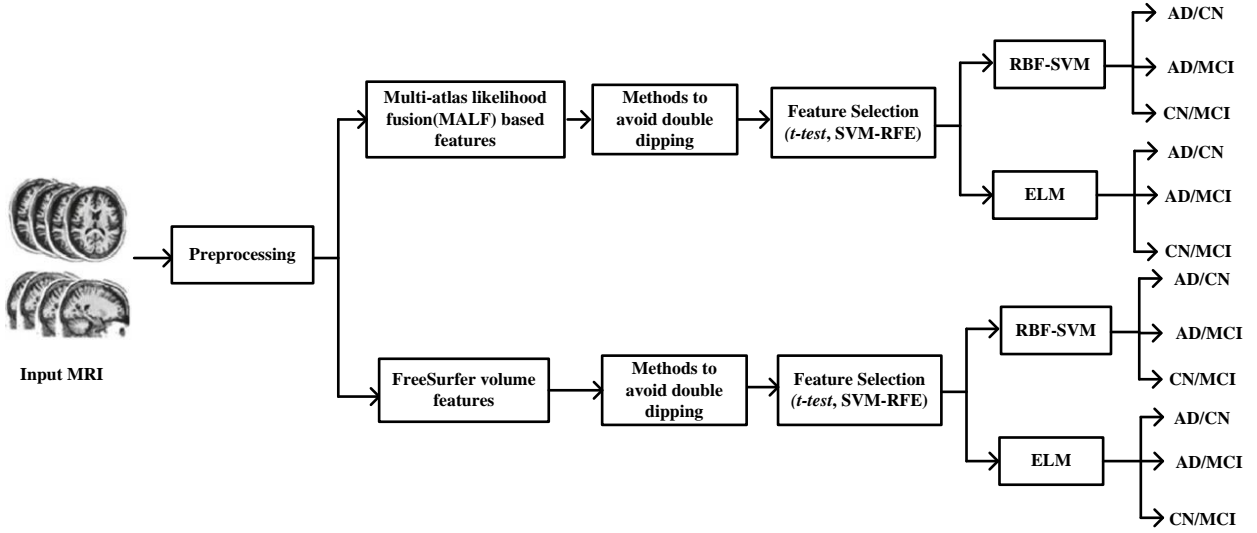


Figure 3-5:Schematic diagram of the proposed approach

### 3.4.1 Support Vector Machine

The Support Vector Machine is an efficient binary classification algorithm which creates a decision surface. The decision surface is a line for two-dimensional data and a hyperplane for higher dimensional data. It can separate two classes by studying the distinctness of the training set which contains samples of both classes. The classifier solves the hyperplane equation in such a way that the distance between the hyperplane and the support vectors is maximal. SVM separates the feature vectors between two classes while maximizing the margin between the data and the decision surface [44] as illustrated in Figure 3-6. The hyperplane equation can be solved in primal or in dual form. In dual form, the hyperplane equation is optimized by Lagrange multiplier [126].

Kernel SVM (KSVM) is very useful to deal with the classification problem of linearly non-separable data. If the data points are not linearly separable, the data is mapped into a higher dimensional space to achieve separability. KSVM applies linear SVM in the higher dimensional space. Several kernel mapping methods such as RBF (Radial Basis Function), Polynomial Kernel, and sigmoid kernel are commonly used. The most frequently used kernel among them is RBF kernel because of its robustness as compared to other kernels (i.e. Polynomial Kernel) [127], [128].

#### 3.4.1.1 SVM Training

Let the training set be  $T = \{(x_i, y_i)\}_{i=1}^n$ , where  $x_i \in R^d$  is the training input data and its corresponding class level is  $y_i \in \{-1, +1\}$ . It is required to find the maximum-margin hyperplane that divides one group (+1) from the other (-1).

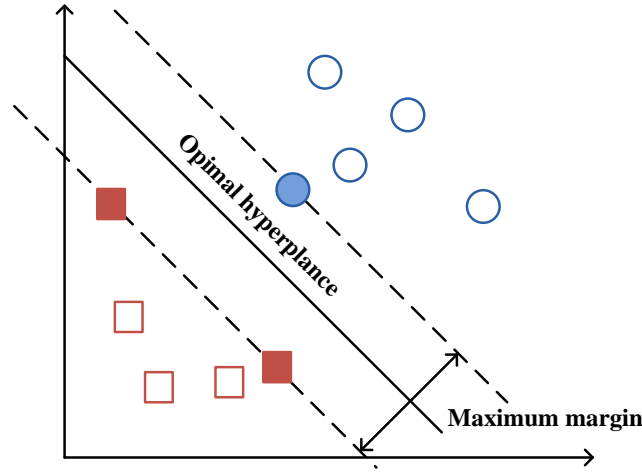


Figure 3-6: Linear Support Vector Machine

The kernel mapping function,  $\Phi: R^d \rightarrow R^f (f > d)$ , embeds the feature set in a higher dimensional space to make it linearly separable.

The kernel mapping can be represented as

$$K = k(x, x') = \langle \Phi(x), \Phi(x') \rangle = \Phi(x)^T \Phi(x') \quad (3-1)$$

$$K(x, x') = \exp(-\gamma \|x - x'\|^2), \gamma > 0 \quad (3-2)$$

The decision boundary at kernel space is denoted by

$$f(x) = w^T \phi(x) + w_0 \quad (3-3)$$

The hyperplane can be represented as

$$f(x) = 0 \text{ or } w^T f(x) + w_0 = 0 \quad (3-4)$$

Where,  $w$  is weight vector which is normal to the hyperplane, and  $w_0$  is a bias term.

When  $f(x) > 0$ , then  $y = +1$  and  $f(x) < 0$ , then  $y = -1$

The primal formulation of SVM with L1 soft margin regularization can be solved as a primal problem,

$$\begin{aligned} \min_{w, w_0} & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t. } & y_i (\langle w, \phi(x_i) \rangle + w_0) \geq 1 - \xi_i, \forall i, \\ & \xi_i \geq 0 \end{aligned} \quad (3-5)$$

Where,  $C$  is the trade-off parameter of training error and margin, and  $\xi_{1:n}$  is a slack vector with non-zero elements. In our study, we have selected the optimized value of  $C$  from the range of -1 to 3 and the optimized value of  $\gamma$  from the range of -4 to 1 based on the performance of the RBF-SVM.

### 3.4.2 Extreme Learning Machine

Extreme Learning Machine is an emerging efficient pattern classification algorithm [129] successfully used in various research fields. It has gained great breakthrough in fields such as image enhancement, video application [130], medical application [97], [131], and ship detection [132]. The ELM has very fast learning speed, better generalization performance compared to gradient-based algorithm such as back-propagation methods, and is less prone to problems like local minima, improper learning rate, and overfitting, etc. ELM is flexible with hidden activation functions. It has the advantage of comparatively superior performance over SVM and its variants [133]. We have applied extreme machine learning rather than using deep learning-based classification because deep learning models need a large number of sample for optimum performance. The ELM algorithm exhibits higher performance when the dataset is small. It is basically a single hidden layer feedforward neural network (SLFN).

#### 3.4.2.1 ELM Training

Again, let the training set be  $T = \{(x_i, y_i)\}_{i=1}^n$ , where  $x_i \in R^d$  is the input training data and its corresponding class level is  $y_i \in \{-1, +1\}$ . The output function of single hidden layer SLFN with  $L$  hidden neurons can be denoted as

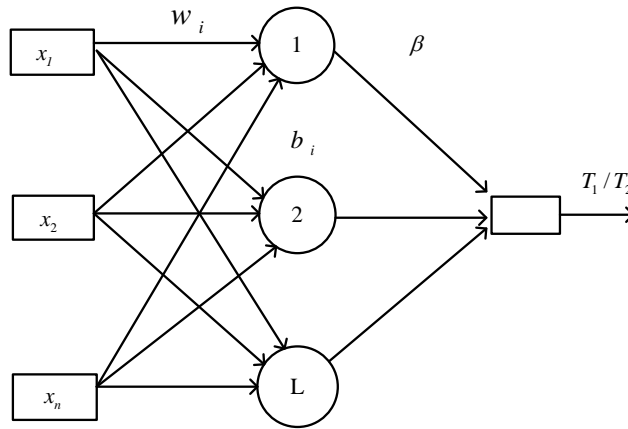


Figure 3-7: ELM SLFN approach

$$f(x_i) = \sum_{k=1}^L \beta_k h_k(a_k, b_k, x_i) = h(x_i) \beta, i = 1, \dots, N \quad (3-6)$$

Where,  $\beta_k$  is the weight matrix of the output,  $h(x_i) = h_k(a_k, b_k, x_i), k = 1, \dots, L$  is the output of the network with respect to its training sample  $x_i$ ,  $h(\cdot)$  is a nonlinear continuous function, and  $a_k, b_k$  are the corresponding parameters of  $k$ -th hidden neuron. The method selects the optimized parameters by minimizing the error function  $\|H\beta - Y\|$ . The steps of a SLFN is shown in Figure 3-7. It assigns the hidden node parameters randomly with the strategy of tuning-free training unlike

other conventional iterative weight updating based neural network. As a result, it is less computationally intensive [134]. The optimization problem can be solved by least squares (LS) algorithm easily [135]. The network output weights are solved by a dual optimization problem [136].

### 3.5 Conclusion

We have extracted features from subcortical and ventricular regions using both five granularity levels based MALF and Freesurfer methods. The MALF based methods are hierarchical and include skull stripping. The skull stripping in MALF is also more accurate as compared to hybrid watershed-based skull stripping employed by FreeSurfer, and brain extraction tool (BET) adopted by FSL. MALF separates various non-brain regions where the other two methods consider those regions as a part of the brain. Accurate separation of non-brain regions is important because it may affect nonlinear registration of the subject image with an atlas which eventually influences parcellation (segmentation) of the brain into subcortical and ventricular regions. Also, at the second hierarchical level, MALF provides several brain structures located in subcortical and ventricular regions at five different granularity levels. The number of structures varies based on the granularity levels. Higher granularity levels provide features from a greater number of regions. It is important to see how such a robust method performs over different protocols and also investigate whether affected regions/structures remain the same or vary while using different protocols. Since FreeSurfer subcortical volume structures are also frequently used by the neuro research community, it is also very important to verify the robustness of this method over protocol difference regarding classification accuracy. As far as methods to measure the classification accuracy, SVM performs better in most of the studies. So, we have used SVM. We have applied another efficient recent supervised method, the ELM classifier, to compare with the performance of SVM because other studies claim superior performance for ELM compared to SVM.

## Data and experimental work

This chapter describes the data used for the two different tools used in our experiment. It also includes various feature selection methods such as the t-test and Support Vector Machine Recursive Feature Elimination (SVM-RFE), and validation technique we have employed.

### 4.1 Dataset

Data used in the preparation of this thesis were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu>). The ADNI was launched in 2003 as a public-private partnership led by Principal Investigator Michael W. Weiner, MD. The primary goal of the ADNI is to test whether serial MRI, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early-onset Alzheimer’s disease AD. For up-to-date information, visit [www.adni-info.org](http://www.adni-info.org).

Alzheimer’s Disease Neuroimaging Initiative (ADNI) is an ongoing, longitudinal, multicenter study [41]. The project has three phases: ADNI-1, ADNI-GO, ADNI-2, and ADNI-3. The ADNI-1 phase has both 1.5 and 3 T MR imaging scans, whereas other phases have only 3T scans. ADNI included other imaging biomarkers such as Diffusion Tensor Imaging (DTI) and Functional MRI(fMRI) after the ADNI-1 phase. This is described in Table 4.1.

Our experiment uses the magnetization-prepared rapid gradient-echo (MP-RAGE) sequence because MPRAGE provides both high tissue contrast and covers whole brain with high spatial resolution with a limited number of scans [137].

The data used in our thesis was acquired by MRI scanners from two different manufacturers: General Electric Healthcare (GE) and Siemens. It consists of two models of GE scanner, the Signa HDX and the Signa Excite, and two models of Siemens, the Symphony and the TrioTim. The Siemens TrioTim model has a 3T magnetic field strength whereas the other scanners are 1.5 T only.

#### 4.1.1 Data for the Multi-Atlas Likelihood fusion (MALF) based algorithms

We ran the MALF algorithm in MRICloud (<https://braingps.mricloud.org/>) which is a service for registered users only. It is a high-throughput cloud-based software as a service (SaaS) which has high computation capability and supported by NSF XSEDE at the User Computational Anatomy Portal [138].

We downloaded 1200 images from the ADNI website, and randomly selected one scan from multiple scans for each subject. These selected subjects have been described in Table 4-2. Since MRICloud software requires the images in a specific orientation, we downloaded raw *dicom* format images and converted to the analyzer format (IMG/HDR) before executing the multi-atlas-based methods. All the downloaded *dicom* scans were not preprocessed. All the pre-processing methods are integrated in the MRICloud pipeline.

#### 4.1.2 Data for the FreeSurfer based Method

CBRAIN is also a cloud based service which facilitates to use multiples neuroimaging tools which need high computing power [139]. CBRAIN facilitates the use of FreeSurfer on their server for registered users. The MRI data used for FreeSurfer is available from the ADNI website. It is ideal to use preprocessed data where several imaging artifacts are corrected. The preprocessed data are in *nifti* format. Therefore, 1200 images have been downloaded in *nifti* format. When downloading, some *nifti* files got corrupted, and we did not include those corrupted files. As a result, not all the images we have used here are the same as the data used in MALF based approach. We randomly selected one scan from multiple scans of each subject. The selected subjects have been briefly described in Table 4.-3.

Table 4-1: Different phases of ADNI project

| Projects | 1.5TMRI | 3T MRI | DTI | fMRI | Weighted Image |
|----------|---------|--------|-----|------|----------------|
| ADNI-1   | Yes     | Yes    | No  | No   | T1/T2/PD       |
| ADNI-GO  | No      | Yes    | Yes | Yes  | T1/T2/PD       |
| ADNI-2   | No      | Yes    | Yes | Yes  | T1/T2/PD       |
| ADNI-3   | No      | Yes    | Yes | Yes  | T1/T2/PD       |

Table 4-2: Imaging parameters used in MALF based approach

| Group | Age       | Scanner models   | Field strength (3T) | Number of subjects |
|-------|-----------|------------------|---------------------|--------------------|
| CN    | 64.3-90.8 | GE HDX           | 1.5                 | 40                 |
| CN    | 70.1-91.8 | GE Excite        | 1.5                 | 83                 |
| CN    | 62.2-95.3 | Siemens Symphony | 1.5                 | 42                 |
| CN    | 56.1-90.2 | Siemens TrioTim  | 3                   | 92                 |
| AD    | 56.4-90.9 | GE HDX           | 1.5                 | 56                 |
| AD    | 55.2-91.3 | GE Excite        | 1.5                 | 74                 |
| AD    | 57.9-88.8 | Siemens Symphony | 1.5                 | 35                 |
| AD    | 56-89.1   | Siemens TrioTim  | 3                   | 67                 |
| MCI   | 55.8-89.4 | GE HDX           | 1.5                 | 51                 |
| MCI   | 55.2-88.2 | GE Excite        | 1.5                 | 72                 |
| MCI   | 56.3-91.9 | Siemens Symphony | 1.5                 | 50                 |
| MCI   | 55.1-93.5 | Siemens TrioTim  | 3                   | 75                 |

Each MPRAGE image undergoes three kinds of preprocessing phases

- **Gradwarp:** This method corrects the image geometry which is distorted by the scanner. In practice, the strength of gradient field doesn't vary linearly with the distance from magnet isocenter. This gradient non-linearity distorts the image geometry. Gradwarp corrects this distortion.
- **B1 non-uniformity:** The method uses calibration of time-varying radio frequency field (B1) parameters to correct the artifacts of an image. This artifact, the non-uniformity in image intensity occurs when the degree of uniformity at head coil and receiver coil varies during Radio Frequency(RF) transmission.
- **N3:** This method normalizes the non-uniform intensities, sharpening the histogram of an image which is already preprocessed through the Grad warp and B1 non-uniformity correction methods.

### 4.1.3 Differences between the data sets

The FreeSurfer implementation requires images in *nifti* format, whereas the MALF implementation that we are using requires images in *dicom* format. We downloaded ADNI MRI images in both formats, but we found that some images were corrupted in one format but not the other. Since the goal of our study is to determine whether protocol differences impact the classification performance, there is no need to compare the feature extraction methods with each other. In fact, it is preferable to perform independent experiments for each feature classification method to reduce the impact of sampling. For these reasons, we have selected subject images randomly and independently for each scanner model from the data available for each technique. Thus we have constructed two different datasets that are as independent as possible. We make no claim regarding the comparison of performance results between these two methods for AD/MCI detection but focus on the impact of protocol differences on the classification performance.

## 4.2 Methods to avoid double dipping

In the ADNI project, longitudinal data is available where MRI scans of a subject were collected again after a specific period (baseline, six months, 12 months, etc.). Meanwhile, some subject progressed to another stage which is related to the degree of brain severity. For example, a normal control may progress to MCI or an MCI cohort may progress to AD. The classification problem needs to avoid double dipping where a test subject shouldn't also be used in training phase. Otherwise, this will introduce bias in performance because the classifier will have prior knowledge of the test subjects during the training period. We applied the following steps to avoid double dipping:

- Eliminate converted patients from the groups. For instance, if a subject convert from MCI to AD at a later stage, then the subject is eliminated from both groups.
- Randomize and rearrange remaining subjects with AD and the remaining subjects from CN separately.
- Split AD set and CN set into 5 folds separately
- Randomly select one fold from AD set and one fold from CN set, concatenate them to create ADCN testing set
- Select remaining four folds from AD set and four folds from CN, concatenate them to create ADCN training set.

A similar approach was implemented while classifying AD vs MCI and CN vs MCI.



Table 4-3: Imaging parameters used in FreeSurfer

| Group | Age       | Scanner models   | Field strength (3T) | Number of subjects |
|-------|-----------|------------------|---------------------|--------------------|
| CN    | 64.2-90.8 | GE HDX           | 1.5                 | 41                 |
| CN    | 70.1-9.8  | GE Excite        | 1.5                 | 62                 |
| CN    | 62.2-95.3 | Siemens Symphony | 1.5                 | 19                 |
| CN    | 56.1-90.2 | Siemens TrioTim  | 3                   | 82                 |
| AD    | 56.4-90.9 | GE HDX           | 1.5                 | 49                 |
| AD    | 55.2-91.3 | GE Excite        | 1.5                 | 60                 |
| AD    | 57.9-88.8 | Siemens Symphony | 1.5                 | 34                 |
| AD    | 56-89.1   | Siemens TrioTim  | 3                   | 54                 |
| MCI   | 55.8-89.4 | GE HDX           | 1.5                 | 40                 |
| MCI   | 55.2-88.2 | GE Excite        | 1.5                 | 67                 |
| MCI   | 56.3-91.9 | Siemens Symphony | 1.5                 | 52                 |
| MCI   | 55.1-93.5 | Siemens TrioTim  | 3                   | 70                 |

Table 4-4: Selected data for MALF and FreeSurfer based approach

| Classification sets | Model of the scanners | Dataset for MALF | Dataset for FreeSurfer |
|---------------------|-----------------------|------------------|------------------------|
| AD vs CN            | GE Signa HDX          | AD=40, CN=40     | AD=41, CN=41           |
|                     | GE Signa Excite       | AD=74, CN=74     | AD=60, CN=60           |
|                     | Siemens Symphony      | AD=35, CN=35     | AD=19, CN=19           |
|                     | Siemens TrioTim       | AD=67, CN=67     | AD=54, CN=54           |
| AD vs MCI           | GE Signa HDX          | AD=51, MCI=51    | AD=40, MCI=40          |
|                     | GE Signa Excite       | AD=72, MCI=72    | AD=60, MCI=60          |
|                     | Siemens Symphony      | AD=35, MCI=35    | AD=34, MCI=34          |
|                     | Siemens TrioTim       | AD=67, MCI=67    | AD=54, MCI=54          |
| CN vs MCI           | GE Signa HDX          | CN=40, MCI=40    | CN=40, MCI=40          |
|                     | GE Signa Excite       | CN=72, MCI=72    | CN=62, MCI=62          |
|                     | Siemens Symphony      | CN=42, MCI=42    | CN=19, MCI=19          |
|                     | Siemens TrioTim       | CN=75, MCI=75    | CN=70, MCI=70          |

We have balanced the data to have the same number of samples for each group to calculate balanced classification performance as shown in Table 4-4.

### 4.3 Feature Ranking by the two-sample $t$ test

The evaluation of the statistical significance of data is widely used to find out the degree of difference between two groups. The two sample  $t$ -test is performed on two random groups where the both the groups are assumed to be normally distributed. The frequently used term to measure the level of significance is called  $p$ -value. The  $p$ -value of a sample less than 0.05 means that the test rejects the null hypothesis that two groups have the same mean. The features can be ordered based on the  $p$ -value. Smaller  $p$ -values indicate more significant features.

In our experiment, one sample is a feature of healthy controls, another sample is the corresponding feature of the AD cohorts. The groups are independent. Therefore, the two-sample  $t$ -test is applied here. The features acquired from different imaging structures or different ROIs are ranked based on its capability to separate the two groups. The feature ranking is done using the  $p$ -value of two sample  $t$ -tests.

#### 4.4 Support Vector Machine Recursive Feature Elimination (SVM-RFE)

We obtain features of every subject after applying the MALF algorithm and FreeSurfer tool to the MRI data. The extracted features from each defined Region of Interest don't necessarily contribute to the classification tasks. Discarding unnecessary data is efficient. Moreover, large feature dimensionality in classification problem is a curse for generalization. To obtain useful training of supervised algorithms, the input data should be non-redundant and highly relevant to avoid overfitting and underfitting. The use of fewer features not only increases the classification performance but also reduces the classification time. The SVM-RFE method employs the support vector machine (SVM) algorithm which is one of the most efficient and well-established methods for binary classification [140]. It applies the RFE based ranking criteria. The features are ranked based on the weights of the SVM primal problem.

SVM RFE has four main steps

- Selecting a subset of feature space or a feature subspace
- Training and testing in the sub-feature space
- Ranking the features based on the coefficients of SVM primal problem
- Drop the features whose rank are below the threshold value.

The algorithm of SVM-RFE is summarized below.

##### 4.4.1 SVM-RFE algorithm:

Let the training dataset  $\{(x_i, y_i)\}_{i=1}^n$ , where  $x_i \in R^d$ , its corresponding class level is  $y_i \in \{-1, +1\}$ .

The surviving training set is

$s_i = s_i(:, j)$ , where  $j=1, \dots, \ell, \ell \leq n$ , indices of surviving of features,

Then, SVM classifier is trained. The coefficients in the dual form are calculated. Consecutively, the weights of the primal form are measured from those coefficients.

$$w = \sum_k \alpha_k y_k x_k \quad \alpha \text{ is a coefficient optimized at the dual form}$$

The ranking value is computed as

$$r_i = (w_i)^2 \text{ for all } i$$

The features having smaller ranking values or weights are computed

$$f = \text{argmin}(c)$$

Then the feature ranking values are updated, and some features are discarded which have smaller ranking values. This process continues until no feature survives.

## 4.5 Evaluation matrices

The performance result can be measured using a confusion matrix, as shown in Table 4.5. The cohorts who are correctly labeled by the classifier are located diagonally [141]. These are true positives (TP) which denotes correctly classified patients and true negatives (TN) which corresponds to correctly labeled healthy controls. The off-diagonal elements are false positives (FP) which represents the number of healthy controls incorrectly classified as patients, and false negatives (FN) which denotes the number of patients wrongly labeled as healthy. Our experiment employs six terms to measure the performance of the classifier. These are accuracy, sensitivity, specificity, precision, f-measure, and g-mean.

The accuracy can be defined as the proportion of subjects correctly labeled compared to the total number of subjects.

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

The sensitivity is the proportion of true positives relative to the total number of patients.

$$\text{sensitivity} = \frac{TP}{TP + FN}$$

Table 4-5: Confusion matrix for a binary classifier

| True Class     | Predicted Class |                |
|----------------|-----------------|----------------|
|                | $P$ (Patients)  | $H$ (Controls) |
| $P$ (Patients) | TP              | FN             |
| $H$ (Controls) | FP              | TN             |

The specificity can be defined as true negative rate relative to the total number of controls.

$$specificity = \frac{TN}{TN + FP}$$

The precision is the positive predictive value.

$$precision = \frac{TP}{TP + FP}$$

F1 measure is the harmonic mean of precision and sensitivity.

$$F_1\ score = 2 * \frac{precision * sensitivity}{precision + sensitivity}$$

Gmean of the geometric mean of sensitivity and specificity.

$$Gmean = \sqrt{sensitivity * specificity}$$

## 4.6 Cross-Validation

The Cross-validation (CV) technique evaluates the performance of the predictive model which makes predictions on unknown or new datasets. This CV portions a dataset and uses a subset of input data for training an algorithm and the remaining subset for testing. At each fold, the training subset is used to train a supervised learning algorithm with the corresponding testing set is used to assess the performance. This process is repeated  $k$  times for  $k$ -fold CV. It randomly partitions data into  $k$  folds, the data from  $k-1$  folds are employed as the training set, and the remaining data from one fold is used for testing. This process is repeated  $k$  times so that every subject is used in  $k-1$  training set and one testing set.

In our thesis, at each round of iteration, all the six measures which are described above are calculated. We ran this 5-cross-validation 30 times. At each run, the data in training and testing fold are shuffled randomly. Then we calculated the average and standard deviation of those six measures.

|             |               |               |               |               |               |
|-------------|---------------|---------------|---------------|---------------|---------------|
| Iteration 1 | Testing Fold  | Training Fold | Training Fold | Training Fold | Training Fold |
| Iteration 2 | Training Fold | Testing Fold  | Training Fold | Training Fold | Training Fold |
| Iteration 3 | Training Fold | Training Fold | Testing Fold  | Training Fold | Training Fold |
| Iteration 4 | Training Fold | Training Fold | Training Fold | Testing Fold  | Training Fold |
| Iteration 5 | Training Fold | Training Fold | Training Fold | Training Fold | Testing Fold  |

Figure 4-1: Illustration of 5-fold cross-validation

## 4.7 Used Tools/Packages

In our thesis, we have used these tools/methods: FreeSurfer, MALF pipeline, SVM-Recursive Feature Elimination, RBF-SVM, and ELM. We ran FreeSurfer in the cbrain cloud [139], MALF pipeline in the MRICloud [138], and all other methods in MATLAB R2017a. We used LibSVM library for running RBF-SVM [142]. We used the code for ELM from [ntu.edu.sg/home/egbhuang/elm\\_codes.html](http://ntu.edu.sg/home/egbhuang/elm_codes.html), and SVM-RFE from [mathworks.com/matlabcentral/fileexchange/50701-feature-selection-with-svm-rfe](http://mathworks.com/matlabcentral/fileexchange/50701-feature-selection-with-svm-rfe). We modified those codes as per our requirement.

## Result and Discussion

We have investigated the robustness of the predictive method using multiple granularity levels across four sets of data having different protocols. The predictive performance is measured using two supervised classifiers, RBF-SVM and ELM. We have also identified the most affected regions due to the progression of the disease across those sets of data. We have also performed a similar investigation using FreeSurfer based subcortical volume features.

### 5.1 Feature setup for classification

We have used a feature selection method after extracting subcortical and ventricular features using MALF pipeline and FreeSurfer, and measured the classification performance. SVM recursive feature elimination (SVM-RFE) has worked as a feature selection method here. SVM-RFE selects the most discriminatory regional features based on weights of the hyperplane. In other words, it selects features based on the separability of the data during classification. This method has been described in chapter 4.

After filtering out less significant features, we train two supervised classifiers, RBF kernel SVM and ELM over four different sets of data having the different protocols (GE Signa HDX, GE Signa excite, Siemens Symphony, and Siemens TrioTim). We have calculated the classification

performance result of AD/CN (AD vs CN), AD/MCI, and CN/MCI in terms of six performance measures (accuracy, sensitivity, specificity, precision, F1 score, and gmean). We have applied 5-fold cross-validation, and run the entire process 30 times, for a total of 150 classification experiments. Finally, we have calculated performance using the mean and standard deviation over all folds of 30 runs.

## 5.2 Classification performance of Multi-atlas-based pipeline

We have analyzed the classification performance using MALF based ventricular and subcortical structures over multiple granularity levels as input feature to the two classifiers RBF-SVM and ELM.

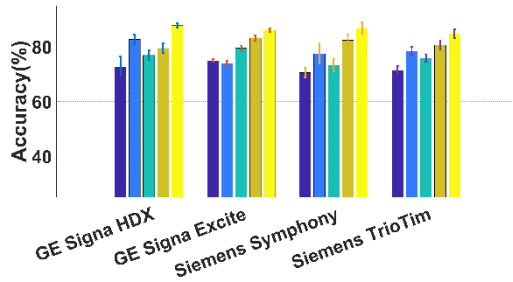
### 5.2.1 Performance of RBF kernel SVM classifier

We have analyzed the performance of RBF SVM over four different scanner models. The performance accuracy while separating AD from CN is shown in Figure 5-1(a). At granularity level 5, the performance accuracy is 87.75%, 85.97%, 86.61%, and 84.61% for the four different scanner models. Across all scanners, the highest granularity level 5 provides better accuracy than lower granularity levels ranging from granularity level 1-4. The accuracy increases consistently for levels 3-5. However, the accuracy fluctuates from level 1 to level 3. For three scanners, GE Signa HDX, Siemens Symphony, and Siemens TrioTim, granularity level 2 outperform levels 1 and 3. But for GE Signa Excite, granularity level 2 performs worse than levels 1 and 3. The accuracy at granularity level 1 is lowest across all scanners except GE Signa Excite where granularity level 2 exhibits the lowest accuracy. All the performance measures for granularity level 5 for all models are shown in Figure 5-2(a). The performance depicts that GE Signa HDX scanner outperforms other scanner models here.

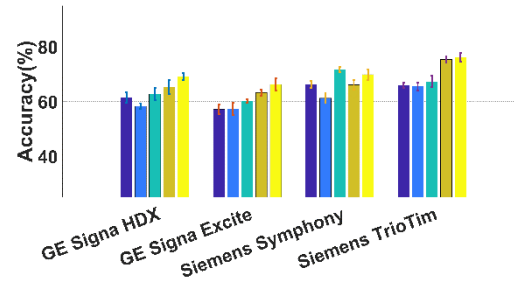
We have measured the AD/MCI classification performance for all scanner models in similar way. The performance accuracy is depicted in Figure 5-1(b), and performance result at granularity level 5 is depicted in Figure 5-2(b). The performance accuracy at granularity level 5 is 69.13%, 66.21%, 69.80%, and 76.07% for the four different scanner models. Granularity level 5 provides better accuracy than the lower granularity levels (granularity levels 1-4) for all scanners except Siemens Symphony where level 3 provides the highest accuracy. The accuracy rises considerably when the granularity level increases from 2 to 5 for all scanner models except Siemens Symphony.

In case of Siemens Symphony scanner, the accuracy shows variation across all granularity levels where level 3 performs slightly better than level 5. For all scanners, accuracy at granularity level 1 is slightly better or equal as compared to granularity level 2. Figure 5-2(b) depicts that the performance result of the Siemens TrioTim scanner outperforms other scanner models.

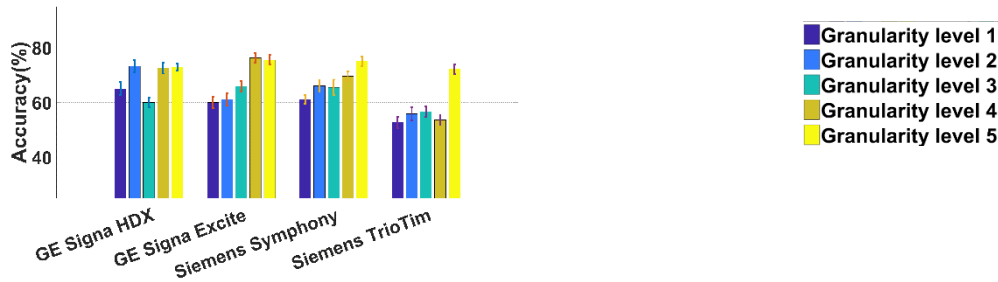
We have also recorded CN/MCI classification performance for all scanner models across all granularity levels. At granularity level 5, the accuracy of the four models is 72.91%, 75.47%, 75.04%, and 72.2% as shown in Figure. 5-1(c). Granularity level 5 provides better accuracy than levels 1-4 for three scanners, GE Signa HDX, Siemens Symphony, and Siemens TrioTim. But in case of GE Signa Excite model, accuracy at granularity level 5 is slightly less than at level 4. Considering GE Signa Excite and Siemens TrioTim, the accuracy increases considerably when the granularity level increases from 1 to 3. But for the other two models, granularity level 2 performs slightly better than levels 1 and 3. The lowest level performs worst across scanners except GE Signa HDX where the accuracy at level 3 is lowest. Figure 5-2 (c) exhibits all the performance measures for granularity level 5. The performance result indicates that GE Signa Excite and Siemens Symphony scanners perform comparably and outperform other scanner models here.



(a) AD vs CN



(b) AD vs MCI



(c) CN vs MCI

Figure 5-1: Performance accuracy of MALF based method using RBF-SVM over four different protocols



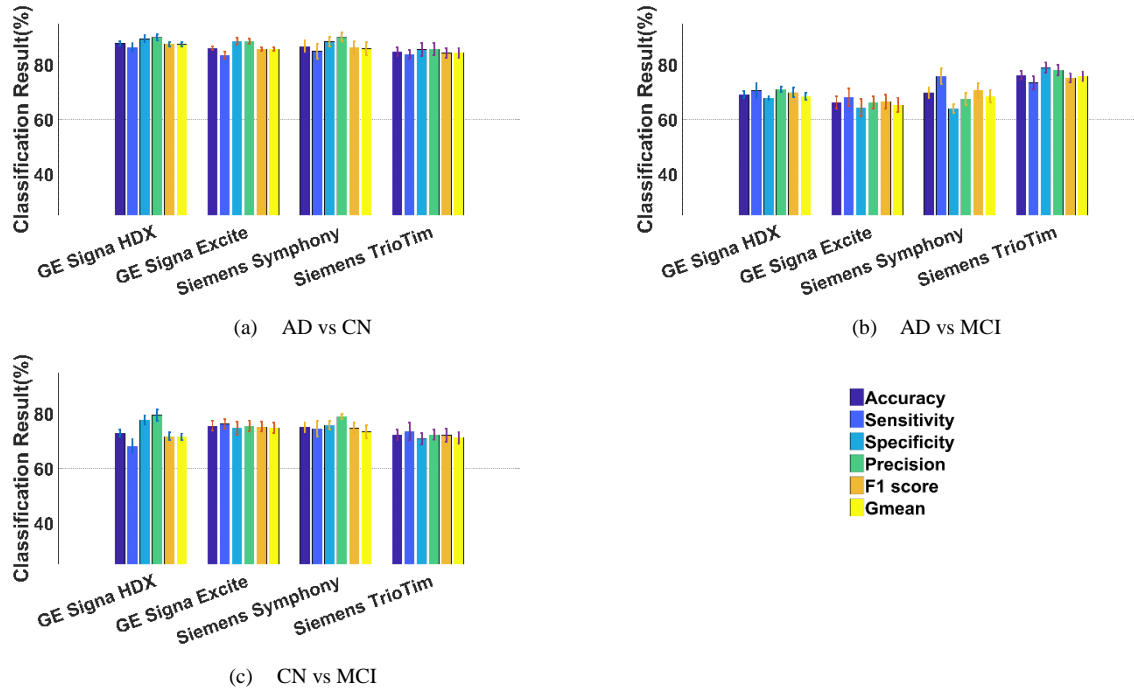


Figure 5-2: Performance result of MALF based features at granularity level 5 using RBF-SVM over four different protocols

### 5.2.2 Performance of ELM classifier

We have analyzed the performance of ELM over granularity levels 3 through 5, and the same four scanner models. We omitted recording the performance results for granularity levels 1 and 2 as the performance of RBF-SVM consistently shows poor result over these granularity levels.

When classifying AD/CN, the accuracy is higher for GE scanners at granularity level 5, than at other levels. These scanners have accuracy of 91.37% for Signa HDX and 87.80% for Signa Excite as shown in Figure 5-3(a). But for the two models of Siemens, accuracy at granularity level 4 is 85.05% and 83.91% which is slightly better than level 5, and significantly higher than level 3. The lowest granularity level provides the lowest performance across all scanners as shown in Figure 5-4(a).

When classifying AD from MCI, the performance at level 5 is higher than levels 3 and 4 for both GE and Siemens scanners across their models as shown in 5-3(b). The accuracy is up to 66%, and 67.04% for GE scanners, and 67.94% and 78.38% for Siemens. The lowest granularity level provides lower accuracy for all scanners. Figure 5-4(b) depicts the performance result indicating that GE Signa HDX, GE Signa Excite and Siemens Symphony scanner models perform comparably with each other. Siemens TrioTim outperforms the other scanner models here.

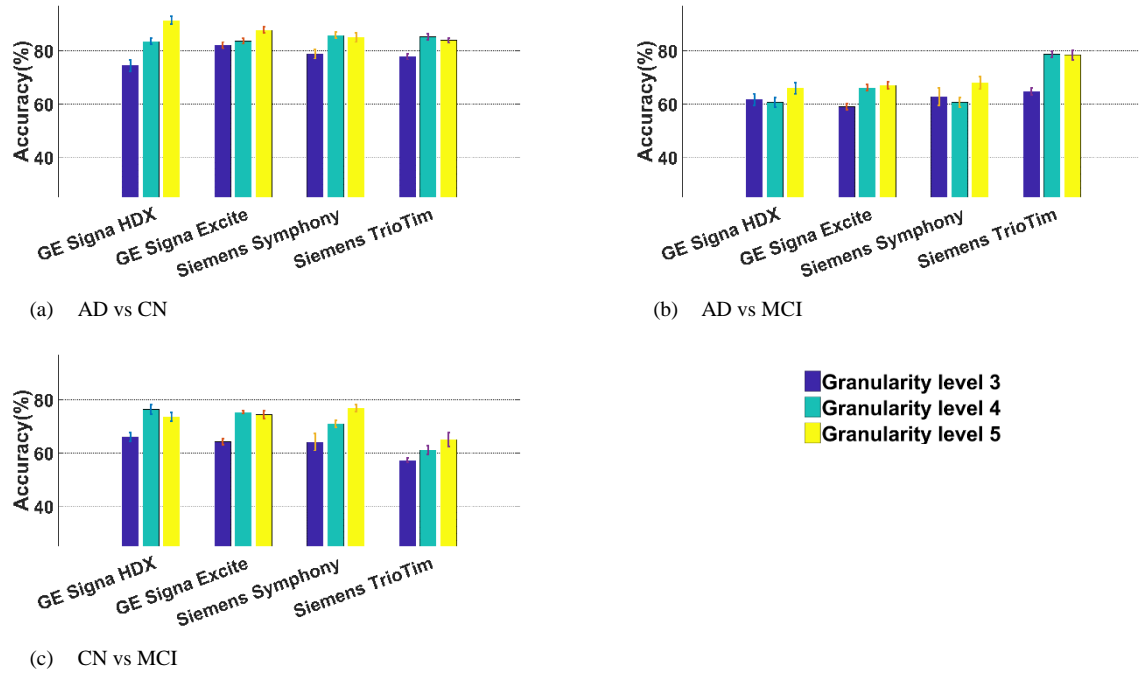


Figure 5-3: Performance accuracy of MALF based method using ELM over four different protocols

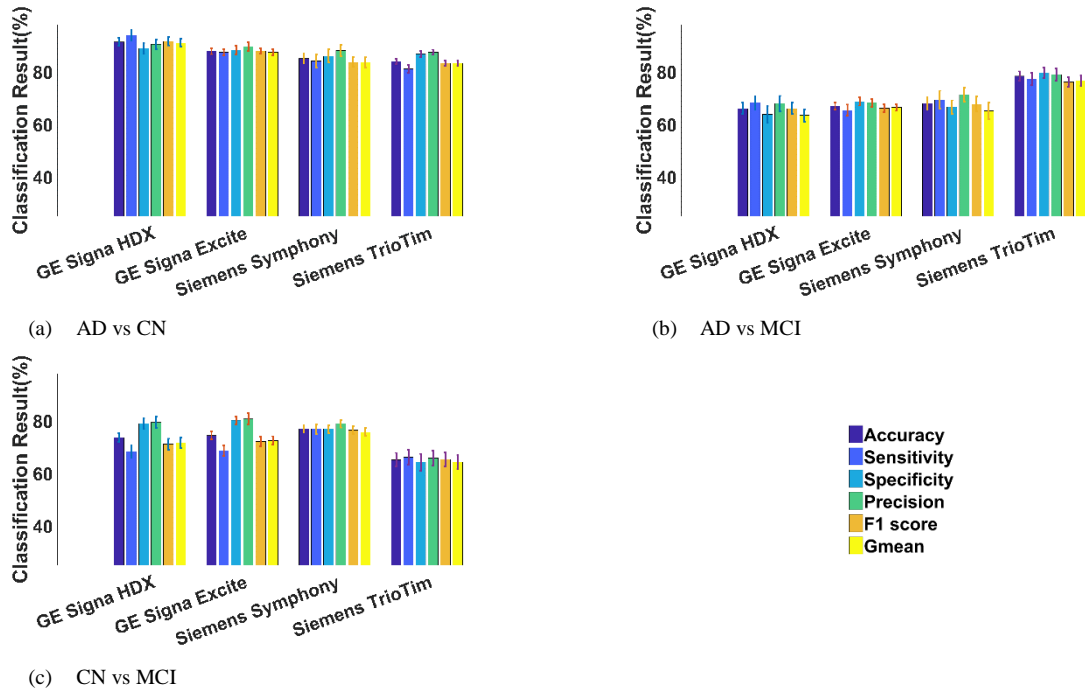


Figure 5-4: Performance result of MALF based features at granularity level 5 using ELM over four different protocols

When classifying CN vs MCI, GE scanners and Siemens Symphony performs similarly across all granularity levels having the highest accuracy about 74% at level 5 as shown in Figure 5-3(c). The accuracy increases from level 3 to level 5 considerably for Siemens scanners, while for GE scanners level 4 performs better as compared to levels 3 and 5. Across all scanners, the lower granularity level 3 performs worse. The performance results show that Siemens TrioTim performs poorly compared to the other three scanner models as shown in Figure 5-4(c).

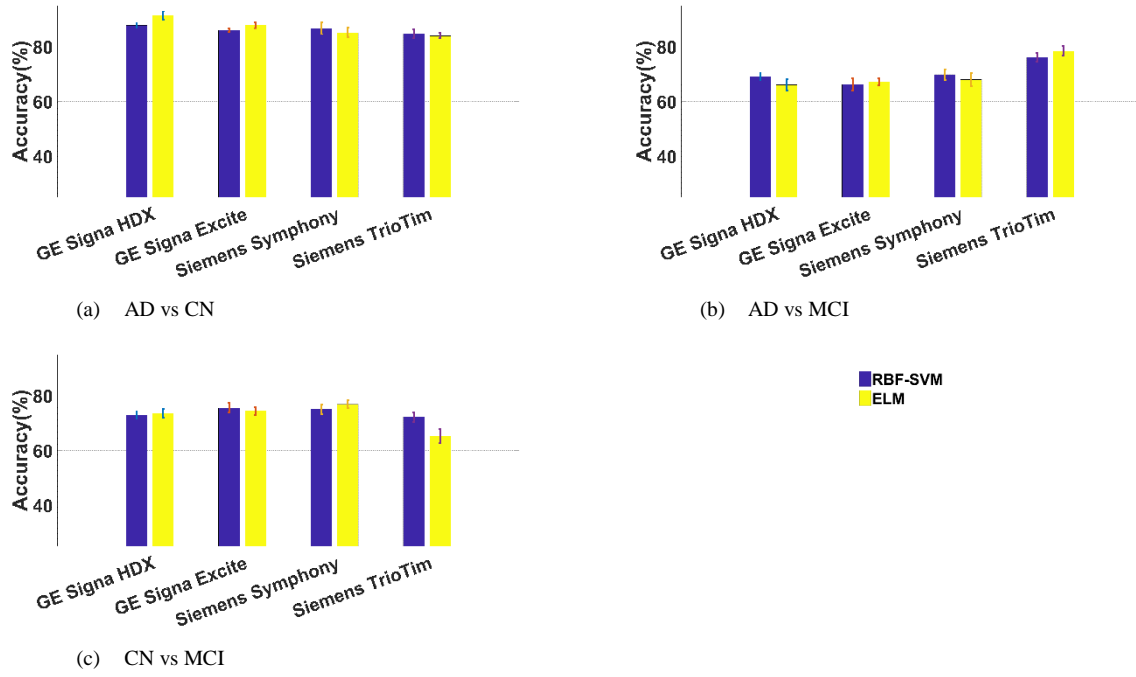


Figure 5-5: Comparison of ELM and SVM at granularity level 5 of MALF based method

We have analyzed three classification sets (AD vs CN, AD vs MCI, and CN vs MCI) across granularity levels and four scanner models using RBF-SVM and ELM. RBF SVM and ELM perform quite similarly to each other for all the scanners as depicted in Figure 5-5(a) through 5-5(c). For both the classifiers, the highest granularity level tends to provide better classification performance, and the lowest granularity level offers poor performance across all scanners. We suggest that this is because the highest granularity level provides information about the microstructure of the brain. Atrophy changes in the microstructures when the disease progresses may contribute to improving the training of the classifier. GE Signa HDX performs better for AD/CN classification, while GE Signa Excite and Siemens Symphony perform better for CN/MCI classification across both classifiers. Siemens TrioTim performs better for AD/MCI classification across both the classifiers, but not for AD/CN or CN/MCI classification despite having 3 T subject images as compared to the other scanners. This indicates that a scanner model having 3 T field strength image data does not necessarily ensure better classification performance. Classification performance also depends on the robustness of the algorithm for measuring brain structures across various scanner models with the same or different field strengths (1.5 T, and 3 T), and on the particular dataset.

### 5.3 Classification performance of FreeSurfer based methods

We have analyzed the classification performance using FreeSurfer defined subcortical volume-based features and the two classifiers RBF-SVM and ELM.

### 5.3.1 Performance using RBF SVM classifier

When classifying AD vs CN using RBF SVM, the highest accuracy has been achieved by the Siemens scanners. The accuracy of Symphony and TrioTim is 86.66%, and 86.16%, respectively. Siemens scanners perform better than GE scanners as depicted in Figure 5-6 (a).

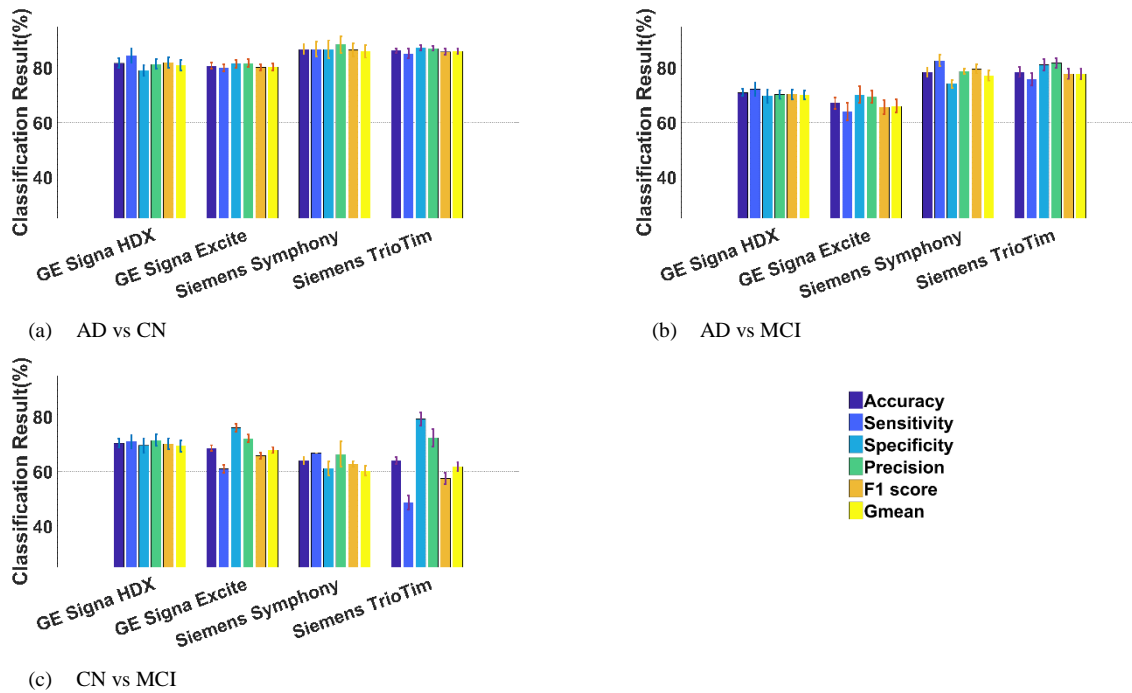


Figure 5-6: Performance result of FreeSurfer based method using RBF-SVM over four different protocols

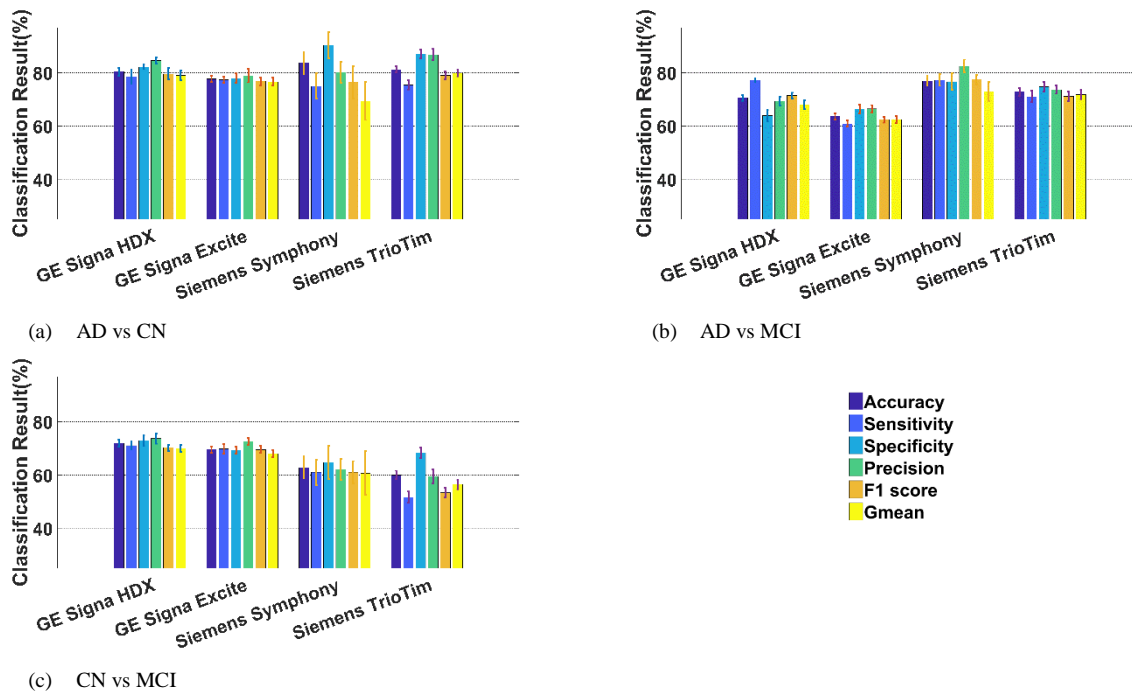


Figure 5-7: Performance result of FreeSurfer based method using ELM over four different protocols

For AD vs MCI classification, Siemens scanners account for the highest accuracy, about 78% across its two models. Figure 5.6(b) shows that Siemens scanners perform better than the GE scanners. For CN-vs MCI classification, GE Signa HDX provides the highest accuracy of 70.20%. Figure 5-6(c) shows that GE scanner performs well when compared with two models of Siemens.

### 5.3.2 Performance using ELM classifier

Figures 5-7(a) through 5-7(c) depict the performance of AD vs CN, AD vs MCI, and CN vs MCI using the ELM classifier. For the classification of AD/CN, Siemens Symphony achieved the highest accuracy with 83.66% which is considerably higher than the other three models of scanners. For AD vs MCI classification, Siemens Symphony again provided the highest accuracy with 76.95% as compare with the other three models. For both the classification methods, RBF-SVM and ELM, Siemens scanners exhibit better performance than GE scanners. However, considering CN vs MCI classification, GE Signa Excite performs better in comparison with other models. In this case, a GE scanner provides better classification performance as compared to Siemens scanners.

Siemens TrioTim performs better for AD/CN and AD/MCI classification across both the classifiers, but not for CN/MCI classification despite having 3 T subject images as compared to 1.5T for the other scanners. For CN/MCI classification, Signa HDX outperforms other scanner models across both the classifiers. This reinforces the results of the MALF based method which showed that a scanner model having 3 T subject image dataset does not necessarily ensure better classification performance even though 3T provides better topological localization of different brain structures with higher contrast as discussed in section 2.6. Classification performance also depends on the robustness of the algorithm for measuring brain structures across different scanner models, and on the dataset we use.

## 5.4 Ranking MALF and FreeSurfer based features

We have ranked the features by applying a feature ranking method which uses a *t*-test to assign scores to features for all the classification sets (AD vs CN, AD vs MCI, and CN vs MCI) across all scanners. The ranking of MALF based subcortical, ventricular structures is depicted in Figures 5-8 through 5-10, and FreeSurfer based subcortical segmented structures are shown in Figures 5-11 through 5-13.

Considering the ranking of MALF structures for AD/CN classification across all scanners, the most affected regions detected are basal forebrain, thalamus, SylvianFissureExt, and Lateral Ventricle as shown in Figures 5-8(a) through 5-8(d). The lateral ventricle ranks highest for GE Signa Excite but the other three scanners detect basal forebrain as the most affected regions. For AD/MCI classification, BasalForebrain is the highest ranked features across all the scanners as shown in

Figures 5-9(a) through 5-9(d). For CN/MCI classification, the highest-ranking feature for GE Signal HDX, GE Signa Excite, Siemens Symphony, Siemens TrioTim are Parietal\_L, SylvianFissureExt\_R, FrontSul\_R, and CentralSul\_L, respectively. The ranks are shown in Figures 5-10(a) through 5-10(d).

When using FreeSurfer, the features from subcortical regions such as hippocampus, amygdala, and Inferior Lateral Ventricles feature rank among the highest for all classification sets across all scanner models are shown in Figures 5-11 through 5-13. Hippocampus ranks highest in some scanners such as GE Signa HDX and GE Signa Excite. In contrast, amygdala ranks highest for Siemens TrioTim. This finding backs the claims of previous studies where some researchers declare that AD/MCI affects the hippocampus and amygdala [32]–[34]. For GE scanners, the highest ranked features are also hippocampus, lateral ventricle, and amygdala. But the Symphony model lists Cerebellum Cortex as the most affected regions while TrioTim shows the inferior lateral ventricles as the most affected region.

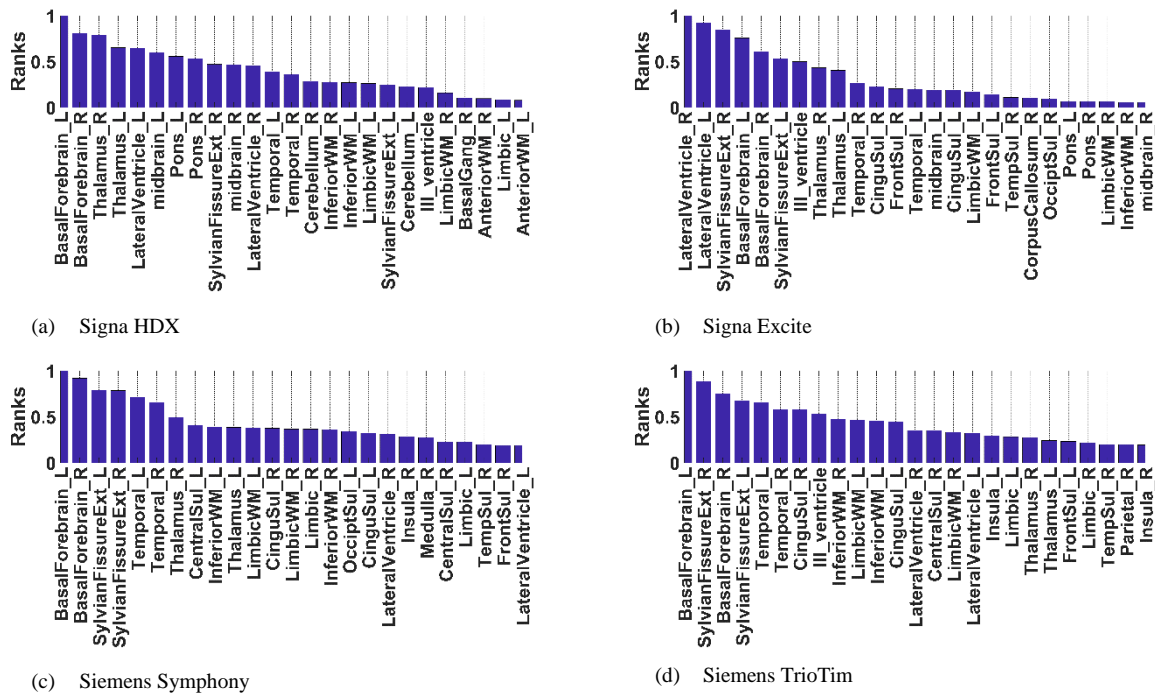
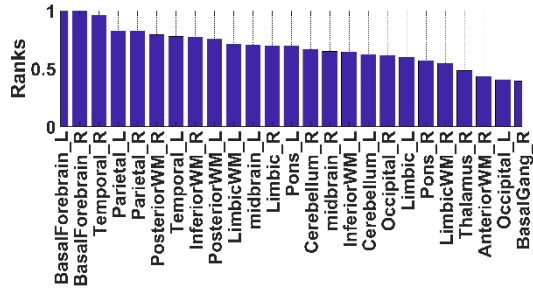
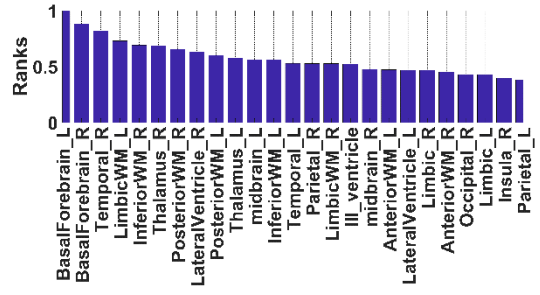


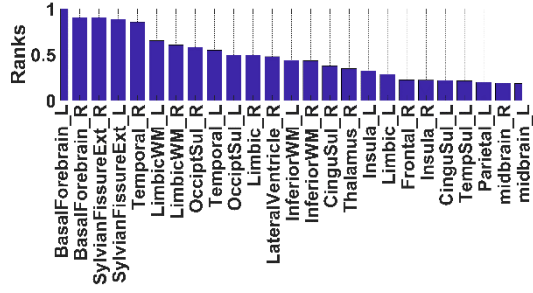
Figure 5-8: Ranking of MALF based features at granularity level 3 while classifying AD/CN



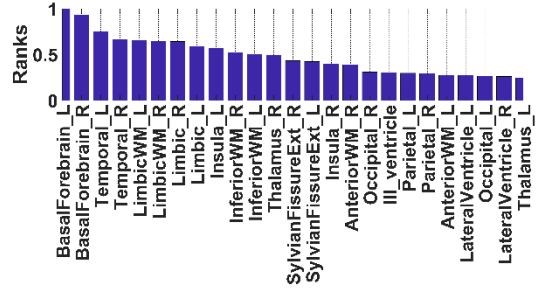
(e) Signa HDX



(f) Signa Excite

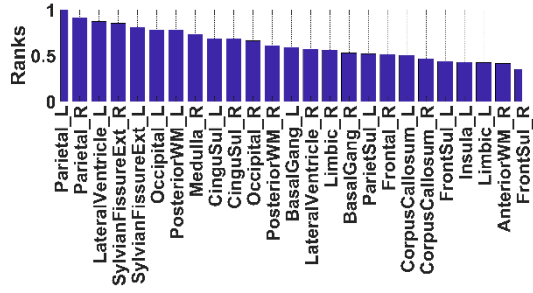


(g) Siemens Symphony

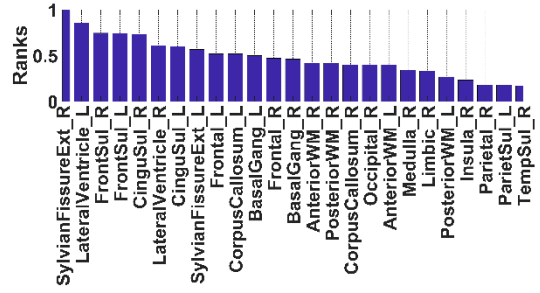


(h) Siemens TrioTim

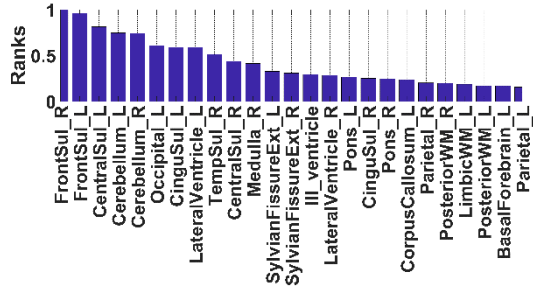
Figure 5-9: Ranking of MALF based features at granularity level 3 while classifying AD/MCI



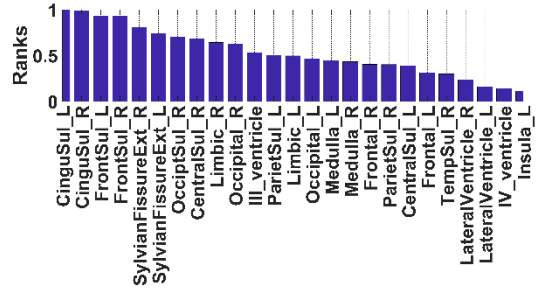
(a) Signa HDX



(b) Signa Excite

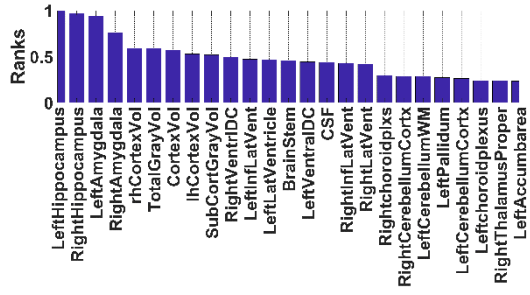


(c) Siemens Symphony

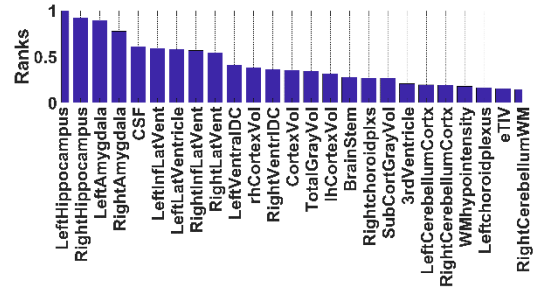


(d) Siemens TrioTim

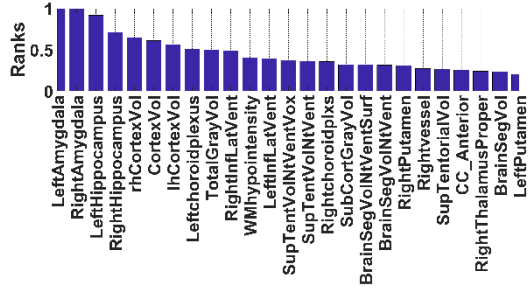
Figure 5-10: Ranking of MALF based features at granularity level 3 while classifying CN/MCI



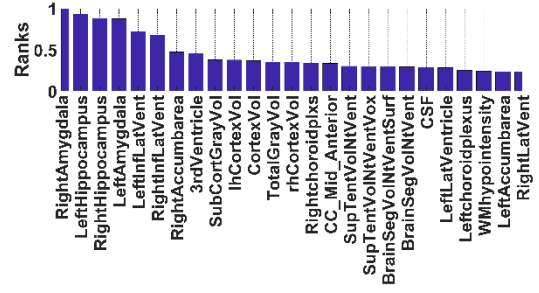
(a) Signa HDX



(b) Signa Excite

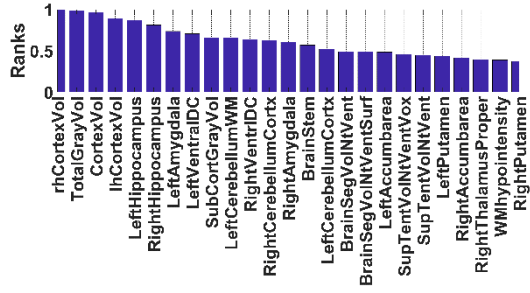


(c) Siemens Symphony

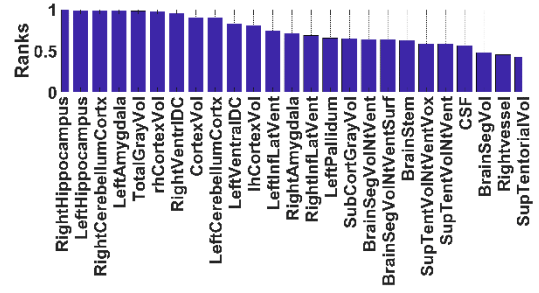


(d) Siemens TrioTim

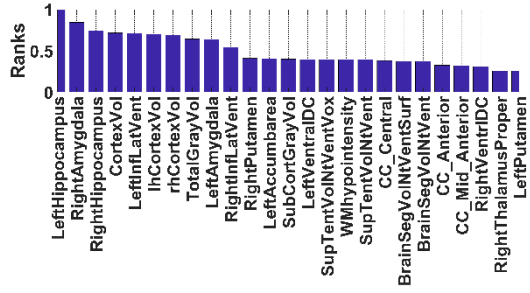
Figure 5-11: Ranking of FreeSurfer based features while classifying AD/CN



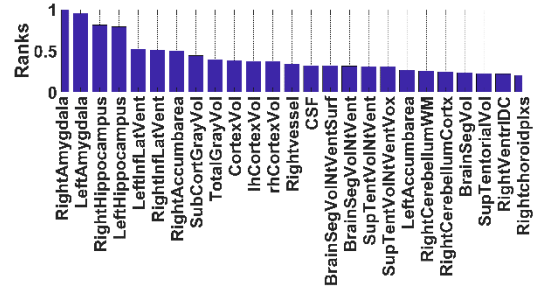
(a) Signa HDX



(b) Signa Excite



(c) Siemens Symphony



(d) Siemens TrioTim

Figure 5-12: Ranking of FreeSurfer based features while classifying AD/MCI



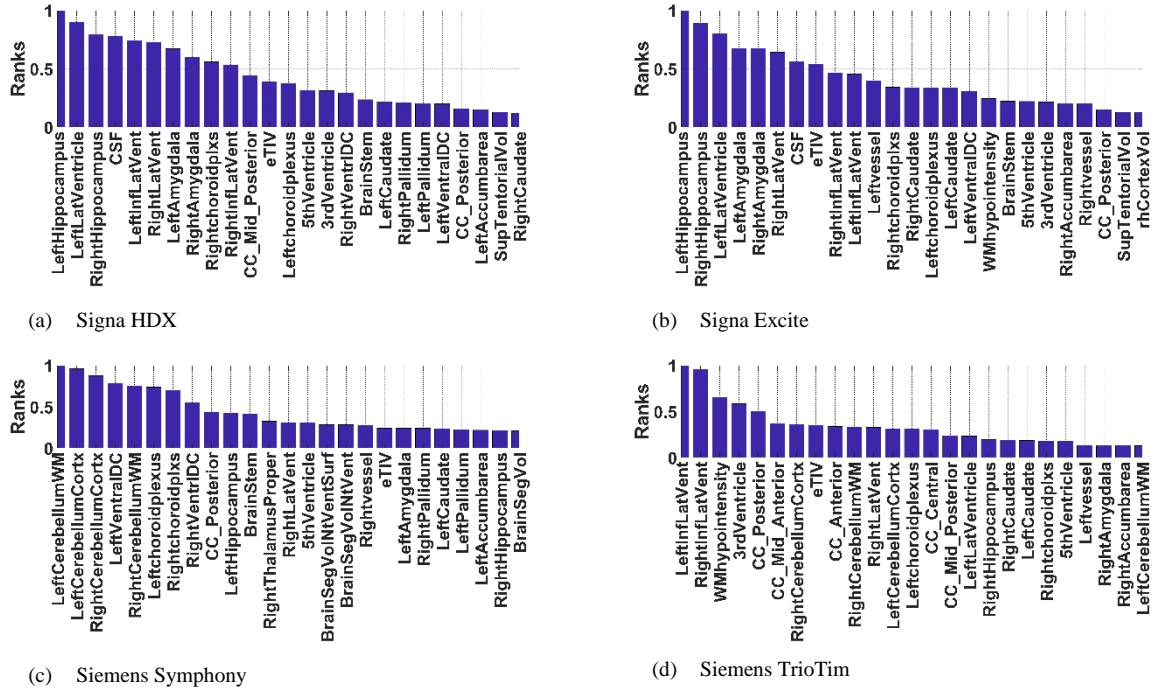


Figure 5-13: Ranking of FreeSurfer based features while classifying CN/MCI

Different scanners perform reasonably well for AD/MCI detection. We have observed minor difference over protocols regarding classification performance and ranking order for both FreeSurfer and MALF based regions. Although there is a minor difference, it does not mean that the data from the different scanners can be merged into a single large dataset because the classifier can adapt to each scanner when it is trained on data from only one scanner. Although the rank order marginally varies, this may not impact the classification performance because other similar significant ROIs can take the same role as the representative features. Also, our classification methods consider several ROIs when classifying one group from other. Despite having different ranking order, these features might still be available in the feature set. As long as these features remain in feature set, it does not impact the classification performance

## 5.5 Conclusion

We have explained the result we have obtained after using RBF-SVM and ELM over four different types of scanner data. We have observed marginal differences in performance result when comparing classification performance of the scanners across three classification tasks: AD vs CN, AD vs MCI, and CN vs MCI. For MALF based features, and for both classifiers, the best performance is achieved by different scanners for each of the three classification tasks: Signa HDX for AD vs CN, Siemens TrioTim for AD vs MCI and the other two scanners for CN vs MCI. For FreeSurfer based features,

Siemens TrioTim performs better for AD vs CN and AD vs MCI classifications across both classifiers, while Signa HDX performs better for CN vs MCI classification across both classifiers. We have also observed that the ranking orders of the most important features vary for all the three classification sets. However, it is difficult to make a claim that these differences occur due to protocol differences because other factors also influence these minor differences.

## Conclusion

### 6.1 Research Summary

We have described the full automated multi-atlas-based brain parcellation tool (MALF) which uses multiple atlases, and extracts subcortical, ventricular structural volume from a subject image at five granularity levels which ranges from 1 to 5. We have also briefly described an efficient automated tool, FreeSurfer, which segments brain into various subcortical regions, and measures volume from those regions. These two robust tools were used to extract features from MRI data collected using four different types of protocols which include scanners having the same or different field strengths. The robustness of these two types of feature extraction methods over protocol differences is discussed. The robustness has been discussed in terms of classification performance and ranking of the most effective ROIs. We have also described the data which are used to obtain the features along with several preprocessing methods and a feature selection approach, SVM Recursive Feature Elimination, where features are selected based its SVM weight vector. A method is explained to avoid double dipping so that features of a subject do not get involved in both training and testing phase during the classification. Supervised classifiers, SVM and ELM have been explained. These are used to disease stage of a subject (i.e., AD, CN, and MCI). We have recorded the classification performance using 5 fold cross-validation technique and running the program multiple times where different subset of data is randomly assigns to a fold at each run. We have also used a  $t$  test based

feature ranking method which ranks all the regions of the brain based on the statistical significance for separating two classes.

When analyzing multi-atlas likelihood fusion-based granularity analysis, it is evident that a higher granularity level provides better classification performance across the protocols. This is because the higher granularity level provides more detailed information of microstructures. That is used to train the supervised classifier. Siemens TrioTim performs better for AD/MCI classification, but not for AD/CN or CN/MCI classification set when using MALF based features and two classifiers, ELM and RBF-SVM. When using FreeSurfer based features, Siemens TrioTim performs better for AD/CN, and AD/MCI classification across both the classifiers, but not for CN/MCI classification. Both classifiers, SVM and ELM, provides similar performance. There is no evidence in our study for superiority of one classifier over the other. We can't claim the superiority of a classifier over another classifier in performance in our study/dataset as some other studies

Considering the ranking of affected ROIs, we have shown that the ranking order of different brain structures/regions varies across different protocols. In many cases, the difference in the rank ordering would not impact disease detection. However, we have overserved some cases where most important regions are significantly different for different protocols. For example, when ranking features for CN/MCI, the important MALF based regions are noticeably different across all protocols and FreeSurfer based regions are significantly different for all Siemens models These differences may impact the detection performance.

## 6.2 The limitations of our study

It is difficult to make strong claims regarding the reason for the differences in classification performance or ranking order of ROIs because these deferences may be affected by several factors. For example, the number of applied atlases may affect the performance result. If the atlas set introduces bias, this may change the extracted features and the classification performance The limitations of our study are explained below.

- **Limited data:** We need to investigate and generalize the effect of protocol differences using a large of number of subjects. A large of number of unique participants having different severity stages of dementia and different protocols is difficult to collect. The publicly available datasets are limited. ADNI data, which we have used, is the largest publicly available dataset for Alzheimer's disease.
- **Unpaired data:** It is ideal to use paired data. Data we have used in our study is not paired where MR images of a subject for all the protocols at a specific time point is not available.

- **Age gap:** The age differences are not considered. It is hard to get data of certain age range (i.e 5-10 years gap) over all the protocols. We have elderly data in broad range. During that time, a subject may undergo different kind of brain severity related disease, which may affect brain tissues, eventually affect the subcortical or ventricular structures of a brain which is crucial features of our study.
- **Scanning Time of a day:** The time of a day when subject images were scanned is not considered. The tissue density in a brain MRI may vary depending on the time of a day. For example Nakamura et al. [143] have shown that brain volume is higher in the morning.

### 6.3 Future Work

We will endeavour to work in large paired dataset to investigate the protocol difference and compare results with the present study. Then, we intend to develop some regression-based correction method to eliminate bias introduced by scanning time of a day, age, gender, and scanner model differences.

## List of symbols

|       |   |
|-------|---|
| AD    | Alzheimer Disease                           |
| ADNI  | Alzheimer's Disease Neuroimaging Initiative |
| MALF  | Multi-Atlas Likelihood Fusion               |
| CAD   | Computer Aided Diagnosis                    |
| CDR   | Clinical Dementia Rating                    |
| VBM   | Voxel Based Morphometry                     |
| HC    | Healthy Controls                            |
| MCI   | Mild Cognitive Impairment                   |
| MRI   | Magnetic Resonance Image                    |
| NC/CN | Normal Controls                             |
| SVM   | Support Vector Machine                      |
| ELM   | Extreme Learning Machine                    |
| RBF   | Radial Basis Function                       |
| ROI   | Region of Interest                          |

## References

- [1] A. Collie and P. Maruff, "The neuropsychology of preclinical Alzheimer's disease and mild cognitive impairment," *Neurosci Biobehav Rev*, vol. 24, no. 3, pp. 365–374, May 2000.
- [2] L. E. Hebert, P. A. Scherr, J. L. Bienias, D. A. Bennett, and D. A. Evans, "Alzheimer disease in the US population: prevalence estimates using the 2000 census," *Archives of neurology*, vol. 60, no. 8, pp. 1119–1122, 2003.
- [3] A. L. Sosa-Ortiz, I. Acosta-Castillo, and M. J. Prince, "Epidemiology of dementias and Alzheimer's disease," *Arch. Med. Res.*, vol. 43, no. 8, pp. 600–608, Nov. 2012.
- [4] M. Sjögren and C. Andersen, "Frontotemporal dementia--a brief review," *Mech. Ageing Dev.*, vol. 127, no. 2, pp. 180–187, Feb. 2006.
- [5] "Frontotemporal Disorders: Information for Patients, Families, and Caregivers | NINDS," 01-Feb-2017. [Online]. Available: /ninds/ninds/product/Frontotemporal-Disorders-Information-for-Patients-Families-and-Caregivers/17-AG-6361. [Accessed: 05-Jan-2018].
- [6] E. R. Danielsen and B. Ross, *Magnetic resonance spectroscopy diagnosis of neurological diseases*. CRC Press, 1999.
- [7] R. A. Sperling *et al.*, "Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimers & Dementia*, vol. 7, no. 3, pp. 280–292, May 2011.
- [8] M. S. Albert *et al.*, "The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimers & Dementia*, vol. 7, no. 3, pp. 270–279, May 2011.
- [9] G. M. McKhann *et al.*, "The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimers & Dementia*, vol. 7, no. 3, pp. 263–269, May 2011.
- [10] P. Celsis, "Age-related cognitive decline, mild cognitive impairment or preclinical Alzheimer's disease?," *Ann. Med.*, vol. 32, no. 1, pp. 6–14, Feb. 2000.
- [11] R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. G. Tangalos, and E. Kokmen, "Mild cognitive impairment - Clinical characterization and outcome," *Archives of Neurology*, vol. 56, no. 3, pp. 303–308, Mar. 1999.
- [12] F. T. Hane, M. Robinson, B. Y. Lee, O. Bai, Z. Leonenko, and M. S. Albert, "Recent Progress in Alzheimer's Disease Research, Part 3: Diagnosis and Treatment," *Journal of Alzheimer's Disease*, vol. 57, no. 3, pp. 645–665, 2017.
- [13] J. Voges *et al.*, "Deep-brain stimulation: long-term analysis of complications caused by hardware and surgery—experiences from a single centre," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 77, no. 7, pp. 868–872, 2006.
- [14] E. Dayan, N. Censor, E. R. Buch, M. Sandrini, and L. G. Cohen, "Noninvasive brain stimulation: from physiology to network dynamics and back," *Nature neuroscience*, vol. 16, no. 7, p. 838, 2013.
- [15] M. D. Fox, R. L. Buckner, H. Liu, M. M. Chakravarty, A. M. Lozano, and A. Pascual-Leone, "Resting-state networks link invasive and noninvasive brain stimulation across diverse psychiatric and neurological diseases," *Proceedings of the National Academy of Sciences*, vol. 111, no. 41, pp. E4367–E4375, 2014.
- [16] I. Arevalo-Rodriguez *et al.*, "Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI)," in *Cochrane Database of Systematic Reviews*, John Wiley & Sons, Ltd, 2015.
- [17] H. M. Hodkinson, "Evaluation of a mental test score for assessment of mental impairment in the elderly," *Age and ageing*, vol. 1, no. 4, pp. 233–238, 1972.
- [18] E. L. Teng and H. C. Chui, "The Modified Mini-Mental State (3MS) examination," *J Clin Psychiatry*, vol. 48, no. 8, pp. 314–318, Aug. 1987.

- [19] E. L. Teng *et al.*, "The Cognitive Abilities Screening Instrument (CASI): a practical test for cross-cultural epidemiological studies of dementia," *Int Psychogeriatr*, vol. 6, no. 1, pp. 45–58; discussion 62, 1994.
- [20] D. Royall, J. Cordes, and M. Polk, "CLOX: an executive clock drawing task," *J Neurol Neurosurg Psychiatry*, vol. 64, no. 5, pp. 588–594, May 1998.
- [21] M.-F. Shadlen, D. Siscovick, A. L. Fitzpatrick, C. Dulberg, L. H. Kuller, and S. Jackson, "Education, Cognitive Test Scores, and Black-White Differences in Dementia Risk," *Journal of the American Geriatrics Society*, vol. 54, no. 6, pp. 898–905, 2006.
- [22] T. Gómez-Isla, J. L. Price, D. W. McKeel Jr, J. C. Morris, J. H. Growdon, and B. T. Hyman, "Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease," *Journal of Neuroscience*, vol. 16, no. 14, pp. 4491–4500, 1996.
- [23] M. Likeman *et al.*, "Visual assessment of atrophy on magnetic resonance imaging in the diagnosis of pathologically confirmed young-onset dementias," *Archives of neurology*, vol. 62, no. 9, pp. 1410–1415, 2005.
- [24] 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*, vol. 32, no. 12, p. 2322.e19, Dec. 2011.
- [25] M. Park and W.-J. Moon, "Structural MR imaging in the diagnosis of Alzheimer's disease and other neurodegenerative dementia: current imaging approach and future perspectives," *Korean journal of radiology*, vol. 17, no. 6, pp. 827–845, 2016.
- [26] H. Park, J. Yang, J. Seo, and J. Lee, "Dimensionality reduced cortical features and their use in predicting longitudinal changes in Alzheimer's disease," *Neuroscience Letters*, vol. 550, pp. 17–22, Aug. 2013.
- [27] M. R. Raman *et al.*, "An MRI-Based Atlas for Correlation of Imaging and Pathologic Findings in Alzheimer's Disease," *Journal of Neuroimaging*, vol. 26, no. 3, pp. 264–268, 2016.
- [28] X. Long, L. Chen, C. Jiang, L. Zhang, and A. D. N. Initiative, "Prediction and classification of Alzheimer disease based on quantification of MRI deformation," *PLOS ONE*, vol. 12, no. 3, p. e0173372, Mar. 2017.
- [29] S. L. Risacher, A. J. Saykin, J. D. West, L. Shen, H. A. Firpi, and B. C. McDonald, "Baseline MRI Predictors of Conversion from MCI to Probable AD in the ADNI Cohort," *Curr Alzheimer Res*, vol. 6, no. 4, pp. 347–361, Aug. 2009.
- [30] Y. Guo *et al.*, "Grey-matter volume as a potential feature for the classification of Alzheimer's disease and mild cognitive impairment: an exploratory study," *Neurosci Bull*, vol. 30, no. 3, pp. 477–489, Jun. 2014.
- [31] E. Busovaca *et al.*, "Is the Alzheimer's disease cortical thickness signature a biological marker for memory?," *Brain Imaging Behav*, vol. 10, no. 2, pp. 517–523, Jun. 2016.
- [32] T. Tapiola *et al.*, "MRI of hippocampus and entorhinal cortex in mild cognitive impairment: a follow-up study," *Neurobiol. Aging*, vol. 29, no. 1, pp. 31–38, Jan. 2008.
- [33] Y. Xu *et al.*, "Usefulness of MRI measures of entorhinal cortex versus hippocampus in AD," *Neurology*, vol. 54, no. 9, pp. 1760–1767, May 2000.
- [34] J. E. Scorzin *et al.*, "Volume determination of amygdala and hippocampus at 1.5 and 3.0 T MRI in temporal lobe epilepsy," *Epilepsy research*, vol. 82, no. 1, pp. 29–37, 2008.
- [35] K. Ota, N. Oishi, K. Ito, and H. Fukuyama, "Effects of imaging modalities, brain atlases and feature selection on prediction of Alzheimer's disease," *Journal of Neuroscience Methods*, vol. 256, pp. 168–183, Dec. 2015.
- [36] R. S. Desikan *et al.*, "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *Neuroimage*, vol. 31, no. 3, pp. 968–980, Jul. 2006.
- [37] R. Min, G. Wu, J. Cheng, Q. Wang, and D. Shen, "Multi-Atlas Based Representations for Alzheimer's Disease Diagnosis," *Human Brain Mapping*, vol. 35, no. 10, pp. 5052–5070, Oct. 2014.
- [38] J. Ramírez *et al.*, "Computer-aided diagnosis of Alzheimer's type dementia combining support vector machines and discriminant set of features," *Information Sciences*, vol. 237, pp. 59–72, 2013.



- [39] C.-Y. Wee, P.-T. Yap, and D. Shen, "Prediction of Alzheimer's Disease and Mild Cognitive Impairment Using Cortical Morphological Patterns," *Human Brain Mapping*, vol. 34, no. 12, pp. 3411–3425, Dec. 2013.
- [40] N. Kriegeskorte, W. K. Simmons, P. S. F. Bellgowan, and C. I. Baker, "Circular analysis in systems neuroscience: the dangers of double dipping," *Nature Neuroscience*, vol. 12, no. 5, pp. 535–540, May 2009.
- [41] C. R. Jack *et al.*, "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods," *Journal of Magnetic Resonance Imaging*, vol. 27, no. 4, pp. 685–691, Apr. 2008.
- [42] J. West, I. Blystad, M. Engström, J. B. Warntjes, and P. Lundberg, "Application of quantitative MRI for brain tissue segmentation at 1.5 T and 3.0 T field strengths," *PloS one*, vol. 8, no. 9, p. e74795, 2013.
- [43] M. M. Adankon and M. Cheriet, "Support vector machine," in *Encyclopedia of biometrics*, Springer, 2009, pp. 1303–1308.
- [44] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.
- [45] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, Dec. 2006.
- [46] Y. Stern, "Cognitive reserve and Alzheimer disease," *Alzheimer Dis Assoc Disord*, vol. 20, no. 2, pp. 112–117, Jun. 2006.
- [47] C. M. Roe, C. Xiong, J. P. Miller, and J. C. Morris, "Education and Alzheimer disease without dementia: support for the cognitive reserve hypothesis," *Neurology*, vol. 68, no. 3, pp. 223–228, Jan. 2007.
- [48] C. Solé-Padullés *et al.*, "Brain structure and function related to cognitive reserve variables in normal aging, mild cognitive impairment and Alzheimer's disease," *Neurobiol. Aging*, vol. 30, no. 7, pp. 1114–1124, Jul. 2009.
- [49] L. G. Apostolova *et al.*, "3D comparison of low, intermediate, and advanced hippocampal atrophy in MCI," *Hum Brain Mapp*, vol. 31, no. 5, pp. 786–797, May 2010.
- [50] B. C. Dickerson *et al.*, "MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease," *Neurobiology of Aging*, vol. 22, no. 5, pp. 747–754, Oct. 2001.
- [51] J. Hardy and D. J. Selkoe, "The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics," *Science*, vol. 297, no. 5580, pp. 353–356, Jul. 2002.
- [52] J. Ashburner and K. J. Friston, "Voxel-based morphometry - The methods," *Neuroimage*, vol. 11, no. 6, pp. 805–821, Jun. 2000.
- [53] M. Bozzali *et al.*, "The contribution of voxel-based morphometry in staging patients with mild cognitive impairment," *Neurology*, vol. 67, no. 3, pp. 453–460, Aug. 2006.
- [54] N. G. Campeau, R. C. Petersen, J. P. Felmlee, P. C. O'Brien, and C. R. Jack Jr, "Hippocampal transverse relaxation times in patients with Alzheimer disease.," *Radiology*, vol. 205, no. 1, pp. 197–201, 1997.
- [55] J. C. Baron *et al.*, "In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease," *Neuroimage*, vol. 14, no. 2, pp. 298–309, Aug. 2001.
- [56] I. C. Wright *et al.*, "A voxel-based method for the statistical analysis of gray and white matter density applied to schizophrenia," *Neuroimage*, vol. 2, no. 4, pp. 244–252, 1995.
- [57] S. Klöppel *et al.*, "Automatic classification of MR scans in Alzheimer's disease," *Brain*, vol. 131, no. 3, pp. 681–689, Mar. 2008.
- [58] J. G. Csernansky *et al.*, "Preclinical detection of Alzheimer's disease: hippocampal shape and volume predict dementia onset in the elderly," *Neuroimage*, vol. 25, no. 3, pp. 783–792, Apr. 2005.
- [59] G. B. Frisoni *et al.*, "Detection of grey matter loss in mild Alzheimer's disease with voxel based morphometry," *Journal of Neurology Neurosurgery and Psychiatry*, vol. 73, no. 6, pp. 657–664, Dec. 2002.
- [60] S. Adaszewski, J. Dukart, F. Kherif, R. Frackowiak, B. Draganski, and A. D. N. Initiative, "How early can we predict Alzheimer's disease using computational anatomy?," *Neurobiology of aging*, vol. 34, no. 12, pp. 2815–2826, 2013.
- [61] M. Termenon and M. Grana, "A Two Stage Sequential Ensemble Applied to the Classification of Alzheimer's Disease Based on MRI Features," *Neural Processing Letters*, vol. 35, no. 1, pp. 1–12, Feb. 2012.

- [62] C. Moller *et al.*, "Alzheimer Disease and Behavioral Variant Frontotemporal Dementia: Automatic Classification Based on Cortical Atrophy for Single-Subject Diagnosis," *Radiology*, vol. 279, no. 3, pp. 838–848, Jun. 2016.
- [63] 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*, vol. 18, no. 3, pp. 984–990, May 2014.
- [64] R. Cuingnet *et al.*, "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database," *NeuroImage*, vol. 56, no. 2, pp. 766–781, May 2011.
- [65] P. M. Thompson *et al.*, "Cortical change in Alzheimer's disease detected with a disease-specific population-based brain atlas," *Cerebral Cortex*, vol. 11, no. 1, pp. 1–16, Jan. 2001.
- [66] M. K. Chung *et al.*, "A unified statistical approach to deformation-based morphometry," *Neuroimage*, vol. 14, no. 3, pp. 595–606, Sep. 2001.
- [67] W. R. Riddle, R. Li, J. M. Fitzpatrick, S. C. DonLevy, B. M. Dawant, and R. R. Price, "Characterizing changes in MR images with color-coded Jacobians," *Magnetic Resonance Imaging*, vol. 22, no. 6, pp. 769–777, Jul. 2004.
- [68] X. Hua *et al.*, "3D characterization of brain atrophy in Alzheimer's disease and mild cognitive impairment using tensor-based morphometry," *Neuroimage*, vol. 41, no. 1, pp. 19–34, May 2008.
- [69] J. C. Lau, J. P. Lerch, J. G. Sled, R. M. Henkelman, A. C. Evans, and B. J. Bedell, "Longitudinal neuroanatomical changes determined by deformation-based morphometry in a mouse model of Alzheimer's disease," *Neuroimage*, vol. 42, no. 1, pp. 19–27, 2008.
- [70] 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*, vol. 29, no. 4, pp. 514–523, Apr. 2008.
- [71] Y. Fan, D. Shen, R. C. Gur, R. E. Gur, and C. Davatzikos, "COMPARE: Classification of morphological patterns using adaptive regional elements," *Ieee Transactions on Medical Imaging*, vol. 26, no. 1, pp. 93–105, Jan. 2007.
- [72] J. Koikkalainen, J. Lotjonen, L. Thurfjell, D. Rueckert, G. Waldemar, and H. Soininen, "Multi-template tensor-based morphometry: Application to analysis of Alzheimer's disease," *Neuroimage*, vol. 56, no. 3, pp. 1134–1144, Jun. 2011.
- [73] M. Liu, D. Zhang, and D. Shen, "View-Centralized Multi-Atlas Classification for Alzheimer's Disease Diagnosis," *Human Brain Mapping*, vol. 36, no. 5, pp. 1847–1865, May 2015.
- [74] M. Liu, D. Zhang, and D. Shen, "Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment," *IEEE transactions on medical imaging*, vol. 35, no. 6, pp. 1463–1474, 2016.
- [75] J. P. Lerch and A. C. Evans, "Cortical thickness analysis examined through power analysis and a population simulation," *Neuroimage*, vol. 24, no. 1, pp. 163–173, Jan. 2005.
- [76] M. S. Panizzon *et al.*, "Distinct genetic influences on cortical surface area and cortical thickness," *Cereb. Cortex*, vol. 19, no. 11, pp. 2728–2735, Nov. 2009.
- [77] S. Li *et al.*, "Abnormal Changes of Multidimensional Surface Features Using Multivariate Pattern Classification in Amnesic Mild Cognitive Impairment Patients," *Journal of Neuroscience*, vol. 34, no. 32, pp. 10541–10553, Aug. 2014.
- [78] H. Park, J. Yang, J. Seo, and J. Lee, "Dimensionality reduced cortical features and their use in the classification of Alzheimer's disease and mild cognitive impairment," *Neuroscience Letters*, vol. 529, no. 2, pp. 123–127, Nov. 2012.
- [79] M. S. Mega *et al.*, "Automated brain tissue assessment in the elderly and demented population: Construction and validation of a sub-volume probabilistic brain atlas," *Neuroimage*, vol. 26, no. 4, pp. 1009–1018, Jul. 2005.
- [80] M. E. Rettmann, X. Han, C. Y. Xu, and J. L. Prince, "Automated sulcal segmentation using watersheds on the cortical surface," *Neuroimage*, vol. 15, no. 2, pp. 329–344, Feb. 2002.
- [81] G. Le Goualher, E. Procyk, D. L. Collins, R. Venugopal, C. Barillot, and A. C. Evans, "Automated extraction and variability analysis of sulcal neuroanatomy," *Ieee Transactions on Medical Imaging*, vol. 18, no. 3, pp. 206–217, Mar. 1999.

- [82] B. Fischl *et al.*, "Automatically parcellating the human cerebral cortex," *Cerebral Cortex*, vol. 14, no. 1, pp. 11–22, Jan. 2004.
- [83] R. S. Desikan *et al.*, "Automated MRI measures identify individuals with mild cognitive impairment and Alzheimers disease," *Brain*, vol. 132, pp. 2048–2057, Aug. 2009.
- [84] P. P. de Magalhaes Oliveira, R. Nitrini, G. Busatto, C. Buchpiguel, J. R. Sato, and E. Amaro, "Use of SVM Methods with Surface-Based Cortical and Volumetric Subcortical Measurements to Detect Alzheimer's Disease," *Journal of Alzheimers Disease*, vol. 19, no. 4, pp. 1263–1272, 2010.
- [85] H. Braak and E. Braak, "Staging of Alzheimer's disease-related neurofibrillary changes," *Neurobiol. Aging*, vol. 16, no. 3, pp. 271–278; discussion 278–284, Jun. 1995.
- [86] C. Fischbach-Boulanger *et al.*, "T1-or T2-weighted magnetic resonance imaging: what is the best choice to evaluate atrophy of the hippocampus?," *European journal of neurology*, 2018.
- [87] C. R. Jack *et al.*, "Steps to standardization and validation of hippocampal volumetry as a biomarker in clinical trials and diagnostic criterion for Alzheimer's disease," *Alzheimers Dement*, vol. 7, no. 4, pp. 474–485.e4, Jul. 2011.
- [88] M. Chupin *et al.*, "Fully Automatic Hippocampus Segmentation and Classification in Alzheimer's Disease and Mild Cognitive Impairment Applied on Data From ADNI," *Hippocampus*, vol. 19, no. 6, pp. 579–587, 2009.
- [89] S. G. Costafreda *et al.*, "Automated hippocampal shape analysis predicts the onset of dementia in mild cognitive impairment," *Neuroimage*, vol. 56, no. 1, pp. 212–219, May 2011.
- [90] J. Ashburner and K. J. Friston, "Unified segmentation," *Neuroimage*, vol. 26, no. 3, pp. 839–851, Jul. 2005.
- [91] S. M. Smith *et al.*, "Advances in functional and structural MR image analysis and implementation as FSL," *Neuroimage*, vol. 23, pp. S208–S219, 2004.
- [92] S. M. Smith, "Fast robust automated brain extraction," *Human Brain Mapping*, vol. 17, no. 3, pp. 143–155, Nov. 2002.
- [93] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, "Improved optimization for the robust and accurate linear registration and motion correction of brain images," *Neuroimage*, vol. 17, no. 2, pp. 825–841, Oct. 2002.
- [94] Y. Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *Ieee Transactions on Medical Imaging*, vol. 20, no. 1, pp. 45–57, Jan. 2001.
- [95] 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*, vol. 44, no. 4, pp. 1415–1422, Feb. 2009.
- [96] C. Salvatore, A. Cerasa, P. Battista, M. C. Gilardi, A. Quattrone, and I. Castiglioni, "Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach," *Front Neurosci*, vol. 9, Sep. 2015.
- [97] R. K. Lama, J. Gwak, J.-S. Park, and S.-W. Lee, "Diagnosis of Alzheimer's Disease Based on Structural MRI Images Using a Regularized Extreme Learning Machine and PCA Features," *Journal of Healthcare Engineering*, 2017. [Online]. Available: <https://www.hindawi.com/journals/jhe/2017/5485080/>. [Accessed: 02-Dec-2017].
- [98] L. Sorensen *et al.*, "Early detection of Alzheimer's disease using MRI hippocampal texture," *Human Brain Mapping*, vol. 37, no. 3, pp. 1148–1161, Mar. 2016.
- [99] A. Chincarini *et al.*, "Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease," *Neuroimage*, vol. 58, no. 2, pp. 469–480, Sep. 2011.
- [100] C. Zu, B. Jie, M. Liu, S. Chen, D. Shen, and D. Zhang, "Label-aligned multi-task feature learning for multimodal classification of Alzheimer's disease and mild cognitive impairment," *Brain Imaging and Behavior*, vol. 10, no. 4, pp. 1148–1159, Dec. 2016.
- [101] S. Alam, G.-R. Kwon, J.-I. Kim, and C.-S. Park, "Twin SVM-Based Classification of Alzheimer's Disease Using Complex Dual-Tree Wavelet Principal Coefficients and LDA," *J Healthc Eng*, vol. 2017, p. 8750506, 2017.
- [102] K. Krupa and M. Bekiesińska-Figatowska, "Artifacts in magnetic resonance imaging," *Polish journal of radiology*, vol. 80, p. 93, 2015.

- [103] K. Farahani, U. Sinha, S. Sinha, L. C. Chiu, and R. B. Lufkin, "Effect of field strength on susceptibility artifacts in magnetic resonance imaging," *Computerized Medical Imaging and Graphics*, vol. 14, no. 6, pp. 409–413, 1990.
- [104] N. Chow *et al.*, "Comparing 3T and 1.5 T MRI for mapping hippocampal atrophy in the Alzheimer's Disease Neuroimaging Initiative," *American Journal of Neuroradiology*, vol. 36, no. 4, pp. 653–660, 2015.
- [105] S. Keihaninejad, R. A. Heckemann, G. Fagiolo, M. R. Symms, J. V. Hajnal, and A. Hammers, "A robust method to estimate the intracranial volume across MRI field strengths (1.5T and 3T)," *Neuroimage*, vol. 50, no. 4, pp. 1427–1437, May 2010.
- [106] K. E. Macdonald *et al.*, "Automated template-based hippocampal segmentations from MRI: the effects of 1.5 T or 3T field strength on accuracy," *Neuroinformatics*, vol. 12, no. 3, pp. 405–412, 2014.
- [107] M. Goodro, M. Sameti, B. Patenaude, and G. Fein, "Age effect on subcortical structures in healthy adults," *Psychiatry Research: Neuroimaging*, vol. 203, no. 1, pp. 38–45, 2012.
- [108] M. Zijlmans *et al.*, "3T versus 1.5 T phased-array MRI in the presurgical work-up of patients with partial epilepsy of uncertain focus," *Journal of Magnetic Resonance Imaging*, vol. 30, no. 2, pp. 256–262, 2009.
- [109] A. Pfefferbaum, T. Rohlfing, M. J. Rosenbloom, and E. V. Sullivan, "Combining atlas-based parcellation of regional brain data acquired across scanners at 1.5 T and 3.0 T field strengths," *Neuroimage*, vol. 60, no. 2, pp. 940–951, 2012.
- [110] C. Di Perri *et al.*, "Signal abnormalities on 1.5 and 3 Tesla brain MRI in multiple sclerosis patients and healthy controls. A morphological and spatial quantitative comparison study," *Neuroimage*, vol. 47, no. 4, pp. 1352–1362, 2009.
- [111] F. Kruggel, J. Turner, L. T. Muftuler, and A. D. N. Initiative, "Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort," *Neuroimage*, vol. 49, no. 3, pp. 2123–2133, 2010.
- [112] Z. Liang *et al.*, "Evaluation of cross-protocol stability of a fully automated brain multi-atlas parcellation tool," *PloS one*, vol. 10, no. 7, p. e0133533, 2015.
- [113] Y.-C. Zhu, C. Dufouil, C. Tzourio, and H. Chabriat, "Silent brain infarcts: a review of MRI diagnostic criteria," *Stroke*, vol. 42, no. 4, pp. 1140–1145, 2011.
- [114] A. J. Ho *et al.*, "Comparing 3 T and 1.5 T MRI for tracking Alzheimer's disease progression with tensor-based morphometry," *Human brain mapping*, vol. 31, no. 4, pp. 499–514, 2010.
- [115] S. J. Teipel *et al.*, "Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment," *Neuroimage*, vol. 38, no. 1, pp. 13–24, Oct. 2007.
- [116] X. Tang *et al.*, "Segmentation of brain magnetic resonance images based on multi-atlas likelihood fusion: testing using data with a broad range of anatomical and photometric profiles," *Frontiers in neuroscience*, vol. 9, p. 61, 2015.
- [117] J. G. Park and C. Lee, "Skull stripping based on region growing for magnetic resonance brain images," *NeuroImage*, vol. 47, no. 4, pp. 1394–1407, 2009.
- [118] A. H. Zhuang, D. J. Valentino, and A. W. Toga, "Skull-stripping magnetic resonance brain images using a model-based level set," *NeuroImage*, vol. 32, no. 1, pp. 79–92, 2006.
- [119] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," *IEEE transactions on medical imaging*, vol. 23, no. 7, pp. 903–921, 2004.
- [120] A. J. Asman and B. A. Landman, "Non-local statistical label fusion for multi-atlas segmentation," *Medical image analysis*, vol. 17, no. 2, pp. 194–208, 2013.
- [121] H. Wang and P. Yushkevich, "Multi-atlas segmentation with joint label fusion and corrective learning—an open source implementation," *Frontiers in neuroinformatics*, vol. 7, p. 27, 2013.
- [122] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *Ieee Transactions on Medical Imaging*, vol. 17, no. 1, pp. 87–97, Feb. 1998.
- [123] F. Segonne *et al.*, "A hybrid approach to the skull stripping problem in MRI," *Neuroimage*, vol. 22, no. 3, pp. 1060–1075, Jul. 2004.

- [124] B. Fischl *et al.*, "Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain," *Neuron*, vol. 33, no. 3, pp. 341–355, Jan. 2002.
- [125] A. M. Dale, B. Fischl, and M. I. Sereno, "Cortical surface-based analysis - I. Segmentation and surface reconstruction," *Neuroimage*, vol. 9, no. 2, pp. 179–194, Feb. 1999.
- [126] J. Shawe-Taylor and S. Sun, "A review of optimization methodologies in support vector machines," *Neurocomputing*, vol. 74, no. 17, pp. 3609–3618, 2011.
- [127] M. López *et al.*, "Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer's disease," *Neurocomputing*, vol. 74, no. 8, pp. 1260–1271, 2011.
- [128] M. M. López *et al.*, "SVM-based CAD system for early detection of the Alzheimer's disease using kernel PCA and LDA," *Neuroscience Letters*, vol. 464, no. 3, pp. 233–238, 2009.
- [129] B. Mirza, Z. Lin, and K.-A. Toh, "Weighted online sequential extreme learning machine for class imbalance learning," *Neural processing letters*, vol. 38, no. 3, pp. 465–486, 2013.
- [130] A. Iosifidis, A. Tefas, and I. Pitas, "Minimum class variance extreme learning machine for human action recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 11, pp. 1968–1979, 2013.
- [131] X. Peng, P. Lin, T. Zhang, and J. Wang, "Extreme learning machine-based classification of ADHD using brain structural MRI data," *PLoS one*, vol. 8, no. 11, p. e79476, 2013.
- [132] J. Tang, C. Deng, G.-B. Huang, and B. Zhao, "Compressed-domain ship detection on spaceborne optical image using deep neural network and extreme learning machine," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 3, pp. 1174–1185, 2015.
- [133] L. Zhang, D. Zhang, and F. Tian, "SVM and ELM: Who Wins? Object recognition with deep convolutional features from ImageNet," in *Proceedings of ELM-2015 Volume 1*, Springer, 2016, pp. 249–263.
- [134] Y. Wang, F. Cao, and Y. Yuan, "A study on effectiveness of extreme learning machine," *Neurocomputing*, vol. 74, no. 16, pp. 2483–2490, 2011.
- [135] G.-B. Huang and L. Chen, "Convex incremental extreme learning machine," *Neurocomputing*, vol. 70, no. 16–18, pp. 3056–3062, 2007.
- [136] R. Fletcher, "Practical Methods of Optimization: Vol. 2: Constrained Optimization.," *JOHN WILEY & SONS, INC., ONE WILEY DR., SOMERSET, N. J. 08873, 1981, 224*, 1981.
- [137] M. Brant-Zawadzki, G. D. Gillan, and W. R. Nitz, "MP RAGE: a three-dimensional, T1-weighted, gradient-echo sequence--initial experience in the brain," *Radiology*, vol. 182, no. 3, pp. 769–775, Mar. 1992.
- [138] S. Mori *et al.*, "MRICloud: delivering high-throughput MRI neuroinformatics as cloud-based software as a service," *Computing in Science & Engineering*, vol. 18, no. 5, pp. 21–35, 2016.
- [139] T. Sherif *et al.*, "CBRAIN: a web-based, distributed computing platform for collaborative neuroimaging research," *Frontiers in neuroinformatics*, vol. 8, p. 54, 2014.
- [140] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1–3, pp. 389–422, 2002.
- [141] H. B. Wong and G. H. Lim, "Measures of diagnostic accuracy: sensitivity, specificity, PPV and NPV," *Proceedings of Singapore healthcare*, vol. 20, no. 4, pp. 316–318, 2011.
- [142] C.-C. Chang and C.-J. Lin, "LIBSVM: A Library for Support Vector Machines," *Acm Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, p. 27, 2011.
- [143] K. Nakamura, R. A. Brown, S. Narayanan, D. L. Collins, D. L. Arnold, and A. D. N. Initiative, "Diurnal fluctuations in brain volume: statistical analyses of MRI from large populations," *Neuroimage*, vol. 118, pp. 126–132, 2015.