February 2020
Master's Degree Thesis

Multi-modal biomarkers study for Alzheimer's disease classification using Extreme Learning Machine

Graduate School of Chosun University

Department of Information and Communication Engineering

Uttam Khatri

Multi-modal biomarkers study for Alzheimer's disease classification using Extreme Learning Machine

February 25, 2020

Graduate School of Chosun University

Department of Information and Communication Engineering

Uttam Khatri

Multi-modal biomarkers study for Alzheimer's disease classification using Extreme Learning Machine

Extreme Learning Machine
을 이용한 알츠하이머병 분류를 위한 멀티모달
바이오마커에 관한 연구

Advisor: Prof. Goo-Rak Kwon

This thesis is submitted to Chosun University in partial fulfillment of the requirements for a Master's degree

October 2019

Graduate School of Chosun University

Department of Information and Communication Engineering

Uttam Khatri

This is to certify that the master's thesis of

Uttam Khatri

has been approved by examining committee for the thesis requirement for the master's degree in Engineering.

Committee Chairperson	**Prof. Jae-Young Pyun**	
Committee Member	**Prof. Bum-Shik Lee**	
Committee Member	**Prof. Goo-Rak Kwon**	

November 2019

Graduate School of Chosun University

ACKNOWLEDGMENT

This thesis present the conclusion of my master's degree at the Faculty of Information and communication, Chosun University. It has been an amazing couple of years, and I have a chance to learn much during this challenging period, yet jubilant time. My task has been highly exciting to deal with, however the work itself was sometimes hard and challenging, the numerous individuals who helped during this period are much helpful and supportive.

First, I would like to express my sincere gratitude to my advisor, Prof. Goo-Rak Kwon for his encouragement, guidance and valuable knowledge shared throughout the development of this thesis. The exposure I have gained during my research would be a valuable treasure of my life.

I would also like to thank and appreciate all my professors, seniors, lab members, friends and family, who have endured my partial absence, yet have been there for me.

Finally, I would like to thank ADNI and its collaborators for their great efforts, large amounts of work and willingness to share their data, without which this thesis and the original work described herein would not be possible.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging.

Table of contents

List of figures

List of tables

요약

Extreme Learning Machine 을 이용한 알츠하이머병 분류를 위한 멀티모달 바이오마커에 관한 연구

우탐 카트리
지도교수: 권구락.
정보통신공학부 조선대학교 대학원

알츠하이머 병 (AD), 치매의 세계에서 가장 흔한 형식이 앞으로 몇년 동안 오를 전망이다. 질병의 치료법은 매우, 질병의 적절한 이해와 정확한 치유 치료의 부족이라고 비싸다. AD 과 그 전구기의 조기 진단(마일드 인지 장애(MCI)질병 경과에 지연 가능성이기에 거기에 관심을 갖기에 초점을 맞춰 많은 것이 매우 중요하다.이전 진단과 체계적인 새로운 방법의 개발. 뇌의 조직의 변화가 AD(뇌 MR 영상에 지대한)의 가장 민감한 기능이고, 하나는 AD 의 중요한 바이오 마커의이라고 생각합니다. 그러나는 MCI 와 AD 의 진단에 민감한 보호가 있는 여러 바이오 마커에 알츠하이머 병이 다른 가능한 바이오 마커, 점점 더 많은 연구들이 집중, 하지만 있다.사람들은 아직도 이 연구에 비해 결합한 복합 분석의 대부분이 오직 MRI 의 볼륨을 사용하는 적절한 바이오 마커 확인했다가 부족하다. 기계 학습 분야에서 같은 영상 데이터에서 형상 배울 수 있는 가능성을 보여 줄 수 있 Multimodal 바이오

마커와 기계 학습 기술, 그리고 특별히 깊은 학습 모델이다.MRI 며 좀처럼 AD 의 자동 분류를 용이하게 한다. 이 논문은 처음으로 알츠하이머 병의 전구기 경도 인지 장애 정보를 준다. 그러고 나서 그것은 가장 중요한 관련 기계 지난 몇년의 연구 학습에 대한 일부의 검토 및 요약 present. 이 지식에 근거하여, 우리는 알츠하이머 병의 영상, 임상적, 그리고 생물적 단위에서 자동 분류 가능성을 점검하기 위한 분석과 실험을 디자인한다.바이오 마커, 개별 기능 설정하고 그것의 합병, perfo 과 학습 업무의 매개 변수에 기능 축소와 변경 방법을 사용하여.다양한 기계 학습 기술을 다듬고 분석하였다. 우리는 SVM-RFE 알고리즘을 통해 선택된 데이터 세트에서 훈련된 익스트림 학습 기계가 AD 와 다른 모든 증상 그룹 사이의 이진 분류기로 학습 패턴을 사용하여 다른 관련 작업에 비해 최상의 결과를 제공했다는 것을 발견한다. 따라서 이 논문은 주로 분류 작업을 위한 다모달 바이오마커와 익스트림 학습 기계에 초점을 맞추고 있다. 나는 이 논문이 이 문제에 대한 추가 연구를 위한 출발점을 줄 수 있기를 바란다. 제안된 방법을 이용한 건강관리(HC), 알츠하이머병(AD), 마일드 인지장애(MCI)의 분류 성과와 함께 MCI 초기 및 후기 분류 결과를 제시한다. ELM 알고리즘을 사용하여 평가할 때, 구별되는 바이오마커의 조합이 AD 대 HC 의 95.15%, MCI 대 HC 의 87.81%, MCI 대 AD 의 85.93%, EMCI 대 LMCI 분류의 81.73%와 함께 우수한 성능을 발휘하는 것으로 확인되었다. 한편, 복수의 바이오마커의 조합이 확인된 수신기 작동 특성(ROC) 곡선에서 곡선 아래의 영역(AUC)은 더 나은 분류 성능에 도달할 수 있다. 제안된 특징 조합과 선택 알고리즘은 효과적으로 AD 와 MCI 환자를 분류하므로 임상 실습에서 AD 분류의 정확성을 도울 수

있다. 또한 우리는 성과를 알츠하이머병 신경영상화 이니셔티브(ADNI) 데이터셋에 대한 교차 검증 방법과 SVM 분류기와 성공적으로 비교한다.

키워드: 알츠하이머병, 다모달 바이오마커, 머신러닝, SVM-RFE, 패턴 인식, 특징 선택 딥러닝, ELM, 서포트 벡터 머신.

Abstract

Multi-modal biomarkers study for Alzheimer's disease classification using Extreme Learning Machine

Uttam Khatri
Advisor: Prof. Goo-Rak Kwon, Ph.D.
Department of Information and Communication Engineering
Graduate School of Chosun University

Alzheimer's disease (AD), the most common type of dementia in the world is expected to rise in the coming years. The treatment of disease is highly expensive, there is lack of proper understanding of disease development and precise curative treatment. Early diagnosis of AD and its prodromal stage (Mild cognitive impairment (MCI) is essential for possible delay on disease progression, and thus there is large number of attention focused in the development of new and systematic methods for earlier diagnosis. Structural changes of the brain are consider to be most sensitive feature of the AD (noticeable on brain MR image), and one of the important biomarker of the AD. Yet there is other possible biomarkers for Alzheimer's disease, more and more researches are focus on multiple biomarkers which have been shown to be sensitive to the diagnosis of MCI and AD, but they still lack to identified the proper biomarkers, most of the multimodal analysis only use the volume of MRI so in this study we combined and compared multimodal biomarkers

and the Machine learning technique, and particularly deep learning model from the machine learning field, which might show potential to learn features from imaging data such as structural MRI, and so that facilitate automatic classification of AD. This thesis first give information on Alzheimer's disease and its prodromal stage mild cognitive impairment. Then it present some reviews and summary over the most relevant and important machine learning research from past few years. Based upon this knowledge, we design and perform analysis and experiments to examine the possibility of automatic classification of Alzheimer's disease from imaging, clinical and biological biomarkers, using methods of features reduction and alteration in the parameters of the learning task with individual features set and merging of it, and performed and analyzed various machine learning technique. We discover that extreme learning machine trained on a dataset that had been selected via SVM-RFE algorithm, with learning pattern as a binary classifier between AD and all other symptomatic groups provided the best results as compare to other related work. Thus, this thesis mainly focus on the multimodal biomarkers and extreme learning machine for classification task. I hope that this thesis can give a jumping off point for further research on this problem. The performance results of classification of healthy controls (HC), Alzheimer's disease (AD) and Mild Cognitive Impairment (MCI) using proposed method is presented along with its results in early and late MCI classification. Obtained results validated that the combination of distinct biomarkers perform well with accuracies 95.15% for AD vs HC, 87.81% for MCI vs HC, 85.93%, for MCI vs AD and 81.73% for EMCI vs LMCI classifications respectively, when evaluated using ELM algorithm. Meanwhile, the area under curve (AUC) from receiver operating characteristic (ROC) curve verified combination of multiple biomarkers could reach a better classification performance. The proposed

features combination and selection algorithm effectively classify the AD and MCI patient therefore may assist the accuracy of AD classification in clinical practice. Furthermore we successfully compare the performance with SVM classifiers with cross validation method for Alzheimer's disease neuroimaging initiative (ADNI) datasets.

Keywords: Alzheimer's disease, Multi-modal biomarkers, Machine Learning, SVM-RFE, Pattern Recognition, feature selection Deep Learning, ELM, Support Vector Machine.

1. Introduction

Alzheimer's disease (AD) is progressive, an irreversible neurodegenerative disorder of the central nervous system and most common form of dementia characterized by unusual accumulation of amyloid plaques and neurofibrillary tangles in the central nervous system, affecting the behavior, thinking and memory patterns of an individual, for which no medication or effective treatment is currently develop. An estimated 5.7 million Americana are living with Alzheimer's disease in 2018. By 2050, this figure is estimated to rise to approximately 14 million [1], this number is in increasing trend worldwide and there is a large number of interest in early detection of the disease, as this may yield the better treatment mechanism. In traditional method diagnosis of Alzheimer's disease relied mainly on clinical investigation and cognitive assessment. Recent studies, however, shows that multiple biomarkers analysis of neuroimaging scans as well as others biomarkers may assist more reliable and precise diagnosis. Thus more and more research has been focusing to find sensitive biomarkers and implementing machine learning approach to perform automatic early diagnosis of Alzheimer's disease. A promising number of ongoing research [2-6] is focus on different biomarkers based techniques, as an effort to early detection of AD-related changes to characterize prominent atrophy patterns during the prodromal stages, when mild symptoms are only evident of the disease. Thus, it is importance to develop a fundamental strategies for timely treatment and progression delay at early stage detection of Alzheimer's disease before clinical manifestation. Which result into the concept of mild cognitive impairment (MCI). MCI, a transitional stage between the healthy (normal) control (HC) and AD, is defined to describe individual who have moderate symptoms of brain deficiency but can able to

perform everyday tasks. Patients in the stage of MCI have high risk of progressing to dementia [7-9]. Some MCI patients are progressive to AD after baseline within a certain time frame, while others remain stable. Repot has been shown that 10% to 15% MCI patient's progress to AD per year and 80% of them will have converted to AD after approximately five-six years of follow-up [8], [10]. It is crucial to find the biomarkers that classify patients who have MCI and later progress to AD (converter MCI) from those who do not converted to AD and healthy control (HC). To identify biomarkers for MCI and AD various machine learning methods have been applied, which also improve their prediction and performances. Different biomarkers have been identified for the detection of mild cognitive impairment (MCI) and AD, including functional and structural neuroimaging measures as well as cognitive score, APOE ε4 allele status and cerebrospinal fluid (CSF) [11-15]. Recent criteria for AD diagnosis [16] suggest that neuroimaging and biological measures may play vital role for the early detection of AD and monitor its prodromal stage.

The Alzheimer's Disease Neuroimaging Initiative (ADNI), leading research project in neuroimaging field, has significantly contributed to the further understanding and analysis of the disease by providing reliable neuropsychological and clinical dataset for research purposes, including a labeled dataset of individual patients from various diagnostic groups consisting of magnetic resonance images, positron emission tomography, CSF, genetics factor and cognitive performance. Recent research has generated very good results on multimodal biomarkers from the ADNI dataset using, machine learning, deep learning methods and artificial neural networks. In the medical field Artificial Neural Networks and machine learning methods have been

successfully utilized to diagnostic assistance and Clinical Decision Support Systems, and there is huge focus in machine learning system for use in cardiology, radiology, oncology, etc. to develop more reliable, cost-effective and simple-to-use systems for assisting clinicians. This kind of Computer-Aided Diagnosis is especially impressive in the context of early diagnosis of disease, which is very crucial in the case of Alzheimer's disease. Imaging modalities seems to be an interesting diagnostic features, as it is consider important biomarker, widely used, and as there are morphology changes in brain that are strongly linked with Alzheimer's disease. A large number of relevant dataset are exist in standardized form in ADNI dataset. An artificial neural networks and machine learning technology is well suited particularly have proven for handling high dimensional data like that of multimodal classification technique [17-18].

This thesis relates the implementation of various machine learning, and feature selection technique in multimodal multiple biomarkers to identify the Alzheimer's disease, first we obtained the imaging as well as non-imaging multimodal biomarkers form the ADNI database, second we combined and compared the different biomarkers to diagnose AD and MCI with HC.

1.1 Overview and Motivation

In diagnostic imaging, Magnetic Resonance Imaging (MRI) is consider as most powerful imaging modalities for assessing neurological disorders. Similarly, CSF, genetic factors (APoE4) and cognitive score are others major non-imaging biomarkers. For Alzheimer's disease (AD), sMRI imaging was shown to be able to detect the onset of neurological disorders. Early detection

provides patients with access to therapies that are more effective in the early stages of the disease. Distinguishing between the different stages of cognitive impairment currently depends on clinical history and some neuropsychological, biological, and molecular genetic examination. sMRI different patterns of atrophy that might aid differential diagnosis. Imaging by different modalities such as CT and MRI provides rich anatomical information. Nevertheless, in the case of AD, sMRI images may also provide evidence of disease progression. In additional, sMRI images include the best-established structural brain imaging measurements of disease progression (i.e., MRI measurements of hippocampal volume) [19]. Therefore, combining sMRI, and other non-imaging biomarkers (CSF, Genetics and Cognitive score) provides better discrimination than MRI [20-26]. Different anatomical regions used in previous works with multi-region approach to classify the subjects. However, integrated specific anatomical regions detected by previous works could obtain better accuracy which attracted our attention to focus on certain offered ROIs in literature which are not tested together. This project focused on the diagnostic value of multi-biomarkers method, particularly on, MRI and other non-imaging biomarkers. sMRI has been used to study neurodegenerative diseases for over two decades. Indeed, a particular application for sMRI is recognizing AD because of the sharp contrast in their atrophy pattern. Therefore, sMRI is a valuable technique for diagnosing AD and for evaluating the efficacy of drugs that aim at modifying the progression of AD. Decreased volume, thickness and area reflects metabolic deficits and neuronal injury [27]. Since the researchers could have more information of atrophy pattern from sMRI images could discriminate between AD and different stages of MCI versus NC. In combination with Genetics, CSF and cognitive score may improve discrimination performance over what they

achieved with only sMRI images. Thus this thesis aim to test the sMRI with non-imaging biomarkers to identify the AD. The present work focused on having a better understanding of the diagnostic accuracy of possible combination of imaging biomarkers in AD, dependently of other available modalities such as CSF, Genetics factor and cognitive examinations. Future work could combine additional information.

1.2 Objectives

This work presented a pipeline using the available image processing tools for each step (skull-striping and segmentation using Freesurfer) to classify individual subjects into four different classes (AD, MCIs, MCIc, and HC). The main biomarker to do this classification is brain sMRI, and other non- imaging biomarkers (CSF, Genetics and Cognitive Score). Thesis objectives are the following:

— To collect and analyze individual cohort data downloaded from a pre-processed Alzheimer's disease Neuroimaging Initiative (ADNI) database, which encompasses patients with CSF, Genetics and structural MRI.

— To classify individual subjects into four classes using different classifiers, features selection techniques; atrophy pattern analysis on sMRI and

— To evaluate the results of classifiers and to compare the results of the multimodal approaches.

1.3 Contributions

The contribution of this work is in atrophy pattern analysis of sMRI and testing the effectiveness of combination of sMRI with other non- imaging biomarkers such as CSF, Genetics and Cognitive Score to discriminate between four classes (AD, EMCI, LMCI, HC) which is not used as a feature in the literature: [28-30]. First step in this thesis extracted the anatomical features of the brain by using Freesurfer [31]. Proposed method uses thickness, volume and area features extracted from Freesurfers pipeline based on desikan –killiany atlas [32]. Beside that this method used feature selection technique and compare the performance of different classifiers. Proposed mehtod additionally attempt the less commonly reported classification task of separating EMCI from LMCI patients.

The obtained outcome is supportive and reliable towards use of machine learning method, and can be intended towards medical CAD system development.

1.4 Thesis Layout

This thesis is composed of seven consecutive section. Following the introduction, section 2 provides theory and background information and describes some useful terms for understanding this thesis. Section 3 present the detail of machine learning and evaluation technique. Section 4 presents a critical literature review of works with their methods, which are compared or applied to the current work. The included methods and their results of classification accuracies are state of the art. Section 5 explains methodology and proposed pipeline, particularly the two important steps of statistics

analysis, features extraction and classification result. This paper details the method of classification of individual subjects and compares the results of biomarkers combination approaches using sMRI, CSF, Genetics and Cognitive scores. Section 6 discussed about the finding of result and section 7 concluded the proposed working pipeline.

2. Theory and Background

2.1 Brain MRI and Alzheimer's disease

Alzheimer's disease (AD) is a characterized by neurodegeneration that causes problems with thinking pattern, memory and behavior which is ultimately fatal. It is the most typical and common form of dementia, it occupy 60 to 80 percent of all dementia case [33]. It is one of the most financially burdening disease for society of developed countries. Report from 2006 Alzheimer's affected 26.6 million individual worldwide, and is forecast to affect 1 in 85 people worldwide by 2050 [34]. The World Alzheimer Report of 2010 [34] shows that there would be 65.7 million individual living with dementia by 2030, and this number will increase to 115.4 million by 2050. The same study also marked out that almost two thirds of individual affected by dementia live in financially middle- and low- income countries, which are expected to largest increase in number of affected people in the coming years [34] as the regions are developing rapidly. This will become challenging for many reasons, one of the reason is that dementia patients heavily rely on informal care in these nation , and it will become exceedingly difficult to maintain the proper treatment and care for such a large number of older segment of these populations and disease prevalence climbs [34]. Dementia has an huge societal cost at present, and occupy 1.01% of the total value of worldwide Gross Products of Nation [34]. It is pointed out that this problem will become worsen in the coming years, with an predicted 85% worldwide societal cost increase by 2030 [34], assuming that without possible background factors (e.g. dementia incidence, macroeconomic factors prevalence of dementia and availability and effectivity of treatment) change. While some of its sign can also show up somewhat similar to normal signs of advanced aging,

it is vital to note that Alzheimer's (and dementia in general) is no longer a normal stage of aging. With the progression of disease over time, dementia symptoms slowly get worse. At present, there is no proper treatment for Alzheimer's disease; the aim is rather to slowdown the progression of the disease, address behavioral problems, improve symptoms and improve life quality of individual. However, current medication can temporarily slowdown the development of disease symptoms if the disease is detected an earlier. While more reliable and effective treatments with ultimately disease prevention or even a proper medication is a very crucial (long term) goal, even early diagnosis can become comparatively better treatment for patient. It is still unknown about the proper cause for Alzheimer's disease, except for the few cases of detectable genetic irregularity. Current research shows, however, that it is strongly associated with neurofibrillary tangles and neuritic plaques in the brain [35]. While Amyloid beta protein which built the neuritic plaques on brain, is known to be strongly associate in the development of the Alzheimer's, there is still lack of proper understanding whether or not it is a major factor, as many researcher believe it to be. It is, however, generally this information is a marker of the disease. In the present the trend are ever-increasing towards the proper diagnosis of disease and track the progression earlier before symptom occurs. Past few years' advancement in research is increase, most importantly proper identification of biomarkers (importantly brain imaging techniques) which allow understanding and diagnosis of AD-related changes months, years and sometimes even decades before appearance of clinical symptoms. The biomarkers for Alzheimer's can be categorized into early biomarkers, which usually measure the amount of amyloid deposition in the brain cell (e.g. CSF amyloid, PET imaging), and later biomarkers, which generally measure neuron degeneration (e.g. sMRI, CSF tau, FDG PET). Brain

scans are generally ignore other causes for disease symptoms, but it can be an important indicator of whether or not Alzheimer's is present. In 1996 article [36] states that "In clinical practice, the diagnosis of Alzheimer's disease is based on typical features of the disease and exclusion of other conditions causing dementia". The main way to identified whether individual was affected by the disease is post-mortem analysis of brain tissue. However, both neurotic plaques and neurofibrillary tangles seems to play a main role in the progression of Alzheimer's disease. Although a large number of research has been performed on Alzheimer's disease, there is still an urgency for an earlier diagnostic system for the Alzheimer's disease.

2.1.1 Mild Cognitive Impairment

It is consider as the beginning phase of Alzheimer's disease, individual which have mild symptom and still shows normal behavior to the everyday life but the brain morphology start to change. There is still confusion about MCI whether it is correspond to different diagnostic phase or to a prodromal stage of Alzheimer's disease. Individual brain morphology start to change in MCI-patients, brain shape and pattern have already been going on for quite some case some time, and syndrome are only just start to appear. It does not (yet) consequence in complication that are severe enough to interfere individual day to day task, which would be acknowledge dementia.

2.1.2 Risk Factors

Although there is still lack of exact cause for dementia but certain factors are strongly associated to the development of Alzheimer's disease. Which are briefly discussed below.

2.1.2.1 Age

Age is one of the strongly and clearly associated with Alzheimer's disease. Individual above age 65 have double the risk of disease development within the five years period and similarly the individual above the age 85 have almost 50% percent of disease development [37].

2.1.2.2 Family History

Family history and background is also associated with Alzheimer's disease the individual are more likely to develop AD whose close family member have Alzheimer's disease. The risk factor will increase with the number of patients in the individual family. Either hereditary or environmental or both factors may have a role to diseases development and progression in families [37].

2.1.2.3 Genetics

Genetics factors are consider another important cause for the development of AD. Generally, two type of genes are responsible for the disease development and progression in Alzheimer's:

• Risk genes (responsible to increase likelihood)

• Deterministic genes (believe to be direct cause of disease)

If the familial AD present (i.e. dominant autosomal form of AD), development of early symptom of Alzheimer's (i.e. MCI) indicate the beginning of Dementia. A huge number of individual are affected by early onset development of Alzheimer's Dementia. Development and progression of MCI to Alzheimer's dementia in individual are variable. However the study shows that the individual with one or two ε4 alleles in their apolipoprotein E (APOE) gene is strongly related to the increasing risk factor for late-onset Alzheimer's dementia. But in opposite, presence of ε2 allele, decrease the risk factor in individual.

2.1.3 Pathophysiology

The proper and precise cause of Alzheimer's is still not known well because of its complex nature and pathogenesis.

2.1.3.1 Biochemistry

Alzheimer's is happen mainly due to the protein misfolding on brain cell by excessive accumulation of amyloid beta (β-amyloid) protein. [1], [35]. Amyloid beta which is short peptide and generate as an abnormal byproduct of protein amyloid precursor protein (APP) in brain cell whose exact function still unknown are thought to be involve in neuron development. These sticky amyloid fragments made clump together and generate plaques (known as

neuritic plaques). These plaques hinder the communication (i.e. signaling) between brain cells, which ultimately cause the death of the neurons.

Amyloid plaques are "a hallmark feature of a pathological diagnosis of AD" [7], and which is the only biomarker in Alzheimer's disease help to proper detection and quantification of the accumulation of amyloid protein in brain cell. The content of protein level can be measure directly through individual plasma and cerebrospinal fluid (CSF). We can also use positron emission tomography (PET) to measure the protein. The clinical benchmark for mild cognitive impairment due to Alzheimer's disease states correlation with the amyloid: "Current evidence suggests that markers of amyloid pathology (i.e., CSF and PET) precede evidence of neuronal injury. This does not prove that Aβ is the initiating factor for the disease. However, it does suggest that these different categories of biomarkers seem to provide different sorts of information about the progress of disease in the brain" [7]. It is also suggested that abnormal formation of tau protein, consider as taupathy, a microtubules-correlated protein present in neurons which actually help to stabilize microtubules in the cell are also consider as cause of disease. The function of tau protein is to keep microtubule straight which help molecules pass freely through it. However in Alzheimer's protein change into twisted strands (i.e. collapses in tangles) which cause to obstruction on transportation of nutrition on brain cell and ultimately leading to the cell death. The change in phosphorylated-tau and tau could link to the beginning of the Alzheimer's with general damage in synapses and neurons. The accumulation of neurofibrillary tangles and β-amyloid plaques in brain cell ultimately lead to loss of synapses and neurons (morphological changes of the brain's structure), which lead to the "memory impairment and other cognitive problems" [38].

2.1.3.2 Neuropathology

The major symptom during the Alzheimer's disease are happen due to the structural and morphological changes in the brain (i.e. brain abnormality), which is consider as the sensitive features for the Alzheimer's. Symptoms usually appears in the hippocampus part of brain initially in the individual affected by disease, the pattern can be notice clearly through volumetric analysis of structural MRI for Alzheimer's diagnosis [36]. The loss of synapse and neuron in patient's brain is clearly noticeable difference, which can be seen through Figures 2-1. Alzheimer's is mainly characterized by degradation of hippocampus and cerebral cortex, which lead to cortical atrophy in parietal, frontal and temporal areas [39]. Compared with healthy control generally ventricles are enlarged in Alzheimer's patient, it is clearly noticeable in Figure 2-1. At the beginning of the disease microscopic alteration is happen in the brain before the first symptom of memory loss, one resent study shows that the mediotemporal lesion are present upto 5.6 years before the clinical diagnosis of disease [40]. The same study shows that there was no atrophy pattern observed in the frontal lobes, this indicate that at the beginning of the disease there was no clear sing of atrophy but with the laps of time at which diagnosis was made it already affected severely. On the series of observation, authors observed the atrophy of frontal lobes only at the time close to diagnosis of disease. This finding also support that reduction in volume of the posterior cingulate cortex may lead to the later progression of the disease.

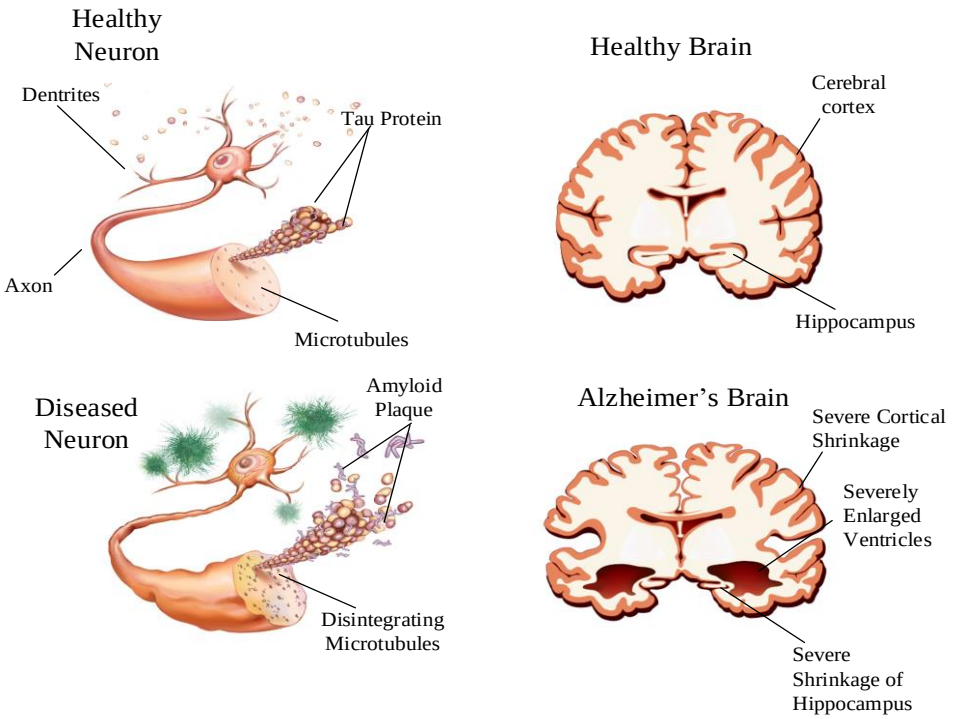


Figure 2-1. Diagram of a normal brain and person with Alzheimer's disease

Magnetic resonance imaging (MRI) facilitate researchers and clinicians to figure out the structural damage in brain linked with Alzheimer's disease. Other imaging modalities such as Pittsburgh compound B PET (PiB PET) clearly show the shapes and location of beta amyloid accumulation in the brain. Even though the technique is more invasive, it need a contrast agent — such as radioactive sugar — which is absorbable in the patient brain. Additionally, this technique is recently developed and not as available easily as like MRI. Study in 1995 [36] observed the specificity of hippocampus volume among 59 patients including mild to moderate Alzheimer's Disease, 9 patients with vascular dementia, 12 patients with idiopathic Parkinson's disease without dementia, 8 patients with Parkinson's and dementia, and 34

elderly control individual using a 1.5-T MR scanner. In their study they point out the significant decrease of hippocampal volume (both hemisphere) in all patient groups as compare to control group and they also find out absolute volume were even lower in the group of subject with Parkinson's and dementia as compare to Alzheimer's group. They hypothesize that "hippocampal atrophy does not seem to be a specific phenomenon of dementia in AD but also occurs in vascular dementia and Parkinson's even when no dementia is present" [36]. Their study however, actually point out the co-occurrence of Alzheimer's pathology in vascular dementia and Parkinson's patients. From their finding researchers concluded that one of the sensitive feature in Alzheimer's disease is hippocampal atrophy, but specificity of hippocampal atrophy seems to confine its use in clinical practice.

2.1.3.3 Biomarkers

Alzheimer's disease is progressive over time, so the biomarker magnitude gain the unusual stage in particular order (given in Figure 2-2). In the given Figure 2-2, which shows the Alzheimer's biomarkers, the curves represents variation caused by five biomarkers studied [41] (in sequential form):

1) Amyloid beta imaging identify through CSF and PET amyloid imaging.

2) Neuron degradation identify by analyzing CSF tau species and synaptic malfunction, assessed through FDG-PET.

3) Neuronal loss and brain atrophy obtained through MRI (most noticeable in caudate nucleus, hippocampus, and medial temporal lobe).

4) Memory loss obtained via cognitive test.

5) General cognitive performance decline measured by cognitive test.

From the above list, first three biomarkers can be recognized prior to diagnosis of disease, while the remaining two are "the classic indicators of dementia diagnosis" [41].

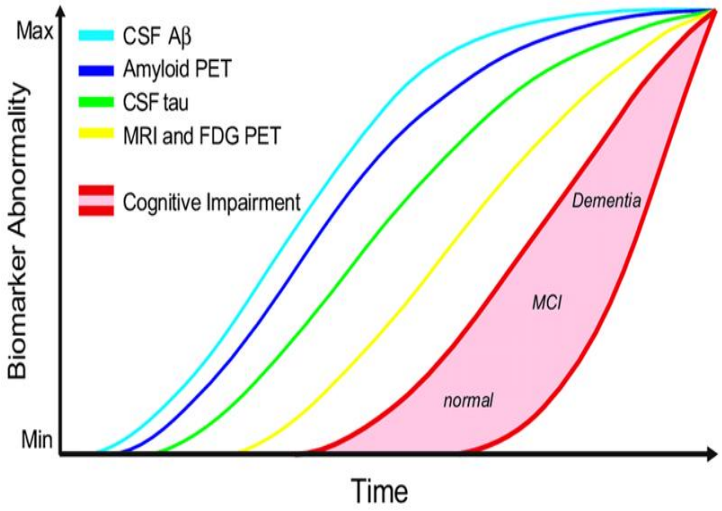


Figure 2-2. Biomarkers over the course of Alzheimer's disease [42]

In order to develop the appropriate and efficient medication, biomarkers plays the vital role and we must incorporate it within diagnostic framework, even though biomarkers primarily suggested for research purpose. Among above mention biomarkers such as tau and beta-amyloid protein (Aβ) which directly indicate the pathology of Alzheimer's; biomarker which have indirect or nonspecific indication of Alzheimer's which track the neuronal injury somewhat shows the specific regional pattern changes in brain. Some biomarkers primarily link to Alzheimer's disease also seen in other brain

disorders. If both of the biomarkers observed in the same subject, which lead to the very strong reason to presume Alzheimer's disease.

3. Machine Learning

Machine learning one of the popular branch of artificial intelligence which include algorithm design and learn pattern automatically through experience. Machine learning algorithm learn to produce intelligent decisions fully based on their capability of complex patterns recognition, which can be applicable to recognize handwritten pattern, stock market analysis, image analysis and medical diagnosis. The main focus of this thesis is medical diagnosis. This is basically a classification task in which the goal could be, for example, to utilize neuroimaging data to identify whether a newly given patient has Alzheimer's.

In this chapter we present the basic overview of machine learning technology that are relevant with this thesis and suitable to image based classification task. Section 3.1 firstly present the details of various classification technique, followed by description of algorithm along with their performance assess in same section. Finally, related literature review with this thesis work using machine learning algorithm for image based classification of Alzheimer's is provided in Section 4.

3.1 Classification Algorithms

An image made up of B numbers of voxels may be denoted by the B-dimensional feature vector, let's say $\boldsymbol{X} = (\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots\ldots\ldots, \boldsymbol{X_B})$. The general objective of a classification algorithm is to assign an each features vector to one of their corresponding $\boldsymbol{K}$ discrete classes $\boldsymbol{C_k}$. In our case we assign $\boldsymbol{K=2}$, because we performed the binary classification between the patient groups, and these two $\boldsymbol{C_1}$ and $\boldsymbol{C_2}$ classes are considered to be disjoint, such that every feature vectors belongs to corresponding classes, of the two classes. A function

Y(X) define the classification algorithm which will returns a corresponding value and which assigned belonging class for features vectors ***X***. During the training phase parameters of the function ***Y(X)*** are optimized according to the set of ***N*** number of training sample, for which the accurate diagnosis are known, which is represented as $(\boldsymbol{X}_i, t_i)|\boldsymbol{X}_i \epsilon \mathcal{R}^B, t_i \epsilon \{-1,1\}_{i=1}^N$. First we train the classifier and assessed the classification performance by using the new dataset. In literature, classification algorithm are huge in number, we only present the detail about the relevant algorithm to this work.

3.1.1 Support Vector Machines

The primary goal of binary support vector machine (SVM) is to construct a hyperplane so that it will maximizes the margin, it is the measure of distance between the possible closest points on either side of that boundary. These points are commonly called as the support vectors, and their main role is to construct the maximum-marginal hyperplane which is illustrated in Figure 3-1.

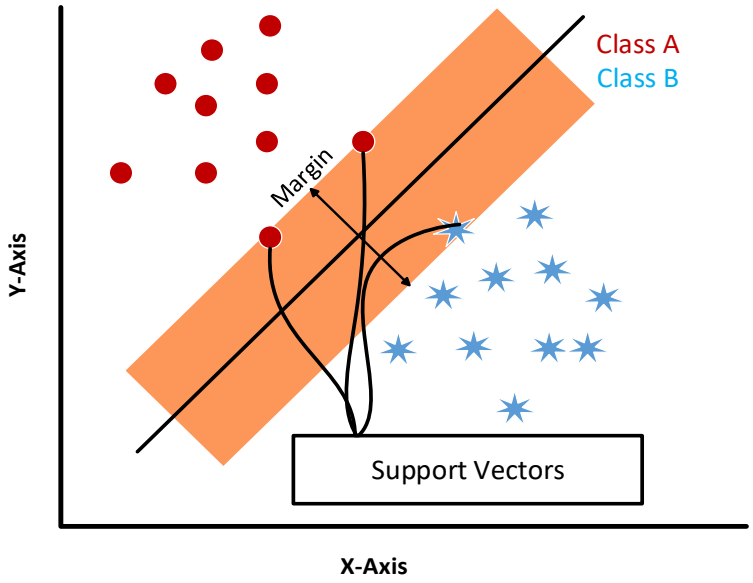


Figure 3-1. Support Vector Machine

The original SVM algorithm was a linear classifier [43], but some modification over linear SVM have seen been made to deal with dataset which are not linearly separable. Provision of soft-margin formulation has been proposed [44] for mislabeled dataset which allows proper classification, author in this article [41] used the concept of kernel tricks to create nonlinear SVM-classifiers [45]. Figure 3-1: 2-D illustration represent the construction of a maximum-marginal hyperplane in SVM-classifier. The decision surface indicated by the arrow one either side of margin maximizes the distance between the support vectors. It is important to scaled both training and testing data before the application of a SVM classifier so that the features which have high variance do not dominate over the feature with lower variance [46-47].

3.1.2 Extreme Learning Machine

Extreme learning machine compose of hidden layer in between an input and an output layer [48]. Whereas weights and biases are required to adjust by gradient-based learning algorithms on traditional feedforward neural networks for all layers, hidden layer biases and input weights are arbitrarily assigns without iterative process, and output weights are compute by solving single hidden layer system [49]. Thus as compare to traditional neural network ELM learn much faster and it is widely use in various regression and classification task as an efficient and reliable learning algorithm [50-53]. Particularly, for N training samples $\{(\boldsymbol{X}^{(j)}, \boldsymbol{I}^{(j)}) | \boldsymbol{X}^{(j)} \epsilon \mathcal{R}^p$ and $\boldsymbol{I}^{(j)} \epsilon \mathcal{R}^q$, and $j = 1,2, \ldots\ldots, N\}$, the output in ELM, $\boldsymbol{o}_j$ with n_h hidden neurons can be represented as shown below:

$$\boldsymbol{o}_j = \sum_{i=1}^{n_h} \beta_i^T a(w_i^T \boldsymbol{X}^{(j)} + b_i = \sum_{i=1}^{n_h} \beta_i^T h_i\left(\boldsymbol{X}^{(j)}\right) = h(\boldsymbol{X}^{(j)})^T \beta, \qquad (1)$$

where $\boldsymbol{X}^{(j)}$ and $\boldsymbol{I}^{(j)}$ are the j^{th} input and target vectors, respectively. The parameter p and q are the input and target vector dimension, respectively. And $\boldsymbol{o}_j \epsilon \mathcal{R}^q$ signifies the output of ELM for the j^{th} training sample, $w_i \epsilon \mathcal{R}^p$ indicates the input weight that link the input nodes to the i^{th} hidden node, b_i represent the bias of the i^{th} hidden node, and $a(.)$ signifies the activation function for the given hidden layer. $\beta = [\beta_1, \ldots\ldots\ldots\ldots, \beta_{n_h}]^T$ is the values of output weights between the output neuron and the hidden layer $h\left(\boldsymbol{X}^{(j)}\right) = [h_1(\boldsymbol{X}^{(j)}, \ldots\ldots\ldots\ldots, h_{n_h}\left(\boldsymbol{X}^{(j)}\right)]^T$ is the output vector of the hidden layer with respect to the j^{th} training sample $\boldsymbol{X}^{(j)}$. $h_i(\boldsymbol{X}^{(j)})$ is the output of the i^{th} hidden layer for the j^{th} training sample.

To obtained the optimal hidden layer weights, $\hat{\beta}$with respect to Ntraining samples which can be considered to solve the following optimization problem:

$$\underbrace{min}_{\hat{\beta}} \quad \lambda \|\boldsymbol{H}\beta - \boldsymbol{L}\|^2 + \|\beta\|^2, \tag{2}$$

where $\boldsymbol{H} = [h(\boldsymbol{X}^{(1)}, \ldots \ldots \ldots h(\boldsymbol{X}^{(N)}]^Tand\boldsymbol{L} = [\boldsymbol{I}^{(1)}, \ldots \ldots \ldots, \boldsymbol{I}^{(N)}]^T$.

Equation 2 represents the optimization problem, and its optimal solution,$\hat{\beta}$; can be analytically obtained as follows:

$$\hat{\beta} = \boldsymbol{H}^T \left(\frac{1}{\lambda}\boldsymbol{I} + \boldsymbol{H}\boldsymbol{H}^T\right)^{-1} \boldsymbol{L}, \tag{3}$$

where λ is a regularization parameter, and $\boldsymbol{I}$represents the identity matrix. After finding the optimal solution $\hat{\beta}$, the output of the ELM on test data $\boldsymbol{X}_{test}$ is determined by

$$\boldsymbol{o}_{test} = h(\boldsymbol{X}_{test})^T \boldsymbol{H}^T \left(\frac{1}{\lambda}\boldsymbol{I} + \boldsymbol{H}\boldsymbol{H}^T\right)^{-1} \boldsymbol{L}, \tag{4}$$

In this proposed method, hidden nodes number was set between 1 and 500, and we selected a sigmoid as an activation function. Beside, we used grid search method to tune the ELM parameter on training dataset in order to achieve optimum cross-validated validation accuracy. Similarly to minimize the random effects during the weight initializations, each parameters of the number of hidden nodes was used hundred times and the average performance was calculated.

3.1.3 Classifier Performance

Classification performance measurement are essential to assess the applicability and reliability of a trained algorithm with independent test data and important for the optimization of parameters during training process. The simplest and widely use performance metric is accuracy, which gives the proportion of dataset that are correctly identified by the classifier. However, this measurement method does not always gives an appropriate evaluation of performance, and beside that other relevant metrics are described detailed in Section 3.1.3.1. The method of cross- validation may be utilized to assess the classifier's generalization performance, as there is lack of abundant and appropriate independent dataset accessible for testing. At single round of cross-validation dataset are partition in to two subsets, so that it may be trained and tested using the different subset of dataset. Generally result are reported as the average over multiple repetition in which different partitions of the dataset are used. Details of the most commonly used cross-validation techniques are provided in Section 3.1.3.2.

3.1.3.1 Performance Metrics

The performance evaluation of a binary classifier can be accessed through a confusion matrix, as present in Table 1. The number of dataset sample correctly labelled by the classifier are presented along the diagonal. These may be categorized into true positives TP, which represent correctly detected patients, and true negatives TN, which represent correctly detected healthy group. The number of incorrectly labelled dataset by classifier may be categorized into false negative FN, which represent the patients incorrectly

identified and classified as healthy, and false positives FP, representing healthy individual falsely classified as patients.

Table 1. Confusion matrix representing binary classifier.

True Class	Predicted Class	
	A (patients)	B (Healthy)
A (patients)	TP	FN
B (Healthy)	FP	TN

The accuracy measure gives the proportion of example which are correctly identified by a classifier, which is shown in equation 5 below:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}, \tag{5}$$

If the class distribution of example dataset is unbalance it may not be a good performance metric. For example, if class B is much smaller than A, a high accuracy measure could be achieved by a classifier which identify all examples as belonging to class A. The sensitivity and specificity are defined by equation 6 and 7 respectively below:

$$Sensitivity = \frac{TP}{TP + FN}, \tag{6}$$

$$and\ Specificity = \frac{TN}{TN + FP}, \tag{7}$$

which give a better assessment of the overall performance evaluation of a classifier. Sensitivity measures the portion of correctly classify patients, and specificity measures the portion of correctly classify Healthy control.

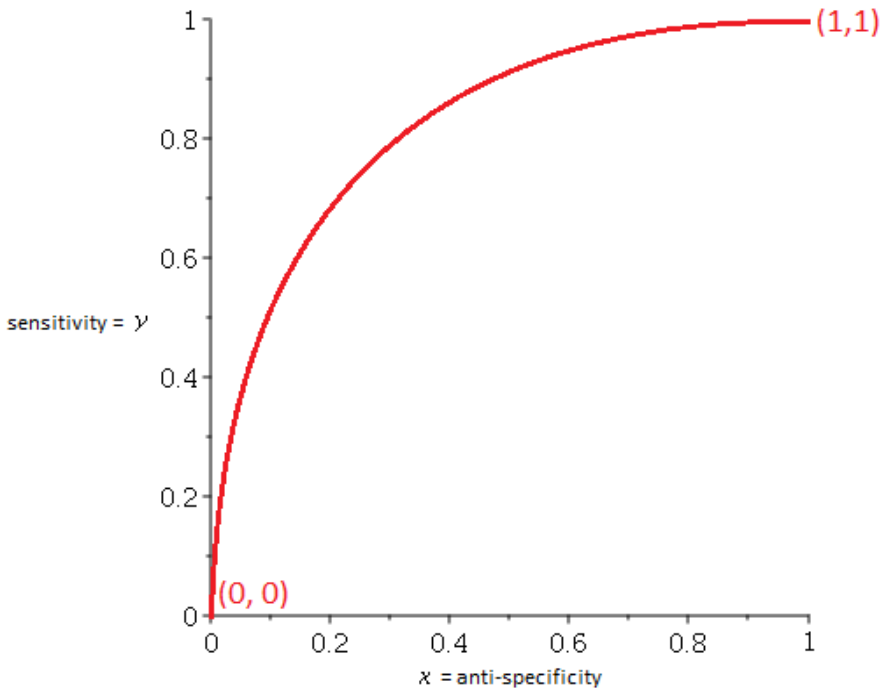


Figure 3-2. Illustration of the ROC (AUC) curve evaluation for a binary classifier. The red solid line represents the relationship between the specificity and sensitivity as classifier discrimination threshold varied. This can be compared with the no-discrimination dashed line [54].

Another important classifier evaluation method is AUC. This can be analyzed by using a receiver operating characteristic (ROC) curve. Which is

shown in Figure 3-2, a ROC (AUC) curve give the relationship between the sensitivity (true positive rate) and 1-specificity (false positive rate) as the discrimination threshold varied for the binary classifier. This curve can be utilized to select the optimal threshold value for a particular application. Consider an example, for earliest stage of disease identification of patients, it may be necessary to define and select a threshold which give a high sensitivity, at the same time a reduced specificity. Overall, the area under a ROC curve (AUC) give the aggregated measure of classifier performance evaluation [55].

3.1.3.2 Cross-Validation

The classifier parameters are optimized according to the training dataset. Therefore an independent test dataset is required for developing a proper assessment of the applicability for the classifier on new dataset. Cross-validation provides a better way to evaluate this generalization performance measure when no availability of such data set. One most commonly used cross-validation method is k-fold, in which the dataset are partitioned randomly into k number of subsets. For training purpose a single cross-validation fold ($k - 1$) subsets are involved to the classifier, and the remaining data subset are involve for testing. The process is repeated until k times, such that for testing each of the subsets is used once, and the average of the folds are presented as a final results. Another alternative for cross-validation method is random sampling with repetition, in which the dataset is partitioned randomly into fixed sizes training and testing sets. For example, 70% of the data were selected on a single round for training, with the remaining 30% data for testing purpose. Then, this process can be repeated, and the averaged result over the repetition represent the final result. Repetition of random sampling has the

advantage because of its independent nature on the proportions of the training and testing sample on the repetition number. However, some overlap between test sets may happen, and the Monte Carlo variation is also exhibits by the method. Which means that repeated use of different partitions of the dataset in analysis will give the variation on results. If two classes A and B are of different sizes, the training and testing sets should be selected accordingly such that both of them contain examples from the each classes in relatively equal proportions to the whole dataset. Which is known as stratified cross-validation, and it has been shown to yield results with a comparatively lower variance than regular cross-validation method [56]. Both the repeated random sampling and k-fold cross-validation generate distribution of an average performance values across the repetitions or folds. The statistical significance of result differences obtained from two classifiers can be measure by performing unpaired t-tests among these distribution. Similarly, permutation test may be useful to assess whether the classifiers results are significantly diverse from chance. Permutation test include cross-validation performance on dataset for which the diagnostic group have been permuted randomly. Under the null hypothesis, the result which involve in a distribution of classification results that the classifier cannot properly predict the clinical group from the dataset. Permutation test between the distribution of obtained results and unpaired t-tests represents whether the obtained results are vary significantly from chance.

4. Literature Reviews of Alzheimer's Classification

4.1 Structural MRI

This part of the review discussed the roles and limitations of sMRI for AD. The two most pathological dementia features are cortical atrophy and vascular changes [57]. As mentioned earlier, MRI is recommended for distinguishing vascular lesions and neurodegenerative dementia. Similarly, MRI images present rich anatomical information, and it is useful to extract some features which can be identified as a saliency criterion of the image [58]. Structural MRI has the ability to visualize specific atrophy patterns in the brain; hence, it is important for the differential diagnosis of AD [59]. From the sMRI point of view, brain atrophy and neuronal loss, as common MRI biomarkers in some special parts of the brain, are key criteria for the diagnosis of AD. Atrophy starts from the entorhinal area and continues in the hippocampus, amygdala, and Para hippocampus along with disease progression [60], [27].Other brain regions, such as the posterior temporal, parietal cortex, and mesial temporal lobe, are affected by progress of AD [48]. When changes in different individual brain regions occur, it is possible to assess the degree of atrophy visually. Volumetric measure is the most common quantitative metric used in AD [27]. Voxel-based morphometry (VBM) is a validated method to assess atrophy over the entire 3D sMRI scan [61]. Moreover, VBM has an alternative, one of which is manual segmentation of ROIs instead of the voxel-based approach. The research [58] proposed a fully data- driven technique to find the anatomical structural patterns in 3D images by using feature-based morphometry (FBM). To classify the subjects based on their image, FBM can promote an image model which is more related to the most probable class. The authors used this technique to obtain high accuracy of multiple kernel

boosting, which is about 95 % accuracy of classification between AD and HC, and about 72 % between MCI convertible and MCI non-convertible. The article [62] presented an automatic classification pipeline for AD identification in structural MRI. Hippocampal area is the ROI in this work. Their proposed method extract the visual features through structural MRI using the image signal approximation by Circular Harmonic Functions. In this method the probabilistic output obtained from classifies on both the amount of CSF and local features were fused, a late fusion scheme could perform the classification task of the MRI scans. The research article [63] applied Sparse Logistic Regression (SLR) to classify 69 AD and 60 HC subjects based on voxel-wise grey matter volumes derived from structural MRI. Penalized Logistic Regression (PLR) and Spatially Regularized Sparse Logistic Regression (SPSLR) are two different formulations of SLR which applied to solve the problem of having large number of voxels in comparison to the number of training subjects. The input features in the classifiers for this work were the standardized grey matter voxel intensities. Their results present about 85 % of overall classification accuracy for AD and HC [63]. The article [27] compared the annual change in CSF and MRI biomarkers for their subjects based on intergroup discrimination, correlation with concurrent cognitive or functional changes and some other biomarkers. This work didn’t use some common feature from imaging to discriminate between different classes of ADNI subjects. Whereas, it compared the Annual change in CSF and MRI biomarkers across clinical groups for the pairwise discrimination between groups. The importance of reviewing this work is to investigate clinically about the ADNI subjects consisting of AD, MCI and HC cohorts with both baseline and 12-month follow-up. There results of this work presents that the annual change did not differ by clinical group in pairwise comparisons. From

the findings, it's possible to indicate that Aβ deposition itself is not directly responsible for clinical symptoms, but rather initiates a pathologic cascade that later results in clinical symptoms. The article [64] assessed the efficacy of adding sMRI scans to a memory test to predict the progression of MCI to AD and found no significant increases in the accuracy of the diagnosis. Therefore, sMRI does not provide adequate diagnostic insight into the progression of MCI to AD [64]. In summary, sMRI is an appropriate biomarker to reflect the disease stage and intensity, which is extremely useful. It must be noted that it is an independent non-invasive measure of neuronal loss and thus provides a supplementary measure based only on anatomy. Numerous publications show that sMRI is a stable biomarker of AD progression.

4.2 Multi-Modality

Recent studies have demonstrated that multi-modal contains complementary information for diagnosis of AD and its classification into different stages with high accuracy.

However, to obtain more reliable classification results, these multiple biomarkers need to be combined to provide an accurate diagnosis. In previous sections, we reviewed some modalities, such as sMRI, CSF, which yielded satisfactory results. In this section, we studied some research that used multi-modality to achieve higher results. Some of the publications above had extra experiments to show the results of combining two multi-modal or combining more clinical and biological biomarkers. Different biomarkers can expose different aspects of pathological changes associated with AD. Some biological biomarkers have been developed for diagnosis of AD. Three CSF biomarkers

are common in state of the art: total τ (T-τ), hyper-phosphorylated τ (P-τ) and the 42 amino acid isoforms of Aβ (Aβ42) [65]. In this section, we reviewed the publications in similar domains, and we selected some relevant research to present here. In these studies, different approaches were applied to prepare the images, to extract the features, to select the classifiers, and to use some extra biomarkers. The results were obtained in several manners, which are studied. Among all studied work, there are a few studies that pointed to the different stages of MCI (EMCI and LMCI). The studies that used MRI images only gave us a good historical review on methods for finding ROIs and different classifiers.

5. Different Biomarkers study on AD Classification

5.1 Material and Method

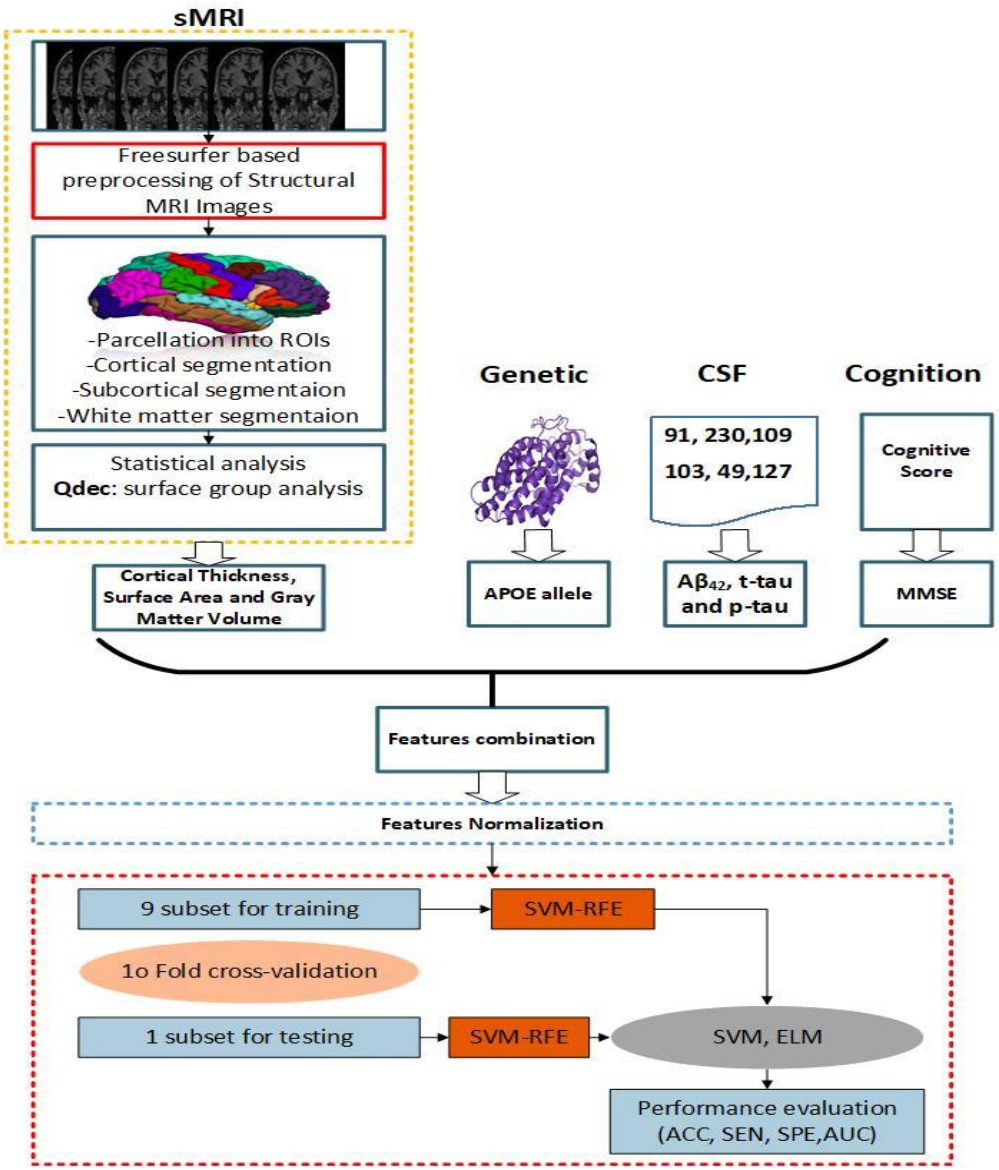


Figure 5-1. Block diagram of the proposed framework

5.1.1 Imaging Data

All individuals used in this analysis were obtained from the ADNI database. The ADNI was initiated in 2003 as public-private partnership, under principal Investigation of Michael W. Weiner, MD. The primary objective of ADNI has been to investigate whether imaging modalities such as MRI, PET, others neuropsychological assessment and clinical and biological markers can be combined to measure the early detection of AD and progression of its prodromal state (i.e. MCI). Demographic information, raw neuroimaging data, CSF measure, APOE genotype, diagnostic information and neuropsychological test scores are publically available on the ADNI data repository (http://adni.loni.usc.edu). Informed consent was obtained for all individual subjects and the study was confirmed by the related institutional review board at each data site (for more information, see http://adni.loni.usc.edu/wp-content/themes/freshnews-dev-v2/documents/policy/ADNI_Acknowledgement_List%205-29-18.pdf).

To prepare this article, we utilized the MRI, CSF and APOE genotype. The resulting study cohort consisted of subject: patients affected by AD, patients with MCI and healthy control. Socio-demographical and clinical information of participants are reported in the Table.1

Table 2. Baseline clinical and socio-demographical information of the cohort.

Group	AD	MCI	HC	EMCI	LMCI
Nos. of Subjects	53	77	57	35	42
Female/male	20/33	34/43	32/25	13/22	21/21
Age	74.4 ± 7.8	74.1 ± 7.2	75.6 ± 5.2	73.9 ± 7.2	74.3 ± 7.2
Education	15.1 ± 3.2	15.9 ± 2.9	15.7 ± 2.8	16.1 ± 2.9	15.8 ± 2.9
MMSE	23.5 ± 1.8	26.9 ± 1.8	29.1 ± 0.9	27.2 ± 1.7	26.6 ± 1.8
CDR	0.7 ± 0.2	0.5	0	0.5	0.5

The entries for age, gender, education and MMSE denote mean and standard deviation for each group. MMSE: mini mental state exam.

5.1.2 Freesurfer Analysis of sMRI

We applied the recon all Freesurfer pipeline (version 6.0.0) to the structural MRI images for cortical reconstruction and volumetric segmentation [31], freely accessible at http://surfer.nmr.mgh.harvard.edu. This pipeline automatically generated reliable volume and thickness segmentation of white matter, gray matter and subcortical volume. Cortical reconstruction and subcortical volumetric segmentation include removal of non-brain, Tailarach transformations, segmentation of subcortical gray matter and white matter regions, intensity standardization and atlas registration. After these steps, cortical surface mesh model was generated and finally the 34 cortical regions

were obtained from the cortical surface parcellation based on sulcal and gyral landmarks for both hemisphere corresponding to the desikan –killiany atlas [32]. For purpose of statistical analysis, smoothing was applied using recon-all with qcache option added. QDEC, a tool within Freesurfer, was utilized to analyze difference in cortical thickness, surface area and gray matter volume between HC, MCI and AD. To control for multiple comparison, statistical significance levels were cluster corrected for both hemispheres using false discovery rate (FDR), $p<0.05$.

5.1.3 Machine Learning Based Prediction and Analysis

An overview of prediction framework developed for this study was shown in Figure 5-1. The framework consist of four major steps: feature extraction, feature selection, feature combination and classification. We used the two machine-learning classification algorithm, SVM and ELM.

5.1.4 Feature Selection

In most studies involving neuroimaging analysis, the number of predictor voxels obtained outnumber the subjects. Thus, a dimensionality reduction technique was necessary in order to obtain the most optimal and relevant features set, discard noise and redundant features, and avoid overfitting and numerical singularities problems, and thus enhancement of the classifier performance. Solving pattern recognition or classification problems with data of high dimensionality is a challenging issue, particularly in neuroimaging

applications with limited samples, and large number of features. The learning models tend to overfit and become less generalizable if input features are redundant or irrelevant to classification. Feature selection is usually performed to identify relevant features, reduce dimensionality of the trained model, and improve generalization of the model [66]. An efficient feature selection algorithm is the essential part of a machine learning approach in case of high dimensional features. We have shown efficiency of the SVM-RFE feature selection algorithm in identifying the early stage of AD [67]. Importantly, feature selection was carried out on the training dataset only. Once identified, the same brain region during training phase were utilized to assess the classifier performance accuracy [18] on the testing data. In this study, SVM-RFE was applied in order to obtain ranked features list which could best differentiated the HC form AD and MCI. As a multivariate wrapper-model-based feature selection algorithm support vector machine-recursive feature elimination (SVM-RFE) method efficiently fits the model and removes the weak features till the specified number of informative features is reached. The ranking principle of SVM-RFE similar to the SVM model. SVM model is trained in every single iteration of the RFE. Then, the feature with lower in rank is removed since it has the no major effect on classification, while the other top ranked features are kept for next iteration in the SVM model. This sequence is repeated until all the weakest features set have been eliminated. Then, the features are graded according to the order of their elimination. The SVM-RFE algorithm described detailed in a previous paper [68]. In this work, after the utilization of SVM-RFE, the most reliable training features that help to maximize cross-validated accuracy were selected for training the classifiers.

5.1.5 Statistical Analysis

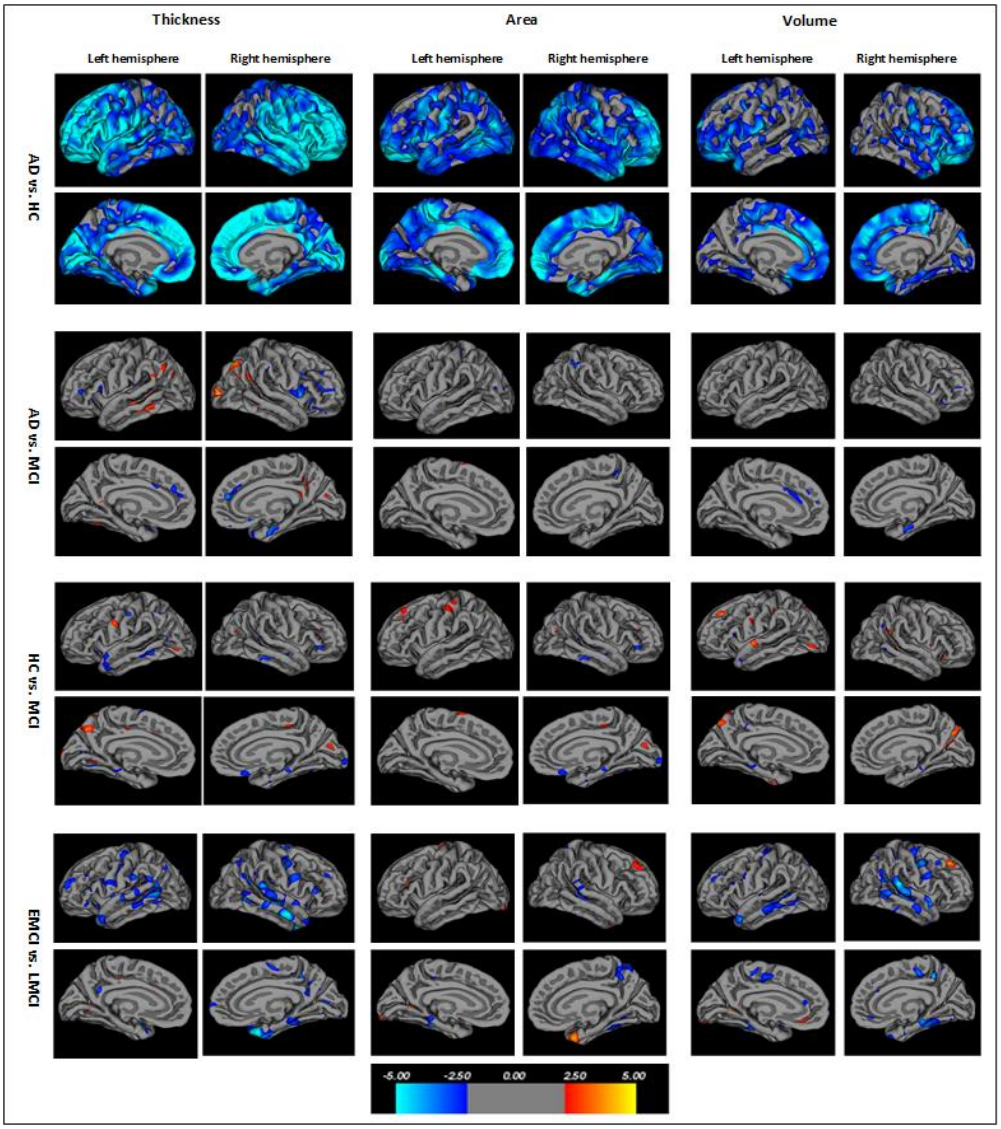


Figure 5-2. Difference on cortical thickness, area and volume for patients with different stage of Alzheimer's disease. The color bar represents the significance level of clusters. The significant threshold are set at $p<0.05$.

The cortical thickness, GM volume and surface area was analyzed using a surface-based group analysis of Freesurfer's Qdec (version 1.5). First, the spatial cortical thickness, GM volume and surface area of the both hemisphere was smoothed with a circularly symmetric Gaussian kernel of 10 mm full width half maximum to provide normal distribution of the results. Then, we employed a general linear model (GLM) analysis with age, gender and education as the nuisance factors in the design matrix to directly compare the three parameters in the both hemisphere of the AD vs. HC, HC vs MCI, AD vs MCI and EMCI vs LMCI groups. The statistical analysis results of cortical thickness, surface area, and gray matter volume are drawn in Figure 5-2. Desikan-Killiany atlas divide the human cerebral cortex into 34 cortical features in each left and right hemisphere. As the number of conflicting features is much, we only present the several top-ranked features with significant differences.

Table 3. Cluster difference in cortical thickness, area and volume in AD patients.

Features	Region	Co-ordinate			Vertex	Value	Size (mm2)
		x	y	z			
AD vs. HC							
Thickness	left insula	**-29.5**	**17.9**	**11.9**	**3165**	**-2.3512**	**3619.47**
	left parahippocampal	**-31.3**	**-41.7**	**-8.7**	**557**	**-2.2303**	**10736.27**
	left cuneus	**-47.9**	**-20.9**	**42.8**	**1039**	**-2.3188**	**2824.28**
	right rostralmiddleforntal	**38.5**	**43**	**7**	**1224**	**-2.1737**	**14.8**
	right superiortemporal	**53.3**	**-5.2**	**-5.2**	**468**	**-2.8676**	**256.28**
	ritht parsopercularis	**50.3**	**10.2**	**8.8**	**1242**	**-3.6071**	**58155.25**
Area	left paracentral	**-15.8**	**-35.5**	**49.4**	**1197**	**-2.2767**	**7018.72**
	left latera orbitoforntal	**-53.3**	**-2**	**7.5**	**489**	**-2.1962**	**3218.23**

	right paracentral	**12.5**	**-37.1**	**55.4**	**1233**	**-2.2555**	**42.55**
	right inferiorparietal	**34**	**-51.9**	**37.4**	**1133**	**-2.1026**	**7213.05**
	right posteriorcingulate	**14**	**-30.4**	**36.6**	**1201**	**-2.0979**	**1416.67**
	right inferiorparietal	**39.5**	**-44.7**	**34.6**	**7**	**-2.0263**	**50.76**
Volume	left caudalanteriorcingulate	**-5**	**18.4**	**26.4**	**2073**	**-3.2264**	**834.61**
	left orbitalis	**-28.7**	**-60.2**	**41.7**	**1103**	**2.3437**	**1202.55**
	left lateral orbitofrontal	**-59.4**	**-6.9**	**9.4**	**412**	**-2.0146**	**635.67**
	right lateralorbitofrontal	**30.4**	**22.9**	**-20.3**	**631**	**-2.7411**	**2025.6**
	right rostralmiddlefrontal	**34.8**	**51**	**4.7**	**198**	**-2.717**	**1098.82**
	right parsopercularis	**46.8**	**15.2**	**8.7**	**136**	**-2.717**	**8077.1**
AD vs. MCI							
Thickness	left inferiorparietal	**-43.6**	**-61.5**	**33**	**939**	**2.7022**	**439.37**
	left fusiform	**-40**	**-53.7**	**-20.3**	**254**	**2.459**	**149.84**
	left superiortemporal	**-47.7**	**-25**	**-9.1**	**210**	**-2.5577**	**328.03**
	right lateraloccipital	**26.1**	**-93.9**	**3.7**	**363**	**3.4944**	**363**
	right inferiorparietal	**37.7**	**-71.7**	**42.8**	**1129**	**3.3654**	**646.24**
	right superiortemporal	**50.7**	**-14.3**	**-2.4**	**722**	**-2.9715**	**342.28**
Area	left superiorparietal	**28.4**	**-51**	**42.7**	**1702**	**-2.5085**	**642.76**
	left precentral	**32.2**	**-21.8**	**60.8**	**109**	**1.8969**	**51.54**
	left paracentral	**14.1**	**-36.7**	**52.4**	**238**	**-1.8959**	**79.68**
	right inferiorparietal	**-35.2**	**-87.2**	**13.7**	**50**	**-1.8099**	**34.53**
	right precentral	**-54.2**	**-1.1**	**7.1**	**51**	**-1.7244**	**21.34**
	right superiorfrontal	**-7.4**	**4**	**66.5**	**34**	**1.6541**	**17.86**
Volume	left superiorparietal	**-28.5**	**-58.8**	**40**	**540**	**3.5868**	**216.1**
	left superiorfrontal	**-8.5**	**41.2**	**30.3**	**48**	**-2.6334**	**33.98**
	left caudalanteriorcingulate	**-4.9**	**11.3**	**32.1**	**314**	**-2.549**	**150.07**
	right entorhinal	**21.5**	**-8.9**	**-29.5**	**207**	**-3.1433**	**55.69**
	right lateralorbitofrontal	**42.8**	**27.8**	**-13.7**	**198**	**-3.1034**	**125.01**
	right superiorparietal	**32.1**	**-44.2**	**41.2**	**115**	**-2.5309**	**40.29**
HC vs. MCI							
Thickness	left insula	**-36.4**	**-9.4**	**-11.7**	**811**	**-3.5988**	**309.94**
	left precuneus	**-6**	**-68.9**	**41.7**	**628**	**2.3682**	**322.02**

	left lingual	**-29.3**	**-44.5**	**-6.6**	**612**	**-2.7175**	**279.16**
	right insula	**35.8**	**-10.6**	**-7.4**	**1154**	**-3.1385**	**436.62**
	right parstriangularis	**38.9**	**31.7**	**1**	**372**	**-2.1264**	**204.87**
	right inferiorparietal	**35.1**	**-72**	**40.9**	**285**	**-2.7121**	**152.2**
Area	left superiorfrontal	**-17.8**	**31.5**	**49.6**	**237**	**1.7293**	**156.88**
	left postcentral	**-52.1**	**-22.7**	**50.7**	**264**	**1.7286**	**122.75**
	left superiorparietal	**-32**	**-49.3**	**47.3**	**141**	**-1.7572**	**66.04**
	right supramarginal	**53.4**	**-47.4**	**35**	**1336**	**2.7163**	**676.85**
	right fusiform	**34.7**	**-12**	**-34.1**	**551**	**2.0892**	**322.84**
	right precuneus	**18.2**	**-77.3**	**27.7**	**482**	**1.7491**	**313.36**
	right precentral	**20.7**	**-30.6**	**53.9**	**328**	**-1.9165**	**115.45**
Volume							
	left supramarginal	**52.1**	**-46.3**	**22.6**	**441**	**2.7521**	**210.23**
	left cuneus	**17.1**	**-69.5**	**16.8**	**628**	**2.611**	**481.74**
	left precentral	**21.2**	**-30.7**	**53.8**	**376**	**-2.1546**	**127.02**
	right parahippocampal	**-23.8**	**-36.1**	**-15.6**	**278**	**-1.8212**	**127.68**
	right superiorparietal	**-9.1**	**-74.3**	**46.5**	**598**	**2.5199**	**303.67**
	right superiorfrontal	-18.1	33.3	39.6	210	2.6035	114.72
EMCI vs. LMCI							
Thickness	left inferiorparietal	**-46.3**	**-60**	**11.1**	**2770**	**-3.7613**	**1427.48**
	left superiortemporal	**-45.5**	**-0.4**	**-20.8**	**1067**	**-2.6742**	**466.24**
	left parahippocampal	**-31.7**	**-40.3**	**-10.1**	**550**	**-2.3925**	**239.85**
	right temporalpole	**29.2**	**9**	**-38**	**1449**	**-5.5983**	**822.82**
	right superiortemporal	**62.3**	**-34.7**	**15.2**	**1292**	**-3.1646**	**549.97**
	right inferiortemporal	**51.6**	**-56.6**	**-3.7**	**882**	**-3.0694**	**507.91**
	right precentral	**48.8**	**-6.5**	**40.5**	**858**	**-2.8315**	**351.93**
Area	left fusiform	**-36.4**	**-29.5**	**-22**	**659**	**-3.0958**	**312.21**
	left lateraloccipital	**-19.4**	**-99**	**-15.1**	**326**	**2.6022**	**250.61**
	left parahippocampal	**-18.8**	**-33.5**	**-14**	**224**	**-2.0347**	**97.59**
	right superiortemporal	**48.1**	**-32.9**	**2.2**	**1515**	**-2.5815**	**579.8**
	right superiorparietal	**10.6**	**-52.7**	**65.1**	**1697**	**-2.2367**	**641.52**
	right postcentral	**33.8**	**-29**	**51.9**	**799**	**-2.0859**	**367.23**
Volume	left bankssts	**-57.9**	**-46.9**	**-1**	**2196**	**-2.6263**	**1162.58**

	left precentral	**-36.1**	**-22**	**51.7**	**2736**	**-2.5339**	**1073.1**
	left rostralmiddlefrontal	**-22.4**	**27.9**	**32.8**	**582**	**-2.3081**	**326.86**
	right superiortemporal	**62.3**	**-34.7**	**15.2**	**3392**	**-4.4677**	**1449.96**
	right precentral	**49.5**	**-6.2**	**41.1**	**4103**	**-3.2472**	**1813.94**
	right fusiform	**33.9**	**-37.8**	**-22.8**	**782**	**-3.1822**	**425.05**

Table. 3 presents the atrophy position and range of clusters involving for the differences in gray matter volume, cortical thickness, and surface area at each vertex between HC, MCI and AD by QDEC analysis. In this table, only the top features which have significant cluster differences for each kind of parameters are provided. From statistical maps shown in Figure 2 and statistical Table 2 we can notice that:

1. The cortical thickness of the left insula, left cuneus, paracentral, right rostralmiddlefrontal, right superiortemporal and right parsopercularis was thinner in AD compared with HC. For HC vs. MCI the cortical thickness of the left precuneus, left lingual, left and right insula, right parstriangularis and right inferiorparietal was thinner. Similarly, for AD vs MCI the cortical thickness of the left inferiorparietal, right lateraloccipital, and right inferiorparietal, left and right superiortemporal was shows the major atrophy. In case of EMCI vs. LMCI the cortical thickness of the left inferiorparietal, left parahippocampal, right temporalpole and right superiortemporal area shows the major difference.

2. In case the surface area, AD have lesser areas than HC in the left and right paracentral, left latera orbitoforntal, right inferiorparietal, right inferiorparietal right posteriorcingulate and right inferiorparietal.For HC vs. MCI the left

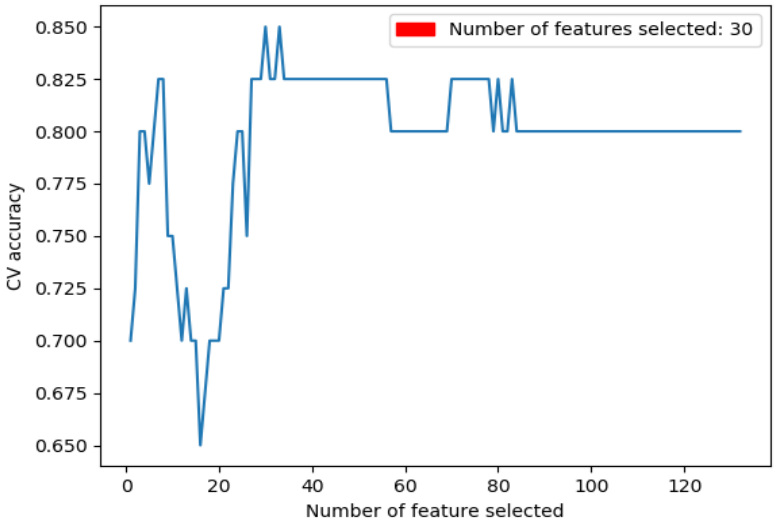
superiorfrontal, left postcentral, left superiorparietal, right supramarginal, right fusiform right precuneus and right precentral show the decrease area. Similarly, for AD vs MCI left superiorpariental,left and right precentral, left paracentral, right inferiorparietal and right superiorfrontal was shows the major atrophy in surface area. In case of EMCI vs. LMCI the left fusiform, left lateraloccipital, left parahippocampal, right superiorparietal and right postcentral shows the major difference.

3. Compared with HC, the volume of gray matter on the left caudalanteriorcingulate, left orbitalis, left lateral orbitoforntal, right lateralorbitofrontal, right rostralmiddlefrontal and right parsopercularis was lessr in AD. In case of AD vs. MCI left superiorparietal, left superiorfrontal, left caudalanteriorcingulate, right entorhinal and right superiorparietal shows the decrease in volume. For HC vs. MCI the left supramarginal, left cuneus, left precentral, right parahippocampal and right superiorparietal shows the major volume atrophy. Similarly, for EMCI vs LMCI the left bankssts, left precentral, left rostralmiddlefrontal, right precentral and right fusiform have decrease volume.

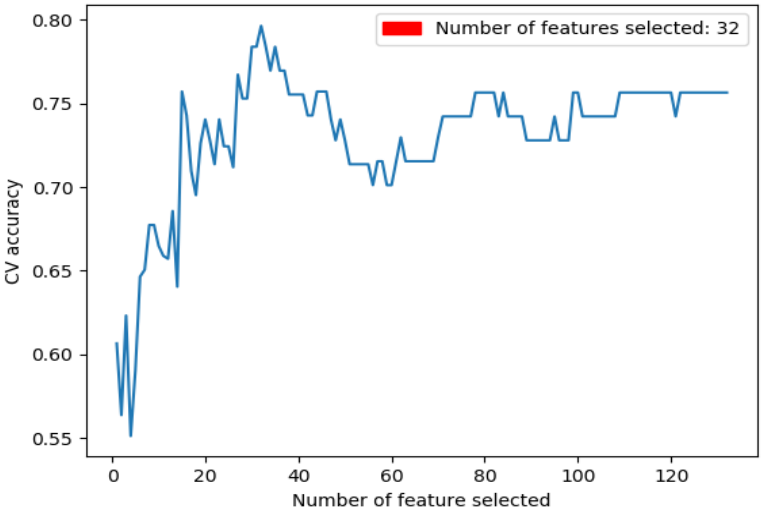
Moreover, from this analysis we can notice that, thickness area and volume of the AD decrease significantly as comparison with HC, similarly, there was significance atrophy between LMCI as comparison with EMCI and there was little difference on atrophy pattern among AD and MCI patients.

6. Experimental Results and Discussion

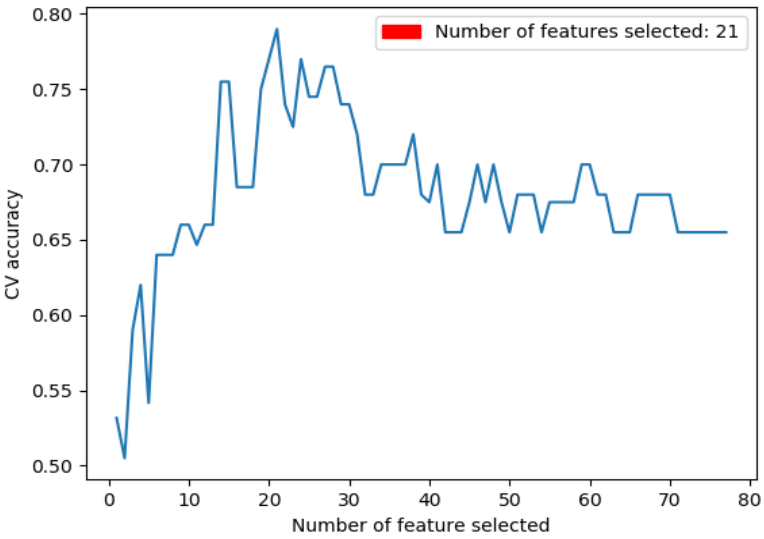
Individual features set is selected by SVM-RFE to identify the different group and perform the classification. Similarly, we combine all the features set from different measures and applied the SVM-RFE method on it to select the optimal features. Multiple measure features set are created by combining the cortical thickness, area and volume form sMRI, three measure form CSF, APOE 4, and MMSE score. Proposed features selection with cross-validation gives the optimal features vector for input to a classifiers. In this proposed method classification performance was quantified by area under the ROC curve.



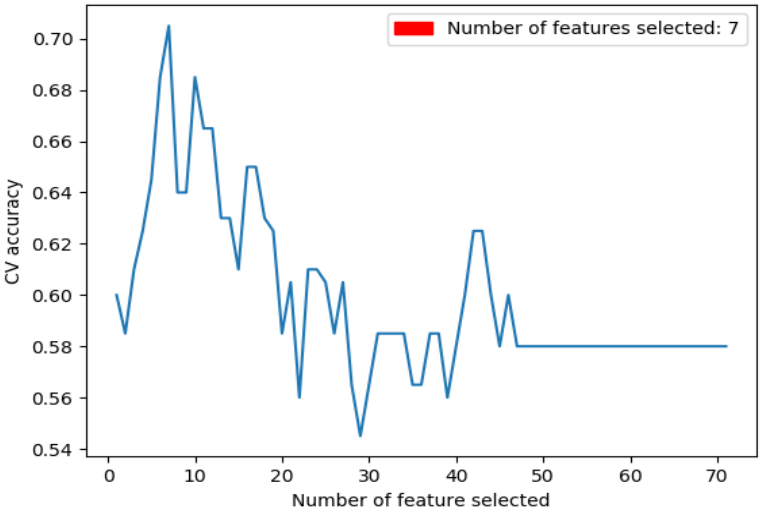
a) AD vs. HC features subset



b) HC vs. MCI features subset



c) HC vs. MCI features subset



d) EMCI vs. LMCI features subset

Figure 6-1. Feature selection for different class group (a-d) using SVM-RFE

Number of features obtained from the Cross-validated SVM-RFE feature selection algorithm are present in the Figure 5-3. The peak value on the x-axis indicate the best number of features for the classification.

Table 4. 10-fold cross-validated classification performance for AD vs. HC

Features Measure	**ELM**				**SVM-RBF**	**SVM-Linear**
	ACC%	**SEN%**	**SPE%**	**AUC**	**ACC%**	**ACC%**
Cortical Thickness	84.72	93.28	83.43	0.83	77.56	80.72
Surface Area	81.34	87.13	90.45	0.87	75.86	76.98
Volume	85.81	94.73	83.78	0.85	78.32	80.13
CSF	86.73	95.33	88.38	0.90	75.40	79.21
APOE+MMSE	88.45	94.15	87.39	0.92	82.03	84.16
Concatenation (All_features_set)	**96.15**	**97.50**	**93.28**	**0.95**	**90.28**	**92.45**

Table 5. 10-fold cross-validated classification performance for AD vs. MCI

Features Measure	**ELM**				**SVM-RBF**	**SVM-linear**
	ACC%	**SEN%**	**SPE%**	**AUC**	**ACC%**	**ACC%**
Cortical Thickness	69.52	84.14	68.02	0.72	63.3	64.15
Surface Area	62.71	68.11	81.09	0.64	60.78	64.37
Volume	65.93	83.24	65.41	0.70	60.54	58.05
CSF	66.89	65.78	80.59	0.71	58.35	60.56
APOE+MMSE	82.31	86.53	85.83	0.80	71.92	73.97
Concatenation (All_features_set)	**85.93**	**93.37**	**87.33**	**0.87**	**80.83**	**77.98**

Table 6. 10-fold cross-validated classification performance for HC vs. MCI

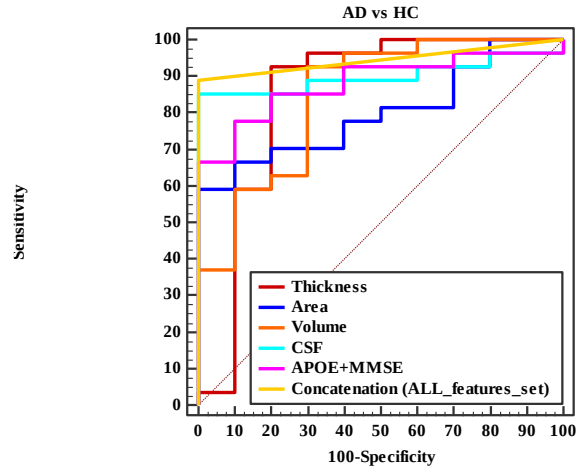
Features Measure	ELM				SVM-RBF	SVM-Linear
	ACC%	SEN%	SPE%	AUC	ACC%	ACC%
Cortical Thickness	76.70	86.15	65.23	0.81	64.17	66.87
Surface Area	73.19	68.5	80.81	0.83	68.74	65.80
Volume	78.91	87.73	73.01	0.85	66.70	71.93
CSF	81.72	75.12	86.32	0.85	72.98	74.90
APOE+MMSE	77.45	94.17	70.89	0.88	69.83	71.30
Concatenation (All_features_set)	**87.81**	**93.07**	**96.10**	**0.94**	**83.70**	**84.70**

Table 7. 10-fold cross-validated classification performance for EMCI vs. LMCI

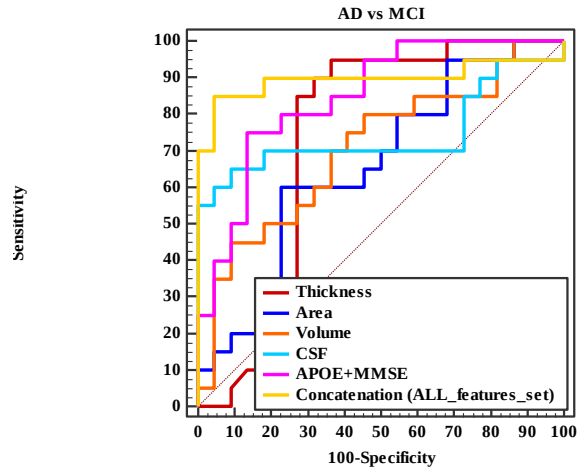
Features Measure	ELM				SVM-RBF	SVM-Linear
	ACC%	SEN%	SPE%	AUC	ACC%	ACC%
Cortical Thickness	64.71	72.50	59.19	0.62	58.31	61.12
Surface Area	62.43	72.85	73.58	0.68	55.03	54.81
Volume	67.95	81.72	85.53	0.72	60.35	61.40
CSF	72.21	80.15	71.12	0.74	63.85	63.25
APOE+MMSE	75.88	89.05	75.28	0.80	64.98	65.57
Concatenation (All_features_set)	**81.73**	**92.3**	**78.73**	**0.88**	**70.93**	**74.95**

To further analyze the effectiveness classification of combining different measures, we calculated the AUCs for all feature concatenation. Figure 5-4 shows that all features (imaging and non- imaging biomarkers i.e. Multi-

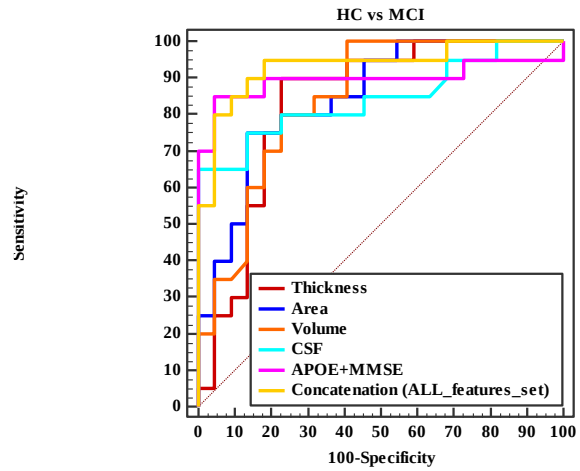
features) combination receiver operating curve for each group of classification.



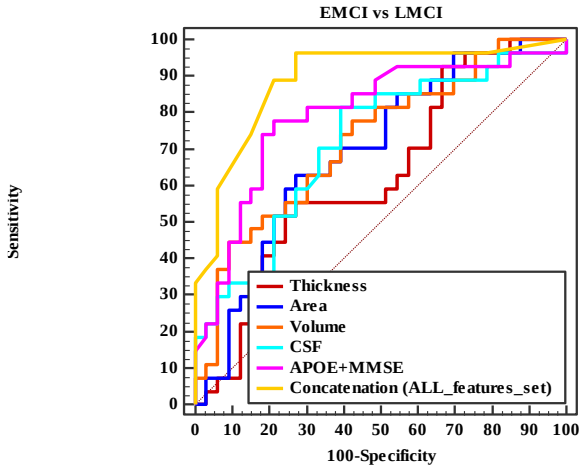
a) ROC curve for AD vs. HC classification



b) ROC curve for AD vs. MCI classification



c) ROC curve for HC vs. MCI classification



d) ROC curve for EMCI vs. LMCI classification

Figure 6-2. ROC curve for different classification group (a-d) on Alzheimer's disease using multimodal biomarkers.

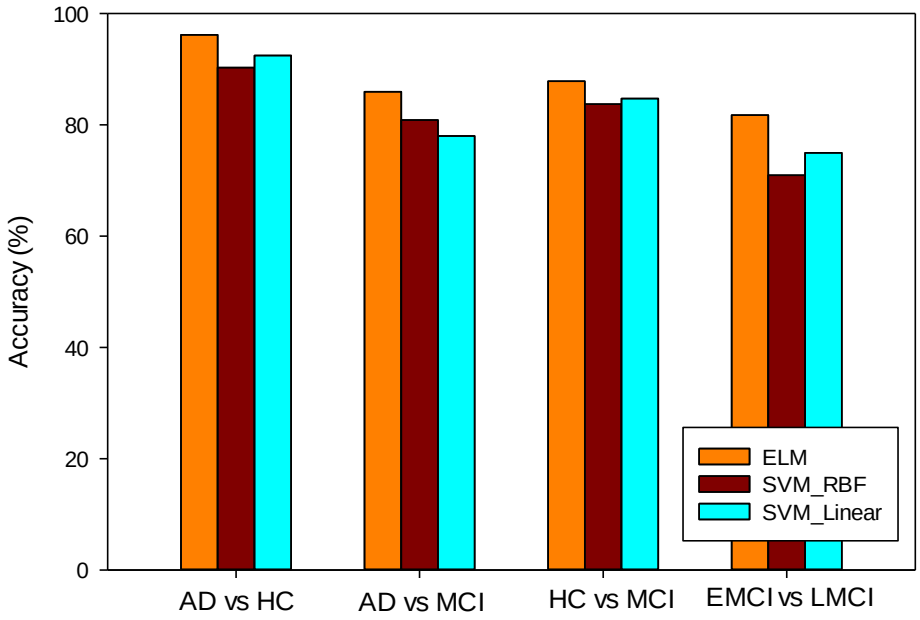


Figure 6-3. Comparison of classifiers performance on all features concatenation

Thus in this study, first performed the statistical analysis and pattern classification to differentiate and identified atrophy pattern for four group (AD, HC, EMCI and LMCI). The individual sMRI were preprocessed using Freesurfer tool. After preprocessing the statistical analysis on sMRI, QDEC was applied and finally, performed the classification task by proposed feature selection and classification method respectively. For the brain atrophy analysis proposed method used sMRI cortical metrics namely three kind of measures (i.e. cortical thickness, surface area, gray matter volume). In case of AD compared with HC, insula, parsopercularis, parahippoccampal and superiortemporal are severely affected cortical thickness, surface area and gray

matter volume. The cortical thickness of the left hemisphere is thinner as compare to that in the left hemisphere. Similarly, for AD-MCI major atrophy are notice on left inferiorparietal, right lateraloccipital. For HC-MCI the super marginal and cuneus show the major atrophy. In case of EMCI-LMCI the superiotemporal, precentral shows the decrease on thickness volume and area. From the difference research we can say that major atrophy of cortical thickness, gray matter volume and surface area mostly appears in the temporal lobe, frontal lobe, cingulate gyrus, occipital lobe and parietal lobe. This phenomenon strongly agree with findings related to the atrophy pattern seen in previous studies [69], [70]. These regions mainly involved in personality expression, motor execution, complex cognitive behavior and decision making [50]. In addition, this research work present the less commonly analysis group EMCI-LMCI. In case of MCI convertible the major atrophy seen on superiotemporal, bankssts, precentral part, inferiorparietal and insula which shows potential to recognize the early progression of the Alzheimer's disease.

At second part, in order to analyze the efficiency of the combination of imaging and non-imaging features, proposed method adapt the individual feature form sMRI, CSF separately to carry out the experiment but in case of genetics and cognitive features proposed method combine them to test the performance and then to compare the accuracy with the accuracy of all features combination. For features selection purpose proposed method used SVM-RFE algorithm and compare the performance of the classifiers on the selected features set. In this experiment SVM-Linear, SVM-RBF and ELM classifier is used for classification. The classification results for the compared methods are presented in Table 4, Table 5, Table 6 and Table 7. As shown in the table, the performance accuracy of feature fusion is noticeably improved as compared

with that of individual features set, and there are distinct degrees of elevation in other indexes, and particularly the specificity index is more noticeable. On the other hand, the performance accuracy based on the non-imaging features is almost the same as imaging features in AD diagnosis but much superior in the MCI diagnosis. For AD-HC: the accuracy obtained by CSF, genetics and cognitive score showing a less increment but in case of MCI-HC, AD-MCI and EMCI-LMCI shows the more increment on accuracy. The experimental result shows the ELM classifier achieves better classification scores compared to SVM classifier for both single modality features set as well as all features concatenation. From the result shown in Table, we can see that suggested method performed better than other methods. For classifying AD and HC, proposed method achieves a classification accuracy of 96.15%, with a sensitivity of 97.50 %, a specificity of 93.28% with AUC value of 0.95%. For classifying MCI and HC, our method achieves a classification accuracy of 87.81%, with a sensitivity of 93.07%, a specificity of 96.10% and AUC value of 0.94%. For classifying AD from MCI, proposed method obtained a classification accuracy of 85.93%, with a sensitivity of 93.37%, a specificity of 87.33%, and AUC value of 0.87%. Similarly for EMCI-LMCI proposed method achieved the outstanding performance as compare to the previous method by combining the multiple features with accuracy of 81.73%, with a sensitivity of 92.50%, a specificity of 78.73% with AUC value of 0.88%.

7. Conclusion

In conclusion, the results demonstrated that the combination of three measure form sMRI, cortical thickness, cortical area, cortical volume, and threes non-imaging measure, CSF, APoE4 and MMSE score improves AD diagnosis, furthermore it is shown that great potential to early identification of the mild cognitive impairment (prodromal stage of the Alzheimer's disease). In this method we proposed the SVM-RFE features selection with ELM classifier for the multiple biomarkers based AD diagnosis which significantly improve the classifier's performance. Moreover the result were shown to be better or satisfactory as compare to the previous literatures, epically for most challenging classification work such as HC versus MCI and EMCI verses LMCI. The added value of combining different anatomical MRI measures should be considered in AD scanning protocols. There is still common practice to only use the specific part or a single measure of whole brain atrophy for AD identification. Obtained results shows that clinical AD diagnosis could benefit from calculating multiple measures from an anatomical MRI scan with others non-imaging biomarkers and incorporate these all in an automated machine learning system. The suggested method in this thesis effectively promotes the diagnosis accuracy of AD and MCI, but this method still has some drawbacks. The next move in future work will include the improvement from the following aspects: firstly, try to optimize parameter obtaining process. Secondly, in order to enhance the effectiveness of the suggested method the dataset can be increased from the following aspects: extending the longitudinal dataset for the better understanding of progression of mild cognitive impairment (MCI); including the multimodal dataset such as PET and function MRI, which can gives the different characteristic information of the Alzheimer's disease.

References

[1] Alzheimer's Association, "2015 Alzheimer's disease facts and figures," *Alzheimers Dement*, vol. 11, no. 3, pp. 332–384, Mar. 2015.

[2] "The influence of Alzheimer disease family history and apolipoprotein E epsilon4 on mesial temporal lobe activation," *J. Neurosci.*, vol. 26, no. 22, pp. 6069–6076, May 2006.

[3] P. M. Thompson and L. G. Apostolova, "Computational anatomical methods as applied to ageing and dementia," *Br J Radiol*, vol. 80 Spec No 2, pp. S78-91, Dec. 2007.

[4] J. L. Whitwell *et al.*, "3D maps from multiple MRI illustrate changing atrophy patterns as subjects progress from mild cognitive impairment to Alzheimer's disease," *Brain*, vol. 130, no. Pt 7, pp. 1777–1786, Jul. 2007.

[5] E. Canu *et al.*, "Microstructural diffusion changes are independent of macrostructural volume loss in moderate to severe Alzheimer's disease," *J. Alzheimers Dis.*, vol. 19, no. 3, pp. 963–976, 2010.

[6] E. M. Reiman *et al.*, "Preclinical evidence of Alzheimer's disease in persons homozygous for the epsilon 4 allele for apolipoprotein E," *N. Engl. J. Med.*, vol. 334, no. 12, pp. 752–758, Mar. 1996.

[7] 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 Dement*, vol. 7, no. 3, pp. 270–279, May 2011.

[8] 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," *Arch Neurol*, vol. 56, no. 3, pp. 303–308, Mar. 1999.

[9] 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 Dement*, vol. 7, no. 3, pp. 280–292, May 2011.

[10] M. Tábuas-Pereira *et al.*, "Prognosis of Early-Onset vs. Late-Onset Mild Cognitive Impairment: Comparison of Conversion Rates and Its Predictors," *Geriatrics (Basel)*, vol. 1, no. 2, Apr. 2016.

[11] J. B. S. Langbaum *et al.*, "Categorical and correlational analyses of baseline fluorodeoxyglucose positron emission tomography images from the Alzheimer's Disease Neuroimaging Initiative (ADNI)," *Neuroimage*, vol. 45, no. 4, pp. 1107–1116, May 2009.

[12] L. Mosconi *et al.*, "FDG-PET changes in brain glucose metabolism from normal cognition to pathologically verified Alzheimer's disease," *Eur. J. Nucl. Med. Mol. Imaging*, vol. 36, no. 5, pp. 811–822, May 2009.

[13] K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, D. Rueckert, and Alzheimer's Disease Neuroimaging Initiative, "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease," *Neuroimage*, vol. 65, pp. 167–175, Jan. 2013.

[14] J. B. Langbaum *et al.*, "Ushering in the study and treatment of preclinical Alzheimer disease," *Nat Rev Neurol*, vol. 9, no. 7, pp. 371–381, Jul. 2013.

[15] H. Hampel, K. Bürger, S. J. Teipel, A. L. W. Bokde, H. Zetterberg, and K. Blennow, "Core candidate neurochemical and imaging biomarkers of Alzheimer's disease," *Alzheimers Dement*, vol. 4, no. 1, pp. 38–48, Jan. 2008.

[16] 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 Dement*, vol. 7, no. 3, pp. 263–269, May 2011.

[17] N. Srivastava and R. Salakhutdinov, "Multimodal Learning with Deep Boltzmann Machines," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 2949–2980, Jan. 2014.

[18] J. B. Colby, J. D. Rudie, J. A. Brown, P. K. Douglas, M. S. Cohen, and Z. Shehzad, "Insights into multimodal imaging classification of ADHD," *Front Syst Neurosci*, vol. 6, p. 59, 2012.

[19] C. DeCarli *et al.*, "The Use of MRI and PET for Clinical Diagnosis of Dementia and Investigation of Cognitive Impairment : A Consensus Report," 2004.

[20] B. Jie, D. Zhang, B. Cheng, and D. Shen, "Manifold Regularized Multitask Feature Learning for Multimodality Disease Classification," *Hum Brain Mapp*, vol. 36, no. 2, pp. 489–507, Feb. 2015.

[21] K.-H. Thung, C.-Y. Wee, P.-T. Yap, D. Shen, and Alzheimer's Disease Neuroimaging Initiative, "Neurodegenerative disease diagnosis using incomplete multi-modality data via matrix shrinkage and completion," *Neuroimage*, vol. 91, pp. 386–400, May 2014.

[22] F. Liu, C.-Y. Wee, H. Chen, and D. Shen, "Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's Disease and mild cognitive impairment identification," *Neuroimage*, vol. 84, pp. 466–475, Jan. 2014.

[23] L. Yuan, Y. Wang, P. M. Thompson, V. A. Narayan, J. Ye, and Alzheimer's Disease Neuroimaging Initiative, "Multi-source feature learning for joint analysis of incomplete multiple heterogeneous neuroimaging data," *Neuroimage*, vol. 61, no. 3, pp. 622–632, Jul. 2012.

[24] D. Zhang, D. Shen, and Alzheimer's Disease Neuroimaging Initiative, "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease," *Neuroimage*, vol. 59, no. 2, pp. 895–907, Jan. 2012.

[25] C. Hinrichs, V. Singh, G. Xu, S. C. Johnson, and Alzheimers Disease Neuroimaging Initiative, "Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population," *Neuroimage*, vol. 55, no. 2, pp. 574–589, Mar. 2011.

[26] O. Kohannim *et al.*, "Boosting power for clinical trials using classifiers based on multiple biomarkers," *Neurobiol. Aging*, vol. 31, no. 8, pp. 1429–1442, Aug. 2010.

[27] P. Vemuri and C. R. Jack, "Role of structural MRI in Alzheimer's disease," *Alzheimers Res Ther*, vol. 2, no. 4, p. 23, Aug. 2010.

[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] O. Ben Ahmed, J. Benois-Pineau, M. Allard, C. Ben Amar, and G. Catheline, "Classification of Alzheimer's Disease Subjects from MRI Using Hippocampal Visual Features," *Multimedia Tools Appl.*, vol. 74, no. 4, pp. 1249–1266, Feb. 2015.

[30] R. Cuingnet *et al.*, "Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database," May 2011.

[31] B. Fischl, “FreeSurfer,” *Neuroimage*, vol. 62, no. 2, pp. 774–781, Aug. 2012.

[32] 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.

[33] “Alzheimer’s Facts and Figures Report | Alzheimer’s Association.”.

[34] A. Wimo, “World Alzheimer Report 2010: The Global Economic Impact of Dementia,” 21-Sep-2010.

[35] P. Tiraboschi, L. A. Hansen, L. J. Thal, and J. Corey-Bloom, “The importance of neuritic plaques and tangles to the development and evolution of AD,” *Neurology*, vol. 62, no. 11, pp. 1984–1989, Jun. 2004.

[36] M. P. Laakso *et al.*, “Hippocampal volumes in Alzheimer’s disease, Parkinson’s disease with and without dementia, and in vascular dementia: An MRI study,” *Neurology*, vol. 46, no. 3, pp. 678–681, Mar. 1996.

[37] “Alzheimer’s Disease Causes & Risk Factors | alz.org.”.

[38] M. W. Weiner *et al.*, “The Alzheimer’s Disease Neuroimaging Initiative: a review of papers published since its inception,” *Alzheimers Dement*, vol. 8, no. 1 Suppl, pp. S1-68, Feb. 2012.

[39] G. L. Wenk, “Neuropathologic changes in Alzheimer’s disease,” *J Clin Psychiatry*, vol. 64 Suppl 9, pp. 7–10, 2003.

[40] C. Bernard *et al.*, “Time course of brain volume changes in the preclinical phase of Alzheimer’s disease,” *Alzheimers Dement*, vol. 10, no. 2, pp. 143-151.e1, Mar. 2014.

[41] “ADNI | Study Design.” .

[42] M. Pawlowski, S. G. Meuth, and T. Duning, “Cerebrospinal Fluid Biomarkers in Alzheimer’s Disease—From Brain Starch to Bench and Bedside,” *Diagnostics*, vol. 7, no. 3, p. 42, Sep. 2017.

[43] V. Vapnik, “Pattern recognition using generalized portrait method,” 1963.

[44] C. Cortes and V. Vapnik, “Support-Vector Networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[45] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “A Training Algorithm for Optimal Margin Classifiers,” in *Proceedings of the Fifth Annual*

Workshop on Computational Learning Theory, New York, NY, USA, 1992, pp. 144–152.

[46] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, “A Practical Guide to Support Vector Classification,” p. 16.

[47] P. Juszczak, D. M. J. Tax, and R. P. W. Duin, “Feature scaling in support vector data description,” p. 7.

[48] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew, “Extreme learning machine: a new learning scheme of feedforward neural networks,” in *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)*, 2004, vol. 2, pp. 985–990 vol.2.

[49] G. Huang, H. Zhou, X. Ding, and R. Zhang, “Extreme Learning Machine for Regression and Multiclass Classification,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 513–529, Apr. 2012.

[50] M. N. I. Qureshi, J. Oh, B. Min, H. J. Jo, and B. Lee, “Multi-modal, Multi-measure, and Multi-class Discrimination of ADHD with Hierarchical Feature Extraction and Extreme Learning Machine Using Structural and Functional Brain MRI,” *Front Hum Neurosci*, vol. 11, p. 157, 2017.

[51] A. Akusok, Y. Miche, J. Karhunen, K. Bjork, R. Nian and A. Lendasse, "Arbitrary Category Classification of Websites Based on Image Content," in IEEE Computational Intelligence Magazine, vol. 10, no. 2, pp. 30-41, May 2015.

[52] Jiuwen Cao and Zhiping Lin, “Extreme Learning Machines on High Dimensional and Large Data Applications: A Survey,” Mathematical Problems in Engineering, vol. 2015, Article ID 103796, 13 pages, 2015.

[53] Y. Chen, E. Yao and A. Basu, "A 128-Channel Extreme Learning Machine-Based Neural Decoder for Brain Machine Interfaces," in IEEE Transactions on Biomedical Circuits and Systems, vol. 10, no. 3, pp. 679-692, June 2016.

[54] S. Spirit, “Receiver Operating Characteristic (ROC) Curves,” *Sem Spirit*.

[55] P. A. Flach, J. Hernández-Orallo, and C. Ferri, “A Coherent Interpretation of AUC as a Measure of Aggregated Classification Performance,” in *ICML*, 2011.

[56] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. (In Proceedings) 14th International Joint Conference on Artificial Intelligence (IJCAI'95), 2:1137-1143,1995.

[57] C. Promteangtrong *et al.*, "Multimodality Imaging Approach in Alzheimer disease. Part I: Structural MRI, Functional MRI, Diffusion Tensor Imaging and Magnetization Transfer Imaging.," *Dement Neuropsychol*, vol. 9, no. 4, pp. 318–329, 2015.

[58] M. Toews, W. Wells, D. L. Collins, and T. Arbel, "Feature-Based Morphometry: Discovering Group-related Anatomical Patterns," *Neuroimage*, vol. 49, no. 3, pp. 2318–2327, Feb. 2010.

[59] M. P. Wattjes, "Structural MRI," *Int Psychogeriatr*, vol. 23 Suppl 2, pp. S13-24, Sep. 2011.

[60] K. A. Johnson, N. C. Fox, R. A. Sperling, and W. E. Klunk, "Brain Imaging in Alzheimer Disease," *Cold Spring Harb Perspect Med*, vol. 2, no. 4, Apr. 2012.

[61] J. Ashburner and K. J. Friston, "Nonlinear spatial normalization using basis functions," p. 13.

[62] J. C. Lambert *et al.*, "Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease," *Nat. Genet.*, vol. 45, no. 12, pp. 1452–1458, Dec. 2013.

[63] A. Rao, Y. Lee, A. Gass, and A. Monsch, "Classification of Alzheimer's Disease from structural MRI using sparse logistic regression with optional spatial regularization," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2011, pp. 4499–4502, 2011.

[64] E. Richard, B. A. Schmand, P. Eikelenboom, and W. A. V. Gool, "MRI and cerebrospinal fluid biomarkers for predicting progression to Alzheimer's disease in patients with mild cognitive impairment: a diagnostic accuracy study," *BMJ Open*, vol. 3, no. 6, p. e002541, Jun. 2013.

[65] T. Tong *et al.*, "Multiple instance learning for classification of dementia in brain MRI," *Med Image Anal*, vol. 18, no. 5, pp. 808–818, Jul. 2014.

[66] A. Khazaee, A. Ebrahimzadeh, and A. Babajani-Feremi, "Identifying patients with Alzheimer's disease using resting-state fMRI and graph theory," *Clin Neurophysiol*, vol. 126, no. 11, pp. 2132–2141, Nov. 2015.

[67] Z. Xiao, Y. Ding, T. Lan, C. Zhang, C. Luo, and Z. Qin, “Brain MR Image Classification for Alzheimer’s Disease Diagnosis Based on Multifeature Fusion,” Computational and Mathematical Methods in Medicine, vol. 2017, Article ID 1952373, 13 pages, 2017.

[68] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene Selection for Cancer Classification using Support Vector Machines,” *Machine Learning*, vol. 46, no. 1, pp. 389–422, Jan. 2002.

[69] A.-T. Du *et al.*, “Different regional patterns of cortical thinning in Alzheimer’s disease and frontotemporal dementia,” *Brain*, vol. 130, no. Pt 4, pp. 1159–1166, Apr. 2007.

[70] B. A. Richards *et al.*, “Patterns of cortical thinning in Alzheimer’s disease and frontotemporal dementia,” *Neurobiol. Aging*, vol. 30, no. 10, pp. 1626–1636, Oct. 2009.