April 2019

Master's Degree Thesis

The comprehensive analysis of Alzheimer's disease using multi-modal features from T1-MRI brain images

Graduate School of Chosun University

Department of Information and Communication Engineering

Yubraj Gupta

The comprehensive analysis of Alzheimer's disease using multi-modal features from T1-MRI brain images

T1-MRI 뇌영상으로 부터 멀티모달 특징들을 이용한 알츠하이머병 포괄적 분석

April 30, 2019

Graduate School of Chosun University

Department of Information and Communication Engineering

Yubraj Gupta

The comprehensive analysis of Alzheimer's disease using multi-modal features from T1-MRI brain images

Advisor: Prof. Goo-Rak Kwon

A thesis submitted in partial fulfillment of the requirements for a master's degree in engineering

August 2019

Graduate School of Chosun University

Department of Information and Communication Engineering

Yubraj Gupta

This is to certify that the master's thesis of

Yubraj Gupta

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**	

August 2019

Graduate School of Chosun University

Table of Contents

List of Figures

List of Tables

Acronyms

AD	Alzheimer's disease
MCI	Mild Cognitive Impairment
MCIs	MCI remains stable within a certain period of time
MCIc	MCI converted to AD within a certain period of time
NC	Normal controls
CAD	Computer aided diagnosis
PET	Position emission tomography
fMRI	Functional magnetic resonance imaging
DTI	Diffusion tensor imaging
RF	Random Forest
SVM	Support vector Machine
ROI	Region of Interest
FS	Freesurfer (v.6.0)
APOE	Apolipoprotein-E
SKF-CV	Stratified k-fold Cross validation
PCA	Principal Component analysis
ANN	Artificial neural network
NACC	National Alzheimer's Coordination Center
ADNI	Alzheimer's Disease Neuroimaging initiative
MMSE	Mini-mental state examination
CDR	Clinical dementia rate
GDS	Global deterioration scale
FAQ	Functional activities questionnaire
NPI-Q	Neuropsychiatric Inventory

Abstract

The comprehensive analysis of Alzheimer's disease using multi-modal features from T1-MRI brain images

Yubraj Gupta
Advisor: Prof. Goo-Rak Kwon
Department of Information and Communication Engineering
Graduate School of Chosun University

Alzheimer's disease (AD) is a common neurodegenerative disease with an often seen prodromal mild cognitive impairment (MCI) phase, where memory loss, behavior issues, and poor self-care are the main complaint. This disease are economically costly for treatment, with a poorly understood about its origin and no curative treatments found till now. Accurate prediction of clinical changes in AD or (in its prodromal stage) including both MCIc (conversion to the AD within some time periods) and MCIs (stable, not converted to AD within some time periods) at future points, is very important for early diagnosis of AD. Moreover, its early prediction can also possibly helps to delay the disease. Therefore, now there is a great interest in the development of new approaches for early prediction of it. Moreover as I know, structural abnormalities of the brain area are a sensitive feature of the disease (visible on MR images), and it is also one of several known biological biomarkers of the AD disease, likewise Apolipoprotein-E (APOE) genotype.

In recent years, several high-dimensional classification techniques have been developed to automatically distinguish among AD, MCIs, MCIc, and normal control (NC) patients based on T1- weighted MR images using single modality of biomarkers. In their experiment, number of subjects were very low, and they have validated their model only by using a single modality of biomarkers, where the obtained result is also not so impressive. Furthermore, recent study suggest that combining more imagining technique into one form can provide more complementary information.

This thesis first presents the information about an AD relevant to our work. Furthermore, it reviews and summarize some of the most relevant machine learning and deep learning research from past few years. Based upon this information, I have built a multimodal technique, which will classify AD with other classification groups, using three different biomarkers (sMRI, APOE, and Cognitive score). For this experiment, I have used both machine learning and its branch deep-learning methods for classification of AD. In addition, machine learning models and its particular technique deep learning might be able to learn important features from high-dimensional dataset like structural MRI, etc.,. Moreover, it enable automatic classification for the AD with other groups. Therefore, I want to apply both method in my thesis, and additionally I also want to do analysis or comparison between these two methods using some statistical tools. In this thesis, I propose to predict future clinical changes of MCI patients by using multi-modal technique.

Here, I have used cognitive score, APOE genotype value with cortical, sub-cortical and white matter features for the classification of each groups using machine learning technique. I have extracted these features using automated

Freesurfer (v.6.0) toolbox. I have also built and validated a 3D deep-learning algorithm for prediction of clinical changes between AD and MCI patients using tensorflow 1.12 version. For this both machine learning and deep learning architecture, I have trained and validated our modal on NACC and ADNI dataset.

Mine main conclusion is that, state of the art methods for the early classification of Alzheimer's disease, using single modality of biomarkers, are not suitable for early prediction of AD or MCIs groups, because of enormous variety of information can be obtained from each biomarkers. Therefore, I proposed a multimodal architecture, which combine three different biomarkers into one form, and helps to classify complex classification groups. Performing stratified 10-fold cross validation with all available database provided by NACC and ADNI organization, our machine learning system achieves area under the receiver operating characteristics (AUC of the ROC) as 96.46% (NACC) and 94.12% (ADNI) for MCIs vs MCIc group. For the same group, deep-learning system achieves 91.62% (NACC), and 93.56% (ADNI) accuracy for MCIs vs MCIc group.

한글요약

T1-MRI 뇌영상으로 부터 멀티모달 특징들을 이용한 알츠하이머병 포괄적 분석

요브라즈 굽타
지도교수: 권구락. 교수
조선대학교대학원
정보통신공학과

알츠하이머 병 (Alzheimer 's disease, AD)은 기억 상실 및 자기 관리가 힘든 경우에 점차 악화되어서 발생할 수 있는 경증인지 손상(Mild cognitive impairment, MCI) 단계에 있는 신경 퇴행성 질환이다. 이 병은 발생 원인이 명확하지 않고, 현재까지 치료 방법이 발견되지 않았으며 경제적으로도 많은 비용이 발생한다. AD 에 대한 임상적 변화에 대한 정확한 예 또는 전기 단계에 포함되는 MCIc(일정 기간 내에 AD로의 변환) 및 MCIs(일정 기간 내에 AD로의 변환되지 않음)에서의 AD의 조기 예측은 매우 중요하다. 조기 예측을 통해서 병을 지연시키는데 도움이 될 수 있다. 이에 따라 조기 예측에 대한 새로운 개발이 높은 관심사가 되고 있다. 또한 뇌 영역에 대한 구조적 이상 발생은 이 병의 민감한

특징이며(MR영상에서 볼 수 있음), Apolipoprotein-E (APOE) 유전자 형 및 몇가지 유전자는 AD질환을 나타내는 생물학적 바이오 마커 중 하나이다.

최근에는 단일 모달리티의 바이오마커를 이용하는 T1- weighted MR 영상을 기반으로 AD, MCIs, MCIc 및 정상 대조군(NC) 환자를 자동으로 구별하기 위한 다양한 고차원 분류 기술이 개발되었다. 기술에 대한 실험과정에서 피험자의 수는 매우 적었으며, 실험에 대한 결과도 인상적이지 못한 단일 양식의 바이오 마커를 사용한 모델의 유효성 확인 정도이다. 최근의 연구는 많은 기술들을 하나의 형태로 결합하여, 보완된 정보를 제공할 수 있다고 제안한다.

본 논문에서는 첫번째로 AD에 관한 정보를 제공한다. 또한, 지난 몇 년간 본 연구과 가장 관련 있는 기계 학습 및 심층 학습 연구를 설명한다. 이 정보를 기반으로, 본 논문에서는 세 가지의 다른 바이오 마커(sMRI, APOE, and Cognitive score)를 사용하여 AD를 다른 그룹과 분류하는 멀티모달 방법을 구축한다. 이 실험을 위해서, AD의 분류를 위한 기계 학습과 심층 학습 방법을 사용한다. 또한, MRI 영상의 데이터 셋에서 기계 학습 모델과 심층 학습의 특정 기술을 통해 중요한 특징을 추출한다. 이를 통해 AD와 다른 그룹에 대한 자동 분류를 가능하게 한다. 따라서 본 논문에서는 두 가지 방법을 모두 적용한다. 또한 통계 프로그램을 사용하여 두 가지 방법에 대한 비교 분석을 실행한다. 본 논문에서는

다중 모드 기법을 사용하여 MCI 환자의 향후 임상 변화를 예측할 것을 제안한다.

기계 학습 기술을 사용하여 각 그룹을 분류하기 위해서, 인지 점수, APOE 유전자형 값, 대뇌 피질, 피질 및 백색질 특징을 이용한다. 자동화된 Freesurfer(v6.0) 프로그램을 사용하여 특징을 추출한다. 또한 Tensorflow 1.12 버전을 사용하여 AD환자와 MCI 환자 간의 임상 변화 예측을 위한 3D 심화 학습 알고리즘을 구축하고 검증한다. 이 기계 학습과 심층 학습 아키텍처를 위하여 NACC 및 ADNI 데이터 셋에 대한 학습 및 검증을 실행한다. 주요 결론은, 단일 마디 형태의 바이오 마커를 사용하는 알츠하이머 병의 조기 분류를 위한 기존 방법은 AD 또는 MCI 군의 조기 예측에 적합하지 않다. 이는 엄청난 다양성의 정보가 각 바이오 마커로부터 얻어 질 수 있기 때문이다. 이에 따라 본 연구에서는 세 가지의 다른 바이오 마커를 하나의 형태로 결합한 멀티모달 아키텍처를 사용하여 복잡한 분류의 그룹을 분석하는 것을 돕는 방법을 제안한다. NACC 및 ANDI에서 제공하는 사용 가능한 데이터베이스를 이용하여 10배수의 교차 유효성 검사를 수행하며, 본 기계 학습 시스템의 영역에서는 ROC상에서 MCIc vs MCIs의 경우 각각 96.46% (NACC)및 94.12% (ADNI)의 결과를 얻었다. 같은 그룹에서, 심층 학습 시스템의 경우 MCIs vs MCIc의 경우 각각 91.62% (NACC)및 93.56% (ADNI)의 정확도를 달성하였다.

Chapter 1: Introduction

1.1 Background and Motivation

Alzheimer's disease (AD), which is one of the most common form of dementia, is initially defined as a clinical-pathologic entity, which was diagnosed definitely at autopsy and in life as possible or probable AD, and for which no cure or successful treatment is available known [1]. In recent years, many effort has been made to find neurological or biological biomarkers for early prediction of AD in a rapid way, allowing early intervention of biomarkers that may prevent or at least delay the onset of AD, as well as its prodromal MCI [2].

Although there is known biomarker for the detection of AD, but diagnosis of AD has traditionally relied mainly on single imaging modality technique or clinical observation or cognitive evaluation. However recent studies indicate that, multimodal technique (combination of more than one imaging modality) performed well compared to single imaging modality methods [3]. Recently, more attention has been shifted to combine more than one biomarkers and thus applying machine learning or deep-learning technique to perform automatic early detection of AD. The National Alzheimer's Coordinating center (NACC), and Alzheimer's Disease Neuroimaging Initiative (ADNI), are one of the world's prominent research projects in the field of AD, and they have contributed significantly more on the research of AD. They also provide reliable clinical data to the researcher for research purposes, which includes a labeled dataset of patients from a different analytic groups consisting of sMRI. More and more researcher has yielded good results on image scans from the

NACC [4], and ADNI [5], [6], dataset using machine and deep learning methods.

Machine learning (ML) technique has successfully applied to clinical decision support systems as well as for diagnostic assistance in the field of medical [7]–[9] and there is also a great interest in leveraging ML technology for use in cardiology, radiology, and oncology, etc. for developing more cost-effective and easy system for supporting physicians, and it is also a suited candidate to make such individual predictions, because it is well equipped to handle high-dimensional data such as those from MRI, or other imaging. Moreover, recently an alternative family of ML methods, knows as deep learning (DL) algorithms, are achieving optimal results in many AD classifications [10].

Recently neuroimaging analysis with the sMRI features for dementia, mainly for the AD and MCI, have shown the promising results [7], [11], [12]. Their report suggests that anatomical feature (e.g., cortical and subcortical thickness, gray matter volumes, white matter volumes) generated from a sMRI imaging can be used to quantify the AD-related brain abnormalities and it will also help to facilitate the early diagnosis of MCI/AD. sMRI seems to be an interesting diagnostic imaging modality [13]–[15], as it is a non-invasive, and it is also a widely used medical image for detection of AD. From a ML point of view, the problem at hand also look interesting, as ANN and deep learning methods are particularly haven proven to be well suitable for dealing with high-dimensional data like that of brain scans.

1.2 Research Question and Methods

1.2.1 Research Question

In this thesis, my aim is to explore the possibility of training a ML, and DL algorithm to perform automatic classification of AD using a sMRI image of subject's brain with the combination of other biomarker's, which helps to increase the classification performance, and it also helps to evaluate the different techniques taken by measuring the generalization capabilities of the trained prototypes by their rate of misclassification on an earlier held out validation part of the dataset.

The work in this thesis mainly focuses on the performance of the multimodal technique compared to single modality method, but it also compared at result obtained by using machine and 3D deep learning technique. Here, I have used, random forest (RF), and softmax classifier for the classification of AD with other's groups. Whereas for deep learning part, I have built a 3D deep learning algorithm from a scratch, which can run on both (GPU and CPU) system with the help of tensorflow (1.12v), for the classification of groups.

Additionally, each of these classification groups are used as foundation for several rounds of test, each of which stance the learning problem in a somewhat different way based upon merging AD or NC with other diagnostic groups. Finally, the model is trained using the above mentioned classifier on all the binary classification groups and later experiments are compared and evaluated.

The research, prototyping and estimation of this thesis is mainly focused on the following question:

RQ 1 *How to combine multiple biomarker's into one form?*

RQ 2 *How classifier can be trained to differentiate between AD groups with other diagnostic groups?*

RQ 3 *Which machine learning methods yield the best results for both (NACC and ADNI) dataset?*

RQ 4 *How does 3D deep learning architecture perform on binary classification problems?*

RQ 5 *Is multi-modal technique is powerful than unimodal technique in terms of performance basis?*

RQ 6 *How does the different outcomes compare with similar kind of research method?*

1.2.2 Methodology

The field of ML is experimental in nature, it often relying on practicality in terms of explain methodology and evaluation of results. Moreover, this study is no exception in this field, and it also takes an exploratory approach to the problems at hand. In this thesis, I am going to deal with practical research and also going to build a modal using training set (ML technique), which later helps to classify test dataset (AD subjects with other groups).

In order to answer the earlier stated research questions, I need to perform experiment on several levels: checking performance of multi and unimodal modality of biomarkers, formulation of the classification problem (i.e.

different schemes for merging classes), using more than one classifier for classification purpose and verifying their result, and method of machine learning and 3D deep learning, and at last comparing result of both ML and 3D DL methods. I also need to combine the different classification groups in order to directly check and compare, which grouping yields the best results, and as well as also to check which experimental group performance is decreased.

As the dataset used in this thesis is a high-dimensional data. Therefore, data dimensional reduction technique is needed before passing them to the classifier, therefore, I have used Principal component analysis (PCA) technique, which will only pass the important feature and block the unwanted ones. This is because the original occurrences in the dataset are a 3D scans with an 8-bit voxel values, which also represent the density. Moreover, a morphological changes in brain (mainly in a few specific area) is a very sensitive feature of Alzheimer's, it would be sensible to assume that the entire area of the brain (and as well as the air, which surrounding the head) is not required for this experiment, I have used BET (Brain extraction tool from FSL toolbox), which deletes non-brain tissue from an image of the whole head, so that I only get the brain region; and I also know that, the relative structural proportions of the cerebral regions which is also known as gray matter, would be all probability suffice as features.

The learning problem constitute in a six different ways, all are a supervised learning classification tasks, which is performed by merging a classes (i.e. diagnostic groups) of different learning variables:

- AD vs NC, AD vs MCIs, AD vs MCIc
- NC vs MCIs, NC vs MCIc
- MCIs vs MCIc

I perform this because these binary classification group sets could reveal or show us where the clearest distinctions or difference are on the Normal-to-Alzheimer-spectrum. Here, I have attempt to solve these classification problem by applying a different ML methods, as these types of training model have different power and weakness. Tree-based approaches (random forests) are one of the best classifier that have achieved comparatively very high performance in the research of automatic diagnostic of AD, and it has also assisted clinicians in several ways. On the other hand, softmax classifier have proven to be very performant in the later years, as DL methods has progressed as a subfield. Deep neural nets have shown unparalleled performance on several problems, and it proved to be a useful in high-dimensional data classification.

The following measures were used to answer the above research questions. For an each research question, I will try to answer by:

RQ 1 *Evaluating the experiment results obtained by using machine and deep learning networks, and also to see which factors or combinations of binary classification shows the greatest performance. Additionally, I also record the insight gained while performing the experiments, which might point to promising new directions.*

RQ1.1 *comparing performance between all experiment results obtained by machine learning with 3D deep learning technique.*

RQ 1.2 evaluating *which method of machine learning will produced a single lowest misclassification score across all the experiments.*

RQ1.3 *comparing the best performance result yielded in this experiment with the presented result in the literature section..*

The methods mentioned below have been used in this thesis:

- Identification of research objectives
- Literature
- Feature extraction methods
- Feature selection process
- Evaluation of machine learning tools, using random forest and softmax classifier using grid search methods
- Development of 3D deep learning algorithm for automatic classification of proposed binary groups
- Training, testing, and validating a different classifier models on the unseen dataset

1.3 Outline

This thesis is subsequently structured as follows:

- Chapter 2 present a brief background about Alzheimer's disease, magnetic resonance imaging, and machine and deep learning technique. After this, I explained about the organizations from where

I have downloaded the dataset for our thesis, and then latter, I present some previous works done on AD classification methods which has achieved best results on classification problems regarding Alzheimer's disease in particularly, using machine and deep learning techniques.

- In chapter 3, I deals with the details of how the research were conducted or designed. The dataset demographic information is presented in this section. In total, I have downloaded 2353 sMRI images from NACC and 158 images from ADNI organization for this thesis. In this section, I have talked about the tools, which I have used in our experiment, like (FSL, Freesurfer), and implementations of machine learning and 3D deep learning techniques.
- Chapter 4 gives the detail explanation about the experiment runs and obtained results. I present experiment results obtained from random forest, softmax of machine learning methods using stratified 10-fold cross-validation technique, and 10-fold 3D deep learning algorithm in this section.
- In Chapter 5, I have discussed about the findings from the earlier chapter, and analyze the result obtained from both machine and 3D deep learning technique for early classification of AD, and I have also answered the relating research questions, which I have stated in the first chapter, and I have also compare our results with some recently published study in AD using ML or DL method.
- Chapter 6 sum up the works and the research findings, and list our main contributions and give an outlook on future work.

Chapter 2: Background

2.1 Alzheimer's Disease

Alzheimer's disease is a most common cause of dementia. It is associated with brain neurodegenerative disease that causes problems such as memory loss and other cognitive abilities, which are serious enough to interfere with daily life [16]. It is one of the most common foundation of dementia, accounting for 60-80% of all cases of the dementia. It is one of the most costly neurodegeneration disease to the society in developed countries. AD affected 26.6 million people globally in 2006, and it is predicted to affect 1 in 85 people globally by 2050 [17]. The World Alzheimer Report of 2010 reported that there will be 65.7 million people suffering from dementia by 2030, and it will increase to 115.4 million by 2050. The same article also suggest that almost 2/3 of people affected by dementia lives in low middle income countries, which are also expected to see major increase in subjects in the coming years, as the regions are developing rapidly [17]. This will be interesting for several reasons, one of which is that dementia suffering patients in these nations largely rely on casual care, a practice that may prove exceedingly difficult to maintain as the elder segment of these inhabitants increase in statistics and disease prevalence climbs. Dementia has an huge societal cost at present, which accounting for the 1.01% of total sum of globally Gross National Products [3]. It is also thought that this problem will become impaired in the coming years, with an projected 85% worldwide growth in societal cost by 2030 [18], [19], supposing that no potential background features (e.g. macroeconomic, dementia occurrence and prevalence of dementia, accessibility and effectivity of treatment) change.

While some of its signs may appear somewhat alike to typical signs of progressive aging, it is important to footnote that AD (and certainly dementia in general) is not a normal part of the aging. As the disease developments over the time, dementia symptoms progressively worsen. At current, there is no curative medicine for AD; the goal is somewhat too slow the development of the disease, recover symptoms, addressing behavioral problems and also to improve quality of life [20]. However, these current treatments can momentarily decelerate the growth of dementia signs if the disease is identified in an early stage. While more effective treatments and eventually prevention or even a treatment is a very important (long term) objective, earlier diagnosis may give the patients comparatively better cure results. The precise cause for AD is still unidentified, except for the rare cases of recognizable genetic differences. Current research specifies, however, that it is associated with neurotic plaques and neurofibrillary tangles in the brain [21]. While Amyloid beta, the protein that neurotic plaques are made up of, is also known to be strongly concerned in the growth of the AD disease, it is still discussed whether or not it is still a causative measure, as many researcher believe it to be. It is, however, usually acknowledged to be a sign of the disease.

The current development is that there is an ever growing ability to see the syndrome and track its growth before signs occur. Recent years have brought some improvements in research, most prominently identification of the biomarkers (particularly brain imaging methods) that allow recognizing and visualization of AD subject-related courses months, years spans, and even I can access decades data before clinical signs were appeared [18]. The biomarkers for an AD is basically based on their behavior, which is typically

measure in an amyloid deposition of a brain (e.g. PET imaging modality, CSF amyloid), and other biomarkers, which typically calculate neurodegeneration volume (e.g. sMRI, FDG-PET, CSF-tau) [22].

The only recognized way to know whether a person was affected by a disease is after a post-mortem brain tissue inspection. However, both brain tissue neurotic plaques and neurofibrillary tangles seems to play a vital role in the growth and in an evolution of AD. Though a great deal of examination has been done on AD, but still there is a need to develop a (non-invasive) analytic tool for this disease.

2.1.1 Mild Cognitive Impairment

MCI is believed to be an early stage of AD or a stage between the expected cognitive decline of normal aging and the more serious decline of dementia, however it is somewhat uncertain whether MCI resembles a different analytic stage, or a prodromal stage of an AD [23]. It can involve problems with memory, language, thinking and judgment that are greater than normal age-related changes. In MCI-subjects, brain pattern changes have already begun for a quite some time, and signs are only just start to appear. However, it doesn't yet prove that these severe are enough to make a major influence on day-to-day work, which would be measured as a dementia. Mild cognitive impairment may increase your risk of later developing dementia caused by AD or other neurological conditions. But some people with mild cognitive impairment never get worse, and a few eventually get better [23].

2.1.2 Age

Age is very clearly associated with the AD. After the age of 65, risk of suffering from an AD doubles in every five years, and it grow almost 50 percent after the age of 85 [24].

2.1.3 Genetics

Genetics may also play an important role in the development of AD. There are two categories of genes that influence Alzheimer's development:

- Risk genes (growth in likelihood)
- Deterministic genes (direct source of disease)

Risk genes increase your likelihood for a disease but do not guarantee you will have it. The strongest risk gene for Alzheimer's is called apolipoprotein E-e (APOE-e4). Studies indicate that this gene may factor into 20-25% of Alzheimer's cases. If you inherit APOE-e4 from one parent, your risk for Alzheimer's increases. Inheriting it from both parent's makes your risks go up even higher, but it is still not a certainty. However, there is a certainty variable about the evolutional development from a MCI to an AD dementia in these peoples. Moreover whereas, for deterministic genes cause a disease or disorder and guarantee you will develop it if you inherit these genes. Research shows that gene variations in three proteins, amyloid precursor protein (APP), presenilin-1 (PS-1) and presenilin-2 (PS-2), will result in AD. The presence of one or two $\epsilon 4$ alleles in the APOE gene is also widely accepted as a growing risk for a late beginning of an AD dementia. On the other hand, $\epsilon 2$ allele decrease this risk.

2.1.4 Neuropathology in Alzheimer's disease

The symptoms that occurs during the development of an AD are caused by the variations in the brain (such as structural abnormalities), which can be used as a sensitive feature of a disease. The development of an AD pathology is stereotyped. In [16] the progression of AD is divided into six phases, based on the supply of neurofibrirally tangles (NFT): the lesions are first placed in the trans-entorhinal cortical area (phase I), after that it is laid out to the entorhinal cortex (phase II), and latter extended to the hippocampus area and also to the limbic lobe (phase III and IV), which involve the association of neocortex (phase V) and finally seen in primary cortex (phase VI). These phases are divided into an entorhinal (I and II), limbic (III and IV) and neocortical phases (V and VI), and these are closely related with the clinical and the cognitive deterioration, reflecting the degeneration of the cortex areas, which are allied with these functions. The hippocampus and medial temporal lobe are thus pretentious by an NFT and neuronal loss can be seen at the earliest stages of a disease. This loss of synapses and neurons, from the AD lead us to clearly visualize the differences in brain soft tissue, as illustrated in Figure 2.1.

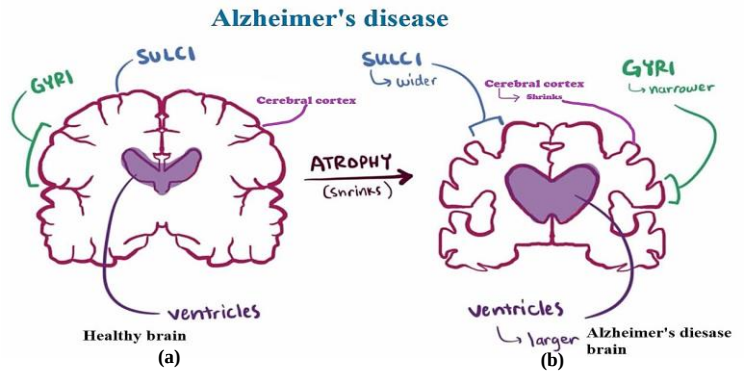


Figure 2.1. Diagram of (a) Healthy and (b) Alzheimer's disease affected brain

The diagnosis of likely AD relies on clinical standards, based on neuropsychological assessment. The important objective behind doing these experiment is to discover biomarkers for early diagnosis of an AD. Moreover, sMRI allows us to visualize brain atrophy which reflects neuronal loss. In recent years, a vast amount of study is performed to extract the diagnostic biomarkers for the AD from sMRI images.

Magnetic resonance imaging helps researcher and clinicians to evaluate the structural changes in the brain associated with AD. There are also other modalities such as the imaging method Pittsburgh Compound B PET (PiB - PET), which clearly shows the shapes of β-amyloid bonds in the brain. This method is more invasive in nature, however, it needs a contrast amount of radioactive sugar, which is engrossed in the brain of the patient.

Laakso et al [25], and Bobinski et al [26], showed that the volumetric of the hippocampal region (as shown in Figure 2.2) is a reliable biomarker for (moderate to severe) phases of the AD disease, where sensitivity and specificity value is above 80%. Moreover, at the stage of MCI, the sensitivity of hippocampus volume is lower. This might be because of initial pathology affects in some specific region of the hippocampus area, NFT and neuronal loss has being dominant in the CA1 and the subiculum region of hippocampus area. The researchers [25], [26] stated that hippocampal atrophy volume is a sensitive feature for AD, but specificity of this region seems to limit its use in the medical practice.

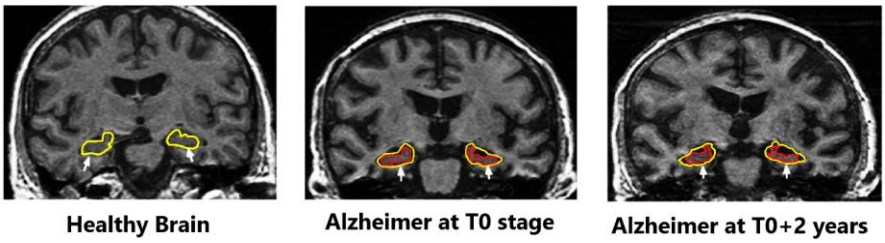


Figure 2.2. Coronal slice of T1 sMRI, showing the atrophy of the hippocampus. The figure shows the R images of normal subject, and the next two panel shows the hippocampal atrophy in an AD at T_0 and T_0 + 2 years.

2.1.5 Alzheimer known Biomarkers

As AD growths over time, its biomarker magnitudes abnormal levels is shown in Figure 2.3. Figure 2.3 shows the biomarkers as indicators of the dementia, the curve shows the changes produced by five studied biomarkers (in chronological order):

1. Amyloid β: imaging noticed in CSF and PET- amyloid imaging.
2. Neurodegeneration noticed by rise of CSF-tau species and synaptic disease, measured via FDG-PET.
3. Brain affected region and neuron loss are measured with sMRI (most noticeable in hippocampus, medial temporal lobe, and caudate nucleus)
4. Memory loss measured by cognitive assessment
5. General cognitive loss measured by cognitive examination

The first three biomarkers of this list can be seen prior to diagnosis of the dementia, while the last two biomarkers are “the classic indicators of dementia diagnosis”. In order to offer suitable therapy, biomarkers are not cooperating with the diagnostic framework, though these are mainly meant for use in research. These include the biomarkers that directly replicate the pathology of

an AD, such as β-amyloid protein and tau; and also these biomarker provide less direct impact or non-specific sign of the disease by racking directories of neuronal injury, which are slightly specific for an AD because of the ROI pattern of abnormalities.

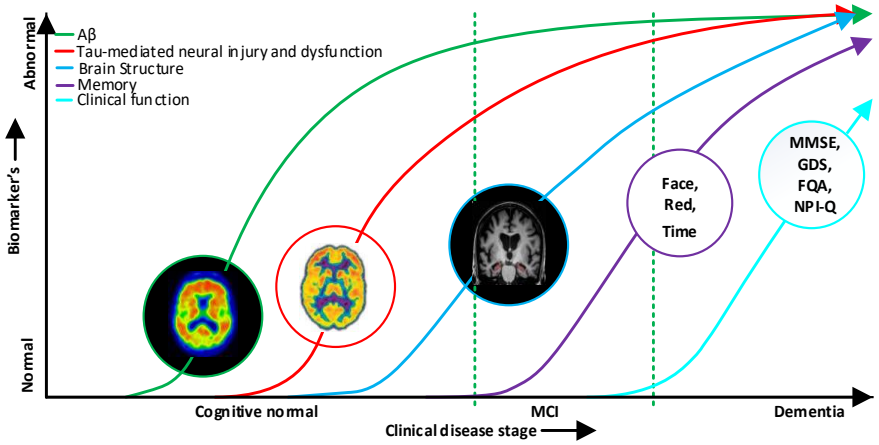


Figure 2.3. Alzheimer's Biomarker's over the course of the disease

When any two biomarkers are found in the same subjects there is a very robust reason to suspect Alzheimer's. And also recently, Jack et al. [18], [27], proposed the A/T/N system, as shown in Table 2.1, in which seven major AD biomarkers are divided into three binary categories based on the behavior of pathophysiology that each subject exhibits.

Table 2.1. A/T/N biomarker grouping

A	T	N
Aggregated Aβ or associated pathological state	Aggregated tau (neurofibrillary tangles) or associated pathological state	Neurodegeneration or neural injury
CSF Aβ42, or Aβ42/Aβ40 ratio	CSF phosphorylated tau	Anatomical MRI
Amyloid PET	Tau PET	FDG-PET, CSF total tau

2.1.6 Dementia due to Alzheimer's Disease

The revised standards for Alzheimer's dementia set out in 2011, which use by both "general healthcare providers without access to neuropsychological testing, advanced imaging, and cerebrospinal fluid measures, and specialized investigators involved in research or in clinical trial studies who would have these tools available" [28]. They report criteria for all cases of dementia including AD dementia. They also suggest to integrate biomarker sign, for probable and possible AD, for use in study settings. Biomarker evidence is "expected to enhance the pathophysiological specificity of the diagnosis of AD dementia", although much of the work lies for the validating of the biomarker of AD dementia.

This diagnostic guideline also propose the terminology for classifying the people with dementia caused by AD:

- Probable AD dementia
- Possible AD dementia
- Probable or possible Alzheimer's dementia with evidence of AD pathophysiological process (proposed for research purposes)

The diagnosis of probable Alzheimer's dementia with the evidence of AD pathophysiological process includes biomarkers for the diagnosis of the AD. These biomarkers include low CSF, $A\beta_{42}$, and positive PET amyloid imaging modality (biomarkers of brain amyloid-β, Aβ); and raised CSF tau (for both total-tau and phosphorylated-tau/p-tau) and "disproportionate atrophy on structural magnetic resonance imaging in medial, basal, lateral temporal lobe, and medial parietal cortex" [29]. Although total-tau and p-tau are treated consistently, p-tau may have extra specificity for the AD than other dementing

syndromes. Biomarker evidence may rise the certainty that the base of the clinical dementia disease is the AD pathophysiological procedure in a persons who meets the core medical criteria for the probable of AD dementia. There is also another category "possible AD dementia with sign of the AD pathophysiological process", which is for a persons who meets clinical conditions for a non- Alzheimer's dementia, but who also meet the neuropathological conditions for Alzheimer's or have biomarker sign of the AD pathophysiological procedure. They write that "Although sophisticated quantitative and objective image analysis approaches do exist, at present, recognized standards for quantitative study of AD imaging tests are lacking. Standard clinical practice in analytic imaging is qualitative in nature. Therefore, quantification of imaging modality of biomarkers must rely on local laboratory specific standards. Quantitative analytic methods are, and will continue to be in development for some time. Therefore, practical use of biomarkers must track best-practice procedures within laboratory-specific circumstances, until standardization has been fully accomplished". At present, the data are insufficient to recommend a scheme that arbitrates among all different biomarker combinations. Further studies are required to prioritize the biomarkers and also to determine their score, and validity in practical and research settings [28].

2.2 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a clinical imaging method, which is used in radiology to form an image of the anatomy and it also physiologically processes the body in both health and disease condition [30]. It uses strong magnetic fields, magnetic field gradients, and radio waves to produce images

of an organs of the body. sMRI is a non-invasive imaging technique that generates three dimensional complete anatomical scans without damaging the radiation. It is used for disease recognition, diagnosis, and treatment observing.

2.2.1 Technology

In most units, MRIs supply a powerful magnets, which generate a strong magnetic field signal that force the proton in the body to aligned with that of field. When a radiofrequency pulse passed through the patient body, the proton start to stimulate, and it spin out of an equilibrium, which is straining beside the pull of the magnetic field. Moreover, when the radiofrequency pulse is turned off, the sMRI sensors are now able to sense the energy released by the realign protons within the magnetic field. Time which take the protons to realign with the magnetic field totally depend on the environment of the chemical molecules. Physicians are capable to tell the alteration between various kinds of tissues based on these pattern of magnetic properties [31]. Whereas, to obtain a sMRI scans, a patient is positioned inside a large magnet and they must remain motionless during the imaging process, so that image is not blur. Contrast agents chemical (containing the element Gadolinium) may be given to a subject intravenously before or during the sMRI process to increase the speed of protons readjust with magnetic field. The faster the protons readjust, the brighter the scans will be.

2.2.2 Imaging

The resultant images have numerous shades of gray color that reflects different thicknesses, based on the fact that these area are with low water, which content fewer hydrogen protons that release signals back to the

radioactive coils. Different machines display this differently (depending on the signal weighting), T1 weighting will produce images of dense bone, air, and other matter containing fewer hydrogen protons, which will be fairly dark, and fat will be light and so on. The voxel pixel number can be from one of 255 shades of gray dependent on signal strength, where 0 implies black and 255 implies white. Figure 2.4 visualize the slices of the brain.

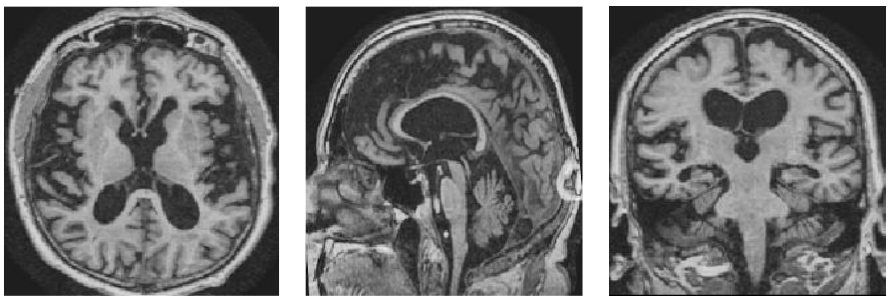


Figure 2.4. Slices of the sMRI brain

2.3 Organizations

This section explain about the organization, which works towards better understanding about the AD, and which also generate and share clinical dataset for the researchers throughout the world, and without their help this thesis could not have been performed.

2.3.1 The Alzheimer's Disease Neuroimaging Initiative

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is an ongoing, longitudinal collaborative research program that aims to "develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of AD". The six-year "ADNI-1" study began in 2004, which includes 400 patients with MCI, 200 patients with early AD and 200 elderly control patients. But later, this research was extended with "ADNI-GO" from 2009 to

2011, adding more 200 subjects that were recognized as having early MCI in order to test biomarkers in an initial phase of disease. The aim of the organizer is to better understand the pathology of AD using recognized biomarkers, to enable earlier diagnosis of AD; which provide clinical test data in order to support new study approaches pertaining to intervention, and to continue and improve the distribution of their database. The initiative effectively develop early stage detection techniques for an Alzheimer's (which includes CSF biomarkers, β-amyloid-42 and total-tau), and standardized techniques for clinical trials (which includes sMRI, FDG-PET and CSF biomarkers), and they also established their database, which contains large amount of brain images.

In a review [7], [32] of the ADNI project, summarize the outcomes of all paper published as of February 2011. Weiner et al. [33] list out the following main accomplishments in the later years:

1) The development of a standardized techniques for clinical, sMRI, PET, and CSF modality of biomarkers is in a multi-center setting.
2) Elucidation of the AD pattern, rate of changes of imaging, and CSF biomarker are some measurement tools for (NC, MCI and AD) subjects. CSF biomarkers are reliable with disease trajectories, as it is predicted by β-amyloid (Aβ) cascade, whereas total-tau mediated neurodegeneration hypotheses for AD patients, whereas brain atrophy and hypo-metabolism level shows the predicted patterns but it also exhibit differing rates of changes depending on region and disease severity.

3) The assessment of alternative approaches of diagnostic classification. Currently, the best classifiers join optimum features from all multiple modalities of biomarkers including (sMRI, FDG-PET, CSF) and clinical examinations.
4) For the development of prominent approaches for the early prediction of an AD. (CSF, Aβ42 and total-tau) and as well as amyloid-PET biomarkers has reflects the earliest stages in the AD pathology, in stable or even non-stable subjects and these biomarkers are the leading candidates for the prediction of an AD in its preclinical phase.
5) The improvement of clinical test efficiency through recognizing the patient's behavior, is most likely to undergo undecided in the future clinical decline and the use of more sensitive result measures to reduce sample sizes.

ADNI organization provide an important step towards the development of improved diagnostic techniques, which possible slow the progression, and later ultimately prevent the AD.

2.3.2 The National Alzheimer's Coordinating Center

In 1999, National Institute on Aging (NIH) establish a National Alzheimer's Coordinating Center (NACC) to facilitate cooperative research, using a data collected from the Alzheimer's Disease Centers (ADCs) whole across the United States (USA). It has developed and it now preserve a large relational dataset of standardized clinical and neuropathological research data. In partnership with the Alzheimer's disease Genetics Consortium (ADGC), the National Centralized Repository for AD and Related Disorders (NCRAD), and the NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS),

NACC provides a valuable resource for both exploratory and explanatory AD research. NACC dataset are freely available for all researchers. The data are contributed by the 39 past and current ADCs supported by the U.S. National Institute on Aging/NIH, where all enrolled subjects undergo a standardized evaluation. The clinic-based population includes subjects with AD and related disorders, as well as cognitively normal subjects and those with MCI.

The NACC database is made up of three main research datasets:

1) The UDS, a longitudinal dataset: From 2005 to the present, ADCs have been contributing data to the Uniform Data Set (UDS), using a prospective, standardized, and longitudinal clinical evaluation of the subjects in the National Institute on Aging's ADC Program. The UDS was expanded with two modules — one to collect detailed clinical information related to frontotemporal lobar degeneration (the FTLD Module, implemented in 2012), and one to collect information on Lewy body disease (the LBD Module, implemented in 2015). Both modules are completed by ADCs on a voluntary basis.
2) The Neuropathology Dataset: The Neuropathology Data Set (NP) contains autopsy data for a subset of both MDS and UDS subjects. Please note that changes in diagnostic criteria and staining methods may limit the available data for certain analyses.
3) Data collected before the UDS, The minimum Dataset: Beginning in 1984 and ending with the 2005 implementation of the UDS, brief, single-record descriptions of ADC subjects were collected retrospectively to form the Minimum Data Set (MDS).* Please see the last table below for more detailed information. “Because of the lack of

detailed, longitudinal, and standardized clinical data in the MDS, the utility of the MDS for research is limited, and combining the clinical data in the MDS with the UDS is generally not recommended".

The dataset includes subject's demographics, health, and physical information, CDR scale score, Geriatric Depression Scale (GDS), and Functional Activities Questionnaire (FQA), etc.

2.4 Machine Learning

Machine learning is characteristically a multidisciplinary field, drawing on the research from a variety of fields, such as artificial intelligence, statistics, philosophy, and neurobiology or it can also be defined as the science of attainment computers to learn and perform task like a humans do, and recover their learning process in autonomous fashion, by supplying them a real-world data with their observations. For an example, a program that studies how to play checkers could recover its performance by its capacity to win with the respect to the class of responsibilities involved with playing checkers through the experience obtained while playing checkers games against itself. A machine learning system generally consist three main parts (as shown in Figure 2.5); training, model, and a training procedure. The learning method can then be explained if the dataset is well trained. Moreover, in order to compute the trained model's actual performance, I have supplied an unseen test data, which is a previously separated dataset. The goal of the learning process is to find out how the training model is performed on unseen test dataset.

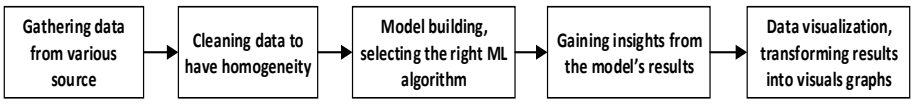


Figure 2.5. Machine learning working pipeline

Machine learning can be divided into three learning paradigms, namely supervised and unsupervised learning, and reinforcement learning. Supervised learning deals with labelled sample data, where inputs are attached with preferred output values. This can be observed as a parallel in the psychological idea of concept learning. Whereas in unsupervised learning, samples are unlabeled and their outcomes are not being of any error to calculate a potential solution. In other hands, Reinforcement learning deals with attempting to reach a specific goal by executing an action in an active environment in order to exploit a reward, without being clearly told if the learner is imminent its goal.

There is a wide variety of machine learning simulations, all of which make different prior assumptions about the likely input-output mappings or data sharing, in supervised and unsupervised learning correspondingly. The models make these different assumptions by the requirement, since the problem they are dealing is ill-posed, and the training data is thus not sufficient for the models to find the right solution by themselves. For an example, a mapping may not occur, or there might not be enough data to rebuild it, or there might be an unavoidable existence of noise that creates a perfect fit model, which is useless in real world. These sets of expectations that the learning procedures makes in order to make learning achievable is called inductive bias. The inductive bias (also called as learning bias) of a learning procedure is the set of expectations that the learner uses to guess the outputs of a given inputs that it has not seen yet. Inductive bias is therefore important because it defines how the learner generalizes outside the observed training samples. The learning process attempts to minimize some amount of error (or loss function), such as

least mean squares or mean squared error. They do this by choosing the best weight that fit the set of training samples, i.e. and, which also helps to minimizing the loss between the observed training instances (and other prior are kept constraints) and the guessed values.

In this section, I will mainly focused on supervised learning techniques like random forest and softmax classifier. Additionally, I will also cover some unsupervised learning techniques for preprocessing purposes, such as PCA.

2.4.1 Deep Learning

Deep learning (DL) is a subfield of machine learning techniques based on learning data representations, as opposed to task specific algorithms. It is an automated method, which deals with high dimensional data (e.g. 3D images) to train a model, and later unseen test dataset is validated to obtained performance of the model. In deep learning process, I don't have to extract features manually, it automatically extract the features from the inputs. DL work on features that are increasingly higher levels through consecutive layers of transformation, essentially structure level of notion of each other, with each layer commerce with a more complex responsibilities, given the inputs from the earlier layer. It also gives deep model, ability to achieve feature learning automatically extract useful features from input. These automatically learned features values are in many cases are better than hand-designed features or from a manually extracted features. Another important theory is about the distributed illustrations, which essentially means that the concepts are represented by the patterns of activity over the several neurons, where each neuron takes portion in the sharing of more than one idea. These concepts are showed in Figure 2. 6. Architectures of this type of algorithm is inspired by a

biological models [34] of the mammal visual cortex, particularly the deceptive processing of information "through stages of transformation and representation", building more complex processing phases upon each other [35]. The key aspect of DL is that these layers of features are not designed by human engineer; they are learned from data using a general-purpose learning procedure.

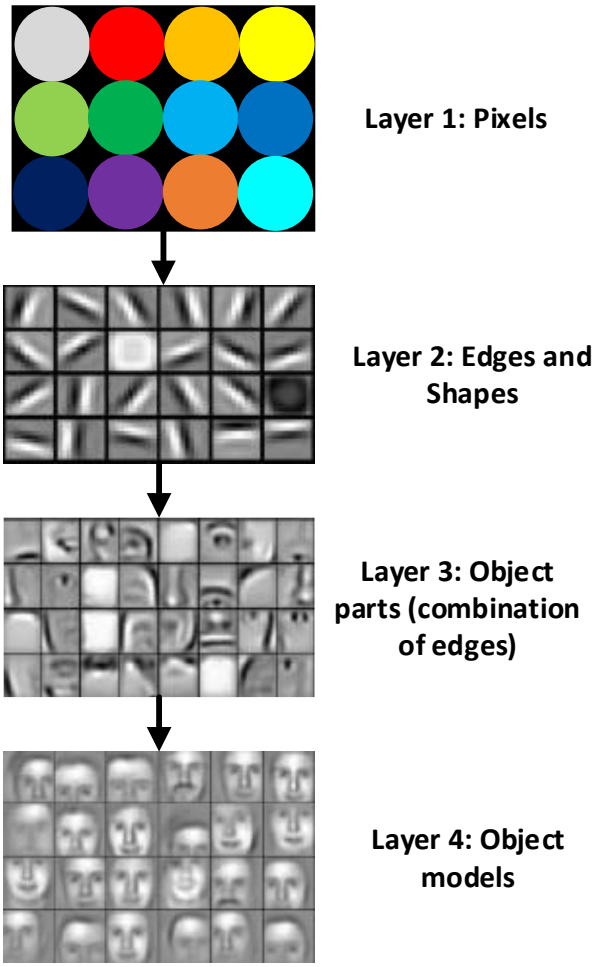


Figure 2.6. Image shows how DNN's learn distributed signs in hierarchical concepts.

Deep learning covers many learning methods such as Deep Belief Nets (DBNs), Deep Auto Encoders (DAE), and Deep Neural Networks (DNNs), are some of the most successful techniques using ANN's idea. DL approaches have shown better performance in later years, particularly with respect to the computer vision problems. As of October 2014, DNN based on GPU architecture, using some form of a dropout regularization technique the researcher get high accuracy on several standardized classification image datasets. DNN has yielded remarkable results on important standardized benchmark datasets, like (MNIST, CIFAR-10, 100, STL-10, ImageNet, etc.). In recent years, DL researchers have produced a methods that helps deep learning techniques to be practically applied, such as sparse initialization, pre-training, fine-tuning, and regularization techniques like dropout are some popular methods [36], [37]. As well as in ML research, advances in hardware (particularly in terms of GPU's) have made these procedure feasible, drastically reducing running periods of training algorithms. As I know, GPU is fit for training the DL systems in a long run for very large datasets. CPU can train a DL model quite slowly. GPU accelerates the training of the model. Hence, GPU is a better choice to train the DL model efficiently and effectively. But to train DL, it required higher capacity of GPU which is not possible for every people to buy. Moreover, In recent years, there are lots of private companies which offer cloud GPU to the researcher for their research using DL, like (Google colab, Google cloud, Amazon AWS, Vast.ai, etc.) are some of the cloud GPU provider. In fact, Ciresan and their team found that, training big and DNN (plain multi-layer perceptron's) using the online backpropagation GPUs are found to be more than forty times faster than training on CPUs in 2010 [38].

2.4.2 Deep learning dominance

Deep learning architectures are essential carried out in order to learn complex functions that can signify high-level abstractions (e.g. computer vision, medical image processing). Deep architectures were, however thought to be more difficult to train. They tended to get trapped in the apparent of local minimum, and these are purely supervised training technique, which might get worse output if deep architectures parameters are not set correctly, according to the 2010 article [39]. The problems could be described by vanishing gradient, and the local optima (which is known to be more serious in deep architectures). Moreover, Because of poor training and generalization mistakes, high error is usually achieved from them, and these deep models were largely unnoticed until an unsupervised pre-training [40] was discovered. Since then, different deep learning algorithm have been proposed and they have successfully applied in many fields, beating state-of-the-arts methods in certain applications.

After that, more methods have appeared using pre-training network [41], many of which are similar to Hinton's original method like a DNN, which is a first pre-trained network, which use unsupervised learning procedure, and then later year a fine-tune is proposed by using a other classical supervised learning techniques (e.g. backpropagation). Although, last few years are seen to decrease, using computer vision for image processing problems, as the models which they used in many problems and it make unnecessary evident by its lack of presence in the state-of-the-art study. A method that has seen to make impact in the later year is a data augmentation [42], [43] of the image datasets, which is basically used for learning, so that I can "inflate" the samples

(i.e. increases the dataset number of training samples). This can be achieved via deformations and the transformation methods. DeepMind published a journal in 2013, specifying their research that deals with a learning DNN's to play the seven Atari 2600 games using the reinforcement learning technique on raw pixel input data. Their model outclassed all previous methods on six games, and they also beat a human skill on three of them. Therefore, these types of model accomplish impressive performance on several types of pattern-analysis problems, recently they have used in a real world tests by major companies in the field of computer technology, such as Siri application developed by Apple and search-by-voice service by Google.

2.4.3 Deep Neural Network architecture

DNN are artificial neural networks with multiple hidden layers [45], [46]. Although "depth" refers the numerous hidden layers of the model, but they are commonly not just deeper, but they are larger in all side of dimensions. This requires more neurons, more layers and more connections. Therefore, it also need more GPU's hardware and computational power, so that these deep model can trained and learned well on high-dimensional samples. After the presence of unsupervised pre-training method, however, now they have quickly become one of the popular machine learning techniques in use. The idea of pre-training network has been an significant development, which addresses the question of how to set initial weights in a neural network [47]. DNN's outputs are mainly based on weight behavior, if the weight value is small than it will take long time to converge and if the weight value is too big, then it raise the likelihood in a local optimum permanently and it converge very fast. The strategy consists of two stages:

- Pre-training: In Pre-trained network one layer is trained at a time like a greedy way method.
- Fine-tuning: Whereas, in this case I train the network with backpropagation method to "fine-tune" the network model.

Different pre-trained model can be used to build a blocks in the pre-training stage, such as a Restricted Boltzmann Machines (RBMs) technique, as Hinton originally use it when he discover this strategy, and he also use various Auto-encoders (AEs) [48]. The act of executing supervised training network after pre-trained network is called "fine-tuning", subsequently it is easier to alter the weights of pre-trained deep networks than to train it with arbitrarily initialized weights. Moreover, because of this method nowadays it has become possible to decrease the complexity to train state-of-the-art method for a computer vision problems, for which previously convolutional neural networks were used.

2.5 Related Work

In this thesis, I primarily focused on behavior of multimodal technique compared to unimodal methods, and I am also going to analyze the result of machine and deep learning architecture. I build the tools needed to test and evaluate different algorithms and approaches for such a classifier. This involves a variety of subtasks, such as collecting a 3D images from the organizer homepage, pre-processing them, manually or automatically extracting features from each images, normalizing the dataset, using dimensionality reduction method which is good for high-dimensional data (because all features are not important for the classification of binary groups), and training a classifier using machine learning techniques or deep learning

techniques. In this chapter, I present the samples of the most interesting and relevant existing work that is related to these tasks. The recent trend for the computer vision difficulties have been to use GPU-based hardware which is implemented with machine learning and as well as with deep convolutional neural network, with a dropout-based regularized, and as well as with performing data augmentation technique on the dataset in order to recover the learning system's performance.

2.5.1 Classification structures for Alzheimer's Disease and for its prodromal phases

Over the recent years, several classification architecture have been used prominently to analyze complex image patterns in neuroimaging samples with a view to classify the AD or MCI subjects with other diagnostic groups. For classification frameworks, feature extraction and classification process are the required components for every machine and deep learning techniques. In this section I will present some previous work done on AD using machine and deep learning methods.

In 2008 Kloppel et al. [49] developed a robust technique that could be generalized through different medical centers, using linear SVM to classify the "grey matter segment of T1-weighted MR scans" and achieved promising AD classification results. In their study, they found relatively lower GM density in the hippocampus of an AD patients, which was a strong sign of hippocampal decline, consistent with previous research [50].

Their model classify up to 96% of pathologically verified Alzheimer's patients properly using a whole brain images. Samples from different centers has achieved comparable good result to the separate examines, allowing SVM

to be trained on sample from one centers, and use it to accurately to differentiate between AD and NC subjects obtained from another centers with different subjects and scanning equipment. Their method also correctly classified 89% of subjects with post-mortem confirmed analysis of the AD or frontotemporal lobar decline to their group, and their method also correctly separated subjects with mild diagnosed probable AD and the age/sex matched controls patients with an 89% of a cases, which was "compatible with published diagnosis rates in the best clinical centers". Later, Morra et al. [51] performed a evaluation of four automated techniques for hippocampal segmentation using different machine learning algorithms in 2010. The methods they compared were "(1) hierarchical Adaboost, (2) support vector machine (SVM) with manual feature selection, (3) hierarchical SVM with automated feature selection (Adasvm), and (4) a publicly available brain segmentation package (Freesurfer)". In their study, they shows that all of their methods were "capable of capturing both disease related effects and correlations between cognition and structure for these well-known, widespread effects".

In 2012 paper, Termenon and Grana et al. [52] used a GM density maps to improve feature vectors for the AD classification, by employing a two-stage classification frameworks, where a relevance vector machine classifier was used in the first stage. The subject that falls into the low confidence interval portion were used as the input samples for the second classifier in the prediction part. SVM, KNN, learning vector quantization, and other relevance vector machine were used as the second-phase classifiers, however, a SVM classifier perform better with the accuracy of 86% for Ad vs NC using GM

voxel value as features compared to mean, standard deviation features. In 2013 Gupta, Ayhan and Maida et al. [53]" used a sparse auto-encoder to learn a set of bases from natural images and then applied convolution to extract features from the Alzheimer's disease Neuroimaging Initiative (ADNI) dataset". They then divide sMRI samples into three groups (AD, MCI, and HC). Their method achieved high diagnostic accuracies, and their method was also very competitive with other methods, despite "being very simple", and without incorporating prior domain-knowledge in the data processing phase. Based on a vast sub-group of ADNI data, Cuingnet et al. [7] offered a comparison of ten sMRI-based feature extraction techniques and their ability to distinguish between clinically relevant subject groups. The ten methods evaluated contain five voxel-based techniques, three methods cantered on cortical thickness and two techniques based on the hippocampus. Optimum sensitivity and specificity values are (81%, 95%) for AD vs HC, (70%, 61%) for S-MCI vs P-MCI and as well as 73%/ 85% for HC vs P-MCI.

In later years, Zhang et al. [54] proposed a multimodal classification approach by employing multiple-kernel SVM based on the biomarkers including sMRI, PET, and cerebrospinal fluid (CSF) to distinguish AD (or MCI) and normal control (NC) subjects. For the binary classification (AD vs NC and MCI vs NC) results, their suggested model get a good accuracy for AD classification whereas for MCI classification they gets an encouraging accuracy. More recently, Cho et al. [55] performed experiment on 72 MCI-C and 131 MCI-NC subjects. Using the incremental learning technique based on spatial frequency, which shows the representation of cortical thickness data. In addition, their proposed method shows better result than the ten-benchmark

methods for MCI-C vs. MCI-NC classification as reported in [7] and get sensitivity as 63% and specificity as 76%. Wolz et al. [56] used four different automated feature extraction techniques (namely hippocampal volume, TBM, cortical thickness, and manifold-based learning) to analyze structural MRI data of 834 ADNI AD, MCI, and Healthy Control (CTL) subjects. The extracted features were used to compare the performance of two classifiers, LDA and SVM, for AD classification and MCI prediction. The best accuracy for AD versus CTL classification was obtained by combining all extracted features and utilizing a LDA classifier; an accuracy of 89% (sensitivity and specificity of 93% and 85%). Similarly, using combined features and the LDA classifier resulted in the highest accuracy of 68% (sensitivity and specificity of 67% and 69%) for classification of MCI-converter and MCI-stable subjects. When different feature types were studied individually, the TBM features showed the best result. The authors also performed feature selection using a stepwise regression method. Moreover, age and gender correction using a linear regression model was applied to remove disease-related effects of age and gender on the classification. On the other hand, Beheshti and Demirel et al. [13] offered a method which reduce the dimensions of GM area in a supervised style. They have used intensity distribution voxels of a GM maps, rather than using strengths of all the voxels of a GM maps, as a features. The optimal hyper number of a bins in the intensity distribution was then selected based on the Fisher criterion of maximization between the AD and HC subjects, and the resulting intensity distribution based features were later used for SVM-based AD classifications. In 2015, Salvatore et.al [57], used MR images as a biomarker for early classification of AD with other groups. They have used PCA function for feature extraction purpose and SVM as a classifier

for classification purpose. They found that hippocampus, entorhinal cortex, basal ganglia, gyrus rectus, precuneus, and cerebellum, are the critical regions known to be strongly involved in the pathophysiological mechanisms of AD. There classifier achieved classification accuracy as 76% for AD vs. CN, 72% for MCIc vs. CN, and 66% for MCIc vs. MCInc using nested 20-fold cross validation technique. In 2018, Basaia et al. [10] built and validated a 3D deep learning algorithm predicting the individuals diagnosis of AD and MCI who will change to AD (c-MCI) based on a single cross-sectional brain structural MRI scan. Convolutional neural networks (CNNs) were applied on 3D T1-weighted images from ADNI homepage. CNN performance was tested in distinguishing AD, c-MCI and s-MCI. High levels of accuracy were achieved in all the classifications, with the highest rates achieved in the AD vs HC classification tests using both the ADNI dataset only (99%) and the combined ADNI + non-ADNI dataset (98%). CNNs discriminated c-MCI from s-MCI patients with an accuracy up to 75% and no difference between ADNI and non-ADNI images.

Here, I have presented some previously published journal paper related to AD using machine learning and deep learning methods. In next section, I am going to define our proposed model using above methods and ideas.

Chapter 3: Materials and Proposed methods

3.1 Overview

This chapter presents the workflow of the experiments to answer research questions stated in Chapter 1, and also to find out how a classifier can be trained to differentiate AD with other diagnostic groups. In this thesis, I have proposed a multimodal method using machine learning technique and a 3D deep learning method to differentiate AD with other diagnostic groups.

In order to answer the research queries, I need to follow these methods:

i) *Compare different kinds of Multimodal fusion technique*: In this thesis, I have used three kind of features (volumetric volumes, APOE values, and Cognitive score) for classification of AD. Therefore, I need to find out the correct way to combine multi-features into one form, so that I can pass them to the classifier for classification process.

ii) *Perform machine learning using several approaches*: Random forest as well as SVM have produced promising results in earlier research. Moreover, the deep learning by neural networks and CNN have proven to achieve very accurate result on computer vision problems in recent years, and they were also appeared to be well suitable for this type of ML problem.

iii) *Train several binary classifiers as well*: In this thesis, I have made six binary classification groups using AD, NC, MCIs, and MCIc dataset, to find out the best separated groups which shows how well our proposed method is performed on this classification problem.

iv) *Compare our results with the several state-of-the-arts methods results*: Here, I have also compared our obtained result with previously published result on similar research. To find out, is our proposed model performance is good or not?

The problem I am dealing during this experiment was one of the supervised learning classification, as the dataset had labeled samples (i.e. a diagnostic group allied with each sMRI brain images). Because of the dataset features, the curse of dimensionality was an issue, so to minimize this issue I have used PCA dimensionality reduction technique.

3.2 NACC and ADNI Standardized sMRI Dataset

The NACC dataset contain 2353 three-dimensional T1-weighted MR images of patients, which were obtained using field strength of 1.5T. I take 498 patients with AD (238 females, 260 males, age ± SD = 74±10 years, MMSE = 19.22±6.68, GDS = 2.26±2.34), 218 patients with MCIs (114 female, 104 male, age ± SD = 73±9 years, MMSE = 27.44±2.67, GDS = 2.56±3.05), 440 patients with MCIc (228 females, 212 males, age ± SD = 76±9 years, MMSE = 26.41±2.59, GDS = 2.35±2.63), and remaining 1197 patients with NC (383 female, 814 male, age ± SD = 66±12 years, MMSE = 29.04±1.27, GDS = 1.03±1.61). Table 3.1 shows the demographic information about NACC dataset. From table 1, I can say that AD group consist low value of MMSE compared to other groups. Likewise NACC dataset, I have also used ADNI dataset. I obtain ADNI dataset from their homepage. In total 158 three dimensional T1-weighted images were downloaded. From 158, 38 patients belong to AD (16 female, 22 male, age ± SD = 77±7 years, MMSE =

21.21±4.15, GDS = 1.76±1.56), 46 patients belong to MCIs (17 female, 29 male, age ± SD = 77±8 years, MMSE = 26.19±2.79, GDS = 1.63±1.52), 36 patients belong to MCIc (10 female, 26 male, age ± SD = 75±6 years, MMSE = 26.91±2.46, GDS = 1.62±1.88), and remaining 38 patients belong to NC groups (13 female, 25 male, age ± SD = 77±5 years, MMSE = 29.05±1.23, GDS = 0.86±1.12). Table 3.2 shows the demographic information about the ADNI dataset.

Table 3.1. Demographics information about NACC dataset. (Mean/Standard deviation) values

Groups	AD	MCIs	MCIc	NC
SEX(F/M)	238/260	114/104	228/212	383/814
AGE	73.33/9.43	73.12/8.92	75.52/8.62	65.98/11.91
EDU	14.44/3.58	14.39/4.08	14.91/3.45	15.89/2.96
MMSE	19.22/6.86	27.44/2.67	26.41/2.59	29.04/1.27
GDS	2.56/3.34	2.26/2.34	2.35/2.63	1.03/1.61
NPI-Q	1.45/0.67	1.68/0.85	1.58/0.81	1.78/0.926

Table 3.2. Demographics information about ADNI dataset. (Mean/Standard deviation) values

Group	AD	MCIs	MCIc	NC
SEX(F/M)	16/22	17/29	10/26	13/25
AGE	77.15/6.93	76.75/7.7	74.76/5.68	76.60/5.043
Subject Weight	73.90/13.18	73.15/13.25	78.44/14.64	74.43/14.08
MMSE	21.21/4.15	26.19/2.79	26.91/2.46	29.05/1.23
GDS	1.76/1.56	1.63/1.52	1.62/1.58	0.86/1.12
FAQ	17.42/6.92	7.15/5.18	3.91/4.06	0.31/1.02
NPI-Q	5.68/4.66	2.82/3.66	2.4/2.3	0.92/1.54

These all sMRI scans were already reviewed for quality and Gradient inhomogeneity correction (gradwarp), B1 non-uniformity correction, and N3 bias field processing (to reduce residual intensity non-uniformity) were applied [33], [58]. Moreover, a phantom based scaling process were associated

with each scans, and later masks were created by the MR Core method, where a pre-processing steps are already included in "Intracranial Space". These dataset also comprises corresponding metadata information for each brain image, which also includes demographics information such as gender info, subject weight at the initial stage, age and diagnostic group. Data used in the preparation of this article were acquired from the NACC and ADNI homepage. For up-to-date info, see www.alz.washington.edu, and www.adni-info.org.

3.3 Selected Features

In this thesis, I have used three kind of features, as shown in Figure 3.1:

(1) Apolipoprotein-E (APOE)

(2) Cognitive scores

(3) Volumetric features: Cortical, subcortical, and White matter

3.3.1 Apolipoprotein-E (APOE)

Apolipoprotein-E (ApoE) [59] is a class of proteins that involved in the digestion of fats in the body. It is also important in AD. The gene, APOE, is mapped of a 19 chromosome in a group with an apolipoprotein C1 and apolipoprotein C2. APOE gene consist four exons and three introns, in total there are 3597 base pairs. The APOE-4 variant is mainly known to be genetic risk factor for the occasional AD in a variety of traditional groups. However, the E4 variant does not correlate with risk in every population. Although 40–65% of the AD subjects have got at least one part of ε4 allele, APOE-4 is not consider as a disease, at least on a third subjects with AD, APOE-4 are found to be negative and in some cases APOE-4 homozygotes never grow the disease. Moreover, those subjects which have got two ε4 alleles have 20 times more

risk of converting into AD. There is also an evidence that APOE2 allele may serve a defensive role in the AD. Thus, the genotype most at risk for AD and at an earlier age is (APOE 4, 4). Using genotype (APOE 3, 3) as a benchmark (with the persons who have this genotype regarded as having a risk level of 1.0), individuals with both (APOE 4, 4) genotype have a chances of 14.9 ratio for developing into AD. People with (APOE 3, 4) genotype face a ratio of 3.2, and individual with (APOE 2, 4), have a ratio of 2.6. Individuals with (APOE 2, 3) have a ratio of 0.6. Individuals with (APOE 2, 2) also have an odds ratio of 0.6. While APOE-4 has found to be greatly increase the chances that a people will develop in AD, a 2002 study determined that, individuals with any two combination of APOE alleles, including high serum total fat and also a high blood pressure in average life are free of any risk factors, which jointly can closely triple the risk that the people will later develop the AD.

For this thesis, APOE values are obtained from NACC and ADNI homepage.

3.3.2 Cognitive Score

The cognitive score [60], [61] is an aggregate measure of the current cognitive strength of different cognitive skills. The cognitive score provides a quick way to discover where you currently stand cognitively and help you assess your progress over time. Cognitive ability examinations assess the abilities involved in thinking (e.g., memory, perception, reasoning, verbal and mathematical skill, and also a problem solving). Such assessments pose questions intended to estimate candidate's potential to use psychological processes to answer work-related questions or to obtain new job knowledge. Here, all subject's has gone through Mini-Mental State Examination test (MMSE), Global Deterioration Scale (GDS), Functional Activities

Questionnaire (FQA), and Neuropsychiatric Inventory Questionnaire (NPI-Q). I obtain these score from the NACC, and ADNI homepage.

3.3.3 Volumetric volumes

Volumetric feature refers to the feature measurements of the volume of selected brain regions and is carried out by summing all voxels within the traced region of interest (ROI). In this study, using Freesurfer toolbox, I have extracted cortical, subcortical and white matter region from each subject brain. In total 130 volumetric attributes are extracted from cortical, subcortical and white matter volume of the brain.

3.3.3.1 Cortical dementia

The cortex of the brain (the word cortical refers to the cortex) is the part most people are familiar with, at least when it comes to appearance. The characteristic twists and turns of the outer layers play an important role in processing information and in functions such as language and memory. Cortical dementia [5], [62] is typically associated with the brain's gray matter. When these outer layers are affected, which is the case with Alzheimer's, frontotemporal dementia, Binswanger's disease and Creutzfeldt-Jakob disease, there are problems with memory, the inability to find the right words, and in understanding what others are saying (aphasia).

3.3.3.2 Subcortical dementia

As the term suggests, these are dementias believed to initially affect structures below the cortex (sub means below) and are more associated with the brain's white matter. Huntington's disease, Parkinson's dementia, and AIDS dementia complex are three examples of conditions classified as

subcortical dementia. It is more common to see changes in personality and a slowing down of thought processes in subcortical dementias. Language and memory functions often appear largely unaffected in the earlier stages of these dementias. In most kinds of dementia there is widespread degeneration in the cerebral cortex – such as the plaques and tangles which are the hallmark of AD. In other kinds of dementia, there is targeted damage to regions lying under the cortex, giving rise to the category known as "subcortical dementias" [62].

3.3.3.3 White matter dementia

White matter dementia (WMD) [34], [63] is a syndrome introduced in 1988 to highlight the potential of cerebral white matter disorders to produce cognitive loss of sufficient severity to qualify as dementia. White matter has a legitimate position in the study of dementia. The neuropathology of white matter disorders is typically diffuse or widespread, thus disrupting many networks simultaneously and producing a multi-domain syndrome that merits the term dementia. Recently, the neuroanatomy of a white matter has seen as the most noticeable form in origin of the concept which is called as “connectome.” This term refers to all networks among the coarsely 100 billion neurons inside the brain, which only not includes synaptic links but also the linking of GM regions by WM tracts. Thus these both are the micro connectivity of GM and the macro connectivity of WM, which are contribute to the connectome study, in essence of providing a structural design of the human brain. Under the sponsorships of the publicly funded Human Connectome Project, a large-scale neuroimaging modality studies with based on diffusion tensor imaging (DTI) are ongoing to map the extraordinary complexity of WM connectivity in the human brains. This method promises

to accompaniment the study of the cortical function with PET and fMRI images by descriptive structural networks between the cortical regions and the distributed neural networks sub portion of a cognition. Cortical, Subcortical, and white matter region of the brain are extracted from the brain using automated Freesurfer toolbox.

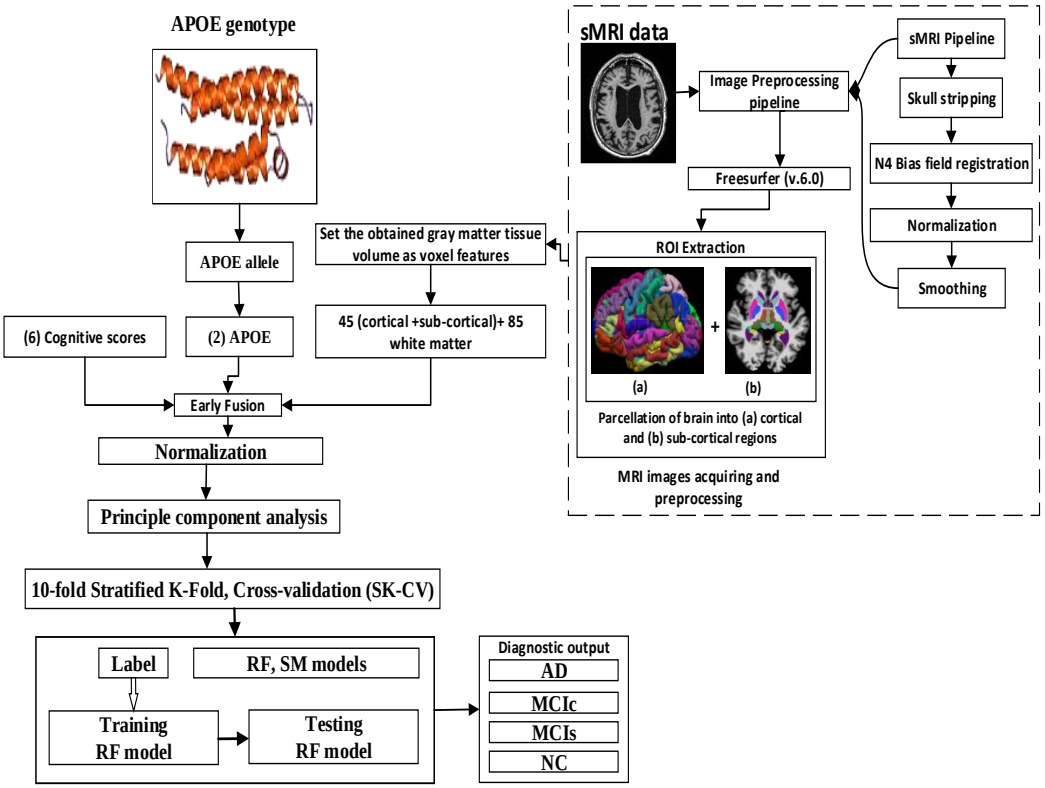


Figure 3.1. Block diagram of proposed machine learning methods

3.4 Feature extraction using Freesurfer (v.6.0)

In this experiment, the features were the cortical thickness, subcortical volume segmentation, and White matter (WM) at each vertex of the cortical, subcortical surface and WM as shown in Figure 3.2. The Freesurfer 6.0

software package was used, which is a fully automatic pipeline for volumetric segmentation and cortical reconstruction. Subcortical volumetric and cortical thickness measures have been widely used for classification purposes. It is known that cortical thickness is a direct index of atrophy and is therefore a potentially powerful candidate in the diagnosis of AD. For processing, the original MRI data (as shown in Figure 3.1), were first subjected into a series of pre-processing steps. Motion correction, T1-weighted image averaging, registration of the volume in the Talairach space, and skull stripping with a deformable template model were performed. In the next step, EM registration, which is a linear volumetric registration, neck removal, CA labeling, intensity normalization, and white matter segmentation were performed. Table 3.3 shows the individual sMRI biomarkers.

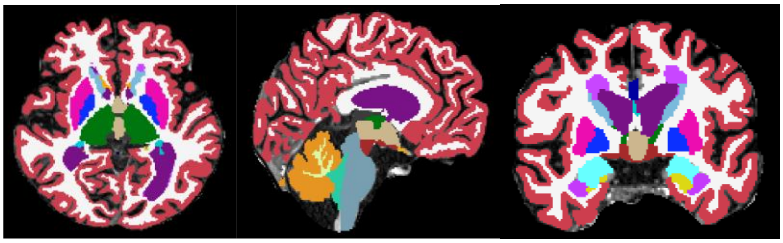


Figure 3.2. Image segmented into Cortical, subcortical, and white matter region

The pial surface was formed for individual hemispheres using training as well as the shape of the corpus callosum, the pons in the Talairach space, and the intensity gradients from the white matter. In the fourth step, the spherical mapping, spherical and contralateral registration, and the cortical parcellation mapping were performed and statistics were obtained. The precise matching of a morphologically homologous cortical location for each patient was obtained by plotting the atlas-based cortical surface on a sphere aligning the

Table 3.3. Overview of extracted individual features from sMRI images using Freesurfer 6.0 toolbox

sMRI biomarkers	Selected ROI region	Toolbox
Cortical Parcellation	Frontal lobe	Freesurfer (v6.0)
	Parietal lobe	
	Temporal lobe	
	Occipital lobe	
	Cingulate lobe	
Subcortical volumetric features	Amygdala	Freesurfer (v6.0)
	Lateral	
	Choroid-plexus	
	Thalamus	
	Caudate-nucleus	
	Hippocampus	
	Putamen	
	Whole brain, etc.	
White matter features	Entorhinal	Freesurfer (v6.0)
	Fusiform	
	Lingual	
	Parahippocampal	
	Precuneus	
	Insula, etc.	

cortical patterns, and later I get the segmented output as shown in Figure 3.2. In the present study, the Freesurfer command was used to automatically produce the cortical thickness, subcortical volume, and white matter volume of a brain segmentation features, as shown in figure below.

3.5 Feature selection

After the feature extraction stage, normalization process is followed, were all features are normalized to zero mean and unit variance to reduce data redundancy and to improve data integrity between each feature, as shown in Figure 3.1. That is, given the data matrix X, where rows signify subjects and columns signify features, the normalized matrix X_{norm} with elements x (i, j) is given by,

$$X_{norm} = \frac{x_{(i,j)} - mean(X_j)}{std(X_j)}, \quad (1)$$

where X_j is the j^{th} column of the matrix (X).

3.5.1 Dimensionality Reduction

It is a type of learning process, where I take higher-dimensional sample or features, like images in matrix form, and later I transform it into a lower-dimensional space using some techniques. Using this method I can reduce any dimensional features into 2D or 1D plane. Moreover, in our case I will use this technique to reduce features of the multimodal process, which I have extracted using Freesurfer toolbox. In our case, I have used Principal component analysis (PCA) [64] as a dimensionality reduction technique.

3.5.1.1 Principal Component Analysis (PCA)

Principal component analysis (PCA) [64] is a statistical process that use an orthogonal transformation method to convert a set of annotations of which are possibly a correlated variables into a set of linearly uncorrelated variables, which is called as principal components. The coefficient gained from a Freesurfer toolbox, has enlarges the dimensionality of the feature space that makes the classifier job more complex. Moreover, it also leads to use excessive computational power and huge memory storage. As a consequence, it is important to lower the dimension of the feature set and get important features to increase the classification outcomes. PCA is applied to find a linear low dimensional reduction space in the dataset. In this instance, the variance of the built data is well preserved. The main notion behind applying PCA is to reduce the dimensionality of the sample features, which results in more

passable and accurate classification. Later, obtain PC are fed to the classifier for the classification purpose. Moreover, these input matrix only possesses PCs. PCA is an unsupervised learning method and a very powerful and reliable tool for data analysis. As explained above, once the specific pattern in the data is found, then they can be compressed into lower dimensions as mentioned above. Here, the number of components was determined by maintaining the variance greater than 99%.

3.6 Classification

I have used two different widely used classifier methods to evaluate the classification accuracy based on the multimodal and unimodal features.

3.6.1 Random Forest (RF)

Random forest [65] algorithm is a supervised method which basically uses ensemble learning technique for classification that operates by building a multitude of decision trees at training mode and outputting the class that is in the mode of classes of the individual trees. The RF algorithm is typically grown using the methodology of classification or regression tree (CART), in which a binary split recursively partition the tree into homogenous or near-homogenous terminal nodes. A good binary split pushes dataset from a parent tree-node to its two daughter tree-nodes, so that in daughter nodes the ensuing homogeneity is improved from the parent node. RF is a group of hundreds to thousands of trees, where each tree-node is grown by using a bootstrap sample from the original data. RF trees are different from the CART as they were developed non-deterministically by using a two-stage randomization method. In addition to the randomization method which is introduced by growing the tree using a bootstrap sample from the original data, the second layer of

randomization is introduced at the node level when growing the tree. Rather splitting a tree node using all variables, RF selects each node of each tree, as a random subset of variables, and only those single variables are used as contestants to find the best split for the node. The purpose of this two-step randomization method is to de-correlate tree-nodes, so that the forest ensemble will have a low variance, and a bagging phenomenon. For our study, I have used in-built RF classifier from Scikit learn 0.19.2 version python library package.

3.6.2 Softmax (SM)

Softmax regression [66] is a generalized form of logistic regression that can be used in multi-class classification problems where the classes are mutually exclusive. The earlier linear regression model produces continuous or unbounded y values. To proceed from arbitrary values $y \epsilon \mathbb{R}^c$ to normalized probability values, $p \in \mathbb{R}^c$ is estimated for a single instance, and exponentiation and normalization are used as follows:

$$p_i = \frac{\exp y_i}{\sum_{c=1}^{c} \exp y_c} \tag{2}$$

where $i, c \epsilon \{1, \ldots, C\}$ range over the classes, p_i refers to the probabilities, and y_i, y_c refer to the value of a single instance. This is called the softmax function, which takes the vector of arbitrary real-valued scores in y and transforms it into a vector of values ranging between 0 and 1 that later sum to 1. A model that converts the un-normalized values at the end of a linear regression to normalized probabilities for classification is called softmax classifier. Here, the softmax layer takes the learned representation p_i and interprets it to the output class. A probability score p_i is also assigned to the output class. The

softmax classifier understands the scores inside the output vector y as the un-normalized log likelihoods for each specific class and then replaces the hinge losses with the cross entropy loss that is represented in the form of,

$$L_i = -\log(p_i) \tag{3}$$

where L_i is the loss of cross entropy of the network. Exponentiation of these quantities yields the (un-normalized) likelihoods, and their division performs the normalization, so that their probabilities sum to 1. In probabilistic terms, the negative log probability of the accurate class is minimized, which can be regarded as performing maximum likelihood estimation (MLE). A fine feature of this observation is that the regularization term R (W) can now be interpreted as the full loss function as approaching from a prior Gaussian value over weight matrix W, where I are executing the Maximum a posteriori (MAP) estimation instead of MLE.

3.7 Proposed methods

3.7.1 Procedure

The procedure can be defined in the following way;

- At first step, data conversion is applied
- After that, resizing of images to the standard template image
- Skull stripping using BET toolbox
- Feature extraction using Freesurfer toolbox
- Dimensionality reduction using PCA
- Merging of several groups (i.e. diagnostic groups)
- Performance measure or compute the output in specific format

After this, a specific program scripts is run for a specific groups to gain classification result. Here, in our case I have applied Machine and deep learning technique for the classification of groups.

3.7.2 Machine learning proposed architecture

As a complete (NACC, and ADNI) dataset is comparatively large, and the samples (i.e. sMRI images) are also in a very high-dimensional stage. Moreover, some modifications steps were needed in order to make the learning practical possible. In ML process, the extracted features value are in high-dimensional space, so to train a high dimensional data, a large computational power is needed, and the gained output performance is not impressive, because it will take long time to converge the data, for this reason, I will use some dimensionality reduction technique to transform high-dimensional feature value into low-dimensional feature value. Therefore, in this experiment, I have applied Principal Component analysis (PCA) as a dimensional reduction process. Here, all classification problems were performed using Ubuntu 16.04 LTS, running python 3.6, and using Scikit learn, (v.0.19.2) library. In this thesis, there were four groups (AD, MCIc, MCIs, and NC) of data, and three different types of features (sMRI imaging modalities (from where I have extracted cortical, subcortical, and white matter volumes for each subjects), APOE genotypes as genetic features, and a cognitive score). Thus, I validated our proposed method using six different types of classification problem, all are binary class problem. At first, I extracted the featured from each and every sMRI using Freesurfer v.6.0 automated toolbox. In total, I obtain 130 features for a single sMRI image, two features from the APOE genotype data, and six features from cognitive score. Moreover, in total there were 138 features for

each subjects. Later, I have applied a dimensionality reduction method using PCA, which select the effective features from all 138 features and send these selected features to the classifier for classification purpose. Figure 3.1 shows the block diagram for the machine learning methods. In this experiment, I have used a random forest and softmax as a classifier. In order to attain unbiased estimates of performance, the set of participants were randomly split into two groups in a 70:30 ratio as training and testing sets, respectively. In the training set, the cross-validation process was applied to obtain optimal hyperparameter values for the criterion, Max_depth, Max_features. The optimal hyperparameter values were find by using a grid-search and ten-fold stratified cross-validation (10-SKF-CV) technique on the training set. For each approach, the obtained optimized value of the hyperparameter was then utilized to train the two classifier using the training dataset, and later the performance of the classifier were evaluated on the remaining 30% of data in the testing dataset. In this mode, I have achieved unbiased estimations of the performance for each classification problem. In our experiment, it should be noted that a number of subjects were not the same in each group. Therefore, only calculating accuracy does not enable comparison of the performances of the different classification experiments. Thus, I considered five measures. For each group, I calculated the accuracy, sensitivity, specificity, precision, and F1-score performance measure values. I have also calculated the area under curve and receiver operating characteristics (AUC of ROC).

3.7.3 3D Deep learning architecture

Before passing 3D MRI images to the 3D deep learning architecture, I have applied some preprocessing steps as mentioned below:

a) Skull Stripping using FSL [67]
b) Normalization of 3D images into MNI standard template image [68]
c) N4-bias field correction using ANT's toolbox [69]

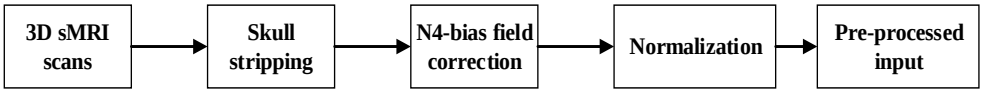


Figure 3.3. 3D Brain MRI Preprocessing module

3.7.3.1 Brain Extraction-Remove the skull from a 3D image

The first step in many MRI analysis sequences is the removal of extra-meningeal tissues from the MRI volume of the whole head. There are many software which can perform skull stripping for MR images, like BET, Brain suite, SPM12, etc. In our case I have used Brain extraction tool (BET) which is also known as FSL. The brain extraction tool (BET) is used to eliminate the skull from a neuroimaging, leaving only the area engaged by actual brain tissue. Moreover, it segments these by utilizing the dark space in the middle of the skull and brain, capture by the CSF region. This toolbox comes from an external program called FSL's toolkit (for more see the FSL homepage) [67]. Fraction intensity threshold value is kept at 0.5 for all images. Figure 3.4 shows the bet output.

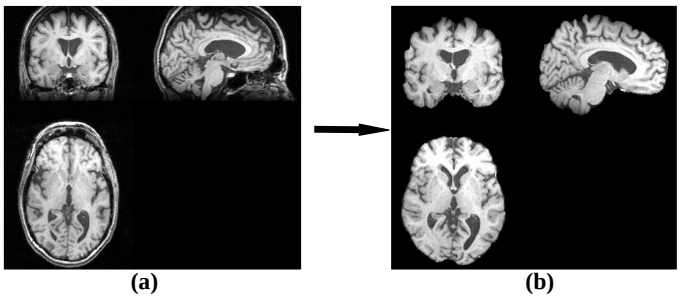


Figure 3.4. BET preprocessed skull stripped images (a) Original image (b) Skull stripped image

3.7.3.2 Normalization of 3D images into MNI standard template image

3D T1-weighted images from both datasets were normalized to the MNI standard space template image using Statistical Parametric mapping (SPM12; MNI space using Statistical Parametric Mapping (SPM12; www.fil.ion.ucl.ac.uk) [68] and the Diffeomorphic Anatomical Registration Exponentiated Lie Algebra (DARTEL) registration method. Briefly, (i) T1-weighted images were segmented to produce GM, white matter (WM) and cerebrospinal fluid (CSF) tissue probability maps in the Montreal Neurological Institute (MNI) space as shown in figure 3.5; (ii) the segmentation parameters obtained from the step (i) were imported in DARTEL; (iii) the rigidly aligned version of the images previously segmented (i) was generated; (iv) the DARTEL template was created and the obtained flow fields were applied to the modulated 3D T1-weighted images of single subjects (generated by the segmentation step) to warp them to the common DARTEL space and then modulated using the Jacobian determinants.

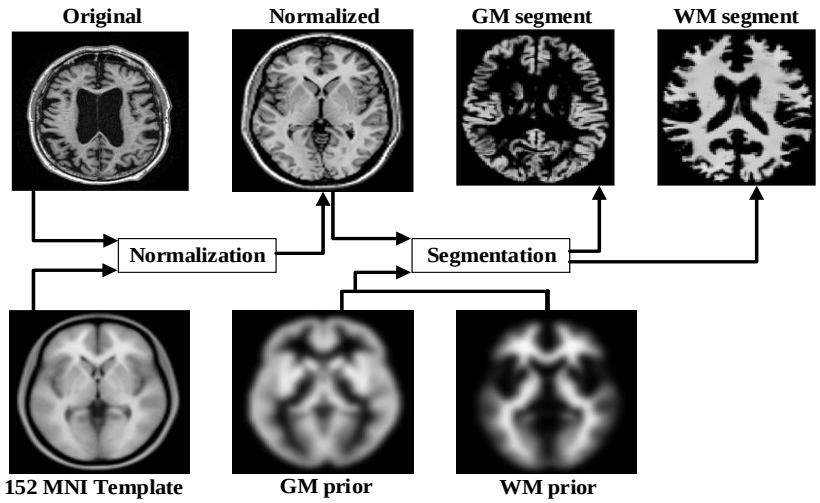


Figure 3.5. GM segmentation using SPM12 toolbox

Since the DARTEL process warps to a common space that is smaller than the MNI space, I performed an additional transformation as follows: (v) the modulated 3D T1weighted images from DARTEL were normalized to the MNI template using an affine transformation estimated from the DARTEL GM template and the a priori GM probability map without resampling.

3.7.3.3 N4 Bias field correction (ANT's toolbox)

N4 is a variant of the popular N3 (non-parametric non-uniform normalization) retrospective bias correction algorithm. Based on the assumption that the corruption of the low frequency bias field can be modeled as a convolution of the intensity histogram by a gaussian, the bias algorithmic protocol is to integrate between deconvolution the intensity histogram by a gaussian, remapping the intensities, and then spatially smoothing this result by a B-spline modelling of the bias field itself. More about N4 bias field correction can be found at http://stnava.github.io/ANTs/ [69].

Therefore, All of the image scans (which had a varying resolution) were first resized into 145x121x145 (the lowest resolution in the dataset) using SPM12 toolbox as mentioned above [68]. After following the above mentioned image preprocessing methods (as shown in figure 3.3), the images were of 145x121x145 matrix size and were loaded onto 64 bit window 10 professional. The machine has 10 core Intel Xeon gold 6128 CPU 3.40 and 3.39-ghz processor, 64 GB of DDR4 SDRAM, and a 2 quadro p4000 8 GB Nividia graphics card with CUDA 8.0 and CuDNN 7.0 (Nvidia). 2D VGG16 convolutional neural network architecture was study from [70], and later I modified it to 3D VGG16 model.

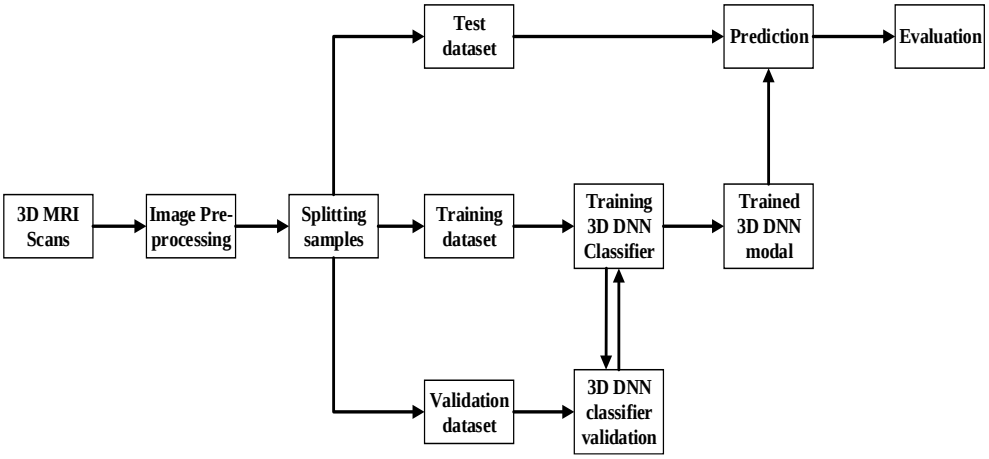


Figure 3.6. Block diagram of the proposed Alzheimer's disease diagnosis framework using 3D DNN.

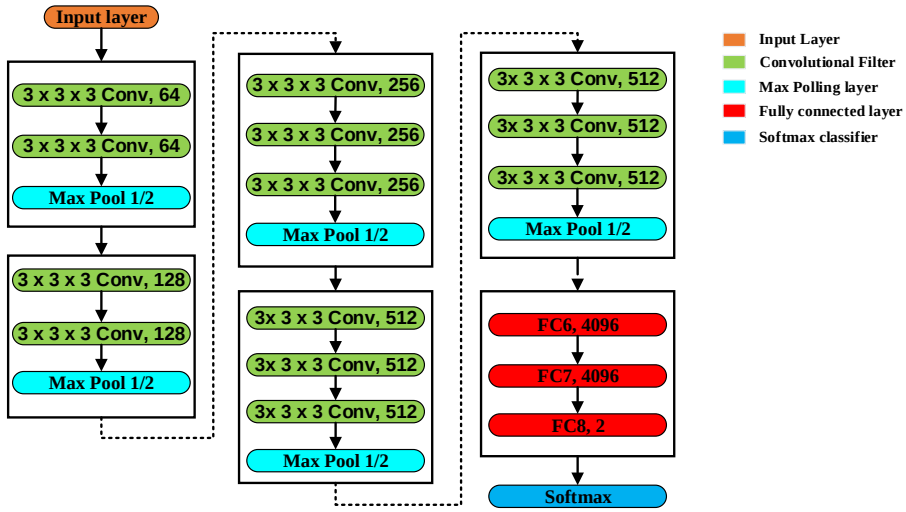


Figure 3.7. 3D deep neural network, VGG16, used in this study. VGG16 network stack 5 VGG modules where each module consists of pooling layers and convolutional filters with rectified linear units as an activation function.

Figure 3.7 shows the 3D (VGG16) deep learning architecture and Figure 3.6 shows the proposed block diagram for the classification of AD using 3D DNN. The network was pretrained on ImageNet, an everyday image dataset

containing 14 million images of 1000 classes, before being fine-tuned using 90% of the (NACC and ADNI) dataset. Here, the dataset was divided into three part with the ratio of 70:20:10, whereas, 70% data is used for training and 20% used for validation and remaining 10% is used for testing purpose. The input of the model is three-dimensional images of 145x121x145 size. Three fully connected layers of size 4096, 4096, and 2 are added to the final concatenation layer. The model is pretrained on ImageNet dataset and further fine-tuned with a batch size of 4 and learning rate of 0.0003 was used for model training. Here, I have used batch normalization to improve the training. The trained algorithm was tested by the accuracy on the held-out NACC and ADNI dataset, and on the independent test dataset. Tensorflow 1.12 (2018; Google) was used for designing 3D DNN and loading pretrained weights. All programs were run in Python 3.6.8 with anaconda installed.

3.8 Statistical Analysis

3.8.1 Area under curve (AUC) analysis

The AUC is a fundamental graph in the evaluation of diagnostic tests and is also frequently used in biomedical research to evaluate classification problem performance and prediction models for decision support, prognosis, and diagnosis. ROC analysis examines the accuracy of a proposed model to separate positive and negative cases or distinguish AD patients from different groups. It is particularly useful in assessing predictive models since it records the trade-off between specificity and sensitivity over that range. In an ROC curve, the true positive rate (known as the sensitivity) is arranged as a function of the false positive rate (known as the specificity) for different cut-off values of parameters. Each point's level of the ROC curve characterizes a

sensitivity/specificity pair, which corresponds to a specific decision threshold. The AUC is a performance measure that shows us how well a factor can differentiate between two binary diagnostic groups (diseased/normal). A result with perfect discrimination has a 100% sensitivity, 100% specificity ROC curve. Moreover, the closer the AUC of ROC curve to the upper left angle, the higher the total accuracy of the test as suggested by [71]. The AUC is commonly used to visualize the performance of binary classes, producing a classifier with two possible output classes. Accuracy is measured using the AUC. Here, an AUC of one signifies a perfect score, and an area of 0.5 represents a meaningless test.

The AUC plot provides two parameters:

i) True positive rate (TPR): the TPR is a performance measure of the whole positive part of a dataset.
ii) False positive rate (FPR): the FPR is a performance measure of the whole negative part of a dataset.

3.8.2 Statistical analysis using Cohen kappa

Cohen's kappa statistic value for each classification problem was calculated. This function computes Cohen's kappa score, which expresses the level of agreement between two annotators or the level of agreement between two different groups in a binary classification problem defined as,

$$k = (p_o - p_e)/(1 - p_e) \quad (4)$$

where, p_o is the empirical probability of agreement on the label assigned to any sample (the observed agreement ratio), and, p_e is the expected agreement when both annotators assign labels randomly. Here, p_e was estimated using a

per-annotator empirical prior over the class labels. The kappa statistic value is always between -1 and 1. The maximum value means complete agreement between two groups, zero or lower value means a low probability of agreement.

3.9 Performance evaluation

The performance of a binary classifier can be understood using a confusion matrix, as shown in Table 3.4. The performance of the system was assessed using the random forest and softmax classifiers for each precise test including binary classification tasks.

Table 3.4. Confusion matrix

True Class	**Predicted Class**	
	G2	**G2**
G1	TP	FN
G1	FP	TN

The diagonal elements of the matrix show the number of correct predictions made by the classifier. The elements can be further divided into true positive (TP) and true negative (TN), thus indicating the correctly identified controls. Similarly, the number of incorrectly classified subjects may be characterized by false negative (FN) and false positive (FP). Accuracy measures the number of examples that were correctly labeled by the classifier, that is,

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

However, for a dataset with unstable class distribution, calculating only the accuracy may result in a misleading estimation of the performance. Therefore, four additional performance metrics should be calculated, namely, specificity, sensitivity, precision, and F1-score. They are defined as follows:

$$Sen = \frac{TP}{TP+FN} \quad (6)$$

$$Spe = \frac{TN}{TN+FP} \quad (7)$$

$$PPV = \frac{TP}{TP+FP} \quad (8)$$

$$F1 - score = \frac{2TP}{2TP+FP+FN} \quad (9)$$

Sensitivity (6) indicates the accuracy of the prediction group, and specificity (7) indicates the accuracy of the prediction of the absence group. Sensitivity measures the success rate for a particular class, i.e., within a class, the percentage of correctly determined subjects (by the classification algorithm) to be in the class. Specificity provides a measure for those not in the class, i.e., it is the percentage of those not in the class that were found not to be in the class. Precision (8) (which is also termed as positive predictive value (PPV)) is the fraction of relevant occurrences among the retrieved occurrences, and F1-score (9) (which is also called F-score or F-measure) is a quantity related to a test's accuracy.

Chapter 4: Results

4.1 Introduction

This section presents the results of the research work which is described in the previous chapters. All experiments were run with a 70%/20%/10% dataset split (training/testing/validation). It was done to ensure that, I have achieved unbiased estimates of the performance for each classification problem, and also to ensure that, there was sufficient test/validation sample available with the dataset differences with fewer samples. Performance of the three selected features (volumetric volume, APOE, and cognitive score) was validated and tested using random forest and softmax classifier and as well as, 3D DNN was validated and tested on patients and controls, with six binary classifications:

1) AD vs HC
2) AD vs MCIs
3) AD vs MCIc
4) NC vs MCIs
5) NC vs MCIc
6) MCIs vs MCIc

4.1.1 Machine learning Results

4.1.1.1 Random forest result

I didn't face the problem of overfitting while using this RF classifier, but the main problem of using RF is that a large number of trees can make the algorithm to slow and ineffective for real-time predictions. Its computational time is high compared to softmax classifier.

4.1.1.2 Softmax result

I didn't face overfitting problem, either in this softmax classifier. While operating this classifier, I kept the following parameter as a constant for all classification groups.

- Epochs: 1000
- Learning rate: 0.001
- Batch size: 16
- Activation function: Adam
- Cross-validation: stratified k-fold (10 fold)

For all classification groups, softmax classifier provide the best performance compared to random forest classifier.

Table 4.1. Classification result of AD vs. NC

NACC	Classifier	Sub (AD/NC)	AUC	ACC	SEN	SPEC	PRE	F1	Cohen Kappa
FS	RF	498/1197	84.44	90.37	**95.41**	89	70.27	80.93	0.7469
	Softmax		91.8	89.39	72.97	98.77	97.12	83.33	0.7578
APOE	RF		86.55	87.93	79.47	91.71	81.08	80.26	0.6918
	Softmax		90.25	89.16	80.64	93.11	84.44	82.5	0.6945
COG	RF		85.79	83.84	72.84	88.75	74.32	73.57	0.6819
	Softmax		93.44	92.84	86.92	95.53	89.86	88.37	0.7089
Combined (ALL)	RF		90.37	93.12	91.85	93.58	83.78	87.63	0.8288
	Softmax		**97.39**	**96.86**	91.25	**99.43**	**98.62**	**94.51**	**0.9256**
ADNI	**Classifier**	**Sub (AD/NC)**	**AUC**	**ACC**	**SEN**	**SPEC**	**PRE**	**F1**	**Cohen Kappa**
FS	RF	38/38	88.2	87.61	78.57	88.89	91.67	84.62	0.7489
	Softmax		92.31	91.3	100	83.33	84.62	91.62	0.8271
APOE	RF		88.56	86.95	84.61	90	91.66	87.99	0.6956
	Softmax		93.78	93.66	96.68	100	100	92.85	0.8566
COG	RF		89.67	89.55	90.9	86.33	**93.55**	89.75	0.6895
	Softmax		90.55	91.24	91.66	81.81	89.75	90.55	0.7866
Combined (ALL)	RF		90.91	91.3	93.66	**100**	100	92.31	0.8862
	Softmax		**96.15**	**95.65**	**100**	90.91	92.31	**96**	**0.9125**

Table 4.2. Classification result of AD vs. MCIs

NACC	Classifier	Sub (AD/MCIs)	AUC	ACC	SEN	SPEC	PRE	F1	Cohen Kappa
FS	RF	498/218	89.91	86.51	84.07	100	100	91.34	0.6583
	Softmax		92.95	90.23	100	75.86	85.91	92.42	0.7891
APOE	RF		91.47	91.55	95.97	84.84	93.46	94.7	0.7456
	Softmax		90.27	87.9	95.55	86.57	90.84	95.55	0.7055
COG	RF		91.64	92.33	89.57	**100**	**100**	88.62	0.7358
	Softmax		94.3	92.09	100	79.52	88.59	93.95	0.8266
Combined (ALL)	RF		94.77	92.56	**100**	89.49	89.54	94.48	0.8316
	Softmax		**95.54**	**93.02**	99.31	85.92	93.46	**96.3**	**0.8805**
ADNI	**Classifier**	**Sub (AD/MCIs)**	**AUC**	**ACC**	**SEN**	**SPEC**	**PRE**	**F1**	**Cohen Kappa**
FS	RF	38/46	85.74	84.62	79.55	100	100	85.71	0.6977
	Softmax		92.12	92.31	93.33	90.91	93.33	93.33	0.8424
APOE	RF		82.14	80.77	70.59	85.55	86.74	82.76	0.6294
	Softmax		90.91	91.85	88.24	100	100	93.75	0.8385
COG	RF		84.72	88.46	85.71	89.47	75.65	80.58	0.7194
	Softmax		88.58	88.46	87.5	90	83.33	92.32	0.7607
Combined (ALL)	RF		90.97	92.31	87.5	94.44	87.5	88.92	0.8194
	Softmax		**95.45**	**96.15**	**93.72**	**100**	**100**	**96.77**	**0.9202**

Table 4.3. Classification result of AD vs. MCIc

NACC	Classifier	Sub (AD/MCIc)	AUC	ACC	SEN	SPEC	PRE	F1	Cohen Kappa
FS	RF	498/440	89.1	89.36	88.68	90.24	92.16	90.38	0.7849
	Softmax		92.95	92.55	97.83	87.5	88.24	92.78	0.8513
APOE	RF		87.8	86.17	83.13	90.9	93.46	88	0.7155
	Softmax		91.42	90.71	95.62	86.01	86.75	90.97	0.8352
COG	RF		88.98	89.36	87.73	91.6	93.46	90.51	0.7844
	Softmax		95.24	95.04	97.93	91.97	92.81	95.3	0.9005
Combined (ALL)	RF		91.81	91.84	92.76	90.77	92.16	92.46	0.8358
	Softmax		**96.67**	**96.45**	**97.97**	**92.81**	**93.46**	**96.35**	**0.9217**
ADNI	**Classifier**	**Sub (AD/MCIc)**	**AUC**	**ACC**	**SEN**	**SPEC**	**PRE**	**F1**	**Cohen Kappa**
FS	RF	38/36	92.31	90.12	100	83.33	84.62	91.67	0.8042
	Softmax		90.87	91.3	92.86	88.89	92.86	92.86	0.8175
APOE	RF		83.62	82.26	90	79.62	85.45	78.26	0.6535
	Softmax		92.86	91.3	100	81.82	85.71	92.31	0.8244
COG	RF		88.46	86.96	100	76.92	89.75	86.96	0.7067
	Softmax		89.29	89.96	100	92.5	87.57	88	0.7418
Combined (ALL)	RF		93.33	91.3	100	89.68	86.67	92.86	0.8189
	Softmax		**96.67**	**95.65**	**100**	**90**	**92.86**	**96.3**	**0.9105**

Table 4.4. Classification result of NC vs. MCIs

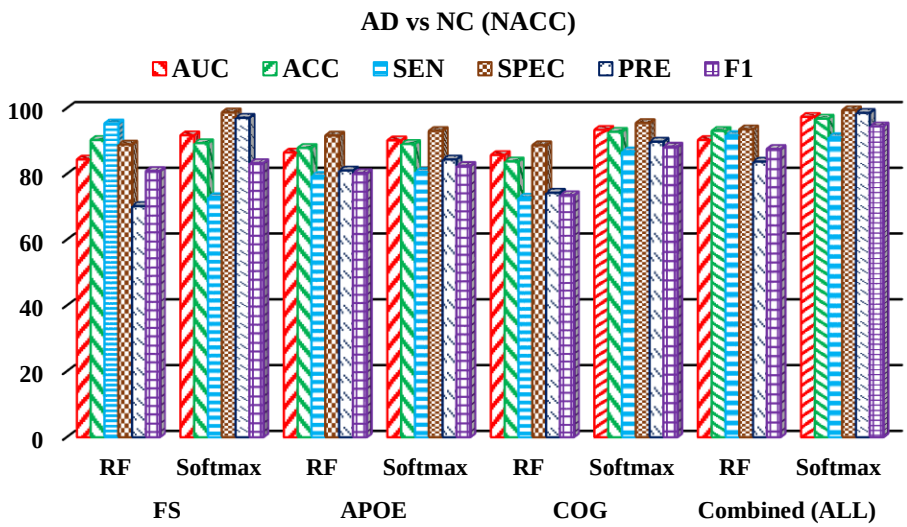
NACC	Classifier	Sub (NC/MCIs)	AUC	ACC	SEN	SPEC	PRE	F1	Cohen Kappa
FS	RF	1197/218	91.56	89.88	100	89.3	87.85	89.82	0.7025
	Softmax		92.64	89.65	60.38	99.37	96.97	74.42	0.6836
APOE	RF		91.76	91.76	78.27	95.64	70.9	79.02	0.6815
	Softmax		90.67	91.75	80	98.28	90.9	85.1	0.7256
COG	RF		89.52	92.94	73.58	70.9	72.22	73.58	0.6852
	Softmax		91.23	92.02	88.48	**100**	**100**	89.8	0.8769
Combined (ALL)	RF		94.1	93.88	**100**	93.52	80.61	75.47	0.7222
	Softmax		**96.135**	**94.59**	84.29	99.72	98.33	**90.77**	**0.8911**
ADNI	**Classifier**	**Sub (NC/MCIs)**	**AUC**	**ACC**	**SEN**	**SPEC**	**PRE**	**F1**	**Cohen Kappa**
FS	RF	38/46	85.62	86.56	93.99	72.76	82.35	87.5	0.697
	Softmax		90.91	92.31	88.24	100	100	93.75	0.8385
APOE	RF		81.82	84.62	78.95	**100**	**100**	88.24	0.6688
	Softmax		91.18	88.46	100	75	82.35	90.32	0.7636
COG	RF		86.36	88.46	83.33	87.55	78.89	90.91	0.7547
	Softmax		89.902	90.65	92.86	86.67	88.77	83.87	0.7067
Combined (ALL)	RF		88.24	89.85	100	89.23	86.47	86.67	0.7923
	Softmax		**95.45**	**96.15**	**100**	93.33	91.67	**95.65**	**0.9222**

Table 4.5. Classification result of NC vs. MCIc

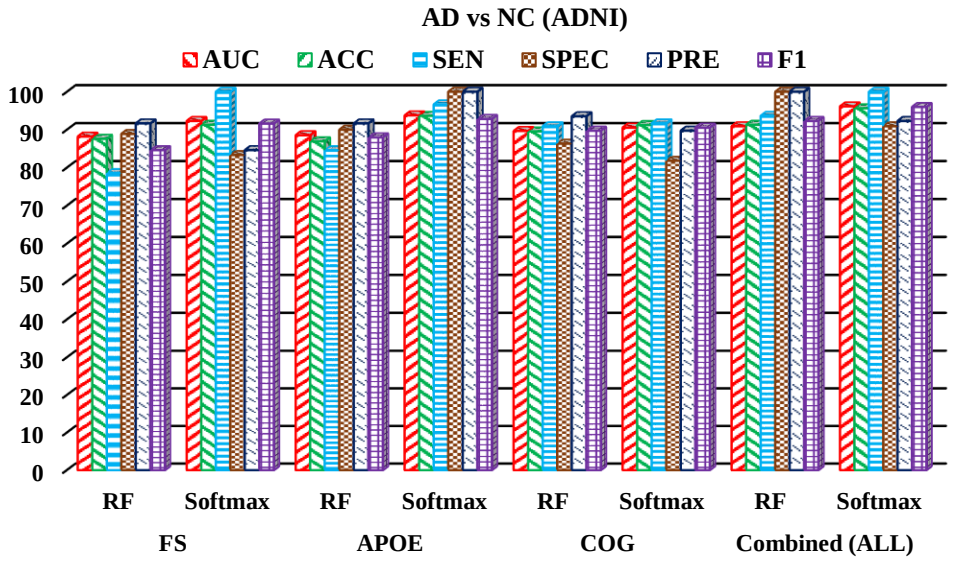
NACC	Classifier	Sub (NC/MCIc)	AUC	ACC	SEN	SPEC	PRE	F1	Cohen Kappa
FS	RF	1197/440	79.06	88.21	**96.34**	86.59	58.96	73.15	0.6615
	Softmax		92.42	90.24	72.15	98.8	96.61	82.61	0.7603
APOE	RF		89.45	90.44	82.7	93.31	82.08	82.39	0.6936
	Softmax		89.52	89.02	70.51	97.61	93.22	80.29	0.6854
COG	RF		87.57	88.41	81.3	90.78	74.62	77.82	0.6588
	Softmax		93.85	90.65	71.95	**100**	**100**	83.69	0.7738
Combined (ALL)	RF		90.16	90.45	93.83	89.78	84.41	86.38	0.7165
	Softmax		**94.74**	**92.68**	81.65	99.3	98.4	**89.58**	**0.8533**
ADNI	**Classifier**	**Sub (NC/MCIc)**	**AUC**	**ACC**	**SEN**	**SPEC**	**PRE**	**F1**	**Cohen Kappa**
FS	RF	38/36	90.87	90.86	89.86	88.89	89.86	90.86	0.8075
	Softmax		89.89	91.3	87.5	100	100	93.33	0.8099
APOE	RF		87.31	86.96	91.67	81.82	84.62	88	0.7376
	Softmax		87.5	90.89	100	88.24	79.85	85.71	0.7965
COG	RF		85.46	87.34	90.91	75	79.65	83.33	0.6875
	Softmax		81.25	86.96	**100**	83.33	82.59	86.92	0.6894
Combined (ALL)	RF		91.15	91.3	92.31	90	92.31	92.31	0.8231
	Softmax		**94.44**	**95.65**	93.33	**100**	**100**	**96.55**	**0.9069**

Table 4.6. Classification result of MCIs vs. MCIc

NACC	Classifier	Sub (MCIs/MCIc)	AUC	ACC	SEN	SPEC	PRE	F1	Cohen Kappa
FS	RF	218/440	87.1	91.92	100	89.47	74.19	85.19	0.7598
	Softmax		88.19	85.86	73.79	98.95	98.7	84.44	0.7197
APOE	RF		88.85	87.87	93.1	85.71	72.97	81.81	0.7085
	Softmax		90.61	88.88	93.33	86.85	89.9	83.58	0.8155
COG	RF		90.78	90.4	95.62	88.88	79.72	86.13	0.7147
	Softmax		92.43	91.41	93.84	90.22	90.74	87.76	0.8565
Combined (ALL)	RF		91.13	94.44	**100**	92.52	82.26	90.27	0.8643
	Softmax		**96.23**	**95.96**	91.57	**99.13**	**98.7**	**95**	**0.9162**
ADNI	**Classifier**	**Sub (MCIs/MCIc)**	**AUC**	**ACC**	**SEN**	**SPEC**	**PRE**	**F1**	**Cohen Kappa**
FS	RF	38/36	85.62	84	88.24	79.52	88.24	88.24	0.6835
	Softmax		91.32	90.56	93.75	88.89	93.75	93.75	0.8264
APOE	RF		87.3	86.96	90.31	80	85.71	87.89	0.7315
	Softmax		91.18	88	100	72.73	82.35	90.32	0.7492
COG	RF		87.55	86.75	85.71	77.78	85.71	88.95	0.7259
	Softmax		89.87	89.53	93.75	77.78	86.58	90.91	0.7331
Combined (ALL)	RF		88.89	91.3	91.55	88.88	90.55	89.88	0.8099
	Softmax		**94.12**	**93.75**	**100**	**89.85**	**94.24**	**93.72**	**0.8276**

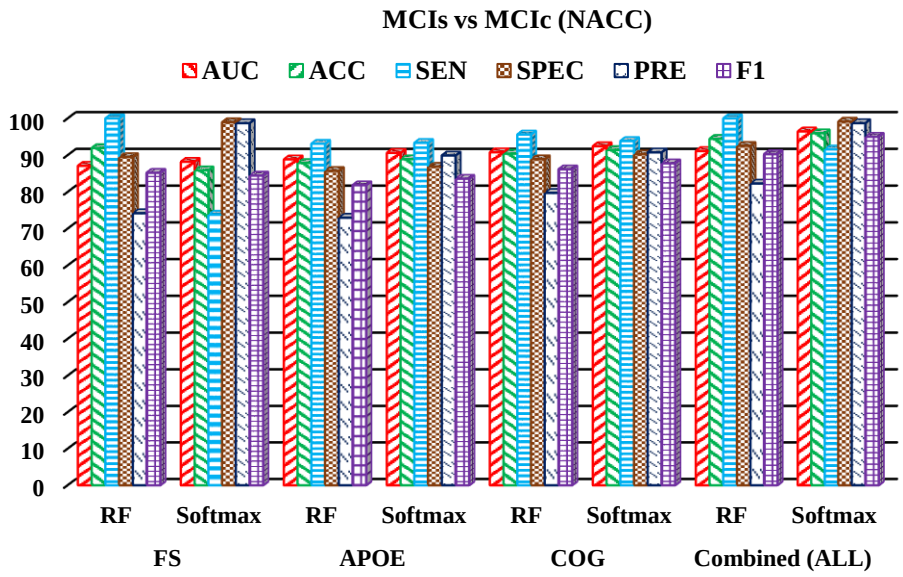


(a)

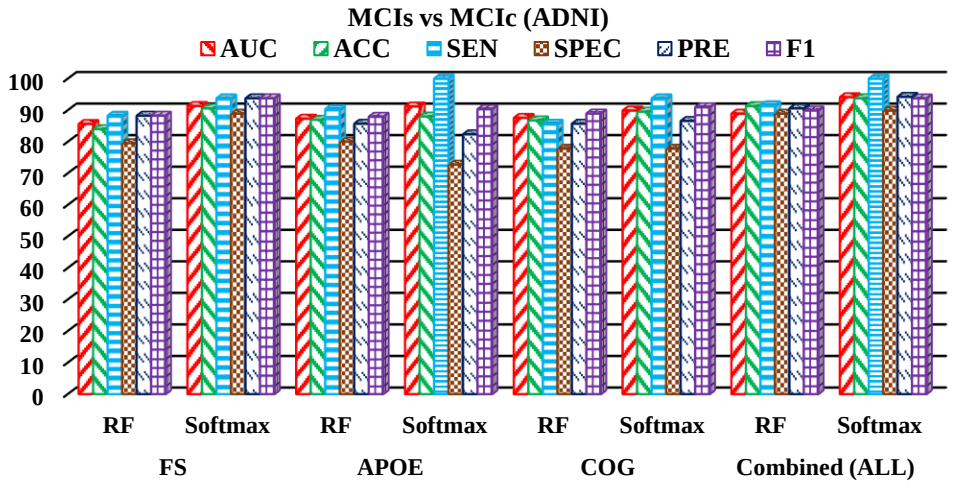


(b)

Figure 4.1. Classification result between AD vs. NC for (a) NACC and (b) ADNI dataset



(a)



(b)

Figure 4.2. Classification result between MCIs vs MCIc for (a) NACC and (b) ADNI dataset

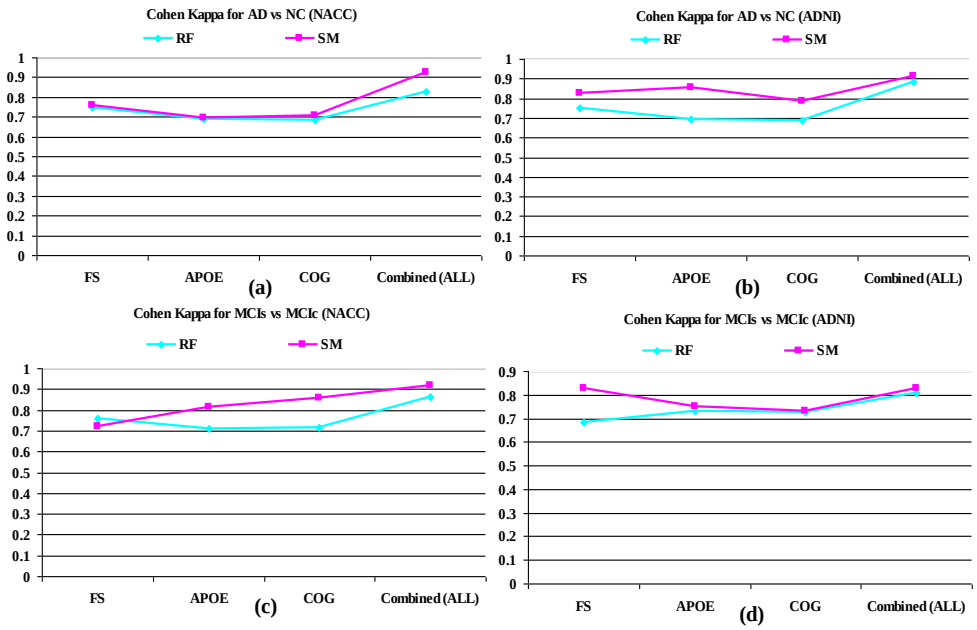
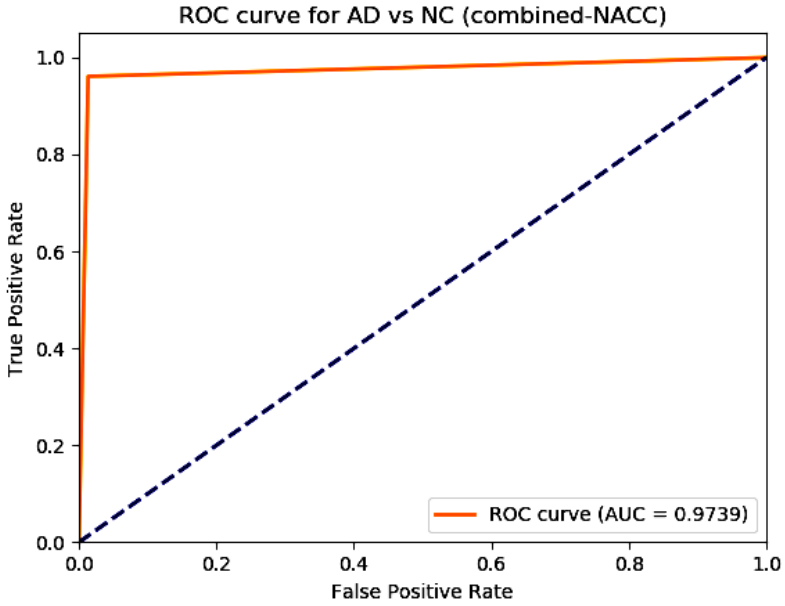
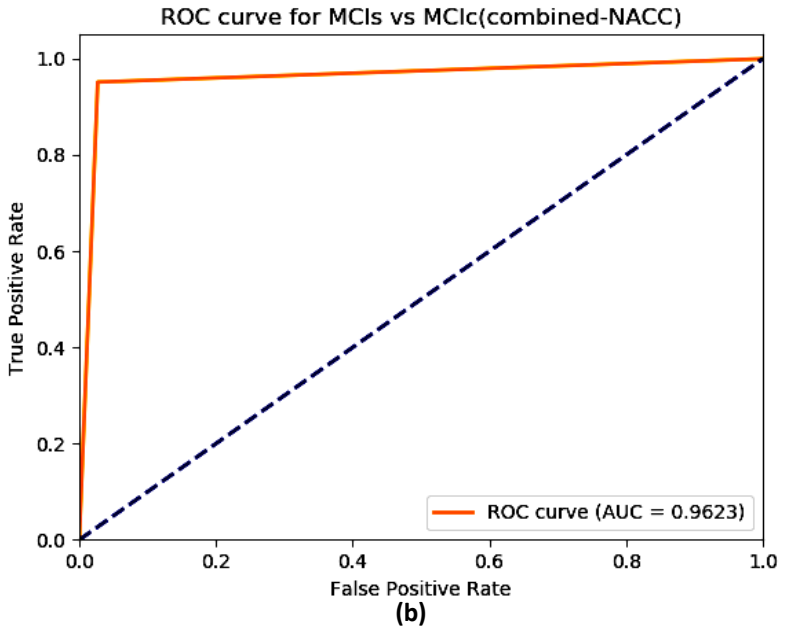


Figure 4.3. Cohen's kappa score graph for (a) AD vs NC (NACC) (b) AD vs NC (ADNI) (c) MCIs vs MCIc (NACC) (d) MCIs vs MCIc (ADNI)

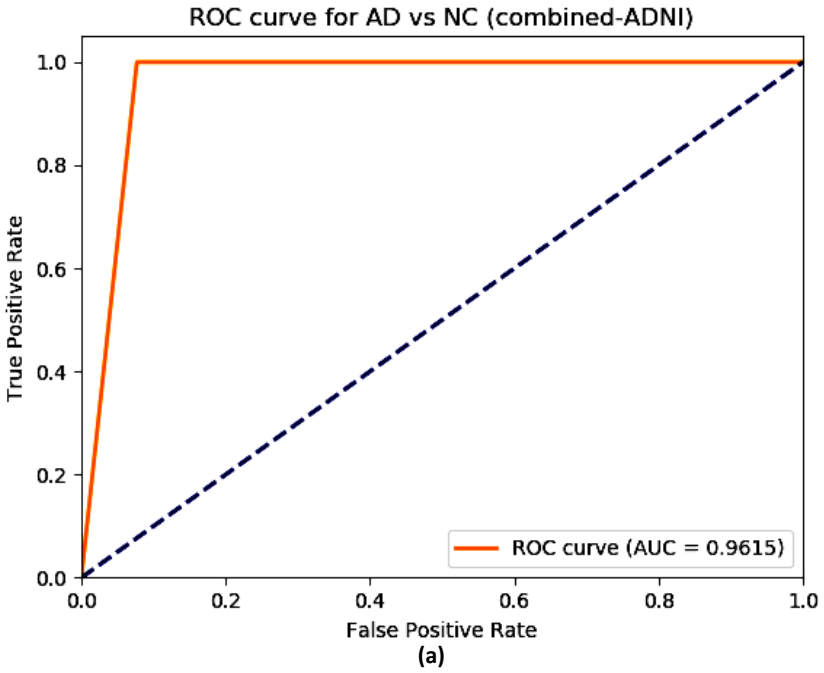


(a)



(b)

Figure 4.4. ROC curve produced by softmax classifier for (a) AD vs. NC and (b) MCIs vs. MCIc using NACC dataset



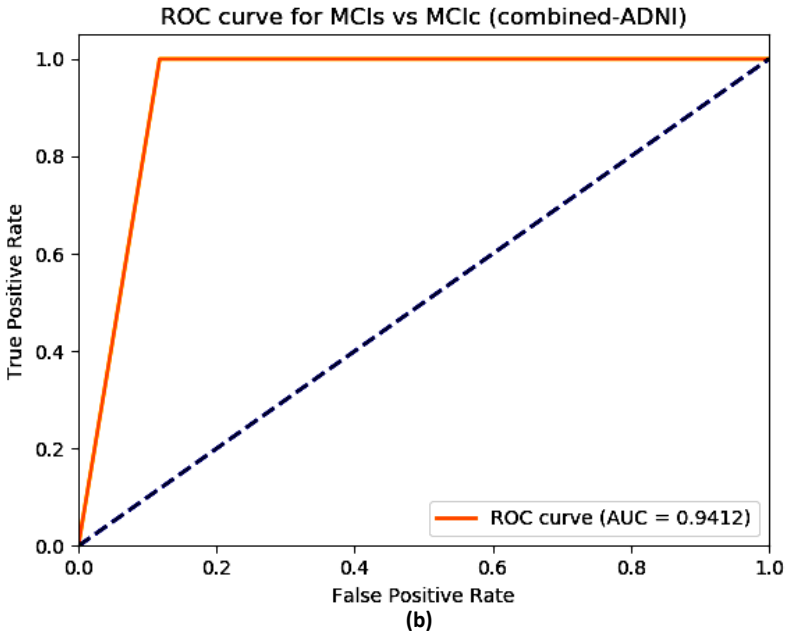


Figure 4.5. ROC curve produced by softmax classifier for (a) AD vs. NC and (b) MCIs vs. MCIc using ADNI dataset

4.1.2 Machine learning classification results

The obtained classification result for ML process are shown from Table 4.1, to 4.6 for AD vs NC, AD vs MCIs, AD vs MCIc, NC vs MCIs, NC vs MCIc, MCIs vs MCIc, respectively. The classification report for AD vs NC and MCIs vs MCIc are shown in Figure 4.1, and 4.2. All programs were executed in 64-bit Python 3.6 environment on Intel(R) Core(TM) i3-7100 at 3.90 Hz and 8 GB of RAM running an Ubuntu 16.04 LTS. The model may be implemented on any computer in which Python 3.6 is compatible.

AD vs. NC: The classification results for AD vs NC (for both NACC and ADNI dataset) are summarized in Table 4.1 and shown in Figure 4.1. For each case, the dataset was divided into two parts in 70/30 ratio. To classify AD from NC group, I have used two types of classifier (random forest and softmax). But softmax classifier outperformed the random forest classifier while predicting AD vs NC group with a high accuracy of 96.86%, and AUC of ROC as 97.39% for NACC dataset, and accuracy of 95.65%, and AUC of ROC of 96.15% for ADNI dataset. Figure 4.3 shows the Cohen's kappa graph, using (NACC and ADNI dataset) for AD vs HC and MCIs vs MCIc groups, and Figure 4.4 and 4.5 visualize the AUC of ROC graph using (NACC and ADNI dataset) for all six classification groups. From Figure 4.3, I can say that softmax classifier has achieved a complete level of an agreement between AD vs NC groups, while comparing with single modality results, which is 0.9256 for NACC and 0.9125 for ADNI dataset, which is very close to 1.

MCIs vs MCIc: For MCIs vs MCIc classification problem, I have got 658 subjects from NACC dataset, and 82 subjects from ADNI dataset. Table 4.6, and Figure 4.2, shows the classification result for MCIs vs MCIc groups. As I

know, this classification group is difficult to classify with each other because both group contain same pattern of brain, there is only slightly different in structure, but our proposed method has achieved 95.96 % of accuracy and 96.23 % of AUC of ROC using NACC dataset for MCIs vs MCIc group, whereas for ADNI dataset our proposed method using softmax classifier achieved 93.75% of accuracy and 94.12% of AUC of ROC. For MCIs vs MCIc, Cohen's kappa value is 0.9162 using NACC dataset, and 0.8276 for ADNI dataset, which are very good compared to single modalities.

Likewise, from Table 4.2, to 4.5 shows the classification result for AD vs MCIs, AD vs MCIc, NC vs MCIs, and NC vs MCIc groups, and Figure 4.4 and 4.5 visualize the AUC of ROC curve for all six proposed groups.

4.2.1 Deep learning Results

In this thesis, I have used 3D VGG16 architecture for the automatic feature extraction from the 3D sMRI images. Moreover, at last I have used softmax classifier for classification of six proposed group. I didn't face the overfitting problem either in this deep learning process. While operating this classifier, I kept the following parameter as a constant for all classification groups.

- Epochs: 15000
- Learning rate: 0.0003
- Batch size: 16
- Activation function: Adam
- Cross-validation: stratified k-fold (5 fold)

Here, before applying 3D VGG16 architecture, all 3D images were passed from image pre-processing step as mentioned in above chapter 3. VGG16

modal consist 13 convolutional layer with 4 max polling layer and 3 fully connected layer.

Table 4.7 Binary classification result on testing datasets

Group	Dataset	ACC	SEN	SPEC
AD vs HC	NACC	98.62	99.98	98.50
	ADNI	99.83	100	100
AD vs MCIs	NACC	92.30	90.89	93.37
	ADNI	91.73	89.88	92.74
AD vs MCIc	NACC	93.40	96.58	96.48
	ADNI	95.87	92.45	90.30
NC vs MCIs	NACC	92.16	93.56	96.62
	ADNI	95.47	94.74	92.44
NC vs MCIc	NACC	95.86	98.86	97.53
	ADNI	96.77	97.60	93.12
MCIs vs MCIc	NACC	91.62	92.85	90.50
	ADNI	93.56	90.75	94.68

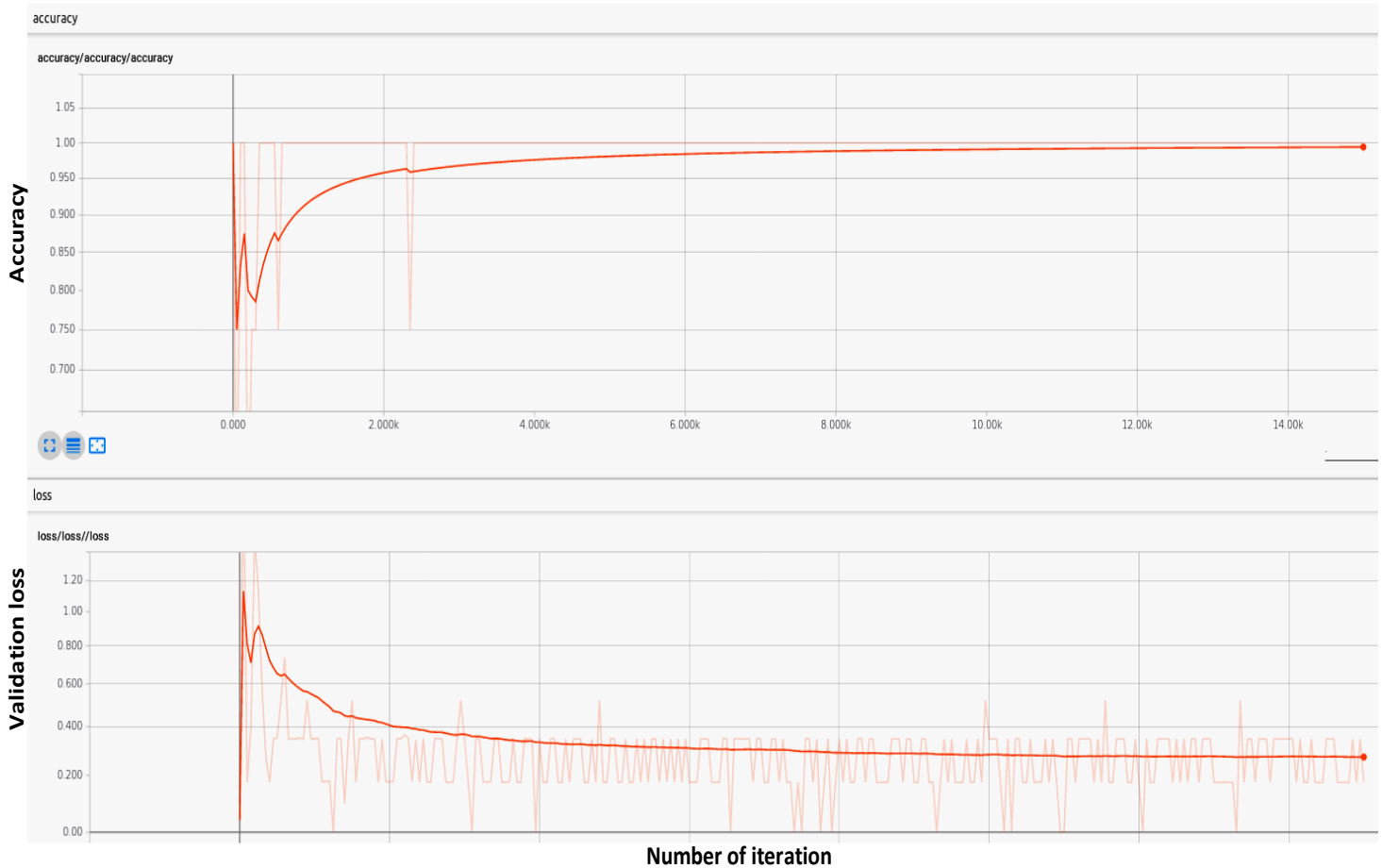


Figure 4.6 AD vs NC training graph for NACC dataset

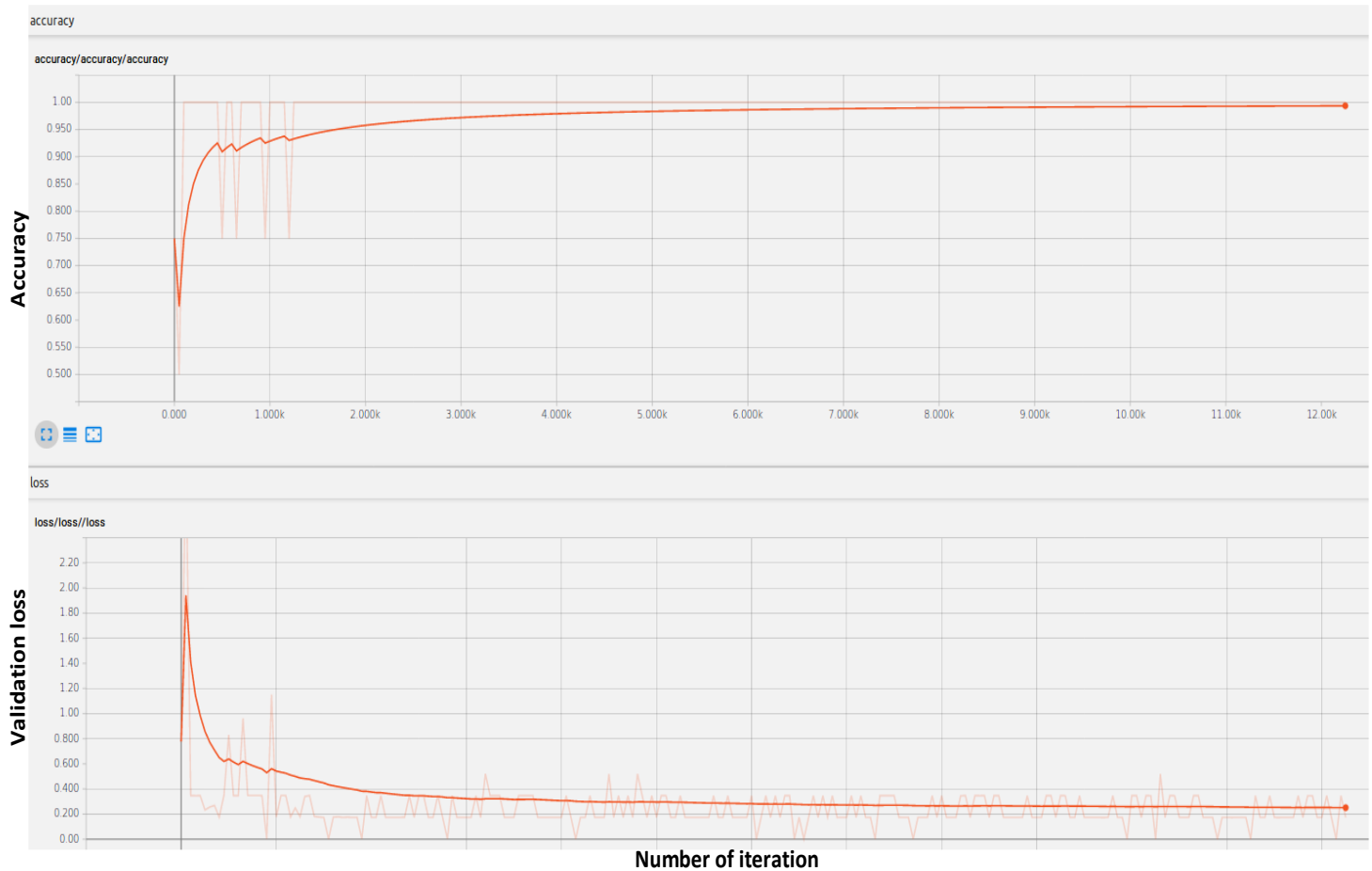


Figure 4.7 AD vs NC training graph for ADNI dataset

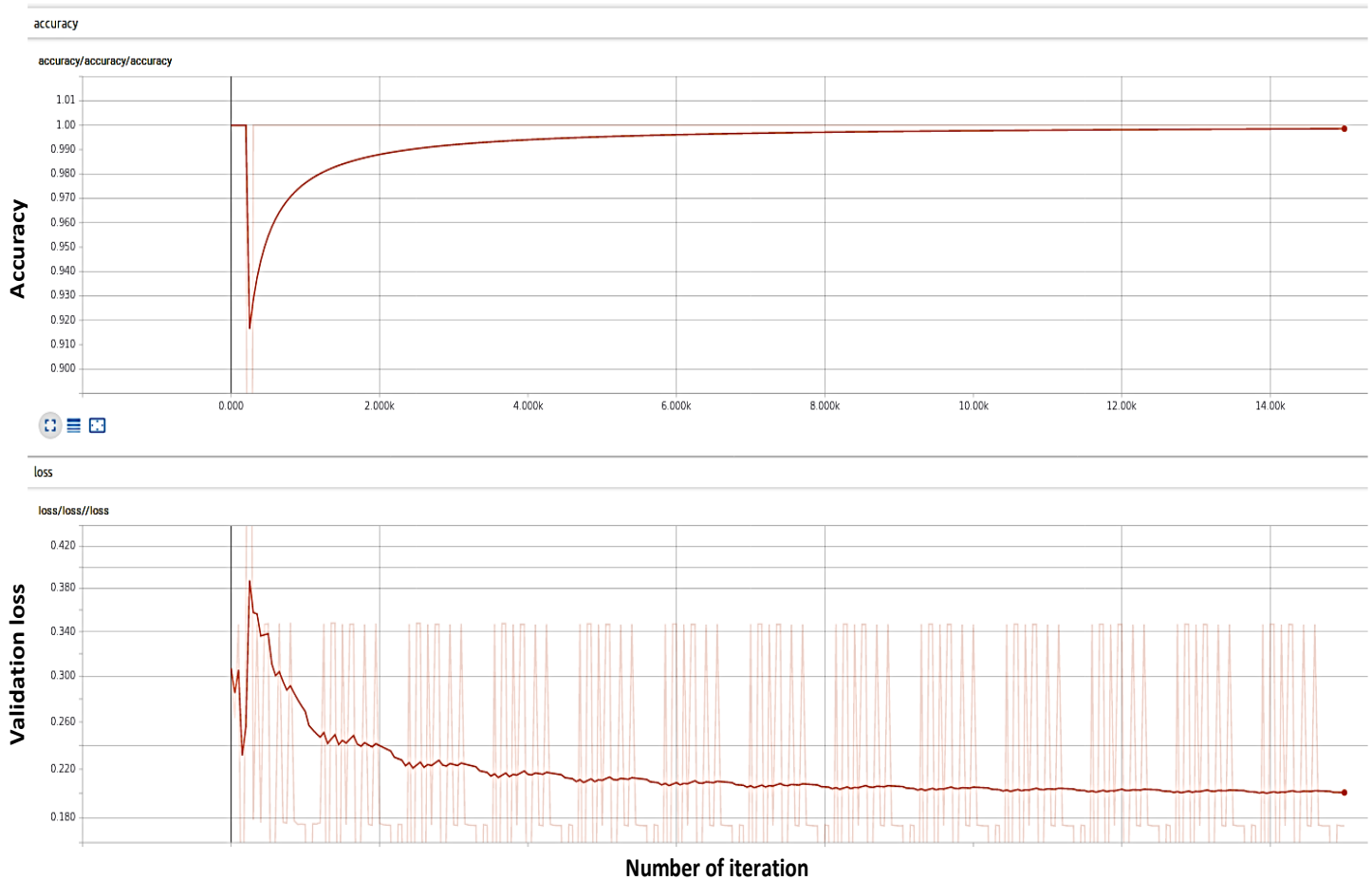


Figure 4.8 MCIs vs MCIc training graph for NACC dataset

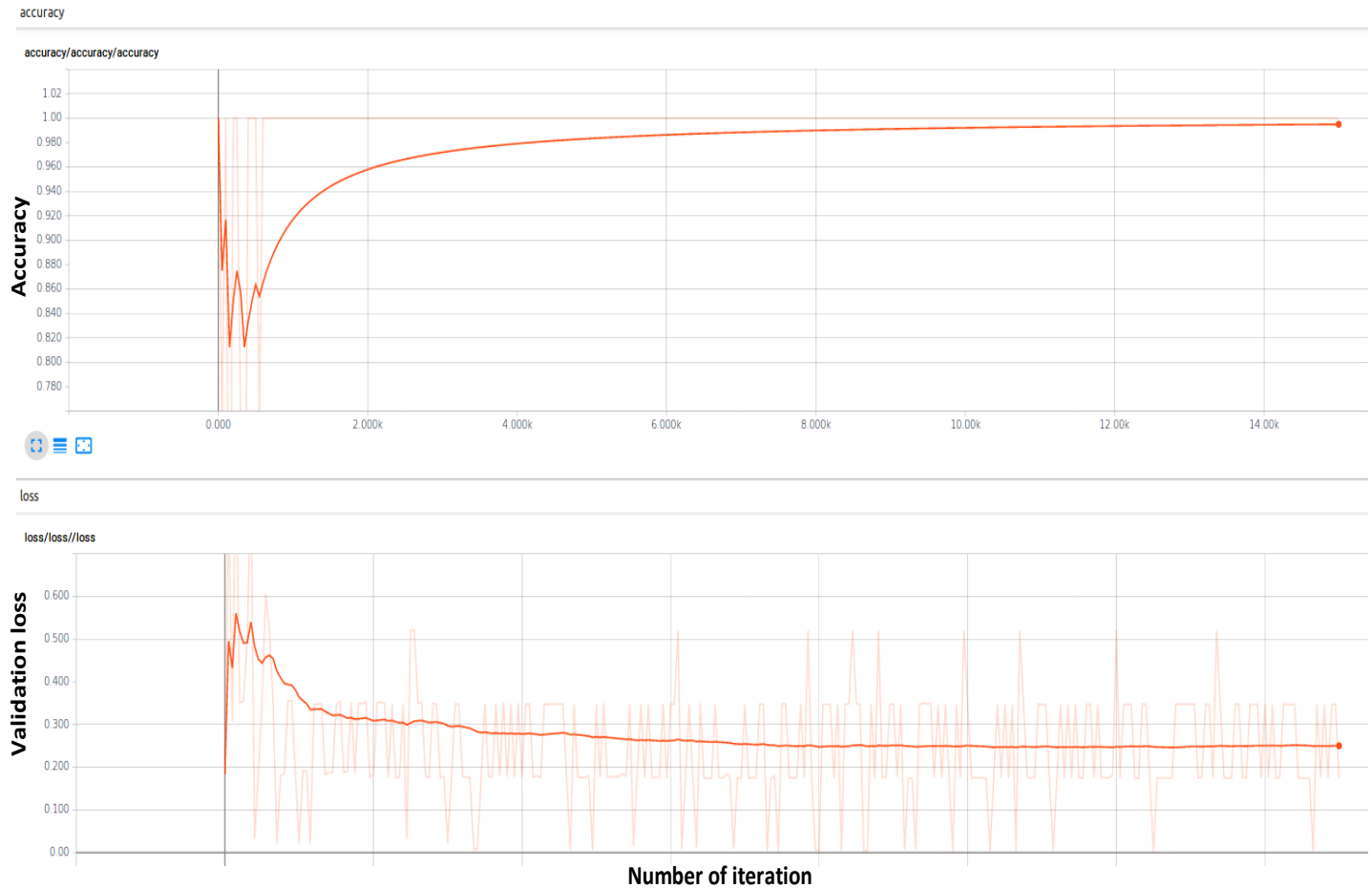


Figure 4.9 MCIs vs MCIc training graph for ADNI dataset

Performance of the 3D DNN was validated and tested on AD and NC subjects, with six binary classifications: AD vs NC, AD vs MCIs, AD vs MCIc, NC vs MCIs, NC vs MCIc, and MCIs vs MCIc. For each classification, the DNN was evaluated firstly on NACC dataset and then same procedure is applied to evaluate ADNI dataset. Each classification include three steps: (i) training, (ii) validation, and (iii) testing. First, MRI data of each classification dataset was randomly split into a large training and validation set (90% of images) and a testing set (10% of images).

For each classification, (i) 3D VGG16 was trained on the training dataset and (ii) validated using a 5-fold cross validation. To improve the performance of our classifier, a so-called transfer learning was applied, i.e., weights of the CNN used to classify ADNI AD vs HC were transferred to the other CNNs and used as (pre-trained) initial weights [72] “Transferring” the learned features reduces training time and increases the network efficiency. CNN was finally used to classify raw images of the testing set (iii). CNN's performance was evaluated by several performance measures, i.e. sensitivity, specificity and accuracy. Sensitivity measures the proportion of true positives correctly identified, whereas specificity refers to the proportion of true negatives correctly identified. The accuracy of a classifier represents the overall proportion of correct classifications.

Table 4.7 reports binary classification performance of the 3D DNN in the testing datasets. The results demonstrated that high levels of accuracy were achieved in all comparisons. Highest accuracy, sensitivity and specificity (higher than 98%) were obtained for the AD vs NC classification group using both NACC and ADNI dataset (Table 4), and Figure 4.6 and 4.7 shows the

training plot for AD vs NC groups using NACC and ADNI dataset. 3D VGG16 architecture were also able to discriminate between MCIs vs MCIc group with an optimal performance (accuracy, sensitivity and specificity values higher than 90%; Table 4), and Figure 4.8 and 4.9 shows the training plot for MCIs vs MCIc group using NACC and ADNI dataset. For both (NACC and ADNI dataset), our 3D VGG16 modal has performed well compared to machine learning technique.

Chapter 5: Discussion

Effective and accurate AD diagnosis is critical for early treatment. Therefore many researchers have devoted their efforts to develop a computer aided system (CAD), which can diagnosis an AD in early phase in an individual basis [32], [49], [55], [73], [74]. In this study, I built and validated a multimodal technique combining three different features into one and using machine learning method, I classify the AD groups with other diagnostic groups. In this thesis, I have also built and validated a 3D deep learning algorithm that predicts the individual diagnosis of AD while comparing with other groups. Moreover, obtained machine learning result shows that each modality is indispensable for achieving good combination and good classification accuracy. 3D DNN result also achieved high classification accuracy for all six classification groups for both NACC and ADNI dataset.

Now, I will answer the research question, which I have stated in chapter 1.

5.1 Answering the research questions

5.1.1 *How to combine multiple biomarker's into one form?*

As my proposed idea is to combine different modality for early classification of AD. Moreover, I propose that, multimodal biomarker can increase the classification accuracy compared to the unimodal biomarkers, so for that, I need to combine all features into one for the classification process. For this, I have used a procedure that is known as multimodal fusion. The simplest technique for utilizing the complementary information provided by a multiple imagining modalities, is to concatenate these features into one single feature vector and later train a classifier on using this vector. In this study, I have performed feature-level fusion, which is also called as early fusion, where

features from the single imaging modalities or features sets are concatenated in order to create a common feature vector. Then, a classifier is trained using this common feature vector in order to form the final classification model.

5.1.2 *How classifier can be trained to differentiate between AD groups with other diagnostic groups?*

In this thesis, I have used three kind of features:

1. First feature is the combination of cortical, subcortical, and White matter of a brain, is feature is also called as volumetric volume. I have used Freesurfer v6.0 to extract these ROI from each subject. In total there were 130 ROI for each subjects.

2. Second feature is take as an APOE genotype value of each patients. As I know, E4/E4 and E3/E4 is an essential APOE biomarker for prediction of AD. I have obtained these biomarkers from NACC, and ADNI homepage.

3. Third feature is related to the cognitive score of each subjects. These feature are the examination score and patient's education level and as well as their age, and also their GDS, and NPI-Q values.

Moreover, at last I combine all these features into one by using early fusion technique before passing it to the classifier. In this thesis, I have applied two kind of classifier (random forest and softmax) for the classification of groups. In order to attain unbiased estimates of performance, the set of participants were randomly split into two groups in a 70:30 ratio as training and testing sets, respectively. In the training set, the stratified 10-fold cross validation process was applied to obtain optimal hyperparameter values for the criterion

(Max_depth, Max_features), and also to minimize overfitting problem. The optimal hyperparameter values were computed by using grid-search and ten-fold stratified cross-validation (10-SKF-CV) technique on the training set. For each approach, the obtained optimized value of the hyperparameter was then utilized to train the training classifier using the training dataset, and later the performance of the classifier was evaluated on the remaining 30% of data in the testing dataset. In this mode, I achieved an unbiased estimates of the performance for each classification problem. From Table 4.1 to 4.6 shows the classification result for six classification problem. Moreover, Figure 4.1 and 4.2 shows the classification result of AD vs NC and MCIs vs MCIs in graph form. Moreover, figure 4.3 shows the Cohen's kappa value for AD vs HC and MCIs vs MCIc group. Figure 4.4 and 4.5 visualize the AUC of ROC curve for all six classification group obtained by using softmax classifier.

Here, for 3D DNN, at first the input were normalized and resized into 182*218*182 matrix, after that a 3D VGG16 architecture is built, which consist 13 convolution layer and 5 max polling layer with three fully connected layer, and one output layer. Performance of 3D DNN was validated and tested on six binary classification problems. Before passing this classification groups for the training, I divided the dataset into three set, training, validation, and testing. In this thesis, the value of epochs, learning rate, batch size, activation function were kept constant for all six classification groups. Here, I have used 5-fold stratified cross-validation technique to minimize the overfitting problem. Table 4.7 shows the classification result for all six binary classification groups, and Figure 4.6 and 4.7 shows the training graph for AD vs NC and MCIs vs MCIc groups.

5.1.3 *Which machine learning methods yield the best results for both (NACC and ADNI) dataset?*

A direct evaluation between machine learning methods reveal that softmax classifier has performed very well for all six binary classification groups of NACC, and as well as for the ADNI dataset. The following parameters were kept constant while applying softmax classifier, epoch, batch size, and learning rate. The softmax classifier computational time is low compared to random forest classifier, this because softmax provide the output's in probabilities range.

5.1.4 *How does 3D deep learning architecture perform on six binary classification problem?*

3D VGG16 architecture perform very well for all six binary classification problems. As I can see from Table 4.7, for AD vs NC group, our modal achieved 98.56% of accuracy for NACC and 99.83% of accuracy for ADNI dataset and whereas for MCIs vs MCIc group, our modal achieve 91.62% of accuracy for NACC, and 93.56% of accuracy for ADNI dataset. As for other classification group, our modal has also performed very well as shown in Table 4.7. Moreover, Figure 4.6 to 4.9 shows the training graph for AD vs NC and MCIs vs MCIc groups.

5.1.5 *Is multimodal technique is powerful than unimodal technique in terms of performance basis?*

As from Table 4.1 to 4.6, I can see that combined features has achieved high level of performance in all six binary classification problem, compared to single modality of biomarker's. Moreover, Cohen's kappa result also shows that multimodal features have achieved high level of agreement between each other while classifying each other, compared to unimodal features. The kappa

statistic value is always between -1 and 1. Moreover in our case, multimodal features values are near to 1 compared to unimodal features.

5.1.4 *How does the different outcomes compare with similar kind of research method?*

However, a direct evaluation between different studies is difficult, as they often use different datasets, different preprocessing methods, and eventual methods for a dimensional reduction, and as well a slightly different measurements and technique regarding evaluations. Whereas in our case, we will attempt to show the typical accuracy achieved.

Table 5.1. Comparison with other state of the art methods

Approach	Year	Dataset	Modalities	AD/NC	ACC	SEN	SPE
Zhang at al. [54]	2011	ADNI	MRI	51/52	86.2	86	86.3
Wolz et al. [56]	2011	ADNI	MRI	198/231	89	93	85
Cuingnet et al.[7]	2011	ADNI	MRI	162/137	-	81	95
Beheshti and Demirel et al.[8]	2016	ADNI	MRI	130/130	92.48	91.07	93.89
Beheshti et al. [11]	2017	ADNI	MRI	136	93.01	89.13	96.80
Aderghal et al. [75]	2017	ADNI	MRI	188/228	85.9	84.3	87.5
Long et al. [9]	2017	ADNI	MRI (AMYG)	227	93.2	88.99	86.32
Islam et al. [76]	2018	ADNI	MRI	347/537	94.97	94.33	95.89
Basaia et al. [10]	2019	ADNI	MRI	294/352	99.2	98.9	99.5
Proposed Method	2019	NACC	Combined	498/1197	96.9	91.3	99.4
	2019	ADNI	Combined	38/38	95.65	100	90.91
	2019	NACC	MRI	498/1197	98.62	99.98	98.50
	2019	ADNI	MRI	38/38	99.83	100	100

Here, I have compared our AD vs NC obtained result with ten latest state-of-arts. Zhang et al. [54] used a multimodal classification of AD based on a

combination of MRI, CSF, and PET data. They reported an ACC of 86.2% in the classification of AD vs NC using the MRI data and by combining all biomarker modality the results, they achieved a high ACC of 93.2%. Wolz et al. [56] used a multi-method technique for early detection of AD. Here, they have used four types of features, HV, CTH, TBM, and Manifold-based learning and they combined all features and they achieved ACC of 89% using LDA and ACC of 86% using SVM classifier. Cuingnet et al. [7] compared a ten widely used method using the ADNI database. Here, they have used three kind of technique to extract features from brain, VBM, CTH, and HV. They reported a SEN of 81% and an SPE of 95% as the best performances by using Voxel-Direct-D-gm. Beheshti and Demirel et al. [8] proposed a CAD system which is composed of four systematic stages and these stage belong to first, global and local differences in the gray matter (GM) of AD patients compared to GM of HC patients using a VBM technique. They have used seven different feature ranking methods and FC as a stopping criterion and achieved ACC of 92.48% for AD vs NC groups. Another study by Beheshti et al. [11], used feature ranking and a genetic algorithm (GA) for selection of optimal features for the classifier. Their method achieved an accuracy of 93.01%, sensitivity of 89.13%, and specificity of 96.80%, for AD vs NC. Aderghal et al. [75], has performed a 2D + ε fusion technique to classify AD vs NC groups. They have used Siamese network with the combination of CNN for classification. There methods yield 85.9% of accuracy, 84.3% of sensitivity, and 87.5% of sensitivity will classifying AD vs NC group. Long et al. [9] used Freesurfer software to segment 3T T1 images into many different parts and later used a multi-dimensional scaling method for feature selection before sending the selected features to the classifier. Their proposed method achieved an

accuracy of 88.88%, sensitivity of 86.32, specificity of 90.91%, and AUC of 93.2% when differentiating sMCI from pMCI using only specific amygdala features. Islam et al. [76], performed a 2D CNN for classification of AD vs NC. They have used ImageNet database for fine-tuning their CNN model. There method achieved 94.97% of accuracy, 94.33% of sensitivity, and 95.89% of specificity while classifying Ad vs NC. Recently, Basaia et al. [10], built and validated a deep learning algorithm for predicting the individual diagnosis of AD and MCI who will convert to AD (c-MCI) based on a single cross-sectional brain structural MRI scan. Convolutional neural networks (CNNs) were applied on 3D T1-weighted images. There method achieved 99.2% of accuracy, 98.9% of sensitivity, and 99.5% of specificity while classifying AD vs NC groups. As can be seen from Table 5.1 that, our proposed system shows highly competitive performance when compared to the other systems reported in this thesis for AD vs NC classifications.

5.2 Takeaways

Overall, both machine learning and deep learning techniques exhibited problems during these experiments. Overfitting was problematic for random forest, while training in training dataset that has utilized PCA for the dimensionality reduction process. Whereas, softmax classifier get stuck in the local minima, as can be understood in many neural networks stopping in the similar place. The approaches of dimensional reduction I utilized were well known methods, it drastically reduce the high-dimensional information of each samples in a lower dimensional space. Here, 10-fold cross validation play an important role to minimize the overfitting problem. The best classifier formed during these experiment was one that could distinguish AD with other

diagnostic groups. In this thesis, softmax classifier outperformed the result of random forest classifier while classifying six binary classification problem. I think that our proposed method has fully answered the research questions. Our 3D DNN modal has also performed well for six binary classification problem. Here, 3D DNN shows a promises for building a model for the automated, individual and early detection of AD and thus accelerating the adoption of structural MRI in routine practice to help assessment and management of patients.

Chapter 6: Conclusion

In this thesis, I have applied machine learning and 3D deep learning technique for classification of AD with other diagnostic groups. Here, I have used two (NACC and ADNI) different types of dataset. I have proposed six binary classification problems: AD vs NC, AD vs MCIs, AD vs MCIc, NC vs MCIs, NC vs MCIc, and MCIs vs MCIc in this experiment.

Recently, Random forests, support vector machines, softmax, and DNN, has shown comparatively good performance on this similar types of problems. As the AD is rapidly becoming global problem and the fields of CAD's, computer vision, ML, and DL makes paces, the problems at hand (i.e. automatic classification of AD from sMRI) appears to play an essential strengths on some newer methods

For machine learning case, I have used three different kind of features (Volumetric volume, APOE genotype, and cognitive score). Here, I have use random forest and softmax classifier for classification of AD. In our experiment, softmax classifier achieved good classification result for all six binary classification groups. From our experiments, I have seen that softmax seems to be a feasible machine learning method to the problem of AD classification by structural MRI which is slightly surprising, given it a plain idea with the help of 10-fold stratified cross validation technique.

Likewise, I have also applied 3D DNN method using 3D VGG16 architecture for classification of AD with other diagnostic group. Here, a pre-trained model is used because it is less time consuming task and can provide high performance in distinguishing only slightly different images. The great

advantage of using pre-trained model relative to a standard CNNs is that it significantly reduces the number of network parameters and thus serves as a form of regularization. For all six binary classification problem, our 3D DNN algorithm achieved good results as can be seen in Table 5.1.

The main thing that this report demonstrate is that, comparatively a simple machine learning methods like softmax classifier are feasible candidates even for apparently complex issue like automatic classification of AD from the brain images, provided that fitting approaches of dimensional reduction are used. In our experiment, I also observed that the studied model is not affected while mixing more than one features into one.

6.1 Future Work

I propose the following potential possibilities for future experiments.

a) *Train on additional data*: In this experiment, I have face the problem of overfitting in 3D deep learning. Moreover, for supplement training data with AddNeuroMed and OASIS, or apply systematic distortions in order to artificially expand the dataset's size, could help to reduce the overfitting problems.

b) *Train to use multiclass classifier*: In this experiment, I have validated our result only on binary class problem, but I suggest to validate on multiclass classification problem, like (AD vs NC vs MCI)

c) *Multi-imaging modality*: In our experiment, I have used only one kind of imaging modalities (sMRI). Moreover, I suggest others to try to use all imaging modalities for classification of AD, because every imaging modalities produce important features, which can help the researcher to predict AD more precisely.

Bibliography

[1] Ferri, C.P. "Global prevalence of dementia: a Delphi consensus study," *Lancet*, vol. 366, no. 9503. pp. 2112-2117, Dec. 2005.

[2] E. M. Reiman, J. B. Langbaum, and P. N. Tariot, "Alzheimer's Prevention Initiative: a proposal to evaluate presymptomatic treatments as quickly as possible," *Biomark. Med.*, vol. 4, no. 1, pp. 3–14, Feb. 2010.

[3] L. J. Bain *et al.*, "Healthy brain aging: A meeting report from the Sylvan M. Cohen Annual Retreat of the University of Pennsylvania Institute on Aging," *Alzheimers Dement.*, vol. 4, no. 6, pp. 443–446, Nov. 2008.

[4] T. Wang, R. G. Qiu, and M. Yu, "Predictive Modeling of the Progression of Alzheimer's Disease with Recurrent Neural Networks," *Sci. Rep.*, vol. 8, no. 1, p. 9161, Dec. 2018.

[5] J. S. Lee *et al.*, "Machine Learning-based Individual Assessment of Cortical Atrophy Pattern in Alzheimer's Disease Spectrum: Development of the Classifier and Longitudinal Evaluation," *Sci. Rep.*, vol. 8, no. 1, p. 4161, Dec. 2018.

[6] M. Hadj-Hamou, M. Lorenzi, N. Ayache, and X. Pennec, "Longitudinal Analysis of Image Time Series with Diffeomorphic Deformations: A Computational Framework Based on Stationary Velocity Fields," *Front. Neurosci.*, vol. 10, Jun. 2016.

[7] R. Cuingnet *et al.*, "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database," *NeuroImage*, vol. 56, no. 2, pp. 766–781, May 2011.

[8] I. Beheshti and H. Demirel, "Feature-ranking-based Alzheimer's disease classification from structural MRI," *Magn. Reson. Imaging*, vol. 34, no. 3, pp. 252–263, Apr. 2016.

[9] X. Long, L. Chen, C. Jiang, L. Zhang, and Alzheimer's Disease Neuroimaging Initiative, "Prediction and classification of Alzheimer disease based on quantification of MRI deformation," *PLOS ONE*, vol. 12, no. 3, p. e0173372, Mar. 2017.

[10] S. Basaia *et al.*, "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks," p. 8, 2019.

[11] I. Beheshti, H. Demirel, and H. Matsuda, "Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm," *Comput. Biol. Med.*, vol. 83, pp. 109–119, Apr. 2017.

[12] C. Ledig, A. Schuh, R. Guerrero, R. A. Heckemann, and D. Rueckert, "Structural brain imaging in Alzheimer's disease and mild cognitive impairment: biomarker analysis and shared morphometry database," *Sci. Rep.*, vol. 8, no. 1, p. 11258, Dec. 2018.

[13] I. Beheshti, H. Demirel, F. Farokhian, C. Yang, and H. Matsuda, "Structural MRI-based detection of Alzheimer's disease using feature ranking and classification error," *Comput. Methods Programs Biomed.*, vol. 137, pp. 177–193, Dec. 2016.

[14] R. Casanova, F.-C. Hsu, and Mark A. Espeland, for the Alzheimer's Disease Neuroimaging Initiative, "Classification of Structural MRI Images in Alzheimer's Disease from the Perspective of Ill-Posed Problems," *PLoS ONE*, vol. 7, no. 10, p. e44877, Oct. 2012.

[15] S. Farhan, M. A. Fahiem, and H. Tauseef, "An Ensemble-of-Classifiers Based Approach for Early Diagnosis of Alzheimer's Disease: Classification Using Structural Features of Brain Images," *Comput. Math. Methods Med.*, vol. 2014, pp. 1–11, 2014.

[16] H. Braak, E. Braak, J. Bohl, and H. Bratzke, "Evolution of Alzheimer's disease related cortical lesions," in *Alzheimer's Disease — From Basic Research to Clinical Applications*, vol. 54, H.-J. Gertz and T. Arendt, Eds. Vienna: Springer Vienna, 1998, pp. 97–106.

[17] C. Wattmo, E. Londos, and L. Minthon, "Risk Factors That Affect Life Expectancy in Alzheimer's Disease: A 15-Year Follow-Up," *Dement. Geriatr. Cogn. Disord.*, vol. 38, no. 5–6, pp. 286–299, 2014.

[18] C. R. Jack *et al.*, "NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease," *Alzheimers Dement.*, vol. 14, no. 4, pp. 535–562, Apr. 2018.

[19] A. Wimo and M. Prince, “Alzheimer’s Disease International World Alzheimer Report 2010 The Global Economic Impact of Dementia,” p. 56, 2010.

[20] C. Hutton, B. Draganski, J. Ashburner, and N. Weiskopf, “A comparison between voxel-based cortical thickness and voxel-based morphometry in normal aging,” *NeuroImage*, vol. 48, no. 2, pp. 371–380, Nov. 2009.

[21] 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.

[22] 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.

[23] 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.

[24] L. Berg, D. W. McKeel, J. P. Miller, J. Baty, and J. C. Morris, “Neuropathological Indexes of Alzheimer’s Disease in Demented and Nondemented Persons Aged 80 Years and Older,” *Arch. Neurol.*, vol. 50, no. 4, pp. 349–358, Apr. 1993.

[25] 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.

[26] T. J. Maher *et al.*, “Phentermine and other monoamine-oxidase inhibitors may increase plasma serotonin when given with fenfluramines,” *THE LANCET*, vol. 353, p. 3, 1999.

[27] C. R. Jack *et al.*, “A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers,” *Neurology*, vol. 87, no. 5, pp. 539–547, Aug. 2016.

[28] 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, pp. S1–S68, Feb. 2012.

[29] I. Baldeiras *et al.*, "Addition of the Aβ42/40 ratio to the cerebrospinal fluid biomarker profile increases the predictive value for underlying Alzheimer's disease dementia in mild cognitive impairment," *Alzheimers Res. Ther.*, vol. 10, no. 1, p. 33, Dec. 2018.

[30] R. Smith-Bindman *et al.*, "Use of Diagnostic Imaging Studies and Associated Radiation Exposure for Patients Enrolled in Large Integrated Health Care Systems, 1996-2010," *JAMA*, vol. 307, no. 22, Jun. 2012.

[31] W. Hollingworth *et al.*, "The Diagnostic and Therapeutic Impact of MRI: an Observational Multi-centre Study," *Clin. Radiol.*, vol. 55, no. 11, pp. 825–831, Nov. 2000.

[32] S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, and C. Davatzikos, "A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages," *NeuroImage*, vol. 155, pp. 530–548, Jul. 2017.

[33] B. T. Wyman *et al.*, "Standardization of analysis sets for reporting results from ADNI MRI data," *Alzheimers Dement.*, vol. 9, no. 3, pp. 332–337, May 2013.

[34] T. M. Schouten *et al.*, "Combining anatomical, diffusion, and resting state functional magnetic resonance imaging for individual classification of mild and moderate Alzheimer's disease," *NeuroImage Clin.*, vol. 11, pp. 46–51, 2016.

[35] T. Serre, G. Kreiman, M. Kouh, C. Cadieu, U. Knoblich, and T. Poggio, "A quantitative theory of immediate visual recognition," in *Progress in Brain Research*, vol. 165, Elsevier, 2007, pp. 33–56.

[36] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Comput. Electron. Agric.*, p. S0168169917313303, Mar. 2018.

[37] C. Käding, E. Rodner, A. Freytag, and J. Denzler, "Fine-Tuning Deep Neural Networks in Continuous Learning Scenarios," in *Computer*

Vision – ACCV 2016 Workshops, vol. 10118, C.-S. Chen, J. Lu, and K.-K. Ma, Eds. Cham: Springer International Publishing, 2017, pp. 588–605.

[38] D. Cireşan, U. Meier, and J. Schmidhuber, "Multi-column Deep Neural Networks for Image Classification," *ArXiv12022745 Cs*, Feb. 2012.

[39] D. C. Ciresan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Deep Big Simple Neural Nets Excel on Handwritten Digit Recognition," *Neural Comput.*, vol. 22, no. 12, pp. 3207–3220, Dec. 2010.

[40] D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent, and S. Bengio, "Why Does Unsupervised Pre-training Help Deep Learning?," p. 36.

[41] G. E. Hinton *et al.*, "Learning multiple layers of representation," *Trends Cogn. Sci.*, vol. 11, no. 10, pp. 428–434, Oct. 2007.

[42] I. J. Goodfellow, Y. Bulatov, J. Ibarz, S. Arnoud, and V. Shet, "Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks," *ArXiv13126082 Cs*, Dec. 2013.

[43] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-Supervised Nets," *ArXiv14095185 Cs Stat*, Sep. 2014.

[44] A. G. Howard *et al.*, "Some Improvements on Deep Convolutional Neural Network Based Image Classification," p. 6.

[45] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[46] Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," *ArXiv12065538 Cs*, Jun. 2012.

[47] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[48] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *ArXiv12070580 Cs*, Jul. 2012.

[49] S. Kloppel *et al.*, "Automatic classification of MR scans in Alzheimer's disease," *Brain*, vol. 131, no. 3, pp. 681–689, Feb. 2008.

[50] G. B. Frisoni, "Radial Width of the Temporal Horn: A Sensitive Measure in Alzheimer Disease," p. 13, 2002.

[51] J. H. Morra, Zhuowen Tu, L. G. Apostolova, A. E. Green, A. W. Toga, and P. M. Thompson, "Comparison of AdaBoost and Support Vector Machines for Detecting Alzheimer's Disease Through Automated Hippocampal Segmentation," *IEEE Trans. Med. Imaging*, vol. 29, no. 1, pp. 30–43, Jan. 2010.

[52] M. Termenon and M. Graña, "A Two Stage Sequential Ensemble Applied to the Classification of Alzheimer's Disease Based on MRI Features," *Neural Process. Lett.*, vol. 35, no. 1, pp. 1–12, Feb. 2012.

[53] A. Gupta *et al.*, "Natural Image bases to represent Neuroimaging data," *International conference on machine learning*, vol. 28, p. 987-994, June 2013.

[54] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 55, no. 3, pp. 856–867, Apr. 2011.

[55] Y. Cho, J.-K. Seong, Y. Jeong, and S. Y. Shin, "Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data," *NeuroImage*, vol. 59, no. 3, pp. 2217–2230, Feb. 2012.

[56] R. Wolz *et al.*, "Multi-Method Analysis of MRI Images in Early Diagnostics of Alzheimer's Disease," *PLoS ONE*, vol. 6, no. 10, p. e25446, Oct. 2011.

[57] C. Salvatore, A. Cerasa, P. Battista, M. C. Gilardi, A. Quattrone, and I. Castiglioni, "Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach," *Front. Neurosci.*, vol. 9, Sep. 2015.

[58] L. M. Besser *et al.*, "The Revised National Alzheimer's Coordinating Center's Neuropathology Form—Available Data and New Analyses," *J. Neuropathol. Exp. Neurol.*, vol. 77, no. 8, pp. 717–726, Aug. 2018.

[59] D. M. Michaelson *et al.*, "APOE ε4- The most prevalent yet understudied risk factor for Alzheimer's disease." *Alz. & Dem.: The Journal of the Alzheimer's Association*, vol. 10, pp. 861-868, Sep. 2014.

[60] J. K. Kueper, M. Speechley, and M. Montero-Odasso, “The Alzheimer’s Disease Assessment Scale–Cognitive Subscale (ADAS-Cog): Modifications and Responsiveness in Pre-Dementia Populations. A Narrative Review,” *J. Alzheimers Dis.*, vol. 63, no. 2, pp. 423–444, Apr. 2018.

[61] L. Rozzini *et al.*, “The Importance of Alzheimer Disease Assessment Scale-cognitive Part in Predicting Progress for Amnestic Mild Cognitive Impairment to Alzheimer Disease,” *J. Geriatr. Psychiatry Neurol.*, vol. 21, no. 4, pp. 261–267, Dec. 2008.

[62] Y. Gupta, K. H. Lee, K. Y. Choi, J. J. Lee, B. C. Kim, and G.-R. Kwon, “Alzheimer’s Disease Diagnosis Based on Cortical and Subcortical Features,” *J. Healthc. Eng.*, vol. 2019, pp. 1–13, Mar. 2019.

[63] X. Tang, Y. Qin, J. Wu, M. Zhang, W. Zhu, and M. I. Miller, “Shape and diffusion tensor imaging based integrative analysis of the hippocampus and the amygdala in Alzheimer’s disease,” *Magn. Reson. Imaging*, vol. 34, no. 8, pp. 1087–1099, Oct. 2016.

[64] I. T. Jolliffe and J. Cadima, “Principal component analysis: a review and recent developments,” *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.*, vol. 374, no. 2065, p. 20150202, Apr. 2016.

[65] A. V. Lebedev *et al.*, “Random Forest ensembles for detection and prediction of Alzheimer’s disease with a good between-cohort robustness,” *NeuroImage Clin.*, vol. 6, pp. 115–125, 2014.

[66] S. R. Bhagya Shree and H. S. Sheshadri, “Diagnosis of Alzheimer’s disease using Naive Bayesian Classifier,” *Neural Comput. Appl.*, vol. 29, no. 1, pp. 123–132, Jan. 2018.

[67] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith, “FSL,” *NeuroImage*, vol. 62, no. 2, pp. 782–790, Aug. 2012.

[68] J. Ashburner and K. J. Friston, “Voxel-Based Morphometry—The Methods,” *NeuroImage*, vol. 11, no. 6, pp. 805–821, Jun. 2000.

[69] N. J. Tustison *et al.*, “N4ITK: Improved N3 Bias Correction,” *IEEE Trans. Med. Imaging*, vol. 29, no. 6, pp. 1310–1320, Jun. 2010.

[70] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *ArXiv14091556 Cs*, Sep. 2014.

[71] M. Greiner, D. Pfeiffer, and R. . Smith, “Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests,” *Prev. Vet. Med.*, vol. 45, no. 1–2, pp. 23–41, May 2000.

[72] E. H. Asl, M. Ghazal, A. Mahmoud, “Alzheimer’s disease diagnostics by a 3D deeply supervised adaptable convolutional network.” *Front. Bioscience*, vol. 23, Jan. 2018.

[73] J. C. Morris *et al.*, “Cerebral amyloid deposition and diffuse plaques in ``normal’’ aging: Evidence for presymptomatic and very mild Alzheimer’s disease,” *Neurology*, vol. 46, no. 3, pp. 707–719, Mar. 1996.

[74] 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.

[75] K. Aderghal, J. Benois-Pineau, and K. Afdel, “Classification of sMRI for Alzheimer’s disease Diagnosis with CNN: Single Siamese Networks with 2D+? Approach and Fusion on ADNI,” in *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval - ICMR ’17*, Bucharest, Romania, 2017, pp. 494–498.

[76] J. Islam, and Y. Zhang, “Deep Convolutional Neural Networks for Automated Diagnosis of Alzheimer’s Disease and Mild Cognitive Impairment Using 3D Brain MRI,” in *Brain Informatics*, vol. 11309, S. Wang, V. Yamamoto, J. Su, Y. Yang, E. Jones, L. Iasemidis, and T. Mitchell, Eds. Cham: Springer International Publishing, 2018, pp. 359–369.

Acknowledgement

This thesis marks the closure of my master's degree at the Faculty of Information and Communication Engineering, Chosun University. It has been an exciting couple of years, and in these years I have learned many new things about medical image processing. My project has been very interesting to work on, and still the work itself was sometimes more challenging, many people that have assisted me in my work have been nothing but helpful and supportive.

I would like to express my gratitude to my advisor Prof. Goo-Rak Kwon for his care, helpful advice, and sociable input and providing me the opportunity to work on this research topic. Under his guidance, the exposure I have gained during my research would be a valuable for my life.

Further, I must express my sincere gratitude to the biggest source of my strength, my family. This achievement would not been possible without their love, support, and continuous inspiration throughout my years of study. I will be thankful forever for your love.

Lastly, I would acknowledge to the center for Dementia and AD, Chosun University, for the financial assistance. And also, I would like to thank NACC, and ADNI and its collaborators for their great efforts, large amounts of research work and willingness to share their dataset, without which this thesis and the original work described herein would not be possible.

*Data used in preparation of this article were obtained from the National Alzheimer's coordination center (NACC) database is funded by NIA/NIH Grant U01 AG016976. NACC data are contributed by the NIAfunded ADCs: P30 AG019610 (PI Eric Reiman, MD), P30 AG013846 (PI Neil Kowall, MD), P50 AG008702 (PI Scott Small, MD), P50 AG025688 (PI Allan Levey, MD, PhD), P50 AG047266 (PI Todd Golde, MD, PhD), P30 AG010133 (PI Andrew Saykin, PsyD), P50 AG005146 (PI Marilyn Albert, PhD), P50 AG005134 (PI Bradley Hyman, MD, PhD), P50 AG016574 (PI Ronald Petersen, MD, PhD), P50 AG005138 (PI Mary Sano, PhD), P30 AG008051 (PI Thomas Wisniewski, MD), P30 AG013854 (PI Robert Vassar, PhD), P30 AG008017 (PI Jeffrey Kaye, MD), P30 AG010161 (PI David Bennett, MD), P50 AG047366 (PI Victor Henderson, MD, MS), P30 AG010129 (PI Charles DeCarli, MD), P50 AG016573 (PI Frank LaFerla, PhD), P50 AG005131 (PI James Brewer, MD, PhD), P50 AG023501 (PI Bruce Miller, MD), P30 AG035982 (PI Russell Swerdlow, MD), P30 AG028383 (PI Linda Van Eldik, PhD), P30 AG053760 (PI Henry Paulson, MD, PhD), P30 AG010124 (PI John Trojanowski, MD, PhD), P50 AG005133 (PI Oscar Lopez, MD), P50 AG005142 (PI Helena Chui, MD), P30 AG012300 (PI Roger Rosenberg, MD), P30 AG049638 (PI Suzanne Craft, PhD), P50 AG005136 (PI Thomas Grabowski, MD), P50 AG033514 (PI Sanjay Asthana, MD, FRCP), P50 AG005681 (PI John Morris, MD), P50 AG047270 (PI Stephen Strittmatter, MD, PhD). This research was supported by the Brain Research Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT & Future Planning (NRF-2014M3C7A1046050). Moreover, this work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2017R1A2B4006533). 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). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association;

Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.