



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

February 2022

PhD Dissertation

**Comparative Study of Brain Networks
Based on Graph Embedding Technique**

Graduate School of Chosun University

Department of Information and Communication Engineering

Ji-In Kim

Comparative Study of Brain Networks Based on Graph Embedding Technique

그래프 임베딩 기술 기반 뇌 네트워크의 비교 연구

February 25, 2022

Graduate School of Chosun University

Department of Information and Communication

Engineering

Ji-In Kim

Comparative Study of Brain Networks Based on Graph Embedding Technique

Advisor: Prof. Goo-Rak Kwon

This thesis is submitted to the Graduate School of
Chosun University in partial fulfillment of the
requirements for the Doctor's degree in engineering.

October 2021

Graduate School of Chosun University

Department of Information and Communication

Engineering

Ji-In Kim


**This is to certify that the Doctor's thesis of
Ji-In Kim**

has been approved by the examining committee for the
thesis requirement for the Doctor's degree in engineering

**Committee Chairperson
Chosun University**

Prof. Jae-Young Pyun (Sign) 

**Committee Member
Sunchon National University**

Prof. Eui-Sung Kang (Sign) 

**Committee Member
Sunchon National University**

Prof. Sang-Hyun Park (Sign) 

**Committee Member
Chosun University**

Prof. Ji-Hwan Moon (Sign) 

**Committee Member
Chosun University**

Prof. Goo-Rak Kwon (Sign) 

January 2022

Graduate School of Chosun University

Table of Contents

Table of Contents	i
List of Table	iii
List of Figures	vi
Acronyms	vii
ABSTRACT	x
초 록	xi
I . Introduction	1
A. Thesis motivation	1
B. Research objectives	2
C. Thesis contribution	4
D. Thesis organization	5
II . Background	6
A. Alzheimer’s disease	6
B. Early-stage Alzheimer’s (mild) disease	7
C. Middle-stage Alzheimer’s (moderate) disease	7
D. Late-stage Alzheimer’s (severe) disease	8
E. Economic impact of dementia	9
F. Diagnosis of Alzheimer’s disease	10
G. Computer aided diagnosis	14

H. Machine learning in computer-aided diagnosis	15
I. Human brain network and graph theory	19
J. Functional connectivity of brain	21
K. Construction of functional brain network using fMRI	23
III. Connectivity based Diagnosis of Alzheimer's Disease	25
A. Method and approach	25
B. Whole-brain connectivity matrix	32
C. Core large-scale networks	33
D. Proposed framework	34
E. Feature selection techniques	37
F. Extreme learning machine	39
IV. Performance Evaluation	43
A. Demographic and clinical findings	44
B. Classification results	44
C. Discussion	59
V. Limitations	60
VI. Conclusion	61
References	62

List of Table

Table 3.1. Summary of subject's demographic status	26
Table 4.1. Confusion matrix	43
Table 4.2. Comparison of performance of binary classification AD against HC with state of the art methods using rs fMRI	47
Table 4.3 Comparison of performance of binary classification MCI against HC with state of the art methods using rs-fMRI	48
Table 4.4 10-fold cross-validation binary mean classification performance for AD against HC using SL-RELM classifier on whole brain network using different feature selection methods	49
Table 4.5. 10-fold cross-validation binary mean classification performance for HC against MCI using SL-RELM classifier on whole brain network using different feature selection methods	49
Table 4.6. 10-fold cross-validation binary mean classification performance for MCI against AD using SL-RELM classifier on whole brain network using different feature selection methods	50
Table 4.7. 10-fold cross-validation binary mean classification performance for AD against HC using SL-RELM classifier on large scale brain network using different feature selection methods	50
Table 4.8. 10-fold cross-validation binary mean classification performance for HC against MCI using SL-RELM classifier on large scale brain network using different feature selection methods	51
Table 4.9. 10-fold cross-validation binary mean classification performance for MCI against AD using SL-RELM classifier on large scale brain network using different feature selection methods	51
Table 4.10. 10-fold cross-validation binary mean classification performance for AD against HC using SL-RELM classifier on Combined brain network using different feature selection methods	52

Table 4.11. 10-fold cross-validation binary mean classification performance for MCI against.. HC using SL-RELM classifier on Combined brain network using different feature selection methods	52
Table 4.12. 10-fold cross-validation binary mean classification performance for AD against MCI using SL-RELM classifier Combined brain network using different feature selection methods	53
Table 4.13 10-fold cross-validation binary mean classification performance for AD against HC using ML-RELM classifier on whole brain network using different feature selection methods	54
Table 4.14. 10-fold cross-validation binary mean classification performance for HC against MCI using ML-RELM classifier on whole brain network using different feature selection methods	55
Table 4.15. 10-fold cross-validation binary mean classification performance for MCI against AD using ML-RELM classifier on whole brain network using different feature selection methods	55
Table 4.16. 10-fold cross-validation binary mean classification performance for AD against HC using ML-RELM classifier on large scale brain network using different feature selection methods	56
Table 4.17. 10-fold cross-validation binary mean classification performance for HC against MCI using ML-RELM classifier on large scale brain network using different feature selection methods	56
Table 4.18. 10-fold cross-validation binary mean classification performance for MCI against AD using ML-RELM classifier on large scale brain network using different feature selection methods	57
Table 4.19. 10-fold cross-validation binary mean classification performance for AD against HC using ML-RELM classifier on Combined brain network using different feature selection methods	57
Table 4.20. 10-fold cross-validation binary mean classification performance for MCI against HC using ML-RELM classifier on Combined brain network	

using different feature selection methods	58
Table 4.21. 10-fold cross-validation binary mean classification performance for AD against MCI using ML-RELM classifier Combined brain network using different feature selection methods	58

List of Figures

Figure 2.1 Economic burden of Alzheimer’s disease	10
Figure 2.2 Brain Connectivity Map, figure obtained from CONN tool	21
Figure 2.3. Schematic representation of brain network construction using fMRI data	24
Figure 3.1. CONN setup, figure obtained from CONN tool	28
Figure 3.2. Denoising GUI of CONN, figure obtained from CONN tool	30
Figure 3.3. CONN analyses (level 1) GUI, figure obtained from CONN tool	31
Figure 3.4. CONN results (level 2) GUI, figure obtained from CONN tool	32
Figure 3.5. Whole brain network	33
Figure 3.6. Classification of Alzheimer's disease using brain network	34
Figure 3.7. Illustration of random walk in node2vec algorithm	36
Figure 3.8. Single hidden layer ELM	40
Figure 3.9. Multiple hidden layer ELM	42

Acronyms

AD	Alzheimer's disease
MCI	Mild cognitive impairment
MCIc	Mild cognitive impairment converted
NC	Normal controls
MRI	Magnetic resonance imaging
fMRI	Functional magnetic resonance imaging
PET	Position emission resonance imaging
ROI	Region of interest
ADNI	Alzheimer's disease neuroimaging initiative
APOE	Apolipoprotein E
CDR	Clinical dementia rate
MMSE	Mini-mental state examination
LASSO	Least absolute shrinkage and selection operator
ELM	Extreme machine learning
RELM	Regularized extreme learning machine
FSASL	Features selection with adaptive structure learning
LLCFS	Local learning and clustering based feature selection
CFS	Pairwise correlation based feature selection
WHO	World health organization
CAD	Computer aided design
FLTD	Frontotemporal lobar degeneration
FT	Fourier transform
VBM	Voxel based morphometry
CSF	Cerebral spinal fluid
GM	Gray matter
WM	White matter

CAE	Convolution auto encoder
SPM	Statistical parameter mapping
GLM	General linear model
GRF	Gaussian random field
ADNI	Alzheimer's disease neuroimaging initiative database
RT	Repetition time
TE	Echo time
GUI	Graphical user interface
BOLD	Blood oxygen level dependent
CNN	Convolution neural network
MLP	Multi layer perceptron
DMN	Default mode network
SN	Saliency network
CEN	Central executive network
DLPFC	Dorsolateral prefrontal cortex
SMN	Sensory motor network
DAN	Dorsal attention network
FEF	Bilateral frontal eye field
IFG	Bilateral inferior frontal gyrus
LPFC	Bilateral lateral prefrontal cortex
FPN	Fronto-parietal network
LAN	Language network
pSTG	Posterior superior temporal gyrus
CER	Cerebellar network
DFS	Depth first search
BFS	Breath first search

FWHM	Full width at half maximum
TP	True positive
TN	Ture negative
FP	False positive
FN	False negative
CV	Cross validation
FC,	Functional connectivity
AUC	Area under curve
LDA	Linear discriminant analysis
ROC	Receiver operating characteristic
AAL,	Automated anatomical labeling
SDFN,	Sliding window-based dynamic functional network;
HMM,	Hidden markov model
GPU	Graphics processing unit
AI	Artificial intelligence

ABSTRACT

Comparative Study of Brain Networks Based on Graph Embedding Technique

Ji-In Kim

Advisor : Prof. Goo-Rak Kwon, Ph.D.

Dept. Info. and Comm. Eng.,

Graduate School of Chosun University

Recent studies suggest the brain functional connectivity impairment is the early event occurred in case of Alzheimer's disease (AD) and its prodromal stage mild cognitive impairment (MCI). This study aims to use functional magnetic resonance (fMR) images in order to classify patients having brain diseases that causes such brain dysfunction. The proposed method detects areas that interact in specific situations and configures them into brain networks with three different functional connectivity. Three different networks, "Large scale brain network", "Whole brain network", and "Combined brain network" are used. The Node2vec graph embedding algorithm transforms these brain networks having graph-based features into feature vector values compatible with general classifiers. Feature selection techniques are used so that a large number of features can be found to accommodate a small data set. Four different feature selection methods namely, LASSO, FSASL, LLCFS, and CFS are used. The selected feature vector values are fed into a single and multiple layer regularized extreme learning machine (RELM), respectively. This study shows that the graph-based function fMR images can be used for classification of brain diseases such as Alzheimer's disease and mild cognitive impairment. It is found that the classification performance changes depending on the size of the brain network. Additionally, experimental results show that even a small network contains information needed to classify patients.

초 록

그래프 임베딩 기술 기반 뇌 네트워크의 비교 연구

김지인

지도 교수: 권구락, 교수, Ph.D.

조선대학교 대학원 정보통신공학과

최근 연구에 따르면 경도인지장애뿐만이 아니라 알츠하이머병도 뇌에 연결 장애를 발생시켜 기능적, 구조적 연결성에 영향을 미친다고 보고되고 있다. 따라서 본 연구에서는 이러한 뇌 기능 연결 장애를 발생시키는 뇌질환의 분류를 위해 fMR 영상을 이용하여 환자의 분류를 하는데 그 목적을 두고 있다.

이를 통해 특정 상황에서 상호작용하는 영역을 검출하여, 이를 3가지의 다른 기능적 연결성을 가진 뇌 네트워크로 구성한다. 각각의 뇌 네트워크는 ‘Large scale brain network’와 ‘Whole brain network’, 이를 결합한 ‘Combined brain network’로 구성된다. 구성된 3가지의 다른 뇌 네트워크를 Node2vec 그래프 임베딩 알고리즘을 통해 그래프 기반의 특징을 일반 분류기와 호환되는 특징 벡터 값으로 변환한다. 변환된 특징 벡터는 적은 데이터셋에 대응하기 위해 많은 수의 특징을 찾을 수 있도록 다중 특징 선택기술을 사용한다. 각각의 특징 선택방법은 LASSO, FSASL, LLCFS, CFS이다. 선택되어진 특징 벡터 값은 RELM을 통해 단일 계층과 다중 계층에서 각각 분류되어진다.

이 연구는 fMR 영상의 그래프 기반 기능이 알츠하이머병과 경도인지장애와 같은 뇌질환을 분류에 사용할 수 있다는 것을 보여주며, 뇌 네트워크의 크기에 따라서 분류 성능이 변함을 알 수 있다. 또한 네트워크의 크기가 작아도 환자를 분류할 수 있는 필요한 정보를 가지고 있음을 알 수 있다.

I. Introduction

A. Thesis motivation

Alzheimer's disease (AD) which commonly appears in elderly people is a progressive neurodegenerative disease [1]-[4]. The neural dysfunction begins far earlier the visible clinical symptoms such as, progressive cognitive impairment are manifested. These symptoms are usually noticed after the age of 65. With the increasing number of population of elderly people, the number of AD patients is increasing requiring more care taker thus, increasing medical expenses [5]. In such scenario, the accurate diagnosis of disease at its early stage can to slow down the effect of disease thus reducing the significant amount of economic burden to the society created by this disease.

In recent years, several computer aided diagnosis (CAD) based studies have been conducted to classify AD and mild cognitive impairment (MCI) from healthy controls. With the availability of recent neuroimaging technology, promising result is obtained in the early and accurate detection of AD [6]-[8]. The study of progression of disease and early detection is carried out by using different imaging models. Structural magnetic resonance imaging (sMRI) [6] is the most commonly used imaging system for study of AD. Similarly, functional magnetic resonance imaging (fMRI) [7] and positron emission tomography (PET) [8] are also used for the study of disease progression and its effect.

Due to broader area of coverage of this technique, several research organizations and researchers are working on CAD based AD/MCI diagnosis. The development of CAD based classification of AD/MCI from HC together the neuroimaging biomarkers will provide the concerned investigator and scientists to focus on treatments on preclinical stage, thus assisting diagnose the disease in its early stage before symptoms appears.

B. Research objectives

Formation of two abnormal proteins plaques and tangles in the prefrontal and medial-temporal lobes are the pathological hallmarks of AD. Structural changes, such as neuronal death and WM degeneration characterized by cortical atrophy take place primarily in medial temporal region and gradually spread over the entorhinal cortex, the limbic system and eventually affecting the neocortical regions through synapses and neural connections. In recent years with the development of advanced neuroimaging technology detailed study of neurological diseases has come under the reach of researchers resulting early and accurate detection of AD.

Conventional diagnosis is carried out based on the neuro physiological examinations using different imaging technology such as MRI, fMRI, PET and SPECT images and series of test on memory impairment, thinking skills and other clinical symptoms [6]-[8]. Studies suggest that memory impairment is most prominent symptoms due degeneration in medial temporal cortex [9]. With the progression, the disease affects gradually in entorhinal cortex, the hippocampus and limbic system and finally at neocortical areas [10]. This results in severe impairment in logical reasoning, planning and cognitive tasks.

Study of medial temporal atrophy usually provides the evidence of progression of AD. Thus the studies are carried out by measuring the atrophy in terms of voxel based, vertex based and region of interest (ROI) based approaches. In AD and MCI subjects, atrophy of medial temporal lobe structures has been discovered in studies carried out based on ROI based MRI volumetric methods [11], [12]. This atrophy in crucial areas of brain such as hippocampus, parahippocampal gyrus and the amygdala contributes to differentiate the MCI and AD subjects from controlled subjects [13], [14]. The voxel based morphometry (VBM) is alternative method to ROI based method to assess patterns of cortical atrophy. VBM based method is less laborious compared to ROI based method thus used as almost universal global volumetric method to measure variances in the regional concentration of grey

matter [15]. Studies based on this method have revealed reduced grey matter volume in different regions of brain in AD and MCI subjects compared to HC. These areas include medial temporal lobe, frontal lobe, and posterior cingulate gyrus [16]. Above mentioned studies are carried out using structural MR images.

On the other hand, fMRI detects the changes in blood oxygenation and flow of brain [17], [18]. Brain activity is mapped in terms of blood-oxygen-level dependent (BOLD) contrast. Blood flow to any particular region of the brain increases with the increase in activity in that region. The fMRI provides measurement on involvement of different brain regions in particular brain activities [19]. Structural MRI primarily focuses to reveal the anatomical information of brain tissues, while the fMRI shows the functional brain activities. Thus, more insight on the abnormalities of functional brain connectivity caused by the progression of MCI and AD [20], [21] can be obtained. Chen et al. [18] performed linear regression analysis to analyze the relationship between changes in network connectivity. The Pearson product moment correlation coefficients of pairwise of 116 ROIs were used as feature. Similarly, Wang et al in [19] used fMRI based feature to classify the AD from HC and MCI. The correlation/anti-correlation coefficients of two intrinsically anti-correlated networks were used as feature with Pseudo-Fisher Linear Discriminative Analysis (pFLDA) classifier. The outcome of all above mentioned studies supports the hypothesis that the cognitive deficiency and decline in AD and its prodromal stage are caused by the connectivity disruptions of the brain networks.

Additionally, various studies show that the connectivity of networks which are active during passive or resting state of the brain are disrupted due to AD [22]. This network includes default mode network (DMN), central executive network (CEN), and salience networks (SN) [23], [24]. Although, changes are often seen in DMN, SN and CEN across the spectrum of AD and MCI Rs-fMRI results have shown that people who have aging or MCI also exhibit the functional connectivity alterations in these large scale network. Similarly, current studies demonstrate that

functional connectivity alterations are visible not only in DMN, but also in SN and motor networks [25]. Thus, other networks including DMN, SN, sensory motor network (SMN), dorsal attention network (DAN), and auditory network and visual network to classify AD from HC and MCI are included in the proposed study. Collection of this widespread brain network is known as core large-scale brain network.

Thus, the primary objective of this study is to extract the features from fMR images in terms of correlation matrix between different ROIs which represents the brain network. The brain network includes whole brain network, core large-scale brain network and combined network. Features of brain network is in the form graph where vertices are brain regions and edges are correlation between these vertices. Since the graph has non-euclidean characteristics. Conventional machine learning algorithms works only on data having Euclidean or grid-like structure. In order to remove the invariances of these structures graph embedding is used. The graph embedding transforms graph data to a vector or set of vectors to overcome. The relevant graph information together with the graph topology, vertex-vertex relationship, is captured by embedding. In this study, node2vec method is used. Next, only the relevant features are selected and finally the multi-layer-regularized extreme learning machine (ML-RELM) classifier is used to classify the AD subjects from NC and MCI.

C. Thesis contribution

Previous studies have revealed the alterations of white matter (WM) and gray matter (GM) microstructure in AD and its prodromal state amnesic MCI. In general, these alterations can be studied comprehensively by modeling the brain as complex network. In this study, a new approach is developed that uses graph based features of brain network using functional magnetic resonance (fMR) images for classification of AD and MCI from HC. For the better understanding of

progression of disease, three different brain network models namely large scale brain network of size 32×32 , whole brain network of size 132×132 and combined brain network of size 164×164 are utilized for the study. Additionally, multiple feature selection techniques is used to cope with the smaller number of subjects with larger number of feature representations. In this study, classification test is performed with two variants of regularized extreme learning machine(RELM), which are single hidden layer RELM and multiple hidden layer RELM.

D. Thesis organization

Thesis is structured as follows. Chapter II presents overview of the AD, Socio-economic impact of AD. Different imaging models such as sMR and fMR image processing and graph feature extraction techniques, feature selection and reduction techniques and theoretical formulation of extreme learning machine classifier will be discussed in chapter III. Experimental results will be discussed in chapter IV. Similarly, limitation of this work is presented in chapter V. Finally the conclusion of the thesis will be presented in chapter VI.

II. Background

A. Alzheimer's disease

Alzheimer's disease (AD) which causes the dementia accounts for 60-80% of dementia [1]. Dementia is mainly responsible for the memory loss and decline other cognitive abilities thus interfering the daily life. Common symptoms of AD appears as difficulty to form memory of newly learned information.

The brain, like the rest of the body, changes as person grows older. Some people notice some slowness in thinking and occasional difficulties with memory. However, significant memory loss, confusion and other severe changes in the way the brain works could be the signs of dementia. AD symptoms usually begin in the brain area that controls learning. As it advances through the brain, AD can lead to increasingly severe signs, including disorientation, mood shifts, and behavioral changes; confusion about events, places, and time; and false suspicions about other people and things.

AD changes typically begin in the part of the brain that affects learning. With the advance of AD through the brain, it results in increasingly severe symptoms including disorientation, mood changes, and behavioral changes; confusion about time, place, and events; suspicions about care-givers; and more severe memory loss and behavior changes; and difficulty in communication, swallowing and other activities including walking. People who have memory loss or other possible symptoms of AD may not recognize they have a problem. Family or friends might be able to spot the signs.

Earlier identification and interventions are improving dramatically, plus treatment options and sources of information are becoming available. In AD, the disease progression is characterized mainly by three stages. Each stage reveals unique symptoms varying in severity. After identifying the current stage, the doctor can predict the next stage and follow up on the symptoms, and recommends a

treatment plan accordingly. The symptoms of the disease worsen over time, and the rate at which it advances varies. Once a person is diagnosed as AD, he or she can live from four to eight years. However, depending on other factors, the person can live up-to 20 years.

In AD, the changes in the brain begin years before any symptoms appear. This period can last for many years and can be divided into three stages: mild AD, moderate AD, and severe AD.

B. Early-stage Alzheimer's (mild) disease

During the early stages of AD, a person can still function independently. He or she may drive, work, and participate in social activities, and yet the person may suffer from memory lapses, such as forgetting familiar words or where everyday objects are.

Symptoms may not be widely evident at this stage, but family members and friends may notice, and a doctor can spot certain symptoms using certain diagnostic tools.

- Common difficulties include :

- Having difficulty to come up with the proper word or name.
- Having difficulty to remember newly introduced people's name.
- Having difficulty to remember the material that was recently read.
- Having difficulty to place valuable object properly.
- Having difficulty to adequate plan or organize.

C. Middle-stage Alzheimer's (moderate) disease

Middle-stage Alzheimer's may continue for several years. As the disease progresses, the person will require more care. A person with mid-stage dementia is likely to confuse words, become frustrated or angry, and behave in unexpected

ways, such as refusing to bathe. The person may also have difficulty expressing thoughts or performing routine tasks without assistance as a result of nerve damage in the brain.

- **Symptoms, include :**

- Having difficulty to remember events or their personal history.
- Feeling moody or withdrawn, especially when socially or mentally challenged.
- Remembering little about themselves, like their address or telephone number.
- Having doubt on their current location or the current day.
- Needing help choosing clothes appropriate for the season or the occasion.
- Having difficulty controlling their bladder and bowels.

Change in sleeping pattern, having sleeping pattern just opposite to normal people. Wandering and getting lost more frequently. Persons suffering from AD will exhibit signs of personality and behavioral changes, such as suspiciousness, delusions, or compulsive, repetitive behaviors, such as hand-wringing and tissue shredding. During the middle stage, the person with AD can still participate in daily activities, if necessary.

D. Late-stage Alzheimer's (severe) disease

At the final stages of the disease, dementia symptoms become severe. Patients are unable to react to their surrounding, to continue the chat, and, at last, to manage the movements. They may still be able to speak, but communicating pain becomes increasingly difficult.

- **At this stage, individuals may :**

Personal care is needed round-the-clock.

The individual loses awareness of the recent past and of their surroundings.

Physical abilities change, including walking, sitting, and eventually swallowing.

- Having more complication in communication.
- Become more prone to infections, particularly pneumonia.
- The individual with AD cannot start the engagement as much during the late stages, however they can still benefit from interaction in a way that is appropriate for him or her, whether it be relaxing music or gentle touch.

E. Economic impact of dementia

Dementia affects someone worldwide every 3 seconds. By 2020, there will be over 55 million people living with dementia worldwide [44]. In 2030, there will be 78 million and by 2050, there will be 139 million [45]. The majority of the increase will be in developing nations, where 60% of people with dementia already live; by 2050, this number will rise to 71%. China, India, and their neighbors in the south Asian and western Pacific region have the fastest-growing elderly populations. Demographic aging manifests itself as a global trend that is a hallmark of improved health care in the past century. Dementia principally affects older people, but there is a growing awareness that cases begin even before the age of 65. As more and more people live longer and healthier lives, the world population is becoming more and more elderly. There are over 10 million new cases of dementia diagnosed each year, which means one new case appears every 3.2 seconds.

In 2020, the WHO published a global status report on the public health response to dementia, which updated these figures. In 2015, the total estimated global cost of dementia was \$818 billion, representing 1.09% of global GDP [46].

Globally, dementia costs are now over US\$1.3 trillion, and they are expected to rise to US \$2.8 trillion by 2050, including costs associated with informal care, direct costs of social care, and direct costs of medical care [47]-[49].

In terms of global dementia costs, direct medical care costs make up roughly

20%, while social sector costs account 40% and informal care costs make up for roughly 40%. Informal care contributes most to dementia costs in the African region and least to those in North America and Western Europe. Costs in the social sector are the opposite [50].

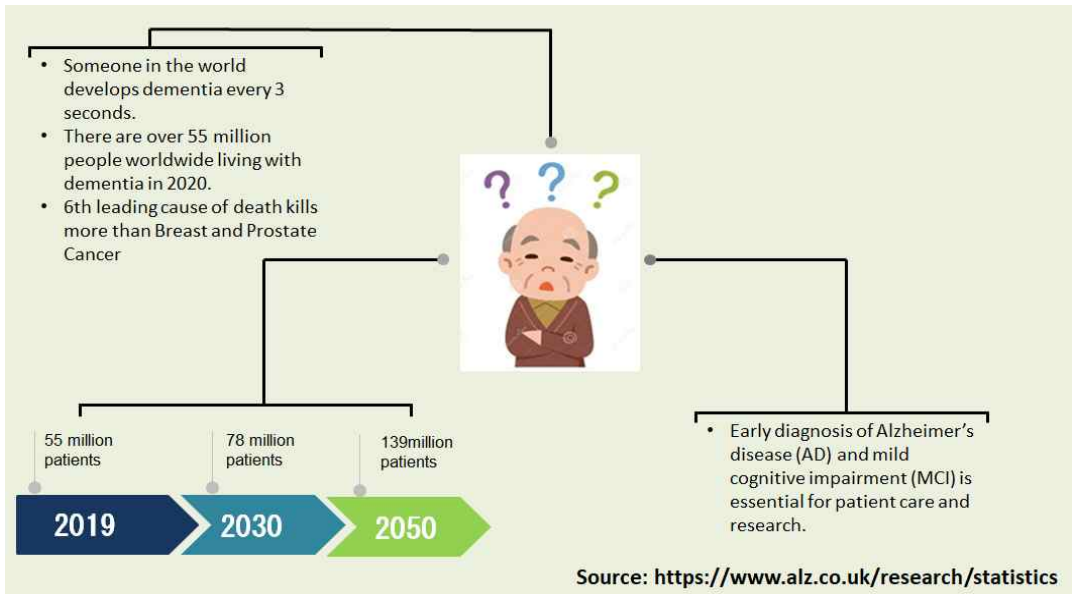


Figure 2.1. Economic burden of Alzheimer's disease.

F. Diagnosis of Alzheimer's disease

Studies have shown that majority of people with dementia do not have a formal diagnosis. It has been estimated that only 20-50% of dementia cases are recognized and documented in primary care in high income countries. This 'treatment gap' appears to be significantly higher in countries with low and middle income economy. A report suggests in India, 90% of cases remain undiagnosed.

Extrapolating these statistics to other countries, it will be found that roughly three quarters of people with dementia do not have a formal diagnosis, and are therefore unable to get treatment, care, and organized support that getting a formal diagnosis can offer. Early diagnosis and early intervention are key to closing the treatment

gap, according to the World Alzheimer Report 2011. As part of the diagnosis of Alzheimer's dementia, doctors conduct tests to assess memory impairment and other thinking skills, evaluate functional abilities, and identify behavioral changes. A series of tests is also carried out to rule out any other causes of impairment.

As part of the diagnosis, a doctor who specializes in brain conditions (a neurosurgeon) or a doctor who treats the elderly (a geriatrician) will review the patient's medical history, medication history, and symptoms, as well as perform various tests.

- Doctor will evaluate :

- Whether there is an impairment of memory or thinking skills
- Whether there is a change in personality or behavior
- The degree of memory or thinking impairment or change
- How thinking problems interfere with a person's daily functioning
- The causes of the symptoms

Doctors may go for additional laboratory tests, brain imaging tests and memory tests. These tests can provide doctors with valuable information about a patient's illness, including ruling out a wide range of other conditions.

1. Ruling out other conditions

Doctors perform a physical examination and check for other medical conditions that may be causing or contributing to symptoms, such as strokes, Parkinson's disease, depression, or other illnesses.

2. Assessing memory problems and other symptoms

To assess symptoms, doctor may ask the patient to answer questions or perform tasks related to their cognitive abilities, such as memory, abstract thinking,

problem-solving, language usage and related skills.

(a) Mental status testing :

Doctors may administer mental status tests to assess thinking (cognitive) and memory skills. These tests are used to evaluate cognitive impairment.

(b) Neuropsychological tests :

A neuropsychologist will evaluate the patient for brain conditions and mental health conditions, including extensive tests to test the patient's memory and thinking skills. These tests allow doctors to determine if the patient has dementia and if they can safely perform daily tasks such as driving and managing money. They give a wide range of information on what the patient is still able to do and what he/she may have lost.

(c) Interviews with friends and family :

A doctor may ask a family member or friend about the patient and their behavior. Doctors seek out details that don't correspond with the patient's previous level of ability.

It is easy to explain how thinking (cognitive) skills, functional abilities, and behaviors have changed over time by talking to a family member or friend. Doctors can diagnose Alzheimer's disease based on a combination of clinical assessment, physical examination, and setting (age and duration of progressive symptoms). It may, however, be necessary to order other tests when the diagnosis is unclear.

3. Laboratory tests

A patient may undergo laboratory tests to rule out conditions, such as a thyroid disorder or a vitamin B-12 deficiency, that cause symptoms similar to those of Alzheimer's dementia.

4. Brain-imaging tests

The cause of Alzheimer's dementia is degeneration of brain cells. This degeneration may be detected in various ways in brain scans, but is not sufficient for diagnosis. Doctors are unable to diagnose the disease with scans since normal age-related changes in the brain overlap with abnormal changes.

(a) Brain imaging tests helps :

To ensure the diagnosis is not caused by other conditions, such as hemorrhages, brain tumors, or strokes.

To differentiate between different types of degenerative brain disease.

To determine the degree of degeneration to apply the brain-imaging technologies most often used to diagnose the condition:

(b) MRI :

MRI is a noninvasive diagnostic test that provides a detailed image of your skull and brain soft tissues by using powerful radio waves and magnets to create a detailed view of brain.

(c) Computerized tomography (CT) :

A CT scan uses X-rays to obtain cross-sectional images of the brain.

(d) PET :

PET scans utilize radioactive tracers to detect various substances in the body. The most commonly used PET scan is the fluorodeoxyglucose(FDG) PET scan, which can identify areas of the brain with decreased glucose metabolism. Metabolic changes can be used to distinguish between different kinds of degenerative brain diseases. PET scans are now available to detect clusters of amyloid protein (plaques), which can cause Alzheimer's disease; however, they are

quite expensive.

G. Computer aided diagnosis

Computer-aided diagnosis(CAD) is the use of a computer-generated output to assist clinicians in making a diagnosis. In general, CAD is used to detect diseases automatically using multiple forms of medical imaging. With CAD output, radiologists can improve the accuracy and consistency of their diagnosis by improving image interpretation.

(a) Computer-aided diagnosis systems typically employ four general schemes :

- i. **Preprocessing** : This step involves processing the data to a sufficient quality so that it can be recognized by the computer. Filters are applied, as well as window-level adjustment techniques, for image contrast. The aim is to reduce noise and artifacts.
- ii. **Segmentation** : In this step, the body is divided into segments using information from an anatomical database. The program uses data from this database to identify whether the areas of interest on the image are masses, microcalcifications, or tissues.
- iii. **Feature extraction** : The region of interest is examined for features such as morphological features, gray levels, and textures.
- iv. **Feature classification** : An algorithm is applied to determine: (a) whether an identified structure is benign or malignant and (b) the difference between true lesions and normal anatomical structures.

(b) The effects of CAD on quality and efficiency of services are as follows [4] :

- i. CAD system can increase the efficiency of the time process in certain cases

as it provides the ability to reduce the time that is required for image diagnosis. Signal processing is used to remove noise or extract parameters using complex mathematical formulas that no human can understand.

- ii. Due to fatigue, boredom, and environmental factors, human observers are prone to making errors when monitoring the status of a patient for extended periods. A computer can, however, be designed to record all episodes and transients in the signal mathematically and consistently.
- iii. CAD can enhance accuracy in image diagnosis by detecting diseases that cannot be seen by human eyes, especially those that are too small or early in their development. Early diagnosis can result in a better patient outcome.
- iv. By using CAD, a highly experienced radiologist can reduce his/her workload, as the system can improve the accuracy and speed of interpretation. Thus, less stress and more time will be available for highly experienced radiologists, resulting in higher productivity.

To summarize, CAD is a quantitative method that is consistently applied to repetitive tasks. Signal analysis alone cannot be used to make a diagnosis and must be complemented by other information that a physician can obtain, such as a patient's general physical appearance, mental state, and family history. Computer-aided diagnosis is, therefore, considered a secondary source of diagnosis. The physician is still the primary source of diagnosis. A second opinion is provided by the CAD system.

H. Machine learning in computer-aided diagnosis

The classification phase in the CAD system is concerned with making inferences from extracted features in order to produce a diagnosis about the input image. This is a kind of supervised learning technique in which the computer makes decisions about the input image based upon what it observes in the input. Various machine learning algorithms are now used to develop high-performance CAD. Similarly,

several CAD systems based on medical imaging are developed to allow detection of the AD at early stages by comparing normal and diseased participants using the analysis of certain features in brain images. The analysis of images can be done using several methods. One of the important method is the statistical parametric mapping (SPM) software that models data of individual voxel. Other method for the analysis of structural MRI images is executed by VBM. Different machine learning algorithms have been introduced to classify AD patients and elderly control individuals. Support vector machines (SVM), artificial neural network (ANN), logistic regression (LR) and other ensemble classifiers are the commonly used classifiers [50]. Among them, SVM and its variants become popular due to its relatively good classification accuracy and ability to combine many features in higher dimensional space [51].

Klöppel et al. [52] applied linear support vector machines to distinguish pathologically proven AD subjects from cognitively normal elderly subjects. They used the grey matter segment of T1-weighted MR scans as a biomarker. This literature is focused on binary classification of individuals (i. e. confirmed AD vs controlled, mild AD vs controlled and AD versus frontotemporal lobar degeneration (FTLD). The classification results demonstrate that their proposed model could classify confirmed AD from controlled with good accuracy compared to FTLD from controlled subjects. Diagnosis approaches described above are based on single feature modality. However, for better analysis different biomarkers can be used which may carry more complementary information. Therefore instead of depending on one feature, one can combine multiple features to improve the classifier performance [53]. The general framework for data fusion can be developed by using the multiple kernels which come from different sources of feature spaces. Fan et al [54] combined structural and functional images in high dimensional pattern classification framework using SVM. MR and PET images were used to identify the complex spatial patterns of brain atrophy and blood flow for 30 individuals. Their combined feature model could obtain better accuracy in contrast to the

maximum classification rate achieved from MRI alone. Multiple kernel SVM simultaneously learns the kernel and related predictor parameters [55, 56]. Hinrichs et al. [57] showed promising evaluations on ADNI data set in terms of goodness of approach with respect to outlier detection. They assigned each feature as one or more kernels. This algorithm utilized alternative minimization to identify outliers and discount their effect on the classifier. Zhang et al. [8] proposed a multimodal classification approach based on multiple-kernel SVM based on the biomarkers including sMRI, PET and CSF to discriminate AD (or MCI) and normal controls (HC) subjects. From the binary classification (i.e., AD vs HC and MCI vs HC) results, their model obtained a good results in terms of accuracy for AD classification and promising accuracy for MCI classification. Liu et al [9] proposed a fourier transform (FT) based multiple kernel learning framework to combine multi-modal features in the primal space. Instead of solving the problem in the dual space, this approach conducts FT on the Gaussian kernel, and then computes the mapping function. More multiple kernel based SVM approaches can be found in [58], [59].

Recent studies [60], [61] indicate that machine learning algorithm such as AdaBoost may be used for automatic feature selection for several imaging applications. Thus it is likely to offer advantage in automatically finding good features for classification as they can select informative features from a potentially very large feature pool [62]. There is no need for experts to choose informative features based on knowledge of every classification problem. AdaBoosts have been used not only as feature selection tool but also as classifier which performs classification in combination with other machine learning algorithms [63], [64]. It iteratively constructs a set of component classifiers. In each iteration, an importance weight is assigned to training samples and misclassified training sample's weight will be increased. At the same time the correctly classified sample gets the decreased weight value. The final result of AdaBoost is obtained according to the voting by all the component classifiers. AdaBoost is very successful because of its

good generalization performance as reported in many studies. Savio et al [65] proposed AdaBoost based approach for the detection of AD using MR images. Authors used the VBM based approach to for the feature extraction processes. Adaptive boosting was performed in conjunction with SVM for each VBM detected cluster. AdaBoostSVM methods improved remarkably the results, mainly the sensitivity of the classification models. Farhan et al. [66] proposed ensemble based AD diagnosis using structural features of brain images. They selected volume of GM, WM, and CSF and size of hippocampus as features. Boosting was performed in conjunction with three different classifiers SVM, Multi-layer Perceptron (MLP), and Decision Tree. Morra et al. [67] presented a comparative analysis between hierarchical AdaBoost, SVM with manual feature selection hierarchical SVM with automated feature selection (Ada-SVM), and a publicly available brain segmentation package (FreeSurfer).

In recent years, there have been developments in hierarchical or deeplearning as a modern branch of machine learning. This technique has led to improvement in performance on variety of problems, such as object recognition, speech recognition and natural language processing. Among various type of deep learning techniques, convolutional neural network (CNN) has been extensively used. This approach of learning is inspired by the human visual perception mechanism however follows same principal of learning principal of classical neural network.

Payan et al. [68] used 3D CNN architecture to distinguish AD and MCI from healthy control subjects which uses 3D MR images as input. The whole process was implemented in two stages. In first stage sparse auto-encoders were used to learn convolution filters. Discriminative features of MR images capturing anatomical variations due to AD are extracted automatically in this stage. Next, the pre-trained 3D convolution auto encoder (CAE) encoding layers are stacked with the fine tuned fully-connected layers. Accuracy rate of 95.39% in distinguishing AD from NC subjects was achieved.

Similarly, Hosseini-Asl,et al [69] proposed AD diagnostic framework using deeply

supervised adaptive (DSA) 3D CNN. Source-domain trained 3D-CAE is used to extract features of brain MRI and deeply supervised target-domain-adaptable 3D-CNN performs task-specific classification. Using a stack of unsupervised CAE with local nodes and shared convolutional weights, they extract local features from 3D images with possibly long voxel-wise signal vectors. Each input image is reduced hierarchically based on the hidden features (activation) of each CAE for training the subsequent layer.

With the advent of artificial intelligence (AI) based methodologies such as convolutional neural networks and simultaneously-acquired multimodal MRI and PET scanning, several recent works have shown their ability to distinguish AD subjects from normal healthy control data for the same age groups. Using state-of-the-art deep learning-based pipelines, which are executed on a GPU-based high performance-computing platform, the data are strictly and carefully preprocessed. Scale and shift invariant low- to high-level features are obtained from a high volume of training images using CNN architecture. However, majority of current CNN based studies are focused only on grey matter volume features of hippocampus and other regions of brain. Although GM volume of the hippocampus has been an important biomarker of medial temporal lobe neurodegeneration, alterations of hippocampal WM pathways is often observed in AD. Several studies also reveal the alterations in widely distributed functional and structural connectivity pairs are prevalent in AD and MCI. Thus, the study of functional connectivity can therefore help researchers and neurologists understand the progression of disease.

I. Human brain network and graph theory

Human brain contains over nearly 86 billion neurons that communicate with each other by way of around 150 trillion synaptic connections. The study of the human brain as a complex system has grown remarkably as neuroscientists strive to understand the information underneath cognition, behavior, and perception. The

human connectome, which is mapping patterns of the connectivity in the human brain has gained increased attention in the field of human neuroscience. As connectome provides comprehensive description of the networks of elements and connections forming the brain network. A brain can be modeled as a graph of nodes connected by edges. Mathematically, a graph G can be well described by the set of vertices A and edges E .

$$G = (A, E). \quad (2.1)$$

In brain network, brain regions are usually represented by nodes and links. Links are represented by edge. Links represents anatomical, functional and effective connections. Anatomical links correspond to WM that connects between pair of brain regions. functional connections correspond to magnitude of temporal correlational in activity and may occur between anatomically unconnected regions. Effective connections represent direct or indirect causal influences of one region on another and may be estimated observed perturbations.

The functional and effective connectivity networks are constructed from the series of brain dynamics simulated on the anatomical network. The functional network represents the patterns of cross correlation between BOLD signals estimated from these dynamics. All networks are represented by connectivity or adjacency matrix. Rows and columns in these matrices denote nodes, while matrix entries denote the links.

The nature of nodes and links in individual brain network is determined by combinations of brain mapping methods, anatomical parcellation schemes and measures of connectivity.

Nodes should ideally represent brain regions with coherent patterns of extrinsic anatomical or functional connections. Type of connectivity, anatomical, functional or effective and measure specific features of connectivity links are also differentiated on the basis of their weight and directionality.

Binary denotes the presence or absence of connections, weighted links also contains the information about the connection strength. Weights in anatomical networks represent the link density while weights in functional and effective networks represent magnitude of correlational or causal inferences.

J. Functional connectivity of brain

A functional network is a measure of how brain regions interact with each other. Non-invasive brain activation can be identified with fMRI, and functional networks refer to the temporal correlation between BOLD signals from spatially distant regions. The map of this connectivities with respect to each ROI is shown in Figure 2.2. In fMRI studies, functional connectivity methods are broadly divided into two groups: model-based and model-free.

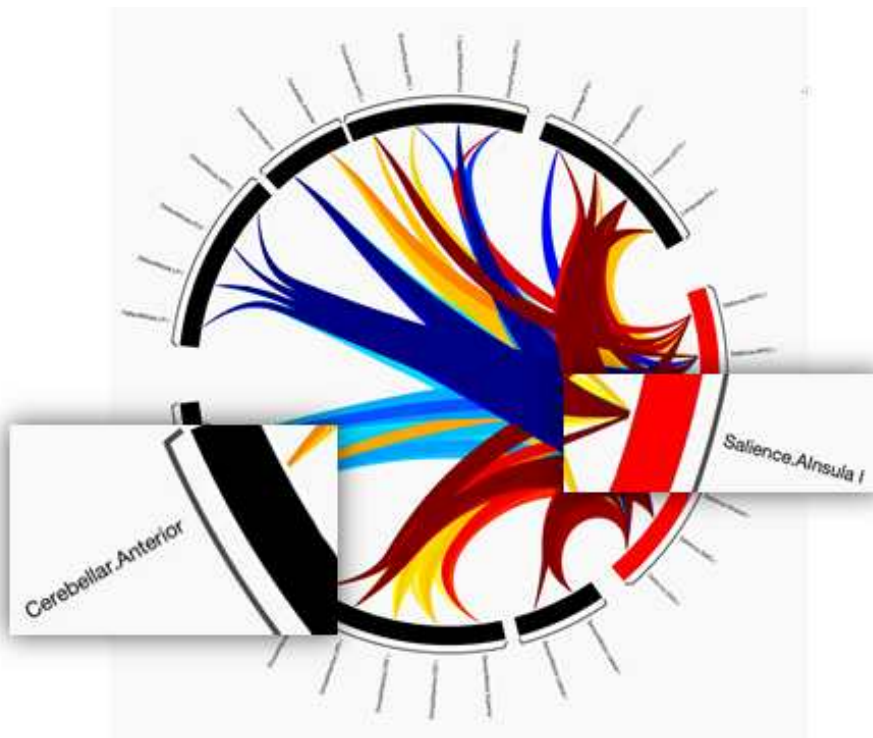


Figure 2.2 Brain Connectivity Map, figure obtained from CONN tool [78].

1. Model-based methods

According to model-based methods, brain connectivity networks are identified by selecting one or more "seed" regions and determining whether they are connected by a linear path using predefined criteria. Although these methods have common use and simple explanation in detecting the functional connectivity, the need for prior knowledge, reliance on seed selection, and inability to detect non-linear interactions limit the discovery of all plausible functional architectures.

2. Cross-correlation and coherence

Functional connectivity is most commonly evaluated by cross-correlation analysis, which measures the correlation between BOLD signals in two brain regions. Using this method to calculate the correlation between two series at any lag involves a high degree of computational complexity. In many fMRI studies, this drawback has been overcome by computing only the correlation with zero lag since the hemodynamic response of blood is short lived. Correlations are also affected by the shape of the hemodynamic response function (HRF), which can vary across individuals and brain regions. In addition, a high correlation can be observed in regions with practically no fluctuations in blood flow. Uncontrolled physiological noise in the brain can also result in high correlations between brain regions.

3. Statistical parametric mapping (SPM)

The SPM is one of the model-based approaches used to identify regional effects in neuroimaging data, including fMRI and PET, by combining the general linear model (GLM) and a Gaussian random field (GRF). By applying GRF theory to the problem of multiple comparisons, statistical inferences can be addressed over a volume of brain data, a similar approach to the Bonferroni correction for discrete data.

4. Model-free methods

Unlike seeds-based methods, model-free methods do not need seeds selection. Moreover, model-free methods may be useful in studies with no temporal or spatial patterns, and to quantify non-linear neuronal interactions.

K. Construction of functional brain network

Figure 2.3, shows the major steps to construct a complex network from fMRI. The acquired fMRI data is initially subjected to a number of pre-processing steps, such as slice timing correction, realignment, image co-registration, and normalization based on segmentation. Depending on the order and choice of preprocessing steps, the complexity of the final graph is affected. In the next step, the entire brain is divided into several cortical and subcortical anatomical units using a parcellation scheme such as anatomical automatic labeling atlas to explore the large-scale brain network.

Afterward, each parcel's time series is extracted by averaging the time series of all voxels within that given area. The cross-correlation is used to determine the pair-wise relations between the time series of brain regions. This cross-correlation represents the functional connectivity network. In the following step, a binary connectivity matrix is computed by thresholding the values of the correlation matrix. Once the connectivity matrix has been computed, it is used as a graph.

Graph theory is a mathematical approach to study complex networks. Network is constructed of vertices which are interconnected by edges. Vertices in our case are brain regions. Graph theory is widely used as tool for identifying anatomically localized sub-networks [26]. These networks are associated with neuronal alterations in different neurodegenerative diseases. In fMRI images, graph represents causal relations or correlations of different nodes in constructed networks. However, the brain network built by graph has non-Euclidean characteristics. Thus, applying machine learning techniques to analyze the brain networks is challenging. Graph

embedding is used to transform graphs to a vector or set of vectors to overcome this problem. Embedding captures the graph topology, vertex-vertex relationship, and other relevant graph information. In the current study, node2vec graph embedding technique is used to transform vertex and edge of brain network graph to feature vector. With the help of this model the networks of brain have been analyzed and classified from fMRI data into AD, MCI and HC.

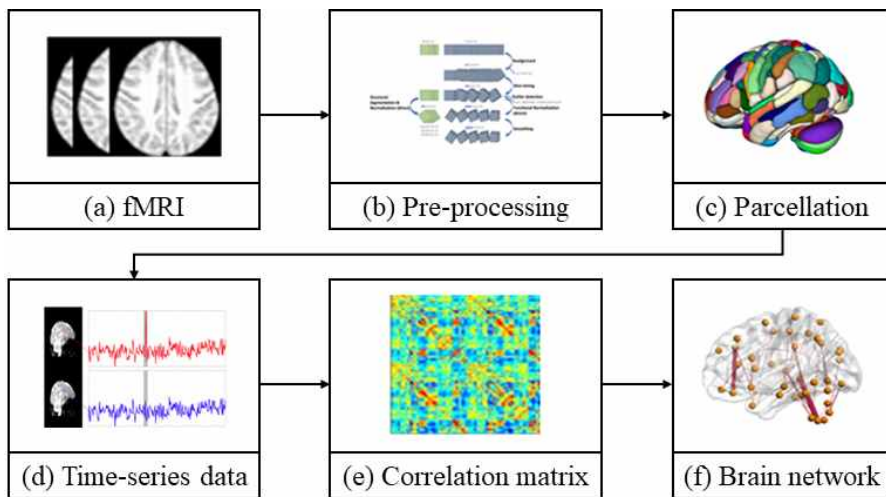


Figure 2.3 Schematic representation of brain network construction fMRI data. (a) fMRI data (b) processing (c) division of the brain into different parcels (d), several time courses are extracted from each region (e) can create the correlation matrix (f). and the corresponding functional brain network are constructed.

III. Connectivity based diagnosis of Alzheimer's disease

A. Method and approach

1. fMRI Dataset

In this study, dataset from Alzheimer's disease neuroimaging initiative database (ADNI, www.loni.ucla.edu/ADNI) [27] is used. The ADNI database was launched in 2004. The database consists of subjects of age ranging from 55-90 years. The goal of ADNI is to study the progression of the disease using different biomarkers. This includes clinical measures and assesses of the structures and functions of brain for the course of different disease states.

All participants were scanned using 3.0-Tesla Philips Achieva scanners at different centers. Same scanning protocol were followed for all participants and the set parameters were ratio of Repetition Time (TR) to Echo Time (TE) i.e., $TR/TE = 3000/30$ ms, 140 volumes, also voxel thickness as 3.3 mm, acquisition matrix size = 64×64 , 48 slices, flip angle = 80° .

Similarly, 3D T1-weighted images were collected using MPRAGE sense2 sequences with acquisition type 3D, field strength = 3 Tesla, flip angle 9.0 degree, pixel spacing X = 1.0547000169754028 mm; Pixel Spacing Y = 1.0547000169754028 mm, slice thickness = 1.2000000476837158 mm; TE = 2.859 ms, inversion time (TI) = 0.0 ms, TR = 6.6764 ms and weighting T1. Subjects are selected as specified in Table 3.1

2. Subjects

Total 95 subjects were selected from ADNI2 cohort. The purpose of ADNI2 is to examine how brain imaging and other biomarkers can be used to measure the development of MCI and AD at its early stage. In addition to its predecessors ADNI1 and ADNI-GO, ADNI2 examines the correlations between brain imaging

results, biomarkers, and clinical, cognitive, and genetic features spanning the entire spectrum of AD as it progresses from normal aging through dementia. A major aim of ADNI2 is to gain a deeper understanding of the sequence of events leading to MCI and Alzheimer's disease, develop new methods for early detection. The ADNI2 selects and categorizes participants in specific group based on certain inclusion criteria. The criteria are well defined in [27]. In this study, subjects were selected according to availability of both MRI and fMRI data. Thus, the subjects with following demographic status out of all available data in ADNI2 cohort were selected in our study.

Table 3.1 Summary of subject's demographic status.

Number of Subjects	HC (n=31)	MCI(n=31)	AD(n=33)
	Mean(SD)	Mean(SD)	Mean(SD)
Age(years)	73.9 ±5.4	74.5±5.0	72.7 ±7.0
Global CDR	0.04±0.13	0.5±0.18	0.95±0.30
MMSE	28.9±1.65	27.5±2.02	20.87±3.6

- 31 HC subjects: 14 males, 17 females; age \pm SD = 73.9 \pm 5.4 years with the mini-mental state estimation (MMSE) score of 28.9 \pm 1.65 and the range was 24-30.
- 31 MCI subjects: 17 males, 14 females; age \pm SD = 74.5 \pm 5.0 with the MMSE score of 27.5 \pm 2.02, and range was 22-30.
- 33 AD subjects: 13 males, 18 females; age \pm SD = 72.7 \pm 7.0 with MMSE = 20.87 \pm 3.6, and the range was 14-26.

3. Data preprocessing

Data processing subordinate for the resting state fMRI were done via CONN toolbox. CONN is a MATLAB-based software for the calculation, demonstration,

and investigation of brain network connectivity using fMRI. Connectivity assessment technique includes seed-to-voxel network maps, ROI-to-ROI connectivity matrix, graph characteristics of network systems, brain interconnection, inherent connectivity, generalized psychophysiological interaction models and other voxel-to-voxel measures. It is available for rs-fMRI and task-related plans. It covers the whole pipeline starting from basic fMRI information to proposition testing, including spatial coregistration, scrubbing, aCompCor technique for management of physiological and movement discomfit, first-level connectivity assessment, and second level arbitrary effect examinations. So, to do connectivity analyses utilizing CONN you will require the following functional and structural data.

Functional data is either a resting state or a task plan that can be examined. Whereas, structural data is defined as at least one anatomical volume for each patient (this is utilized generally for plotting graphs in addition to determining the white/CSF/gray masks utilized in the aCompCor confound elimination technique).

4. Data processing

The functional and structural data of AD, MCI and HC patients were preprocessed and examined utilizing the CONN tool to produce a connectivity matrix that gives data about the association between ROIs present in a specific region of the brain. The preprocessing and examination of AD, MCI and HC patients were done independently.

The process for computing fMRI measures involves five steps namely setup, preprocessing, denoising, analysis, and results exploration. The steps associated with the calculation of fMRI measures.

(a) Setup

In this step, the essential experiment data, ROIs (seeds), second-level models, and temporal covariates are characterized. Provides anatomical and functional

preprocessing steps, such as trimming, realignment, anomaly identification, co-registration. Upon determining the number of subjects, designating the structural and functional fMRI data to the appropriate subjects is included in this progression. Figure 3.1 shows the Conn GUI for setup.

(b) Pre-processing

The fMRI preprocessing pipeline portrays standard and progressed preprocessing steps in fMRI. These measures are intended to reduce the impact of components that are notable in affecting the nature of functional and anatomical MRI data.

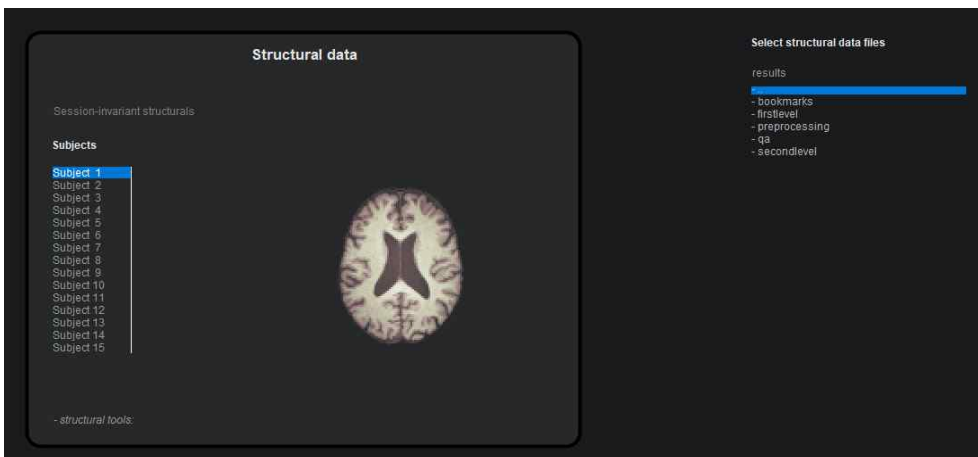


Figure 3.1 CONN setup, figure obtained from CONN tool [78].

MRI data includes the effects of patient movement in the scanner, transient and spatial image distortions due to the sequential nature of the imaging protocol, and anatomical contrasts between patients. A preprocessing pipeline designed specifically for volume-based examinations is used in this study. It involves the following steps: unwrapping and functional realignment; slice timing rectification; anomaly detection; and functional smoothing. CONN toolbox was used to process the fMRI and sMRI images [27]. The default preprocessing pipe line was used to process the images. This pipeline starts with realignment and unwrapping of

functional data, slice timing correction followed by outlier identification, direct segmentation and normalization and finally the functional smoothing. In functional realignment and unwrap step, CONN toolbox uses SPM12 realign [28] and unwrap procedure [29] to realign the functional data. B-spline interpolation was used to co-register and re-sample all scans to a reference image. In the slice timing correction SPM slice timing correction (STC) procedure corrects the temporal misalignment between different slices of functional data [30].

Similarly, CONN uses Artefact Detection Tools (ART) toolbox(https://www.nitrc.org/projects/artifact_detect/) to identify the outlier scans.

ART tool box identifies the outlier scans from the observed global bold signals together with the amount of subject motion in the scanner. Global bold signals that exceeding 5 standard deviation from global mean and from wise displacement above 0.9mm are identified as outlier scans.

Next, functional and anatomical data are normalized into standard MCI space. The functional and anatomical data then segmented into GM, WM, CFS classes using SPM12 unified segmentation [31]. Outlier detection step is followed by normalization and segmentation step while SPM12 is used to normalize the functional and anatomical data to normalize in MNI space and segment into GM, WM and CSF. For the functional data, mean BOLD signal is taken as difference image and for structural data, T1 weighted volume is taken as reference image [30], [33]. Fourth-order spline interpolation was used to resample both functional and anatomical data. A default $180 \times 216 \times 180$ mm bounding box was set. similarly, 2 mm isotropic voxels was set for functional data and 1 mm for anatomical data.

Spatial convolution operation with Gaussian kernel was performed to smooth the functional data to increase the BOLD signal-to-noise ratio. The size of Gaussian was taken as 8mm full width at half maximum (FWHM). Additional purpose of use the smoothing filter was to reduce the effect of residual irregularity in functional and gray anatomy across subjects

(c) Denoising

It is likely that the BOLD sign will still contain some noise or non-neural inconstancy after the functional information has been processed due to a mixture of physiological, aberration, and residual patient-movement effects. As a result of these residual variables, fcMRI investigations present extremely solid and perceptible inclinations in all measures of functional connectivity. As a result, conventional preprocessing procedures have supported more conservative approaches than those normally seen with activation-based fMRI examinations.

In this study, CONN's default denoising pipeline is utilized. It is composed of two stages: band-pass temporal filtering and linear regression of potential perplexing impact on the BOLD image. In order to address potential perplexing effects, an anatomically-based noise rectification strategy is applied and evaluated subject-movement parameters are used, as well as outliers or scanning parameters detected from cerebral WM and cerebrospinal regions. Figure 3.2 shows CONN's denoising GUI.

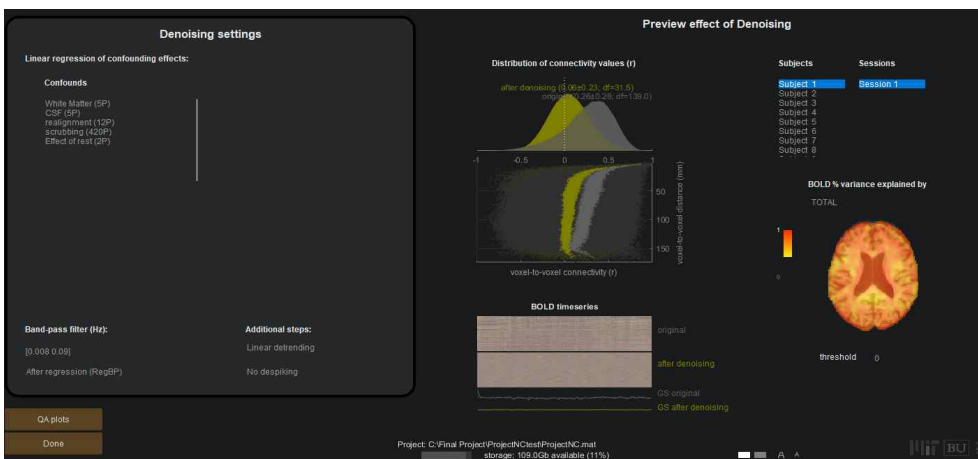


Figure 3.2 Denoising GUI of CONN, figure obtained from CONN tool [78].

(d) Analyses (Level 1)

Based on the regressors specified during the denoising step and the condition

regressors specified in the Setup tab, a model is constructed to fit the data observed in each voxel.

Correlation analysis is then performed by averaging across all ROIs, comparing those to other ROIs in the brain, and evaluating correlation maps using each voxel as a seed. In the setup process, several ROIs were generated: GM, WM, atlases, and networks. ROIS on atlases and networks provide a blueprint of the brain by parcelling it into different portions, sometimes referred to as seeds.

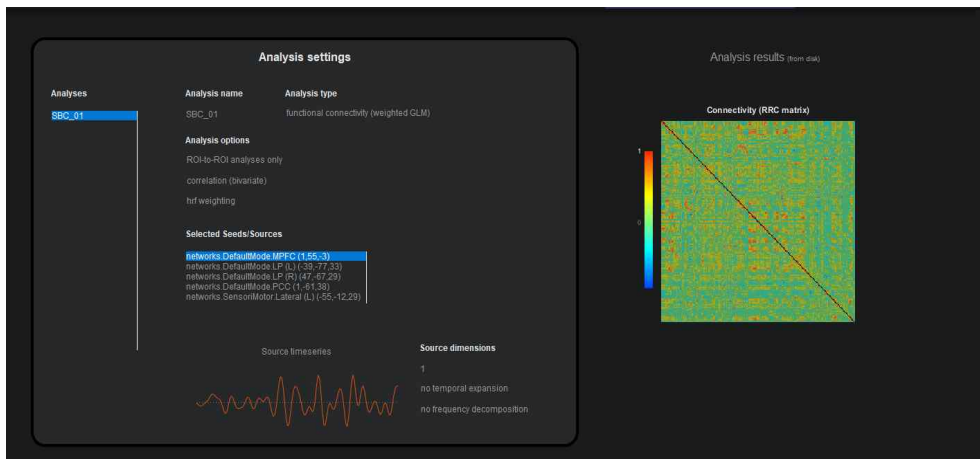


Figure 3.3 CONN analyses (level 1) GUI, figure obtained from CONN tool [78].

(e) Analyses (Level 2)

In CONN, a general linear model (GLM) is used for all second-level examinations, which allow analysts to make inferences about the population by examining a subset of patients. A GLM model description, parameter estimation, and inference testing framework are integrated into the package to help scientists examine and visualize the properties of seed-based maps over multiple patients, or ROI-to-ROI networks over a large number of patients. Figure 3.4 shows the CONN results (level 2) GUI.



Figure 3.4 CONN results (level 2) GUI, figure obtained from CONN tool [78].

B. Whole-brain connectivity matrix

In this study, the whole brain was parcellated into 132 structurally homogeneous ROIs, per the FSL Harvard-Oxford atlas for the GM and subcortical regions. ROI-to-ROI connectivity(RRC) matrix was constructed by computing the Fisher-transformed bivariate correlation coefficients between the time-series of each pair of ROIs. Connectivity metrics between ROIs describe the connectivity between all pairs of ROIs within a pre-defined region.

Connectivity metrics between ROIs describe the connectivity between all pairs of ROIs within a pre-defined region. RRC matrix elements represent the relationship between two ROIs. They are defined as bivariate Fisher-transformed correlation coefficients between the BOLD time series of the ROIs. Which is expressed as,

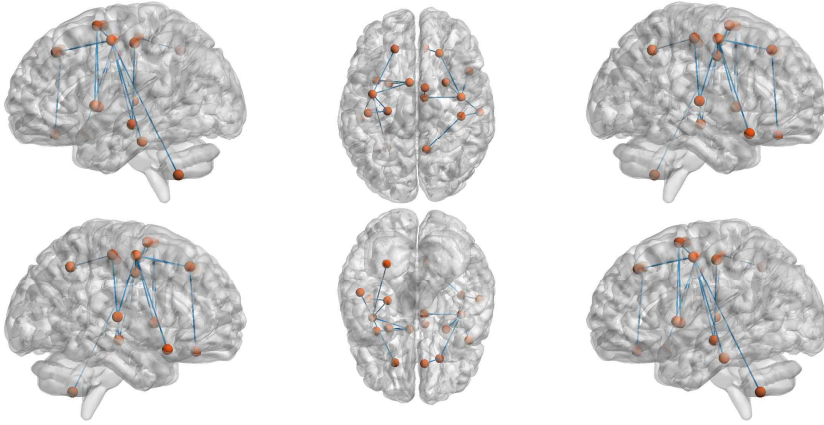


Figure 3.5 Whole brain network.

$$r(i,j) = \frac{\int R_i(t)R_j(t)}{\left(\int R_i^2(t)dt \int R_j^2(t)dt\right)^{1/2}}, \quad (3.1)$$

$$Z(i,j) = \tanh^{-1}(r(i,j)), \quad (3.2)$$

where R is the BOLD time series within each ROI (for simplicity all time series here are considered centered to zero mean), r is a matrix of correlation coefficients, and Z is the RRC symmetric matrix of Fisher-transformed correlation coefficients.

C. Core large-scale network construction

In total, 32 ROI seeds, whose shapes and locations are defined in HCP Atlas, were used for assessment of eight RSNs. The network consists of DMN, SAL, frontoparietal network (FPN), DAN, SMN, language network (LAN), VIS, and cerebellar network (CER).

In order to identify ROI-to-ROI connectivity patterns, a pair of extracted mean BOLD signal time courses of each ROI was used to compute bivariate Pearson's

correlation coefficients between the previously predefined seed regions of rs-networks. Finally, Fisher's transformation was used transform the obtained correlation coefficients to normally distributed scores.

D. Proposed framework

This proposed method consists of the following four major functional steps as shown in Figure 3.6.

- i. Construct the brain networks including whole-brain network and large-scale brain network interms of graph.
- ii. Convert graph to feature vector using node2vec graph embedding.
- iii. Perform the feature selection task.
- iv. Perform the classification using regularized extreme learning machine (RELM).

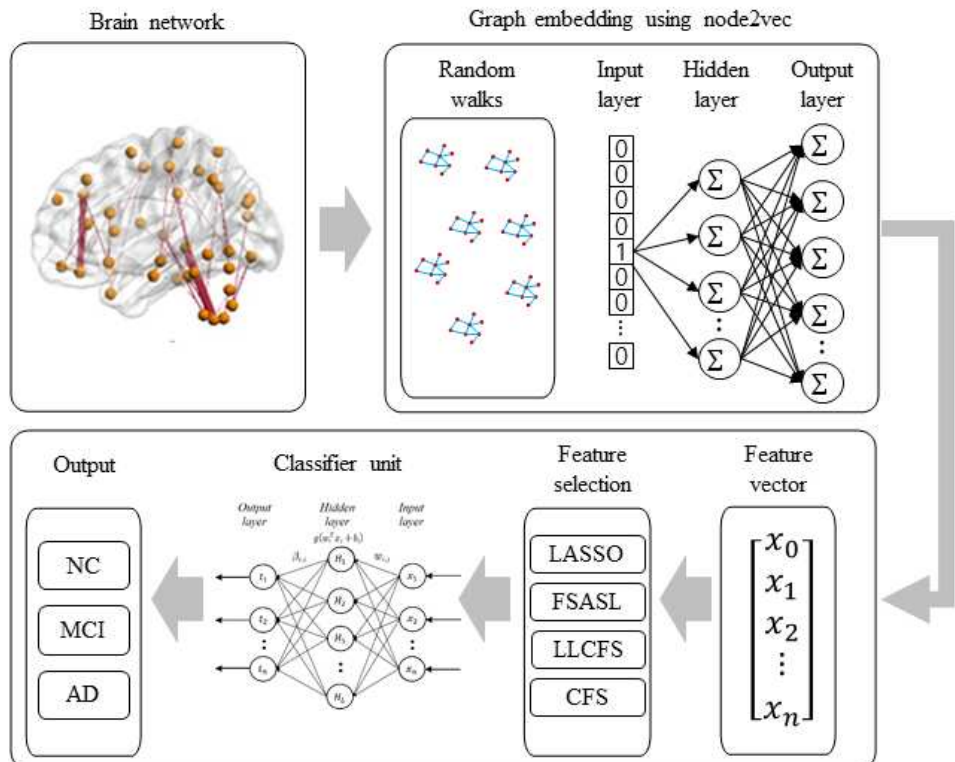


Figure 3.6 Classification of Alzheimer's disease using brain network.

1. Construction of brain networks

For the construction of network from fMR images, all the steps were followed to process the fMR images and structural MR data and examined utilizing the CONN tool. All the processes were described in detail for construction of whole- brain network and large-scale brain networks in sub sections II.J, III.B-D.

2. Graph-embedding

Graphs are complex data structures, consisting a finite set of vertices and set of edges which connect a pair of nodes. One of the possible solutions to manipulate prevalent pattern recognition algorithms on graphs is embedding the graph into vector space. Indeed, graph embedding is a bridge between statistical pattern recognition and graph mining. The node2vec [34] algorithm was employed as graph embedding tool in this study. The node2vec algorithm aims to learn a vectorial representation of nodes in a graph by optimizing a neighborhood preserving objective. It extends the previous node embedding algorithm Deepwalk and it is inspired from the state of art word embedding algorithm word2vec.

In word2vec, given a set of sentences also known as corpus, the model learns word embedding by analyzing the context of each word in the body. The word2vec uses the neural network with one hidden layer to transform words into embedding vectors. This neural network is known as Skip-gram. This network is trained to predict the neighboring word in the sentence. It accepts the word at the input and is optimized such that it predicts the neighboring words in a sentence with high probability. node2vec applies the same embedding approach to train and predict the neighborhood of a node in graph. However, word is replaced by the node and the bag of nodes is used instead of corpus. The sampling is used to generate this bag of nodes from a graph.

- Generally, the graph embedding consists of three steps :

- **Sampling** : Samples are taken from a graph using random walks. This

random walk results in bag of nodes of neighborhood from sampling. The bag of nodes acts as a collection of contexts for each node in the network. The innovation of node2vec with respect to Deepwalk is the use of flexible biased random walks on the network. In Deepwalk, random walk is obtained by a uniform random sampling over the linked nodes, while node2vec combine two different strategies for the network exploration: depth-first search (DFS) and breadth-first-search (BFS). For current random walk position at node v and traversed position at previous step at node t and neighboring nodes x_1 , x_2 and x_3 , the sampling of next node x is determined by evaluating the unnormalized transition probabilities π_{vx} on edge (t,v) with the static edge weight w_{vx} as shown in Figure 3.7.

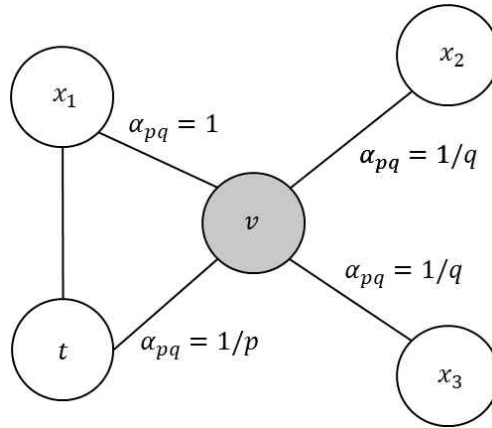


Figure 3.7 Illustration of random walk in node2vec algorithm.

This unnormalized transition probability is estimated based on search bias α defined by two parameters p and q such that $\pi_{vx} = \alpha_{pq}(t,x) \cdot w_{vx}$ where,

$$\alpha_{pq}(t,x) = \begin{cases} \frac{1}{p}, & \text{if } d_{tx} = 0 \\ 1, & \text{if } d_{tx} = 1 \\ \frac{1}{q}, & \text{if } d_{tx} = 2 \end{cases} . \quad (3.3)$$

Here d_{tx} denotes the shortest path distance between nodes t and x . The parameter p determines the likelihood of sampling the node t again during random walk. When the value of p is high revisit of the node possibility is low. Similarly the parameter q allows to differ between local and global nodes. If $q > 1$, the random walk has the likelihood of sampling the nodes around the node v is high.

- **Training skip-gram** : The bag of nodes generated from the random walk is fed into the skip-gram network. Each node is represented by a one-hot vector and maximizes the probability for predicting neighbor nodes. The one-hot vector has size same as the size of the set of unique words used in the text corpus. For each node only one dimension is equal to one and remaining are zeros. The position of dimension having one in vector defines the individual node.
- **Computing embedding** : The output of the hidden layer of the network is taken as the embedding of the graph.

E. Feature selection techniques

1. Least absolute shrinkage and selection operator (LASSO)

LASSO [35] is a powerful method which is used to remove insignificant features. Two major tasks of this method are regularization and feature selection. This method minimizes residual sum of squares of the model using ordinary least square regression (OLS) by placing a constraint on the sum of the absolute values of the model parameters. LASSO computes model coefficients β by minimizing the following function:

$$RSS_{LASSO}(\beta_v, \beta_0) = \underset{\beta}{\operatorname{argmin}} n \left[\sum_{i=1}^n (y_i - (\beta_i x_i + \beta_0))^2 + \alpha \sum_{j=1}^k |\beta_j| \right], \quad (3.4)$$

where x_i is the graph embedded feature input data, a vector of k values at observation j and n is the number of observations. y_i is the response at observation i . α is a non-negative user defined regularization parameter. This parameter controls the strength of penalty. When α is sufficiently large then coefficients are forced to be zero which leads to produce few relevant features. If α approaches 0 the model becomes OLS with more relevant features.

2. Features selection with adaptive structure learning (FSASL)

FSASL is an unsupervised method which performs data manifold learning and feature selection [36]. This method first utilizes the adaptive structure of the data to construct the global learning and the local learning. Next, the significant features are selected by integrating both of them with $L_{2,1}$ -normregularizer. This method utilizes the sparse reconstruction coefficients to extract the global structure of data for global learning. In sparse representation, each data sample x_i can be approximated as a linear combination of all the other samples, and the optimal sparse combination weight matrix.

For local learning, this method directly learns a Euclidean distance induced probabilistic neighborhood matrix.

$$\begin{aligned}
 & \min_{W, S, P} (\| W^T X - W^T X S \|^2 + \alpha \| S \|_1) \\
 & + \beta \sum_{i,j}^n (\| W^T x_i - W^T x_j \|^2 P_{ij} + \mu P_{ij}^2) + \gamma \| W \|_{21} \\
 & s.t. S_{ii} = 0, P 1_n = 1_n, P \geq 0, W^T X X^T X = I,
 \end{aligned} \tag{3.5}$$

where α is used to balancing the sparsity and the reconstruction error, β and γ are regularization parameters for global and local structure learning in first and second group and the sparsity of feature selection matrix in the third group. Similarly, S is used to guide the search of relevant global feature and P defines the local neighborhood of data sample x_i .

3. Local learning and clustering based feature selection (LLCFS)

LLCFS is clustering based feature selection method [37]. This method learns the adaptive data structure with selected features by constructing the k-nearest neighbor graph in the weighted feature space. The joint clustering and feature weight learning is performed by solving the following problem.

$$\min_{Y, \{W^i, b^i\}_{i=1}^n, z} \sum_{i=1}^n \sum_{c'=1}^c \left[\sum_{x_j \in \square_{x_i}} \beta (Y_{ic'} - x_j^T W_{c'}^i - b_{c'}^i)^2 + (W_{c'}^i)^T \text{diag}(z^{-1}) W_{c'}^i \right] \quad (3.6)$$

$$s.t. \mathbf{1}_d^T z = 1, z \geq 0,$$

where z the feature weight vector and \square_{x_i} is the k-nearest neighbor of x_i based on z weighted features.

4. Pairwise correlation based feature selection (CFS)

CFS selects features based on the ranks attributes according to an empirical evaluation function based on correlations. Subsets made of attribute vectors are evaluated by evaluation function, which are associated with the labels of class, however autonomous among each another [38]. CFS accepts that unrelated structures express a low correspondence with the class and hence they are ought to be overlooked by the procedure. Alternatively, additional features must be studied, as they are typically hugely correlated with one or additional amount of other features.

F. Extreme learning machine

ELM is single layer feedforward neural networks [39]-[43] as shown in Figure 3.8. This neural network is implemented using Moore-Penrose generalized inverse to set its weights [39]. This learning algorithm doesn't require iterative gradient-based

back-propagation to tune the hidden nodes [40]. Thus this method is considered as effective solution with extremely reduced complexity.

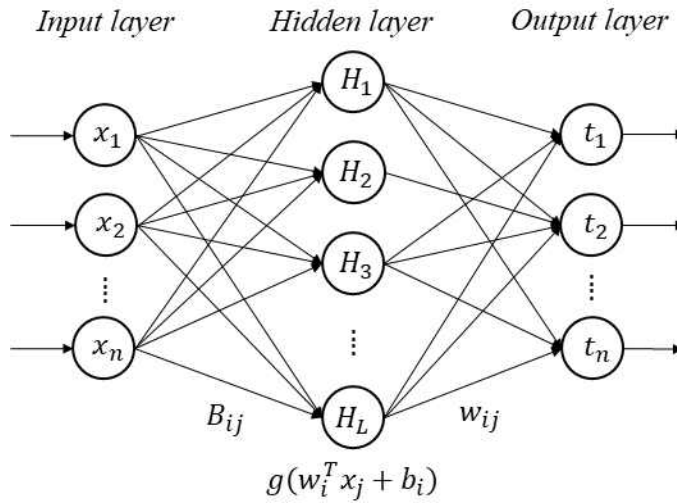


Figure 3.8 Single hidden layer ELM.

ELM with L number of hidden nodes and $g(x)$ as activation function is expressed as,

$$Y_L(x) = \sum_{i=1}^L \beta_i h_i(x) = h(x) \beta_i . \tag{3.7}$$

Here x is input feature vector. $h_i(x)$ is the input to output node from hidden layer node output. $\beta = [\beta_1, \dots, \beta_L]^T$ is the weight matrix of i_{th} node. The input weight w_i and the hidden layer biases b_i are generated randomly before the training samples are fed to input layer. For N training samples $(x_j, t_j)_{j=1}^N$. The loss function of ELM is expressed as,

$$\| H(w_1, \dots, w_N, b_1, \dots, b_N) \hat{\beta} - T \| = \min_{\beta} \| H \hat{\beta} - T \| , \tag{3.8}$$

with,

$$H(w_1, \dots, w_{\tilde{N}}, b_1, \dots, b_{\tilde{N}}) = \begin{bmatrix} g(w_1 x_1 + b_1) & \dots & g(w_L x_1 + b_L) \\ \vdots & \dots & \vdots \\ g(w_1 x_N + b_1) & \dots & g(w_L x_N + b_L) \end{bmatrix}. \quad (3.9)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix} \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_m^T \end{bmatrix}$$

Here, H represents the hidden layer output matrix and T represents output label of training data matrix. The matrix β is estimated as,

$$\beta = H^+ T. \quad (3.10)$$

Here, H^+ represents the Moore-Penrose generalized inverse of the matrix H . Since ELM learning approach requires no back-propagation, this method is best suited for the binary and multiclass classification of big data and neuroimaging features. However the decrease in computation time comes with the expense of increase in the error in the output, which ultimately decreases the accuracy. Thus, the regularization is performed by adding a constant such that generalization performance is improved with more robust constraint. The output weight of the regularized ELM can be expressed as,

$$\beta = \left(\frac{I}{C} + H^T H \right)^{-1} H^T T. \quad (3.11)$$

The output weights β is calculated by using ridge regression based approximation. Where C is a non-negative user defined regularization parameter. I is the identity matrix. In this work, the RELM is constructed in multi-layer manner. Each layer is connected to subsequent layer in feedforward fashion as

shown in Figure 3.9. The overall training procedure is described in Algorithm 1.

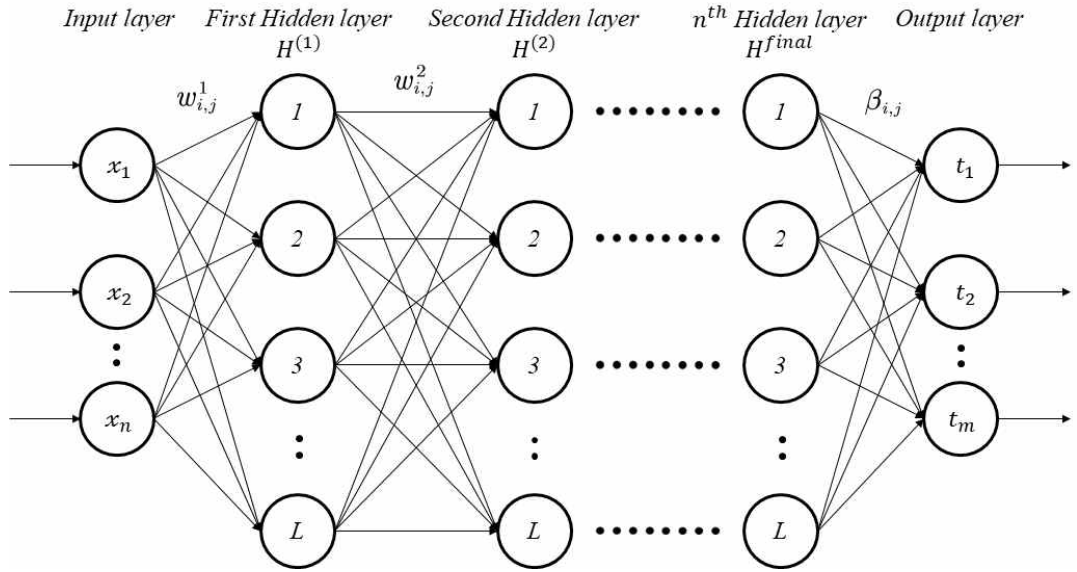


Figure 3.9 Multiple hidden layer extreme learning machine.

Algorithm 1. Multi layer extreme learning machine.

Input : feature matrix X , Output matrix T , regularization C for all layers, input weights w , biases and activation g and the number of layers n

Output : hidden layer feature representation H^{final} and output weight β

Step1 : Let $X^{(1)} = X$, calculate $H^{(1)} \leftarrow g(w^{(1)}x + b^{(1)})$

Step2 : $X^{(2)} = H^{(1)}$

For $i = 2 : n - 1$ do

Step3 : calculate $H^{(i)} \leftarrow g(w^{(i)}x + b^{(i)})$

Step4 : $X^{(i+1)} = H^{(i)}$

Step5 : Let $i = n$, calculate H^{final} and β

$$H^{final} \leftarrow g(w^{(n)}x + b^{(n)})$$

$$\beta = \left(\frac{I}{C} + (H^{final})^T (H^{final}) \right)^{-1} (H^{final})^T T$$

IV. Performance evaluation

Performance of the classifier is tested using RELM classifiers for each specific test including the binary test. Confusion matrix is constructed to visualize the performance of the binary classifier in a form of a as shown in Table 4.1. Correct numbers of prediction of classifier are placed on the diagonal of the matrix. These components are further divided into true positive (TP), true negative (TN), which represent correctly identified controls. Similarly, the false positive (FP) and false negative (FN) represent the number of wrongly classified subjects.

Table 4.1 Confusion matrix.

		Predicted Class	
		C ₁	C ₂
Accurate class	C ₁	TP	FN
	C ₂	FP	TN

The proportion of subjects which are correctly classified by the classifier is expressed as the accuracy,

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.1)$$

However, for dataset with unbalanced class distribution accuracy may be a good performance metric. Thus two more performance are used. These metrics are known as sensitivity and specificity are used.

$$SEN = \frac{TP}{TP + FN} \quad (4.2)$$

$$SPE = \frac{TN}{TN + FP} \quad (4.3)$$

The sensitivity (SEN) measures the rate of true positives (TP) while the specificity (SPE) measures rate of true negatives (TN).

A. Demographic and clinical findings

A significant group difference in age in AD versus, AD versus MCI and MCI versus HC was not found. However significant group difference was found in MMSE ($P < 0.01$) and CDR ($P < 0.01$) in all group combinations. The gender proportion on both AD and HC is male dominant. AD has 54.83% and HC has 45.16% male dominance. In addition, all three subject groups are married dominant, HC (80.9%), MCI (72.1%) and AD (95.1%). Table 3.1 shows the detailed descriptions and analysis of these variables.

B. Classification results

In this section, the performance is evaluated using the SL-RELM and ML-RELM classifiers for each specific test. It is observed that the performance of the proposed algorithm with by comparing the test result of three different models namely Large scale brain network, whole brain network and combined network. The size of large scale brain is of 32×32 , whole brain network is 132×132 and combined network is of 164×164 . ML-RELM classifier is used for respective test comprising the binary classification and four different feature selection methods. The performance of the binary classifier is presented in terms of accuracy, sensitivity and specificity. Accuracy measures the proportion of correctly classified subjects by the classifier. While the sensitivity measures the rate of TP and the specificity measures rate of TN thus signifying appropriately recognized controls. Correspondingly, all the erroneously classified matters can be symbolized by false positive and false negative. Feature selection and classification algorithms are evaluated on data set using a 10-fold cross validation (CV). First, the subjects were divided into 10 equally sized subsets: each of these subsets (folds), containing 10% of the subjects as test set and remaining 90% for training set. Then feature ranking was performed on the training sets. Different feature selection algorithms were used to rank the features. The ML-RELM classifier was trained using these

top-ranked features. Separate feature selection was performed for each training and test to avoid the feature selection bias during 10-fold cross validation. Cross validated average classification accuracy and standard deviation was calculated for specific feature using k -top most ranked features, where k ranges from 1 to 100. Finally, calculated the mean accuracy and standard deviation of highest ranked features for different feature selection was calculated as depicted in Table 4.4-Table 4.12. Additionally, mean sensitivity and specificity together with corresponding standard deviation were calculated amongst corresponding values estimated for highest ranked features. Bold values in each table indicate the maximum value of accuracy, sensitivity and specificity. Tables 4.4-4.6 show the binary classification results using SL-RELM classifier with four different feature selections for whole brain network. Tables 4.7-4.9 show the classification results using same classifier feature selections for large scale brain network. Similarly, Tables 4.10-4.12 show the classification results for combined brain network. Results obtained through the feature selection methods are compared in regards to the performance metrics such as accuracy, sensitivity specificity and f-measure. Table 4.4 summarizes the AD against HC classification on whole brain network. The FSASL feature selection method outperforms all other methods in terms of accuracy and sensitivity with the highest mean accuracy of 86.51%, mean sensitivity 85.25% and mean specificity of 88.08%. Similarly, the classification results of HC versus MCI and AD versus MCI using SL-RELM are shown in Tables 4.5 and 4.6. As shown in Table 4.5, the highest mean accuracy is 96.14 (± 1.71) for HC against MCI classification and 95.19 (± 2.63) for MCI against AD classification. Additionally the F-score is high in all three classifications (0.92) for HC against AD, 0.99 for HC against MCI, 1 for AD against MCI using FSASL and LASSO feature section method. Similarly, the comparison of classification of HC, MCI and AD with Large scale brain network classifier with different feature selection methods are shown in Tables 4.7, 4.8 and 4.9. Similar to previous network, better results in terms of accuracy, sensitivity and specificity was obtained using the

FSASL feature selection technique. As shown in Table 4.7, the accuracy of 95.42% specificity of 94.5% and sensitivity of 96.41% and F-score of 0.97 was obtained for AD against HC. In Table 4.8 the highest mean accuracy, specificity, sensitivity and F-score are obtained as 96.47%, 95.33%, 97.66% and 0.97 for HC against MCI classification. Similarly, Table 4.9 shows the classification performance of AD against MCI. The highest mean accuracy, sensitivity, specificity and F-score are 98.38%, 97.16%, 99.66% and 1. Tables 4.10, 4.11 and 4.12 show the results and comparison of HC, MCI and AD with combined brain network. As shown in Table 4.10, the accuracy of highest 85.82% sensitivity of 85.0% and specificity of 88.0% and F-score of 0.93 were obtained for AD against HC using the FSASL feature selection method. In Table 4.11 the highest mean accuracy, sensitivity, specificity and F-score are obtained as 96.75%, 97.75%, 95.83% and 0.94 for HC against MCI classification using LASSO feature selection. Similarly, Table 4.12 shows the classification performance of AD against MCI. The highest mean accuracy, sensitivity, specificity and F-score are 86.35%, 85.08%, 87.5% and 0.86. From all these results, it is clearly evident that the use of FSASL feature selection method is the ideal choice for the classification using ML-RELM classifier for the graph embedded data when the network is big. Three sizes 32×32 for large scale brain network, 132×132 for whole brain network and 164×164 for combined brain network were used for experiments. FSASL generates better results for two small sized brain networks combined brain network and whole brain network. While for large sized brain network LASSO generates better results in terms of accuracy, sensitivity and specificity.

Table 4.2 Comparison of performance of binary classification AD against HC with state of the art methods using rs-fMRI.

Dataset	Feature measures	Classifier	Accuracy (%)	Reference
AD:77, HC :173	Combination of FC matrices, FC dynamics, ALFF	AUC	85	de Vos et al., 2018 [71]
AD: 12, HC: 12	Difference between DMN and SN map	LDA	92	Zhou et al., 2010 [72]
AD: 34, HC: 45	Graph measures	Naïve Bayes	93.3	Khazaei et al.,2017 [22]
AD: 15, HC: 16	Averaged voxel intensities of core regions in resting state networks: DMN, DAN, VAN.	Multivariate ROC	95	Wu et al., 2013 [73]

Table 4.3 Comparison of performance of binary classification MCI against HC with state of the art methods using rs-fMRI.

Dataset	Feature measures	Classifier	Accuracy (%)	Reference
MCI: 31, HC: 31	functional activity co-variations of ROIs	SVM	62.90	Eavani et al., 2013 [74]
MCI: 31, HC: 31	group sparse representation	SVM	66.13	Wee et al., 2014 [75]
MCI: 31, HC: 31	SDFN	SVM	70.97	Leonardi et al.,2013 [76]
MCI: 31, HC: 31	Deep auto encoder and HMM	SVM	72.58	Suk et al., 2016 [77]
MCI: 89, HC: 45	Graph measures	Naïve Bayes	93.3	Khazae et al.,2017 [22]

Abbreviation: FC, functional connectivity; AUC, area under the curve; DMN, default mode network; SN, salience network; LDA, linear discriminant analysis; ROC, receiver operating characteristic; ROI, region of interest; AAL, automated anatomical labeling; SDFN, sliding window-based dynamic functional network; HMM, hidden markov model.

Table 4.4 10-fold cross-validation binary mean classification performance for AD against HC using SL-RELM classifier on whole brain network using different feature selection methods (132 × 132).

Feature selection method	Performance metrics	Accuracy	Sensitivity	Specificity	F-measure
LASSO	Mean (%)	82.06	78.58	85.58	0.86
	Standard deviation	2.67	2.751	4.25	
FSASL	Mean (%)	86.51	85.25	88.00	0.92
	Standard deviation	3.670	6.18	4.75	
LLCFS	Mean (%)	85.24	78.66	91.91	0.85
	Standard deviation	4.06	7.59	5.65	
CFS	Mean (%)	86.28	82.33	90.08	0.86
	Standard deviation	3.27	6.51	4.88	

Table 4.5 10-fold cross-validation binary mean classification performance for HC against MCI using SL-RELM classifier on whole brain network using different feature selection methods (132 × 132).

Feature selection method	Performance metrics	Accuracy	Sensitivity	Specificity	F-measure
LASSO	Mean (%)	90.64	83.33	98.08	0.995
	Standard deviation	2.05	4.27	3.19	
FSASL	Mean (%)	96.14	95.16	97.08	0.97
	Standard deviation	1.71	2.74	1.89	
LLCFS	Mean (%)	85.40	81.0	89.83	0.95
	Standard deviation	4.03	4.33	6.67	
CFS	Mean (%)	89.09	86.33	92.00	0.89
	Standard deviation	4.10	6.54	4.12	

Table 4.6 10-fold cross-validation binary mean classification performance for MCI against AD using SL-RELM classifier on whole brain network using different feature selection methods (132×132).

Feature selection method	Performance metrics	Accuracy	Sensitivity	Specificity	F-measure
LASSO	Mean (%)	90.05	93.33	86.67	0.96
	Standard deviation	2.50	3.19	3.98	
FSASL	Mean (%)	95.19	94.16	96.16	1
	Standard deviation	2.63	3.62	2.81	
LLCFS	Mean (%)	86.86	87.16	86.5	0.79
	Standard deviation	5.51	6.67	6.66	
CFS	Mean (%)	87.91	88.41	87.58	0.93
	Standard deviation	2.87	6.81	6.16	

Table 4.7 10-fold cross-validation binary mean classification performance for AD against HC using SL-RELM classifier on large scale brain network using different feature selection methods (32×32).

Feature selection method	Performance metrics	Accuracy	Sensitivity	Specificity	F-measure
LASSO	Mean (%)	84.06	81.58	86.75	0.81
	Standard deviation	3.48	4.48	5.32	
FSASL	Mean (%)	95.42	94.5	96.41	0.97
	Standard deviation	2.14	2.58	2.48	
LLCFS	Mean (%)	85.01	81.66	88.41	0.93
	Standard deviation	3.86	5.37	5.29	
CFS	Mean (%)	88.38	84.25	92.41	0.91
	Standard deviation	2.36	4.25	2.55	

Table 4.8 10-fold cross-validation binary mean classification performance for HC against MCI using SL-RELM classifier on large scale brain network using different feature selection methods (32×32).

Feature selection method	Performance metrics	Accuracy	Sensitivity	Specificity	F-measure
LASSO	Mean (%)	90.12	83.0	97.16	0.97
	Standard deviation	1.89	3.89	2.69	
FSASL	Mean (%)	96.47	95.33	97.66	0.97
	Standard deviation	1.46	2.122	1.61	
LLCFS	Mean (%)	87.02	82.25	91.75	0.82
	Standard deviation	4.37	4.02	6.55	
CFS	Mean (%)	88.38	84.25	92.42	0.91
	Standard deviation	2.36	4.25	2.56	

Table 4.9 10-fold cross-validation binary mean classification performance for MCI against AD using SL-RELM classifier on large scale brain network using different feature selection methods (32×32).

Feature selection method	Performance metrics	Accuracy	Sensitivity	Specificity	F-measure
LASSO	Mean (%)	84.95	86.75	83.08	0.84
	Standard deviation	4.81	5.188	5.18	
FSASL	Mean (%)	98.38	97.16	99.66	1
	Standard deviation	1.51	2.69	1.05	
LLCFS	Mean (%)	88.83	90.91	87.0	0.91
	Standard deviation	4.60	3.89	8.30	
CFS	Mean (%)	88.07	87.66	88.5	0.97
	Standard deviation	4.18	7.70	6.22	

Table 4.10 10-fold cross-validation binary mean classification performance for AD against HC using SL-RELM classifier on Combined brain network using different feature selection methods (164 × 164).

Feature selection method	Performance metrics	Accuracy	Sensitivity	Specificity	F-measure
LASSO	Mean (%)	84.88	81.83	88.0	0.93
	Standard deviation	1.76	3.68	4.12	
FSASL	Mean (%)	85.82	85.0	86.91	0.86
	Standard deviation	2.88	5.29	4.332	
LLCFS	Mean (%)	82.58	82.41	82.91	0.88
	Standard deviation	2.83	3.75	5.43	
CFS	Mean (%)	70.15	70.66	69.33	0.73
	Standard deviation	7.37	6.28	11.26	

Table 4.11 10-fold cross-validation binary mean classification performance for HC against MCI using SL-RELM classifier Combined brain network using different feature selection methods (164 × 164).

Feature selection method	Performance metrics	Accuracy	Sensitivity	Specificity	F-measure
LASSO	Mean (%)	86.35	85.08	87.5	0.86
	Standard deviation	3.00	5.037	4.79	
FSASL	Mean (%)	88.19	91.58	84.91	0.93
	Standard deviation	3.10	4.77	3.35	
LLCFS	Mean (%)	82.5	81.66	83.16	0.86
	Standard deviation	4.02	6.56	5.14	
CFS	Mean (%)	70.55	65.83	75.25	0.96
	Standard deviation	6.01	5.77	7.61	

Table 4.12 10-fold cross-validation binary mean classification performance for MCI against AD using SL-RELM classifier on Combined brain network using different feature selection methods (164×164).

Feature selection method	Performance metrics	Accuracy	Sensitivity	Specificity	F-measure
LASSO	Mean (%)	96.75	97.75	95.83	0.94
	Standard deviation	1.52	2.22	3.04	
FSASL	Mean (%)	90.12	91.16	89.25	0.94
	Standard deviation	3.64	5.58	4.39	
LLCFS	Mean (%)	78.57	81.0	76.0	0.78
	Standard deviation	3.06	5.93	3.98	
CFS	Mean (%)	74.03	73.58	74.5	0.73
	Standard deviation	5.13	9.77	9.74	

Similarly, Tables 4.13-4.21 show the binary classification results using ML-RELM classifier. Similar to the SL-RELM, the mean accuracy and the standard deviation of highest ranked features are calculated for different feature selection methods. Table 4.13-4.15 show the classification results for whole brain network. As shown in Table 4.13 highest accuracy of **85.42%** was obtained using FSASL for whole brain network while the highest sensitivity and specificity of 87.16% and 83.58% was also obtained by FSASL. For HC against MCI LASSO generated highest accuracy of 82.5% with sensitivity and specificity of 82.08% and 83.06% as shown in Table 4.14. Similarly, for AD against MCI, as shown in Table 4.15 LASSO generated highest accuracy of 88.66% with mean sensitivity of 90.66% and specificity of 86.75%.

Tables 4.16-4.18 show the classification result of large scale brain network. For large scale brain network FSASL generates highest accuracy for all three classification schemes. Table 4.16 show the classification of AD against HC, where

highest accuracy, sensitivity and specificity was 96.59%, 97.0% and 96.08%. Table 4.17 shows classification of HC against MCI. Here, the FSASL generates highest accuracy of 96.69%, sensitivity of 95.91% and specificity of 97.5%. Similarly, Table 4.18 show the classification results of AD against MCI. Here the highest accuracy is 96.09%, sensitivity of 95.33% and specificity of 96.75%.

Tables 4.19 to 4.21 show the classification results of ML-RELM of combined brain network. Table 4.19 show the classification results of AD against HC. Here, the highest accuracy was obtained using FSAL feature selection method. As shown in Table, the highest accuracy is 89.65%, sensitivity is 91.54% and specificity is 86.9%. Classification of HC against MCI is shown in Table 4.20. The highest accuracy of 88.76% was generated by FSASL together with sensitivity and specificity of 92.5% and 84.86% accordingly. Finally, Table 4.21 show the classification result of AD against MCI. Here LASSO generates highest accuracy of 96.14% with sensitivity and specificity of 94.91% and 97.5%.

Table 4.13 10-fold cross-validation binary mean classification performance for AD against HC using ML-RELM classifier on whole brain network using different feature selection methods (132 × 132).

Feature selection method	Performance metrics	Accuracy	Sensitivity	Specificity	F-measure
LASSO	Mean (%)	81.41	81.08	82.0	0.80
	Standard deviation	3.01	5.76	5.34	
FSASL	Mean (%)	85.42	87.16	83.58	0.79
	Standard deviation	4.67	4.89	5.77	
LLCFS	Mean (%)	81.631	84.66	78.58	0.83
	Standard deviation	3.29	6.56	3.40	
CFS	Mean (%)	69.90	71.92	67.66	0.61
	Standard deviation	5.37	9.96	9.410	

Table 4.14 10-fold cross-validation binary mean classification performance for HC against MCI using ML-RELM classifier on whole brain network using different feature selection methods (132×132).

Feature selection method	Performance metrics	Accuracy	Sensitivity	Specificity	F-measure
LASSO	Mean (%)	82.5	82.08	83.16	0.83
	Standard deviation	3.85	7.88	4.40	
FSASL	Mean (%)	78.0	76.0	80.41	0.87
	Standard deviation	4.34	5.66	5.40	
LLCFS	Mean (%)	69.95	71.0	68.75	0.69
	Standard deviation	4.65	8.44	7.77	
CFS	Mean (%)	65.14	60.83	69.5	0.79
	Standard deviation	6.41	7.22	9.31	

Table 4.15 10-fold cross-validation binary mean classification performance for MCI against AD using ML-RELM classifier on whole brain network using different feature selection methods (132×132).

Feature selection method	Performance metrics	Accuracy	Sensitivity	Specificity	F-measure
LASSO	Mean (%)	88.66	90.66	86.75	0.90
	Standard deviation	2.29	3.11	3.65	
FSASL	Mean (%)	80.92	78.5	83.33	0.80
	Standard deviation	4.21	5.19	5.86	
LLCFS	Mean (%)	87.07	90.08	84.25	0.85
	Standard deviation	3.47	5.85	4.58	
CFS	Mean (%)	70.71	0.73.5	67.91	0.53
	Standard deviation	9.51	9.67	12.85	

Table 4.16 10-fold cross-validation binary mean classification performance for AD against HC using ML-RELM classifier on large scale brain network using different feature selection methods (32×32).

Feature selection method	Performance metrics	Accuracy	Sensitivity	Specificity	F-measure
LASSO	Mean (%)	82.21	80.42	84.41	0.77
	Standard deviation	5.31	5.79	6.11	
FSASL	Mean (%)	96.59	97.0	96.083	0.97
	Standard deviation	1.82	2.46	2.54	
LLCFS	Mean (%)	84.36	84.41	84.16	0.88
	Standard deviation	2.60	5.528	5.32	
CFS	Mean (%)	87.14	84.91	89.25	0.97
	Standard deviation	3.39	6.66	4.25	

Table 4.17 10-fold cross-validation binary mean classification performance for HC against MCI using ML-RELM classifier on large scale brain network using different feature selection methods (32×32).

Feature selection method	Performance metrics	Accuracy	Sensitivity	Specificity	F-measure
LASSO	Mean (%)	90.90	88.58	93.5	0.96
	Standard deviation	2.74	3.81	3.53	
FSASL	Mean (%)	96.69	95.91	97.5	1
	Standard deviation	1.75	2.404	2.078	
LLCFS	Mean (%)	86.66	83.08	90.25	0.91
	Standard deviation	4.096	6.56	3.95	
CFS	Mean (%)	87.35	86.08	88.58	0.97
	Standard deviation	3.343	5.21	4.28	

Table 4.18 10-fold cross-validation binary mean classification performance for MCI against AD using ML-RELM classifier on large scale brain network using different feature selection methods (32×32).

Feature selection method	Performance metrics	Accuracy	Sensitivity	Specificity	F-measure
LASSO	Mean (%)	89.28	93.17	85.25	0.91
	Standard deviation	4.06	5.56	6.17	
FSASL	Mean (%)	96.09	95.33	96.75	0.97
	Standard deviation	1.08	2.33	1.59	
LLCFS	Mean (%)	86.92	87.16	86.66	0.90
	Standard deviation	5.07	9.08	4.63	
CFS	Mean (%)	86.98	89.58	84.41	0.82
	Standard deviation	4.03	5.72	4.95	

Table 4.19 10-fold cross-validation binary mean classification performance for AD against HC using ML-RELM classifier on Combined brain network using different feature selection methods (164×164).

Feature selection method	Performance metrics	Accuracy	Sensitivity	Specificity	F-measure
LASSO	Mean (%)	82.11	86.5	77.83	0.65
	Standard deviation	7.16	6.71	9.56	
FSASL	Mean (%)	89.35	91.54	86.90	0.92
	Standard deviation	3.14	6.24	1.96	
LLCFS	Mean (%)	87.23	86.83	87.66	0.93
	Standard deviation	2.66	4.64	3.215	
CFS	Mean (%)	68.17	64.16	72.0	0.76
	Standard deviation	6.019	6.03	9.64	

Table 4.20 10-fold cross-validation binary mean classification performance for HC against MCI using ML-RELM classifier Combined brain network using different feature selection methods (164 × 164).

Feature selection method	Performance metrics	Accuracy	Sensitivity	Specificity	F-measure
LASSO	Mean (%)	85.42	91.75	79.08	0.82
	Standard deviation	2.98	4.62	6.61	
FSASL	Mean (%)	88.76	92.5	84.86	0.86
	Standard deviation	3.10	4.77	3.35	
LLCFS	Mean (%)	82.46	83.66	81.25	0.82
	Standard deviation	3.62	5.34	5.11	
CFS	Mean (%)	73.76	74.33	73.166	0.66
	Standard deviation	6.88	10.58	5.39	

Table 4.21 10-fold cross-validation binary mean classification performance for MCI against AD using ML-RELM classifier on Combined brain network using different feature selection methods (164 × 164).

Feature selection method	Performance metrics	Accuracy	Sensitivity	Specificity	F-measure
LASSO	Mean (%)	96.14	94.91	97.5	1
	Standard deviation	2.975	4.63	3.33	
FSASL	Mean (%)	93.01	92.26	93.92	0.97
	Standard deviation	3.22	3.74	4.37	
LLCFS	Mean (%)	80.74	84.88	76.78	0.77
	Standard deviation	5.412	7.92	4.82	
CFS	Mean (%)	69.75	69.04	70.59	0.74
	Standard deviation	3.370	4.23	5.26	

The number of hidden layer nodes influences the performance of the ML-RELM classifier. In experiments performed, it is found that 1000 number of hidden layer generated the best performance in terms of accuracy. Similarly, the parameters p and q were set to correspond localized random walks. With the smaller value of p and larger value of q , the random walk is easy to sample to the high-order -order proximity. Thus, p and q were selected randomly and performed graph embedding with. $p = 0.1$ and $q = 1.6$.

C. Discussion

Several studies based on rs-fMRI have been carried out for the classification of AD and MCI from HC subjects. Binary classification in combination of different classifier with different feature measure reported the accuracy ranging from 85% to 95% for AD against HC and 62.90% to 72.58% to and MCI against HC as shown in Tables 4.2 and 4.3. These studies used the same MCI and HC subjects from the ADNI2 cohort. One can clearly notice that the number of subjects directly influences the accuracy. As the number of subjects increase the accuracy is decreased. As reported in previous section the highest accuracy for the classification of AD from is obtained in proposed work is 93.957% using the combination of FSASL and RELM in large scale network. If the results are compared for MCI against HC, the results obtained in current study outperform all the state of art methods. However, it is not fair to compare performance with other studies directly because each work employ different datasets, preprocessing pipelines, feature measures, and classifiers. Majority of works including [70]-[77] have used subjects less than or nearly equal to 30 in each subject class. The main reason behind small number of dataset is the availability of fMRI data in ADNI2 cohort. All of these studies performed classification and made conclusion. Likewise, the study was conducted using ADNI2 cohort with nearly equal number of subjects with previous studies and the cross validation was also done using these dataset.

V. Limitations

While this study is focused on the stage diagnosis of AD progression using fMRI alone using ADNI2 cohort, the major limitation of this study is the limited sample size of ADNI2(AD=33, MCI=31, and HC=31). In this context, the entire population is not represented adequately with the dataset used in this study. Thus, the generalization of our results to other groups cannot be guaranteed.

VI. Conclusion

It is widely accepted that the early diagnosis of AD and MCI plays an important role to take preventive action and to delay the future progression of AD. Thus the accurate classification task of different stages of AD progression is essential. In this study, it is demonstrated graph based features from fMR images can be used for the classification of AD and MCI from HC. The proposed approach was tested on three different network modes ranging from large scale network, whole brain network and combined network. Better classification accuracy was obtained on large scale network and on combined network. This result suggests the large scale network is composed of low number of nodes and edges however, these nodes and edges carry distinct features required to classify the AD from healthy and MCI subjects. Additionally, multiple feature selection techniques were used to cope with the smaller number of subjects with larger number of feature representations. The appropriate amount of features is extracted from standard ADNI cohort that lead to maximal classification accuracies as compared to all other recent researches. Among different feature selection methods FSASL worked better for big network size as well as small network excluding few exceptions.

References

- [1] American Psychiatric Association and American Psychiatric Association (1994) Task Force on DSM-IV., “Diagnostic and statistical manual of mental disorders,” DSM-IV, vol xxv, 4th edn. American Psychiatric Association, Washington, DC.
- [2] D. Schmitter, A. Roche, B. Maréchal, D. Ribes, A. Abdulkadir, M. Bach-Cuadra, and Alzheimer’s Disease Neuroimaging Initiative, “An evaluation of volume-based morphometry for prediction of mild cognitive impairment and Alzheimer’s disease,” *NeuroImage: Clinical*, vol. 7, no. 1, pp. 7–17, 2015, doi:10.1016/j.nicl.2014.11.001.
- [3] Alzheimer’s association, “2016 Alzheimer's disease facts and figures,” *Alzheimer's and Dementia*, vol. 12, no. 4, pp. 459-509, Apr. 2016.
- [4] F. Liu, L. Zhou, C. Shen, and J. Yin, “Multiple kernel learning in the primal for multimodal alzheimer’s disease classification,” *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 3, pp. 984–990, May. 2014, doi:10.1109/JBHI.2013.2285378.
- [5] W. Wong, "Economic burden of Alzheimer disease and managed care considerations." *The American journal of managed care*, vol. 26, no. 8, pp. S177-S183, Aug. 2020.
- [6] A. M. Fjell, K. B. Walhovd, C. Fennema-Notestine, L. K. McEvoy, D. J. Hagler, D. Holland, and Alzheimer's Disease Neuroimaging Initiative, “CSF biomarkers in prediction of cerebral and clinical change in mild cognitive impairment and Alzheimer’s disease,” *Journal of Neuroscience*, vol. 30, no. 6, pp. 2088–2101, Feb. 2010, doi:10.1523/JNEUROSCI.3785-09.2010.
- [7] R. K. Lama and G. R. Kwon, “Diagnosis of Alzheimer’s Disease Using Brain Network,” *Frontiers in Neuroscience*, vol. 15, article 605115, Feb. 2021, doi: 10.3389/fnins.2021.605115.
- [8] 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, doi:10.1016/j.neuroimage.2011.01.008.

- [9] M. W. Vernooij and M. A. van Buchem, “Neuroimaging in Dementia,” Springer, pp. 131–142, Jan. 2020, doi:10.1007/978-3-030-38490-6_11.
- [10] D. Chan, N. C. Fox, R. L. Scahill, W. R. Crum, J. L. Whitwell, G. Leschziner, and M. N. Rossor, “Patterns of temporal lobe atrophy in semantic dementia and Alzheimer’s disease,” *Annals of neurology*, vol. 49, no. 4, pp. 433–442, 2001.
- [11] J. S. Phillips, F. Da Re, L. Dratch, S. X. Xie, D. J. Irwin, C. T. McMillan, and M. Grossman, “Neocortical origin and progression of gray matter atrophy in nonamnestic Alzheimer’s disease,” *Neurobiology of Aging*, vol. 63, pp. 75–87, Mar. 2018, doi:10.1016/j.neurobiolaging.2017.11.008.
- [12] A. T. Du, N. Schuff, J. H. Kramer, H. J. Rosen, M. L. Gorno-Tempini, K. Rankin, and M. W. Weiner, “Different regional patterns of cortical thinning in Alzheimer’s disease and frontotemporal dementia,” *Brain*, vol. 130, no. 4, pp. 1159–1166, Apr. 2007, doi:10.1093/brain/awm016
- [13] C. D. Good, R. I. Scahill, N. C. Fox, J. Ashburner, K. J. Friston, D. Chan, & and R. S. Frackowiak, “Automatic differentiation of anatomical patterns in the human brain: Validation with studies of degenerative dementias,” *NeuroImage*, vol. 17, no. 1, pp. 29–46, Sep. 2002, doi:10.1006/nimg.2002.1202.
- [14] F. Shi, B. Liu, Y. Zhou, C. Yu, and T. Jiang, “Hippocampal volume and asymmetry in mild cognitive impairment and Alzheimer’s disease: Meta-analyses of MRI studies,” *Hippocampus*, vol. 19, no. 11, pp. 1055–1064, Nov. 2009, doi: 10.1002/hipo.20573.
- [15] J. Ashburner and K. J. Friston, “Voxel-based morphometry - The methods,” *NeuroImage*, vol. 11, no. 6 I, pp. 805–821, 2000, doi:10.1006/nimg.2000.0582.
- [16] M. A. Trivedi, A. K. Wichmann, B. M. Torgerson, M. A. Ward, T. W. Schmitz, M. L. Ries, and S. C. Johnson, “Structural MRI discriminates individuals with Mild Cognitive Impairment from age-matched controls: A combined neuropsychological and voxel based morphometry study,” *Alzheimer’s and Dementia*, vol. 2, no. 4, pp. 296–302, Oct. 2006, doi:10.1016/j.jalz.2006.06.001.
- [17] G. B. Karas, P. Scheltens, S. A. Rombouts, P. J. Visser, R. A. van Schijndel, N. C. Fox, and F. Barkhof, “Global and local gray matter loss in mild cognitive

- impairment and Alzheimer's disease," *NeuroImage*, vol. 23, no. 2, pp. 708–716, Oct. 2004, doi: 10.1016/j.neuroimage.2004.07.006.
- [18] G. Chen, B. D. Ward, C. Xie, W. Li, Z. Wu, J. L. Jones, and S. J. Li, "Classification of Alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional MR imaging," *Radiology*, vol. 259, no. 1, pp. 213–221, Apr. 2011, doi: 10.1148/radiol.10100734.
- [19] K. Wang, T. Jiang, M. Liang, L. Wang, L. Tian, X. Zhang, and Z. K. Liu, "Discriminative analysis of early Alzheimer's disease based on two intrinsically anti-correlated networks with resting-state fMRI," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 4191, pp. 340–347, 2006, doi:10.1007/11866763_42.
- [20] E. Challis, P. Hurley, L. Serra, M. Bozzali, S. Oliver, and M. Cercignani, "Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI," *NeuroImage*, vol. 112, pp. 232–243, May 2015, doi: 10.1016/j.neuroimage.2015.02.037.
- [21] B. Jie, D. Zhang, W. Gao, Q. Wang, C. Y. Wee, and D. Shen, "Integration of network topological and connectivity properties for neuroimaging classification," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 2, pp. 576–589, Feb. 2014, doi:10.1109/TBME.2013.2284195.
- [22] A. Khazaei, A. Ebrahimzadeh, and A. Babajani-Feremi, "Identifying patients with Alzheimer's disease using resting-state fMRI and graph theory," *Journal of the International Federation of Clinical Neurophysiology*, vol. 126, no. 11, pp. 2132–2141, Nov. 2015, doi:10.1016/j.clinph.2015.02.060.
- [23] M. D. Greicius, G. Srivastava, A. L. Reiss, V. Menon, and M. E. Raichle, "Default-mode network activity distinguishes Alzheimer's disease from healthy aging: Evidence from functional MRI," 2004. [Online]. Available: www.fmridc.org
- [24] V. Menon, "Large-scale brain networks and psychopathology: A unifying triple network model," *Trends in Cognitive Sciences*, vol. 15, no. 10, pp. 483–506, Oct. 2011, doi:10.1016/j.tics.2011.08.003.

- [25] S. H. Joo, H. K. Lim, and C. U. Lee, “Three large-scale functional brain networks from resting-state functional MRI in subjects with different levels of cognitive impairment,” *Psychiatry Investigation*, vol. 13, no. 1, pp. 1-7, Jan. 2016, doi: 10.4306/pi.2016.13.1.1.
- [26] F. Vecchio, F. Miraglia, and P. M. Rossini, “Connectome: Graph theory application in functional brain network architecture,” *Clinical neurophysiology practice*, vol. 2, pp. 206–213, Oct. 2017, doi:10.1016/j.cnp.2017.09.003.
- [27] “Alzheimer’s Disease Neuroimaging Initiative,” ADNI Site, last modified 2017, accessed Sep 17, 2021, <http://adni.loni.usc.edu/>
- [28] W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols, “Statistical parametric mapping: the analysis of functional brain images,” Elsevier, 2007.
- [29] J. L. R. Andersson, C. Hutton, J. Ashburner, R. Turner, and K. Friston, “Modeling geometric deformations in EPI time series,” *NeuroImage*, vol. 13, no. 5, pp. 903–919, May. 2001, doi:10.1006/nimg.2001.0746.
- [30] R. N. A. Henson, C. Buechel, O. Josephs, and K. J. Friston, “The slice-timing problem in event-related fMRI,” *NeuroImage*, vol. 9, no. 6, p. 125, Jan. 1999, doi: 0.1016/j.neuroimage.2011.06.078
- [31] J. Ashburner and K. J. Friston, “Unified segmentation,” *NeuroImage*, vol. 26, no. 3, pp. 839–851, Jul. 2005, doi:10.1016/j.neuroimage.2005.02.018.
- [32] Y. Behzadi, K. Restom, J. Liao, and T. T. Liu, “A component based noise correction method (CompCor) for BOLD and perfusion based fMRI,” *NeuroImage*, vol. 37, no. 1, pp. 90–101, Aug. 2007, doi:10.1016/j.neuroimage.2007.04.042.
- [33] A. Nieto-Castanon, “Handbook of functional connectivity Magnetic Resonance Imaging methods in CONN,” Hilbert Press; Jan. 2020.
- [34] A. Grover and J. Leskovec, “Node2vec: Scalable feature learning for networks,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 855–864, Aug. 2016, doi:10.1145/2939672.2939754.
- [35] R. Tibshirani, “Regression Shrinkage and Selection via the Lasso,” 1996.

- [36] L. Du and Y.-D. Shen, “Unsupervised Feature Selection with Adaptive Structure Learning,” Apr. 2015, [Online]. Available: <http://arxiv.org/abs/1504.00736>
- [37] H. Zeng and Y. M. Cheung, “Feature selection and kernel learning for local learning-based clustering,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1532–1547, Aug. 2011, doi:10.1109/TPAMI.2010.215.
- [38] M. A. Hall, “Correlation-based Feature Selection for Machine Learning,” Apr. 1999.
- [39] J. Cao, K. Zhang, M. Luo, C. Yin, and X. Lai, “Extreme learning machine and adaptive sparse representation for image classification,” *Neural networks*, vol. 81, pp.91-102, Sep. 2016, doi:10.1016/j.neunet.2016.06.001.
- [40] W. Zhang, H. Shen, Z. Ji, G. Meng, and B. Wang, “Identification of mild cognitive impairment using extreme learning machines model,” in *International Conference on Intelligent Computing*, pp. 589–600, 2015.
- [41] X. Peng, P. Lin, T. Zhang, and J. Wang, “Extreme learning machine-based classification of ADHD using brain structural MRI data,” *PLoS ONE*, vol. 8, no. 11, Nov. 2013, doi:10.1371/journal.pone.0079476.
- [42] M. N. I. Qureshi, B. Min, H. J. Jo, and B. Lee, “Multiclass classification for the differential diagnosis on the ADHD subtypes using recursive feature elimination and hierarchical extreme learning machine: Structural MRI study,” *PLoS ONE*, vol. 11, no. 8, Aug. 2016, doi:10.1371/journal.pone.0160697.
- [43] E. Cambria, G. B. Huang, L. L. C. Kasun, H. Zhou, C. M. Vong, J. Lin, and J. E. Liu, “Extreme learning machines [trends & controversies],” *IEEE intelligent systems*, vol. 28, no. 6, pp. 30–59, Dec. 2013, doi:10.1109/MIS.2013.140.
- [44] UCSF Weill Institute for Neurosciences: Memory and Aging Center. What is dementia? Accessed Jun. 2020, memory.ucsf.edu/what-dementia
- [45] “2020 Alzheimer’s disease facts and figures,” *Alzheimer’s and Dementia*, vol. 16, no. 3, pp. 391–460, Mar. 2020, doi:10.1002/alz.12068.
- [46] Cure Alzheimer’s Fund. Stats & costs. Accessed April 30, 2020. curealz.org/the-disease/stats-and-costs/
- [47] A. Deb, J. D. Thornton, U. Sambamoorthi, and K. Innes, “Direct and indirect

- cost of managing alzheimer’s disease and related dementias in the United States,” Expert Review of Pharmacoeconomics & Outcomes Research, vol. 17, no. 2, pp. 189–202, Mar. 2017, doi:10.1080/14737167.2017.1313118
- [48] B. Pyenson et al., “The real-world Medicare costs of Alzheimer disease: considerations for policy and care,” Journal of managed care & specialty pharmacy, vol. 25, no. 7, pp. 800–809, Jul. 2019, doi:10.18553/jmcp.2019.25.7.800.
- [49] A. S. Kelley, K. McGarry, R. Gorges, and J. S. Skinner, “The burden of health care costs for patients with dementia in the last 5 years of life,” Annals of Internal Medicine, vol. 163, no. 10, pp. 729–736, Nov. 2015, doi:10.7326/M15-0381.
- [50] P. Lin, Y. Zhong, H. M. Fillit, E. Chen, and P. J. Neumann, “Medicare expenditures of individuals with Alzheimer’s disease and related dementias or mild cognitive impairment before and after diagnosis,” Journal of the American Geriatrics Society, vol. 64, no. 8, pp. 1549–1557, Aug. 2016, doi:10.1111/jgs.14227.
- [51] B. Magnin, L. Mesrob, S. Kinkingnéhun, M. Pélégrini-Issac, O. Colliot, M. Sarazin, and H. Benali, “Support vector machine-based classification of Alzheimer’s disease from whole-brain anatomical MRI,” Neuroradiology, vol. 51, no. 2, pp. 73–83, Feb. 2009, doi:10.1007/s00234-008-0463-x.
- [52] S. Klöppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, and R. S. Frackowiak, “Automatic classification of MR scans in Alzheimer’s disease,” Brain, vol. 131, no. 3, pp. 681–689, Mar. 2008, doi:10.1093/brain/awm319.
- [53] F. R. Bach FRANCISBACH, A. Rakotomamonjy, F. R. Bach, S. Canu, and Y. Grandvalet RAKOTOMAMONJY, “SimpleMKL Alain Rakotomamonjy Stéphane Canu Yves Grandvalet,” 2008. [Online]. Available: <http://www.mloss.org>.
- [54] Y. Fan, S. M. Resnick, X. Wu, and C. Davatzikos, “Structural and functional biomarkers of prodromal Alzheimer’s disease: A high-dimensional pattern classification study,” NeuroImage, vol. 41, no. 2, pp. 277–285, Jun. 2008, doi:10.1016/j.neuroimage.2008.02.043.

- [55] G. R. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. I. Jordan, “Learning the Kernel Matrix with Semidefinite Programming,” *Journal of Machine Learning Research*, vol. 5, pp. 27-72, Jan. 2004.
- [56] F. R. Bach, G. R. G. Lanckriet, and M. I. Jordan, “Multiple Kernel Learning, Conic Duality, and the SMO Algorithm.”, *Proceedings of the twenty-first international conference on Machine learning*, Jul. 2004, doi:10.1145/1015330.1015424.
- [57] C. Hinrichs, V. Singh, G. Xu, and S. Johnson, “MKL for robust multi-modality AD classification,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009, vol. 5762 LNCS, no. PART 2, pp. 786–794. doi:10.1007/978-3-642-04271-3_95.
- [58] X. Li, X. Chen, Y. Yan, W. Wei, and Z. J. Wang, “Classification of EEG signals using a multiple kernel learning support vector machine,” *Sensors*, vol. 14, no. 7, pp. 12784–12802, Jul. 2014, doi:10.3390/s140712784.
- [59] P. Vemuri, J. L. Gunter, M. L. Senjem, J. L. Whitwell, K. Kantarci, D. S. Knopman, and C. R. Jack Jr, “Alzheimer’s disease diagnosis in individual subjects using structural MR images: Validation studies,” *NeuroImage*, vol. 39, no. 3, pp. 1186–1197, Feb. 2008, doi:10.1016/j.neuroimage.2007.09.073.
- [60] R. Wang, “AdaBoost for Feature Selection, Classification and Its Relation with SVM, A Review,” *Physics Procedia*, vol. 25, pp. 800–807, Dec. 2012, doi:10.1016/j.phpro.2012.03.160.
- [61] J. Cao, L. Chen, M. Wang, H. Shi, and Y. Tian, “A Parallel AdaBoost-back-propagation Neural Network for Massive Image Dataset Classification,” *Scientific Reports*, vol. 6, no. 1, Dec. 2016, doi:10.1038/srep38201.
- [62] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, Aug. 1997, doi:10.1006/jcss.1997.1504.
- [63] R. E. Schapire, Y. Freund, P. Bartlett, and S. Lee, “Boosting the margin: a new explanation for the effectiveness of voting methods,” *Ann. Statist*, vol. 26, no. 5,

- pp. 1651-1686, Oct. 1998, doi:10.1214/aos/1024691352.
- [64] Z. Tu, “Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering,” in Proceedings of the IEEE International Conference on Computer Vision, 2005, vol. 1, pp. 1589–1596. doi:10.1109/ICCV.2005.194.
- [65] A. Savio, M. García-Sebastián, M. Graña, and J. Villanúa, “Results of an AdaBoost approach on Alzheimer’s disease detection on MRI,” in International Work-Conference on the Interplay Between Natural and Artificial Computation, pp. 114–123, 2009, doi:10.1007/978-3-642-02267-8_13.
- [66] 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,” Computational and Mathematical Methods in Medicine, vol. 2014, article. 862307, Sep. 2014, doi:10.1155/2014/862307.
- [67] J. H. Morra, Z. 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 Transactions on Medical Imaging, vol. 29, no. 1, pp. 30–43, Jan. 2010, doi: 10.1109/TMI.2009.2021941.
- [68] A. Payan and G. Montana, “Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks,” Computer Vision and Pattern Recognition, arXiv preprint arXiv:1502.02506, Feb. 2015.
- [69] C.-Y. Wee, P.-T. Yap, D. Zhang, L. Wang, and D. Shen, “Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification,” Brain Structure and Function, vol. 219, no. 2, pp. 641–656, Mar. 2014, doi: 10.1007/s00429-013-0524-8.
- [70] E. Hosseini-Asl, G. Gimel’farb, and A. El-Baz, “Alzheimer’s disease diagnostics by a deeply supervised adaptable 3D convolutional network,” arXiv preprint arXiv:1607.00556, Jul. 2016.
- [71] w, “A comprehensive analysis of resting state fMRI measures to classify individual patients with Alzheimer’s disease,” Neuroimage, vol. 167, pp. 62–72, Feb. 2018, doi:10.1016/j.neuroimage.2017.11.025.

- [72] J. Zhou, M. D. Greicius, E. D. Gennatas, M. E. Growdon, J. Y. Jang, G. D. Rabinovici, and W. W. Seeley, “Divergent network connectivity changes in behavioural variant frontotemporal dementia and Alzheimer’s disease,” *Brain*, vol. 133, no. 5, pp. 1352–1367, Apr. 2010, doi:10.1093/brain/awq075.
- [73] X. Wu, J. Li, N. Ayutyanont, H. Protas, W. Jagust, A. Fleisher, and K. Chen, “The receiver operational characteristic for binary classification with multiple indices and its application to the neuroimaging study of Alzheimer’s disease,” *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 10, no. 1, pp. 173–180, Feb. 2012, doi:10.1109/TCBB.2012.141.
- [74] H. Eavani, T. D. Satterthwaite, R. E. Gur, R. C. Gur, and C. Davatzikos, “Unsupervised learning of functional network dynamics in resting state fMRI,” in *International conference on information processing in medical imaging*, vol. 7917, pp. 426–437, 2013, doi:10.1007/978-3-642-38868-2_36.
- [75] C.-Y. Wee, P.-T. Yap, D. Zhang, L. Wang, and D. Shen, “Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification,” *Brain Structure and Function*, vol. 219, no. 2, pp. 641–656, 2014.
- [76] N. Leonardi, J. Richiardi, M. Gschwind, S. Simioni, J. M. Annoni, M. Schlupe, and D. Van De Ville, “Principal components of functional connectivity: a new approach to study dynamic brain connectivity during rest,” *NeuroImage*, vol. 83, pp. 937–950, Dec. 2013, doi:10.1016/j.neuroimage.2013.07.019.
- [77] H.-I. Suk, C.-Y. Wee, S.-W. Lee, and D. Shen, “State-space model with deep learning for functional dynamics estimation in resting-state fMRI,” *NeuroImage*, vol. 129, pp. 292–307, Apr. 2016, doi:10.1016/j.neuroimage.2016.01.005.
- [78] “NeuroImaging Tools&Resources Collaboratory-CONN : functional connectivity toolbox,” NITRC Site, last modified 2007, accessed Sep 21, 2021, https://www.nitrc.org/frs/?group_id=279