



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](#)

August 2022

Master Dissertation

Efficient Deep Learning Methods for Supporting Diagnosis of Alzheimer's Disease and Mild Cognitive Impairment

Graduate School of Chosun University

Department of Information and Communication
Engineering

Fazal Ur Rehman Faisal

Efficient Deep Learning Methods for Supporting Diagnosis of Alzheimer's Disease and Mild Cognitive Impairment

알츠하이머 병 및 경도 인지 장애 진단 지원을 위한 효율적인
딥러닝 방법

26, August, 2022

Graduate School of Chosun University
Department of Information and Communication
Engineering

Fazal Ur Rehman Faisal

Efficient Deep Learning Methods for Supporting Diagnosis of Alzheimer's Disease and Mild Cognitive Impairment

Advisor: Prof. Kwon, Goo-Rak


A dissertation submitted in partial fulfillment of the requirements
for a Master Degree


April, 2022


Graduate School of Chosun University
Department of Information and Communication
Engineering

Fazal Ur Rehman Faisal

페이잘 파잘 얼 레흐만 석사 학위 논문을 인준함

위원장 조선대학교 교수 변재영 (인) 

위원 조선대학교 교수 김찬기 (인) 

위원 조선대학교 교수 권구락 (인) 

2022년 05월

조선대학교 대학원

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND ACRONYMS	iv
ABSTRACT	vii
초록	ix
1 Introduction	1
1.1 Motivation	1
1.2 Research Objectives	2
1.3 Contributions	3
1.4 Thesis Scheme	4
2 Background	6
2.1 Alzheimer’s Disease	6
2.1.1 Mild Cognitive Impairments	8
2.1.2 Neuropathology	8
2.1.3 Biomarkers	10
2.1.4 Risk Factors	11
2.1.4.1 Age	12
2.1.4.2 Family History	12
2.1.4.3 Genetics	12
2.2 Magnetic Resonance Imaging (MRI)	13
2.2.1 Technology	13
2.2.2 Imaging	14
2.3 Database Organization	15
2.3.1 Alzheimer’s Disease Neuroimaging Initiative (ADNI) . .	15

2.4	Machine Learning	15
2.4.1	Deep Learning	17
2.4.1.1	Convolutional Neural Network (CNN)	19
2.5	Related Work	22
2.5.1	Machine Learning based Alzheimer’s Disease Classification	23
2.5.2	CNN approaches for Alzheimer’s Disease diagnosis	24
3	Proposed Method	27
3.1	Overview	27
3.2	Machine Learning method for Alzheimer’s Disease Classification	27
3.2.1	sMRI Dataset	27
3.2.2	Selected Features	28
3.2.2.1	Volumetric volumes	28
3.2.2.2	Cortical dementia	29
3.2.2.3	Subcortical dementia	29
3.2.3	Extraction of Features	30
3.2.4	Selection of Features	31
3.2.4.1	Principal Component Analysis (PCA)	31
3.2.4.2	Recursive Feature Elimination (RFE)	32
3.2.5	Classification	32
3.2.5.1	Random Forest (RF)	32
3.2.5.2	Support Vector Machine (SVM)	34
3.2.5.3	K-Nearest Neighbors	35
3.2.5.4	Naive Bayes	35
3.2.5.5	Softmax Classifier	36
3.2.6	Methodology	36

3.2.6.1	Architecture	38
3.2.6.2	Implementation Details	38
3.3	Deep Learning based Alzheimer’s Disease Classification	39
3.3.1	Dataset	39
3.3.2	Image Patch Generation	40
3.3.3	Deep Learning Architectures	41
3.3.3.1	The VGG-Net Model	42
3.3.3.2	ResNet Model	42
3.3.3.3	Proposed Model	43
4	Results	50
4.1	Performance Metrics	50
4.2	Machine Learning Results	52
4.2.1	Results Summary	55
4.3	Deep Learning Results	55
5	Discussion	59
6	Summary and Conclusion	65
	REFERENCES	66
	ACKNOWLEDGEMENTS	81

LIST OF ABBREVIATIONS AND ACRONYMS

AD	Alzheimer's disease
MCI	Mild Cognitive Impairment
eMCI	Early stage Mild Cognitive Impairment
IMCI	Late stage Mild Cognitive Impairment
NC	Normal Control
CNN	Convolutional Neural Network
DNN	Deep Neural Network
RF	Random Forest
SVM	Support vector Machine
NB	Naive Bayes
KNN	K-Nearest Neighbors
MRI	Magnetic Resonance Imaging
PET	Position emission tomography
ROC	Receiver operating characteristics curve
Adam	Adaptive Moment Estimation
CV	Cross validation
ROI	Region of Interest
PCA	Principal Component analysis
RFE	Recursive Feature Elimination
ADNI	Alzheimer's Disease Neuroimaging initiative

LIST OF FIGURES

2.1	Healthy brain and Alzheimer’s disease effected brain illustration.	9
2.2	Progression of Alzheimer’s Disease biomarkers [36].	11
2.3	MRI scans depicted various stages of Alzheimer’s disease. . . .	14
2.4	Working process of Machine Learning.	16
2.5	Convolutional Neural Network structure [48].	20
3.1	a) Cortical ROIs, b) Subcortical volume extracted features [78]. .	30
3.2	Support Vector Machine	35
3.3	Machine learning based Proposed Method Architecture.	40
3.4	Basic Architecture of VGG-Net [78].	43
3.5	The fundamental representation of residual network [84].	44
3.6	The layout of the proposed model.	44
3.7	Demonstration of the standard convolution architecture [86]. . .	46
3.8	Demonstration of the depthwise convolution [86] architecture . .	48
4.1	Performance comparison for AD vs eMCI between classifiers. . .	54
4.2	Performance comparison for AD vs lMCI between classifiers. . .	54
4.3	Performance comparison for AD vs HC between classifiers. . . .	55
4.4	Class 2 (AD), Class 1 (MCI), and Class 0 (CN) Receiver operating characteristics (ROC) curve	57
4.5	Testing Data Confusion matrix	57
4.6	Training and validation accuracy curve	58
4.7	Loss curve for training and validation data	58
5.1	Model comparison for FLOPs, parameters and accuracy	64

LIST OF TABLES

3.1	Statistics representation of all subjects.	28
3.2	A Comparative Study of Machine Learning based Classification techniques	37
3.3	Representation of ADNI subject’s demographics.	41
4.1	Confusion Matrix	50
4.2	AD vs eMCI classification results	52
4.3	AD vs lMCI classification results	53
4.4	AD vs HC classification results	53
4.5	Comparison of multiclass classification model performances . .	56
5.1	Number of Parameters compared between proposed and published methods.	62
5.2	Classification results of experiment and published methods. . . .	63

ABSTRACT

Efficient Deep Learning Methods for Supporting Diagnosis of Alzheimer's Disease and Mild Cognitive Impairment

Fazal Ur Rehman Faisal

Advisor: Prof. Kwon, Goo-Rak

Department of Information and
Communication Engineering

Graduate School of Chosun University

A common neuro-degenerative chronic condition, Alzheimer's disease (AD) distinguished by memory loss, poor self-care, and behavioral difficulties. This disease is extremely expensive to treat, with little known about its origins and due to the unavailability of curative treatments. Correct diagnosis of Alzheimer is vital for treatment implementation and progression, neuroimaging is the most promising arenas towards timely detection of AD.

Several high-dimensional classification techniques based on T1-weighted MR images have recently been developed to automatically distinguish between patients with AD, late mild cognitive impairment (IMCI), early mild cognitive impairment (eMCI), and normal control (NC) patients. When working with machine learning techniques, they frequently encounter issues with data overfitting and, as a result, accuracy. The time-consuming nature of machine learning motivates researchers to investigate alternatives to machine learning, such as deep learning. Ever since, image classification using CNN-based algorithms has been widely employed in medicine field. Unfortunately,

constructing an effective classification algorithm capable of providing decent outcomes is neither practical nor reasonable to use on embedded devices.

This thesis first demonstrates Machine Learning based techniques for reducing model complexity, that also attributed for an overfitting problem. For that, an improvised feature selection method (a method that combines Principal Component Analysis and Recursive Feature Elimination to simultaneously reduce dimension size and select best features) has been proposed to reduce model complexity. In this study, subcortical and cortical features from ADNI-based structural magnetic resonance imaging (sMRI) images were utilized. Following experiments have been performed to examine the model's performance: AD vs IMCI, eMCI vs AD, AD vs CN.

The second experiment included a deep neural technique for detecting significant AD biomarkers from structural MRI scans (sMRI), as well as brain scan classification into normal subjects (CN), AD, and mild cognitive impairment (MCI) groups. The CNN technique was employed in this experiment to convert medical imaging scans into higher-level information by merging characteristics from different levels of CNN. The proposed methodology employs fewer parameters, which simplifies the computing challenge. In comparison to other cutting-edge AD classification algorithms that are already available, the suggested strategy produces better results for standard evaluation measures.

초록

알츠하이머 병 및 경도 인지 장애 진단 지원을 위한 효율적인 딥러닝 방법

페이잘 파잘 알 레흐만

지도 교수: 권구락

정보통신공학과

조선대학교 대학원

일반적으로 신경 퇴행성 만성 질환인 알츠하이머병(AD)은 기억상실, 자기 관리의 문제 및 행동 장애로 인해 발생한다. 이 질병은 발병원인을 찾기 힘들며, 치료비용이 매우 비싸 치료 요법을 사용하기 어렵다. 알츠하이머의 정확한 진단은 치료를 진행함에 있어서 매우 중요하며, 뇌 영상은 알츠하이머의 발병을 적시에 확인하기 위해 필수적인 부분이다.

최근 T1의 강조된 자기공명영상을 기반으로 하는 여러 가지의 고차원적인 분류 기술이 AD 환자, 후기 경도인지장애, 초기 경도인지장애 및 정상대조군 환자들을 자동으로 판별하기 위해 개발되었다. 머신러닝 방법을 사용할 때는 데이터의 과적합과 정확도의 문제가 결과에서 자주 발생한다. 머신 러닝의 많은 시간을 소비하는 특성은 연구자들이 딥러닝과 같은 머신 러닝의 대체를 위한 조사를 할 수 있는 동기를 부여한다. 이에 따라, CNN 기반의 알고리즘을 이용한 영상 분류가 현재 의학 분야에서 많이 활용되고 있다. 하지만, 적절한 결과를 제공할 수 있는 효과적인 분류 알고리즘들은 임베디드 장치에서 사용하기에는 실용성이 낮고 합리적이지 않다.

본 논문에서는 첫 번째로, 과적합 문제에 대한 원인이 되는 모델의 복잡성을 줄이기 위한 기계 학습 기반의 기술을 보여준다. 이를 위해 모델의 복잡성을 줄이기 위한 즉석에서 특징을 선택하는 방법(주성분 분석과 재귀 특징제거

방법을 결합하여 차원의 크기를 줄이는 동시에 최상의 특징을 선택하기 위한 방법)을 제안한다. 이 연구에서는 ADNI 기반의 구조적 자기공명영상(sMRI) 영상의 피질하 및 피질 특징을 활용한다. 모델의 성능을 조사하기 위해 다음의 실험을 수행한다: AD vs IMCI, eMCI vs AD, AD vs CN.

두 번째 실험에서는 구조적 자기공명영상의 스캔에서 중요한 알츠하이머 병의 바이오마커를 검출하기 위한 심층신경 기술과 알츠하이머병과 정상대조군, 경도인지장애 그룹으로 분류하는 방법이 포함된다. 이 실험에서 CNN 방법을 사용하여 다양한 수준의 CNN의 특성을 병합하여 의료영상 스캔에서 더 높은 수준의 정보로 변환된다. 제안된 방법은 적은 양의 매개변수를 사용하여 컴퓨팅 문제를 단순화한다. 이미 사용 가능한 다른 AD 분류 알고리즘과 비교하여 제안된 방법은 표준 평가 측정에서 더 나은 결과를 제공한다.

1 Introduction

1.1 Motivation

A common neurological condition Alzheimer's disease (AD) cited as an important issue in the 21st century. Over 50 million individuals were diagnosed with Alzheimer so far, and the number is increasing rapidly [1]. There are some common symptoms through which AD can be characterized such as: Impaired memory, mental confusion, and visual difficulties [2].

Many attempts have been made so far to detect and diagnose the condition at an early stage. In this regards, many well-known biomarkers for AD has already been identified, which are mostly related with a clinical observation, or cognitive evaluation [3]. Whereas, in recent studies the position emission tomography scans and magnetic resonance imaging (MRI) proved to become a widely utilized structural and molecular biomarkers [4] for AD diagnosis. In distinctive, MRI is a widely used biomarker that is an effective technology for analyzing functional and anatomical brain alterations which are linked with AD. The anatomical and functional brain alterations are acknowledged as essential part in the evolution of AD [5]–[7].

Due to the rapid developments in neuroimaging techniques, integrating high-dimensional neuroimaging data has become more difficult in disease findings [8], [9]. As a consequence, curiosity in cognitive machine learning (ML) approaches for integrative framework has increased [10], [11]. These methods were utilized to produce the required outcome from a collection of training examples, which included spatial intensities, tissue volume, and shape descriptor characteristics. Therefore, many studies looked into the possibility of machine learning-based

approaches on MRI data to determine AD [12]–[15].

1.2 Research Objectives

In the field of digital health care, Machine Learning (ML) is gaining popularity owing to its special capacity to incorporate data on a wide scale [16], [17]. However, suitable architectures or time-consuming and computationally expensive preparatory processes such as feature engineering must be established in order to use ML techniques. Deep Learning (DL) approaches, in contrast, are being investigated by the researchers as an alternative to standard ML methods. DL, in particular, are a subtype of feature representation techniques given that they may efficiently identify the correct description from real data without the need for feature selection beforehand [18]. Because of its effectiveness in image analysis and classification, deep learning has gotten interest in medical image analysis [19], [20]. These achievements have aroused researchers' intention of improving CNN-based solutions for Alzheimer diagnosis. Considering benefits of ML and DL algorithms in medical image analysis. The objectives of this research is identified as follows:

- The ADNI database was utilized to collect and evaluate pre-processed sMRI scans.
- Examine the idea of employing a machine and deep learning algorithms to automatically classify AD using a sMRI scan of subject's brain.
- For ML, the purpose is to resolve the accuracy and data overfitting problem by employing features selection procedures that help to categorize individual subjects into four classes using different classifiers.

- For DL, the aim is to provide a computationally less expensive and automated model capable of effectively classifying brain scans into AD, MCI, and healthy groups.

1.3 Contributions

The efficacy of machine learning (ML) algorithms for AD detection have been hotly contested. However, researchers are constantly thinking about how to identify the critical features for enhancing the efficacy of AD diagnosis using machine learning models. This work contributes:

- Extraction of cortical and subcortical features of the brain by using Freesurfer.
- A combination of feature reduction and selection algorithm based on Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) has been presented, which can help to overcome the issues of overfitting and ultimately enhancing the model accuracy.
- Comparison of various ML classifiers performance (Naïve Bayes, SVM, Random Forest(RF), Softmax classifier, and KNN) for the identification of AD, early mild cognitive impairment (eMCI), late mild cognitive impairment (lMCI) and cognitive normal (CN) groups.

On the other hand, deep learning (DL) demonstrated great performance in a number of medical imaging tasks, such as CT scans, MRIs, and X-rays [21]. Despite the fact that previous approaches have yielded positive diagnostic outcomes, minimal effort has made to enhance the convolution technique for the

determination of AD. In this context, this work present a better CNN architecture for Alzheimer detection using MRI scans. The contributions in DL methods are four-fold:

- This thesis investigates the computational cost of the convolutional layer, builds a convolutional structure, and focuses in depth on the convolutional structure. This minimizes the computational complexity and computational costs while maintaining classification accuracy.
- Each block is represented with the same resolution and size in a CNN model. Its usefulness was established by the fact that it improved an automatic CNN classifier’s classification performance.
- The proposed model extracts important characteristics from the data without the necessity for pre-processing, and it performs well.
- The proposed method was tested against two cutting-edge classification models: VGG and ResNet50. Our proposed model achieved higher accuracy results with fewer parameters as compared to others.

1.4 Thesis Scheme

The thesis study is arranged and organized in the format of five chapters, which are outlined below:

- i) **Chapter 2** provides a formal background information about the AD, dataset and data base organization. Lastly, it will describe the advanced technique for diagnosis and the previous work regarding those techniques.

- ii) **Chapter 3** presents the pipelined implementation of the proposed model which consist of ML and DL methods.
- iii) **Chapter 4** presents the results and the evaluation of the proposed models.
- iv) **Chapter 5** highlights the summary and conclusions of thesis.

2 Background

2.1 Alzheimer's Disease

A progressive neurological disease Alzheimer is acknowledge as the most widespread type of disability. Linked with brain neurodegeneration disorder that causes issues including memory damage and other cognitive impairments extreme enough to interfere with regular activities [22]. This is one of most prominent forms of dementia, contributing for 60-80% among all dementia cases [23]. In developed countries, AD is one of the most expensive neurodegenerative diseases. In 2006, 25 million people were suffered by Alzheimer's, and by 2050, 1 in 85 people will be affected [24]. According to the same article [24], over 2/3 of dementia sufferers live in a under develop nations, which are likely to witness a significant increase in dementia cases in the next years as the regions develop swiftly. It will be challenging for a variety of reasons, the first is that dementia patients in these nation rely significantly on informal care, making it extremely difficult to sustain effective treatment and care for such a huge number of older people as disease prevalence rises [25]. Financially, dementia has a massive social impact at the moment, accounting for 1.01 percent of global Gross National Product [26]. This issue is predicted to deteriorate in the coming years, with a projected 85% rise in worldwide societal expenses by 2030 [27], assuming no changes in other background factors (e.g. macroeconomic, dementia occurrence and prevalence, accessibility, and treatment efficacy).

While some of its symptoms may resemble those of progressive aging, it is vital to remember that AD (and certainly dementia in general) is not natural part of the aging process. Dementia symptoms grow as the disease progresses.

AD currently has no cure; the goal is to slow the disease's progression, restore symptoms, treat behavioral issues, and improve overall quality of life [28]. However, if dementia symptoms are detected early enough, these existing medicines can temporarily slow the progression of the condition. While more effective treatments, prevention, and finally treatment are all significant (long-term) goals, early diagnosis may provide patients with better cure results. Except in rare cases of observable genetic variations, the exact cause of AD remains unknown. However, new study suggests that it is connected with neurotic plaques and neurofibrillary tangles inside the brain [29]. While Amyloid beta, a protein that composes neurotic plaques, is known to have a role in the course of AD, many studies investigate whether it is still a causal element. It is, however, often recognized as a symptom of the condition [30].

The capability to detect the disease and follow its evolution before symptoms arise is constantly increasing. Recently, some advancements in research have seen, the most notable of which is the finding of biomarkers (especially brain imaging technologies) that allow for the recognition of AD-related courses over decades [27]. The biomarkers for Alzheimer's disease are mostly based on their behavior, which is generally evaluated in amyloid deposition in the brain (e.g., PET scans, CSF amyloid), and other biomarkers, which commonly compute neurodegeneration volume (e.g. sMRI, FDG-PET, CSF-tau) [31].

The only sole technique to determine whether or not someone has been impacted by a illness is a post-mortem brain tissue study. However, both neurotic plaques and neurofibrillary tangles in brain appear to have a role in the pathogenesis of Alzheimer's disease. Despite the fact that extensive study has been done on AD, still there is a need of a (non-invasive) diagnostic tool.

2.1.1 Mild Cognitive Impairments

MCI are presumed to be an initial phase of Alzheimer or a transitional period between both the projected cognitive impairment of aging and the extreme deterioration of dementia, though it's unclear whether MCI is a wide range of analytical stage or a prodromal disease stage [32]. Memory, language, reasoning, and judgment impairments that are more severe than normal age-related declines are common. Brain pattern changes in patients with MCI have been going on for ages, and indicators are just now starting to emerge. However, it has yet to be demonstrated that these indicators are severe enough to interfere with day-to-day work, a state that would be categorized as dementia. MCI may maximize your odds of early onset dementia later life due to Alzheimer's disease or even other neurological disorders. However, many people with moderate cognitive impairment never worsen, while others improve over time [33].

2.1.2 Neuropathology

Variations in the brain (such as structural abnormalities) cause symptoms which could be used to identify a disease. The course of AD is predictable. The progression of Alzheimer's disease is divided into six phases based on the supply of neurofibrillary tangles (NFT): lesions are first placed in the trans-entorhinal cortical area (phase I), then to phase II (entorhinal cortex), then to hippocampus and limbic lobe (phase III and IV), then to the neocortex (phase V), and finally to the primary cortex (phase VI) [22]. These phases are classified into entorhinal (I and II), limbic (III and IV), and neocortical (V and VI) phases, and they are all linked to clinical and cognitive impairment, reflecting degradation of the cortical areas associated with these processes. An NFT makes the hippocampus

and medial temporal lobe vulnerable, and neuronal death can be detected even in the early stages of the disease. As a result of the loss of synapses and neurons caused by AD, we were able to clearly see variations in brain soft tissue, as describe in Figure 2.1.

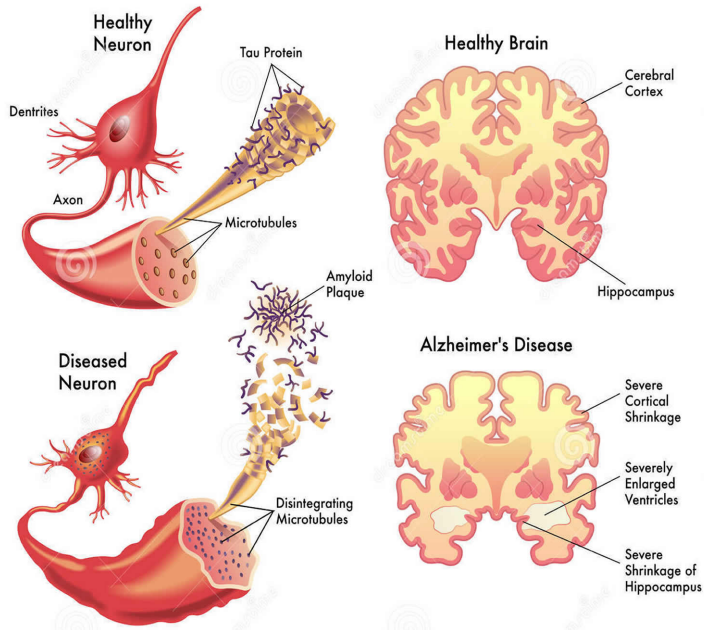


Figure 2.1: Healthy brain and Alzheimer's disease effected brain illustration.

Clinical standards, based on neuropsychological examination, are used to diagnose AD. The main goal of these studies is to find indicators for the detection of AD. Furthermore, sMRI allows us to see brain atrophy, which is a reflection of neuronal death. In recent years, numerous studies have been conducted to identify diagnostic biomarkers for AD from sMRI images [34].

Researchers and doctors can use magnetic resonance scans to examine the alterations in the structural part of brain linked with AD. Other imaging

techniques include the Pittsburgh Compound B PET (PiB - PET) neuroimaging method, vividly reveals the forms of β -amyloid bonds in the brain. This approach is more invasive, but it does require a minimal quantity of radioactive sugar to be enmeshed in the patient's brain [35].

2.1.3 Biomarkers

Whenever AD advances, the magnitudes of its biomarkers approach abnormal levels, as seen in Figure 2.2. The figure shows biomarkers as dementia indicators, and the curve shows the changes induced by the five biomarkers examined (in chronological order):

- Amyloid β imaging has been observed in the CSF and on PET.
- Increased CSF-tau species and synaptic pathology, as determined by FDG-PET, indicate neurodegeneration.
- sMRI is used to determine the damaged brain region and neuron loss (particularly noticeable in the hippocampus, medial temporal lobe, and caudate nucleus).
- Cognitive testing measures memory loss.
- The cognitive examination measures general cognitive decline.

The first three biomarkers on this list can be noticed before a dementia diagnosis, but the last two are "typical signs of dementia diagnosis" [36]. Biomarkers, which are primarily intended for use in research, do not cooperate with the diagnostic framework in order to provide appropriate therapy. This

included genetic markers that explicitly simulate AD pathology, including β -amyloid protein and tau, and also biomarkers that have a less direct affect or are non-specific symptoms of the disorder, such as racking directories of neuronal damage, which are slightly precise for an AD due to the ROI structure of anomalies [36].

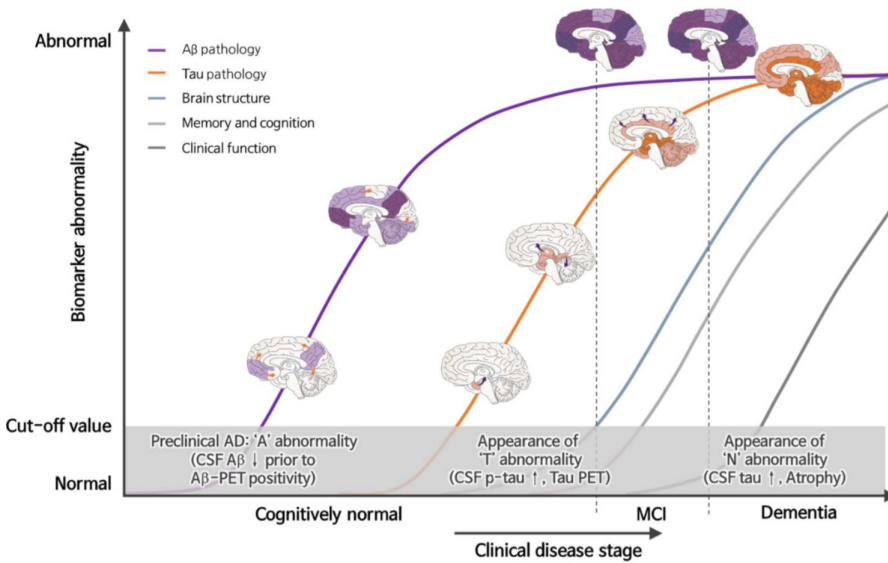


Figure 2.2: Progression of Alzheimer's Disease biomarkers [36].

2.1.4 Risk Factors

Although no definite cause for dementia has been identified, certain factors are significantly connected to the AD progress. Which will be briefly described further on.

2.1.4.1 Age A substantial risk for AD. After the age of 65, the likelihood of developing Alzheimer’s disease increases every five years, reaching roughly 50% by the age of 85 [37].

2.1.4.2 Family History AD is also linked to family history and background. Individuals who have a close relative with AD are more likely to be effected. The risk factor rises in proportion to the number of sufferers in a family. In families, inherited, environmental, or both attributes may have a job in the development and progression of diseases [38].

2.1.4.3 Genetics Genetics might have a role in the course of AD. There are two categories of genes that impact Alzheimer’s disease progression:

- Genes associated with risk (growth in likelihood).
- Genes that are deterministic (believe to have the direct impact on disease).

If there is a family history of Alzheimer’s disease (i.e. dominant autosomal form of AD), the onset of early Alzheimer’s symptoms (i.e. MCI) indicates the dementia. Early onset AD affects a large number of people. Individuals’ MCI development and progression to Alzheimer’s dementia differ. The study does reveal, however, that having one or two $\epsilon 4$ genotypes in the apolipoprotein E (APOE) genome is substantially associated with an elevated risk of late-onset dementia. The existence of the $\epsilon 2$ gene, on the other hand, reduce the risk factor in an individual [39].

2.2 Magnetic Resonance Imaging (MRI)

MRI is a medical imaging system that offers an image of the anatomy as well as physically processes the body in both healthy and sick stages. It creates pictures of body organs using powerful magnetic fields and radio waves. sMRI is indeed a non-invasive scanning technique that generates three-dimensional anatomical scans without exposing the patient to harmful radiation. It is used to spot diseases, diagnose them, and track their progress [40].

2.2.1 Technology

MRIs provide powerful magnets in most units, which provide a strong magnetic field signal that forces the proton in the body to coordinate with fields. The proton begins to stimulate when a radio frequency pulse passes through the patient's body, and it spins out of equilibrium, straining against the magnetic field's pull. Furthermore, once the radio frequency pulse is turned off, the sMRI sensors may measure the energy released by the realign protons within the magnetic field. A time taken for protons to create a magnetic field is totally determined by the chemical molecules in their vicinity. On the basis of these patterns of magnetic characteristics, physicians can discern the difference between distinct types of tissues [40]. sMRI scans, on the other hand, require a patient to be positioned inside a huge magnet and to remain stationary throughout the imaging process in order for the image to be clear. To increase the speed with which protons readjust to the magnetic field, contrast agents chemical (containing the element Gadolinium) may be given intravenously to a subject before or during the sMRI process. The brighter the scans are, the faster the protons adjust.

2.2.2 Imaging

The resulting photos exhibit many shades of gray color that reflect varying thicknesses because these areas have less water, there are fewer number of hydrogen protons that emit feedback signal to the radioactive detectors. T1 weighing will provide pictures of thick bone, air, and other substances with less hydrogen protons, which will be dark, and fat will be lighter, and so on. Depending on signal strength, the voxel pixel number can be one of 255 shades of gray, with 0 indicating black and 255 indicating white. Figure 2.3 depicts brain slices.

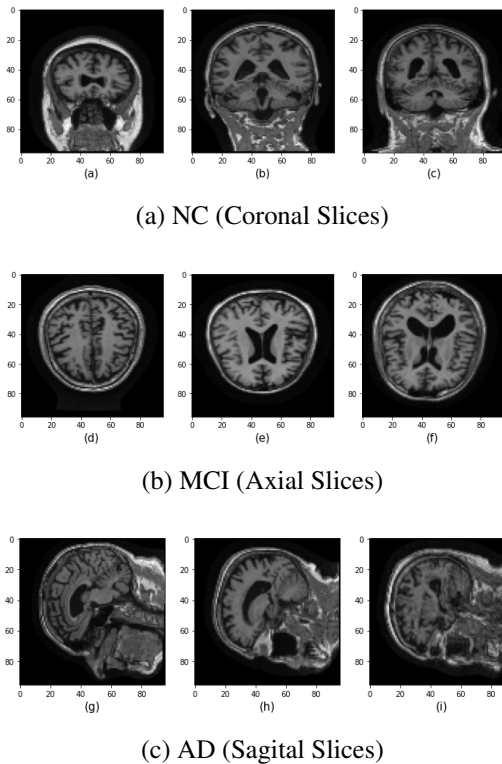


Figure 2.3: MRI scans depicted various stages of Alzheimer's disease.

2.3 Database Organization

This section describes the organization that aims to improve understanding of Alzheimer and shares information for researchers all over the world; without their assistance, this thesis would not have been possible.

2.3.1 Alzheimer’s Disease Neuroimaging Initiative (ADNI)

The ADNI is a long-term research project whose goal is to “create clinical, imaging, and genetic indicators for the identification of AD.” The six-year “ADNI-1” trial began in 2004 with 400 MCI patients, 200 early AD patients, and 200 old control patients. However, from 2009 to 2011, this study was expanded by “ADNI-GO,” which included 200 more participants who were identified as having early MCI and were tested for biomarkers in the early phase of the disease. ADNI organization marks a big step towards a development of improved diagnostic procedures that could assist to delay the progression of AD and, eventually, prevent it. The ADNI website has more detailed information (<http://adni.loni.usc.edu/>).

2.4 Machine Learning

An interdisciplinary topic that relies on studies from several disciplines such as artificial intelligence, statistics, philosophy, and neuroscience. It can also be defined as the science of teaching computers to learn and perform tasks in the same way that humans do, and then recovering their learning process in an autonomous manner by providing them with real-world data and observations. A ML algorithm’s learning system is describe in Figure 2.4.

Furthermore, ML paradigms are classified into three types: supervised,

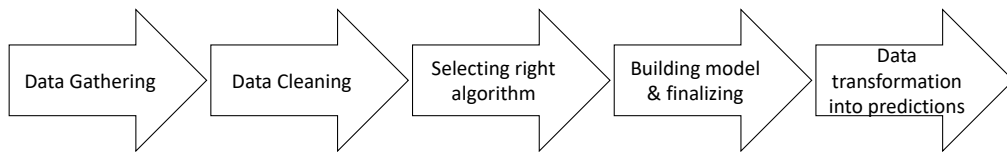


Figure 2.4: Working process of Machine Learning.

unsupervised, and reinforcement learning. Supervised learning is concerned with tagged sample data, in which inputs are linked to desired outputs. In the psychological concept of concept learning, this might be seen as a parallel. Unsupervised learning, on the other hand, uses unlabeled data with no errors in order to determine a potential solution. Reinforcement learning, on the other hand, is concerned with an aim to attain a specific goal by completing an action in an active environment to get a reward, without being explicitly alerted if the learner is on the verge of reaching its goal [41].

There are several ML simulations in both supervised and unsupervised learning, which together set various prior assumptions about the expected input-output mappings or data sharing. Because the problem with which the models are dealing is inadequately and the training data is inadequate for the models to find the correct response on their own, they make these various assumptions out of necessity. For example, the mapping may not happen, or there may not be enough data to reassemble it, or there may be inevitable noise, resulting in a suitable model that is worthless in the real world. Inductive bias refers to the set of expectations that learning techniques create in order to facilitate learning. A learning procedure's inductive bias (also known as learning bias) is the set of assumptions that the learner use to guess the outputs of unknown inputs. As a result, inductive bias is crucial since it governs how well the learner generalizes

beyond the given training set. The learning process, such as least mean squares or mean squared error, tries to reduce some quantity of error (or loss function). They accomplish this by selecting the best weight for the set of training samples, i.e. and, which also assists in reducing the loss between observed training instances (and other prior constraints) and estimated values [42].

This work concentrate on supervised learning approaches like Support Vector Machine (SVM) classifier, Naive Bayes, Random Forest (RF), K-Nearest Neighbors (KNN) and Softmax classifier etc. along-with unsupervised learning methods for preprocessing including PCA and RFE.

2.4.1 Deep Learning

Deep learning(DL) allows computational models with several processing units to learn and understand information at various levels of abstraction, mimicking how humans access and perceive multimodal data and indirectly capturing sophisticated data patterns. DL models are indeed a subcategory of representation learning techniques because they can autonomously identify the optimal representation from original data without the requirement for feature selection beforehand. This is attained by the use of a rigid hierarchy with increasing complexity, along with sequentially making nonlinear alterations to the raw data. Higher-level attributes are less susceptible to input data misinterpretation than low-level features, resulting in higher levels of abstraction [18].

DL neural network models, also known as neural nets, try to imitate the human brain by combining input data, weights, and bias. These components work together to identify, categorize, and characterize items in data.

DL models are composed of many layers of connected nodes, each of which

improves and alters categorization. The advancement of calculations over the network is referred to as forward propagation. Both input and output layers of a deep network are the visible layers. Before producing final predictions or categorizations in the output layer, the deep learning model processes the data in the input layer.

Backpropagation is another approach that uses techniques such as gradient descent to generate generalization error and then alters the function's weights and biases by traversing the layers backwards to train the model. Forward and backpropagation act in tandem to enable a neural network to foresee and correct for mistakes. The algorithm improves and gets more accurate with time.

The preceding illustrates the most fundamental type of deep learning model. DL approaches, on contrary, are exceedingly sophisticated, with several types of neural networks employed to handle specific difficulties or data. As an example:

- **Convolutional neural networks (CNNs):** Can discern features or patterns in an image, allowing tasks like as item identification and authentication to be done. For the first time, a CNN outperformed a human in an object identification challenge in 2015.
- **Recurrent neural networks (RNN):** Employs a sequential or time-series data. Natural language and speech recognition applications typically use RNN. .

DL encompasses a wide range of learning approaches, including Deep Belief Nets (DBNs), Deep Auto Encoders (DAE), and Deep Neural Networks (DNNs). In recent years, DL techniques have outperformed traditional approaches, particularly in the area of computer vision. As of October 2014, researchers were

able to achieve good accuracy on numerous standardized classification image datasets utilizing DNN based on GPU architecture and some type of dropout regularization technique. On key standardized benchmark datasets, such as, DNN has produced outstanding results (MNIST, CIFAR-10, 100, STL-10, ImageNet, etc.). DL researchers have recently created methodologies to help in the practical implementation of DL techniques including sparse initialization, pretraining, fine-tuning, and regularization methods like dropout [43], [44].

2.4.1.1 Convolutional Neural Network (CNN)

CNN is among the most extensively applied architecture, with applications spanning from image recognition to mobile vision, object identification, and surveillance. It takes inspiration from biological organisms optic nerves and analyses data using a huge number of neuron connections to attain great precision. These networks necessitate a large amount of computing and data storage, posing a challenge in terms of both computational performance and energy economy. The CNN follows a specialized computing pattern that is inefficiently performed by conventional computation processors. Several CNN accelerators based on various hardware platforms have been in recent years in this area [45]–[47].

A CNN is typically made up of two parts: feature extraction and classification. Edges, lines, and corners are examples of input features that are unaffected by position and distortions. The feature extractor takes the input and extracts the features, mapping them to a low-dimensional vector output feature map. The collected characteristics are fed to the classifier through a sequence of such computational levels, as well as optional sub-sampling layers. The classifier is a

fully connected framework that determines input's class. The animation exhibits the workflow of CNN processing an image as input and categorizing objects based on values.

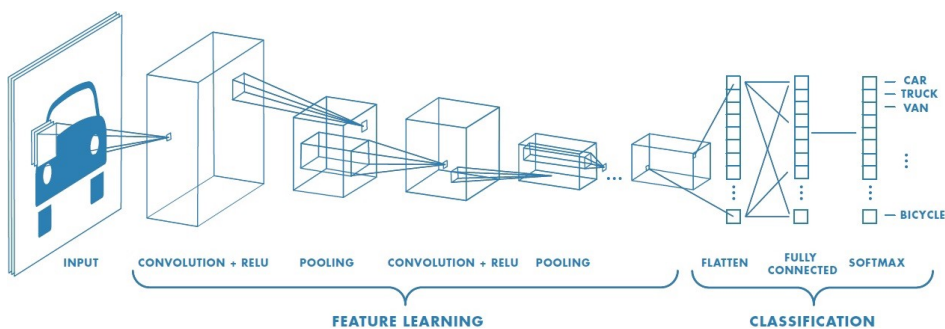


Figure 2.5: Convolutional Neural Network structure [48].

CNN's layers are classified into three types:

- Convolutional layer
- Pooling layer
- Fully Connected layer

A CNN model's first layer is the convolutional layer. While pooling layers can be introduced after the convolution layer, the last layer is the fully-connected layer. The CNN becomes more complicated with each layer, detecting more areas of the image. Earlier layers concentrate on basic elements such as colors and boundaries. As the CNN layers analyze the visual input, it learns to distinguish larger portions or attributes of the item, ultimately recognizing the object. [49].

Convolutional Layer: The convolutional layer is by far the most significant component of a CNN as it is where the most of the computing occurs. It takes data input, a filters, and a feature map, among many other things. Assume the input is a color picture composed of a 3D matrix of pixels. This means that the input will have three components: height, width, and depth, that correspond to the RGB color space of an image. A feature detector, also referred as a kernels or a filter, will search the image's representations for the existence of the feature. This method is titled as convolution.

The feature estimator is a two-dimensional weighted matrix that duplicates a part of the image. The filter size is typically a 3x3 array, that also determines the size of the receptive field. After applying the filter to a section of the image, the dot product between the input data and the filter is determined. This dot product is sent to the array. The filter then advances one step, and the process is repeated until the kernel has completely swept across the image. A feature map, activation map, or convolved feature is the end result of a sequence of dot products created by the input and the filter.

Pooling Layer: Downsampling is another term for layer pooling. That is a dimension reduction method for reducing the number of features in the data. The pooling approach, similar to convolution operation, applies a filter to the whole input except using weights. Instead, the kernels employs a clustering method to generate the output matrix from the receptive field values. Pooling is classified into two types:

- **Max pooling:** As it moves over the input, the filter chooses the pixels with the greatest value to send to the output array. This method is used more often than ordinary pooling.

- **Average pooling:** As the filter sweeps over the input, the average value within the receptive field is computed and transferred to the output array.

While a pooling layer eliminates a lot of information, it still offers several advantages to the CNN. They contribute to the reduction of complexities, the improvement of efficiency, and the avoidance of overfitting.

Fully Connected Layer: The term "fully-connected layer" is self-explanatory. As explained previously, the image pixels of the input image are not directly connected to the output layer in partially linked layers. Each node in the output layer is directly linked to a module in the preceding layer in the fully-connected layer.

This layer performs categorization tasks based on the features gathered by the preceding layers and associated filters. Whereas convolutional and pooling layers utilize relu functions to categorize inputs, FC layers use a softmax function to provide a probability range from 0 to 1.

2.5 Related Work

Several techniques for neuroimaging classification have been developed in recent years to enhance classification performance, taking into consideration the advantages of ML and DL models. Some of the recent studies regarding ML and deep convolutional neural networks for neuroimaging analysis are examined below:

2.5.1 Machine Learning based Alzheimer’s Disease Classification

Many ML-based techniques for multi-class and binary class classification for the early detection of AD have been utilized in recent times. For example, in their study, Kim et al. [50] developed a automated classification method based on cortical thickness. Whereas, Long et al. [51] examined localized morphological abnormalities in the brain and observed that distortion in the amygdala and hippocampus aided in the diagnosis of progressive MCI. However, their study stated that diffusive structural changes in the whole-brain gray matter were responsible for diagnosing mild or moderate AD. Following that, a linear support vector machine was used to classify these individuals (SVM). Guo et al. [52], on the other hand, proposed a method for saving data during feature extraction. The proposed approach uses a multi-kernel SVM to classify brain area and subgraph features from functional magnetic resonance imaging. This method retains both meaningful features information and brain region vulnerability to alteration. In contrast to Guo et al. [52], Khedher et al. [53] proposed a technique for identifying AD that employed independent component analysis to retrieve information from the WM, GM, and CSF regions, followed by an SVM classifier. Furthermore, Tong et al. [54] constructed a multiple instance learning-based approach for alzheimer classification by isolating the regions of MRI voxel patches and mapped them to graphs. An SVM classifier was then used to distinguish the gap amongst patients with ad and NC (Normal Control) individuals.

In later years, Zhang et al. [55] used multiple-kernel SVM to develop a multimodal classification system based on diagnostic markers such as sMRI, PET, and cerebrospinal fluid (CSF) to differentiate AD (or MCI) and normal

control (NC) participants. Their proposed model has a good accuracy for AD classification and an encouraging accuracy for MCI classification for binary classification (AD versus NC and MCI vs NC) results. Salvatore et al. [56] employed MR images as a marker for early classification of Alzheimers in comparison to other groups in 2015. For feature extraction, they employed the PCA function, and for classification, they used the SVM classifier. They discovered that the hippocampus, entorhinal cortex, basal ganglia, gyrus rectus, precuneus, and cerebellum are key regions involved in AD pathophysiology. The classifier achieved classification accuracy of 76% for AD vs. CN, 72% for MCIc vs. CN, and 66% for MCIc vs. MCInc using the nested 20-fold cross validation technique. In addition, Baron et al. [57] introduced a voxel-based feature extraction technique and used empirical voxel features for diagnosis in their study. In another study, Gupta *et al.* [58] employed machine learning algorithms (SVM, k-nearest neighbor (KNN), and Random Forest to distinguish atrophic conditions (AD, NC/healthy control (HC), asymptotic Alzheimer’s disease (aAD), and mild Alzheimer’s disease (mAD)) (RF).

2.5.2 CNN approaches for Alzheimer’s Disease diagnosis

DL has highlights the potential in medical imaging diagnostics [59]–[61], where it was originally used for segmentation methods or extraction of features, followed by conventional machine learning techniques like SVM and boosting. Silva et al. [62], for instance, proposed a cnn architecture for extraction of features from MRI scans, proceeded by SVM, KNN, and Random Forest techniques for AD diagnosis. Correspondingly, Liu et al. [63] proposed a novel fully convolutional learning technique for Alzheimer’s and Parkinson’s disease

classification that combined supervised and unsupervised methods learning. DL has lately succeeded in the field of computer vision for extracting visual information, and their potential for AD diagnosis has been investigated [64]. For instance, Wang et al. [65] applied brain removal methods to select hippocampus-containing sections before feeding them to a deep convolutional neural network for assessment. To anticipate the AD and NC classes, a patch-based ensemble model was created [66]. In addition, 2-D and 3-D deep learning models for the diagnosis of AD were constructed in [67]. Unlike others, Basaia et al. [68] processed the data by warping, resizing, and rotating the MRI pictures at multiple viewpoints before feeding it into a deep neural network for disease classification.

In a recent study, Korolev et al. [69] presented two 3D cnn models inspired from VGGNet and ResNet, indicating how manual extraction of features is not necessary for MRI images classification. Their 3D models for medical image classification, such as 3D-VGG and 3D-ResNet, are widely used in research. To categorize ADNI data, Ehsan Hosseini-Asl et al. [6] and [70] used 3D CNN. A 3D deep neural network was used to extract features from MRIs and identify biomarkers for many kinds of AD. Similarly, Abrol et al. [71] created 3D CNNs using the ResNet architecture and tested them on a range of classification tasks. Using ADNI data, they created a training dataset for cross-validation and a smaller test set. Despite the encouraging results, popular assessment frameworks were performed, leaving the possibility that the algorithm was overfitted on the training examples. [72] provides a deep learning-based algorithm for AD versus NC assessment, in which a discrete volumetric estimations model with CNN models is used to extract deep properties of the left and right hippocampus discrete volume (RHM and LHM). Recently, [2] proposed a diagnostic model for AD using a densely connected 3D CNN and an attention-driven technique

to integrate high-level features with spatial data collected from MRI. Lately, the depthwise separable convolution was proposed by J Liu et al. [73], which replaces the depthwise separable convolution for the standard convolution. They used AlexNet and GoogLeNet transfer learning methods to train their concept, which significantly reduced the processing cost and parameters.

In this section, previously published journal papers employing ML and DL methodologies to study AD have been presented. In the next section, proposed model using the methods and ideas mentioned above will be explained in detail.

3 Proposed Method

3.1 Overview

The procedure of the experiments covered in Chapter 1 is presented in this chapter. First phase of thesis is based on machine learning (ML) methodologies, where ML offer an effective classification strategy for distinguishing Alzheimer's disease (AD) from other diagnostic categories. This section is divided into two sections. One component of the section is dedicated for the suggested ML framework. Whereas, the rest of the section discuss about the deep neural-based model.

3.2 Machine Learning method for Alzheimer's Disease Classification

3.2.1 sMRI Dataset

There are almost 6000 people in the ADNI dataset, ranging in age from 18 to 96. Preprocessed images of all 313 patients in this study that met the ADNI protocol were chosen from this group. The statistics of the subjects used in this thesis are described in Table 3.1.

For an unbiased examination, the data was split in (70/30 ratio). 70% of the data was allocated to training, while 30% was allocated to testing. Moreover, Gradient inhomogeneity correction, non-uniformity correction, and bias field processing were already applied to all sMRI images [47], [74]. Furthermore, each scan was coupled with a phantom-based scaling procedure, and later masks were constructed using the MR Core approach, which already includes

Table 3.1: Statistics representation of all subjects.

Class	Subjects	Age	Gender M/ F
AD	78	76.5 ± 5.7	42/36
HC	78	79.6 ± 5.4	35/43
LMCI	79	72.5 ± 6.7	41/38
EMCI	78	74.5 ± 6.9	40/38

AD: Alzheimer’s disease; LMCI: Late Mild Cognitive Impairment; EMCI: Early Mild Cognitive Impairment; CN: Normal control.

pre-processing processes in ”Intracranial Space”. These datasets additionally provide corresponding metadata information for each brain image, including demographics such as gender, beginning subject weight, age, and diagnostic group. The ADNI 1 group provided the data for this experiment.

3.2.2 Selected Features

For the extraction of features, FreeSurfer software was used to have two volumetric features.

- Cortical Features
- Subcortical Features

3.2.2.1 Volumetric volumes

The term ”volumetric feature” refers to the ability to calculate the volume of certain brain locations by accumulating the voxels inside the traced region

of interest (ROI). In this work, in total 873 extracted features from cortex and subcortical volume of each subject's brain were utilized that are extracted using the Freesurfer toolbox [75].

3.2.2.2 Cortical dementia

The cortex is the part of the brain with which most people can identify, at least visually. The outer layer's unique twists and turns play a significant role in information processing and the procedures like language and memory. Cortical dementia [76], [77] is usually linked to gray matter in the brain. When the outer layers of the brain are weakened, as in Dementia, Binswanger's disease, and Creutzfeldt-Jakob disease, challenges with remembering, selecting appropriate words, and understanding what others are saying develop (aphasia).

3.2.2.3 Subcortical dementia

As the name indicates, these dementias are hypothesized to affect parts of the brain underneath the cortex are predominantly affiliated with white matter. Huntington's disease, and Parkinson's dementia are all examples of subcortical dementia. Personality changes and a slowing of mental functions are more common in subcortical dementias. In the early stages of many dementias, language and memory abilities generally appear to be intact. The cerebral cortex degenerates widely in most types of dementia, resulting in tangles, which are the markers of AD. Other types of dementia involve targeted injury to areas beneath the cortex, giving birth to the term "subcortical dementia" [77].

3.2.3 Extraction of Features

The features of this experiment were cortical thickness and subcortical volume segmentation at each vertex of the cortical, subcortical surface. As illustrated in Figure 3.1, the Freesurfer 6.0 software [75] suite is employed, which is a fully automated volumetric segmentation and cortical reconstruction pipeline. Subcortical volumetric and cortical thickness measurements have been used routinely for classification. Where, cortical thickness is a direct indicator of atrophy, making it a potentially powerful candidate for AD diagnosis. The original MRI data were first submitted to a variety of pre-processing processes before being processed. Such as Motion correction, T1-weighted image averaging, registration of the volume in the Talairach space, and skull stripping with a deformable template model. Moreover, EM registration (linear volumetric registration), neck removal, CA labeling, intensity normalization, and white matter segmentation are also included.

In this study, freesurfer software was utilized to generate the cortical thickness and subcortical volume of a brain attributes automatically as show in the Figure 3.1.

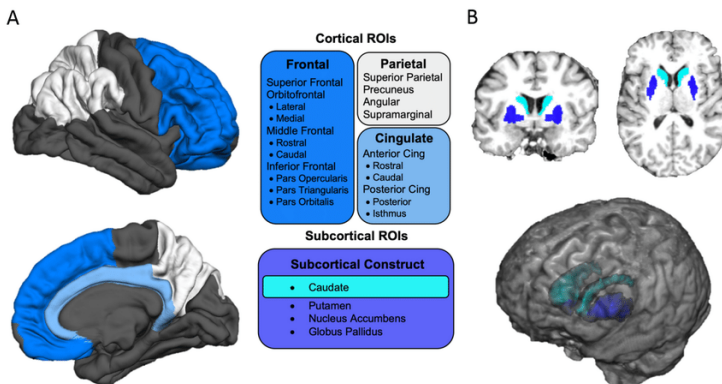


Figure 3.1: a) Cortical ROIs, b) Subcortical volume extracted features [78].

3.2.4 Selection of Features

Following the feature extraction stage, the obtained information was normalized to zero mean and unit variance using a standard scalar function. The purpose of normalization is to remove data inconsistencies that make analysis more difficult. The standardized sequence of components $x(i, j)$ is given by:

$$X_n = \frac{x(i, j) - \text{mean}(X_j)}{\text{std}(X_j)} \quad (3.1)$$

3.2.4.1 Principal Component Analysis (PCA)

It's a type of learning procedure in which a higher-dimensional sample or feature has been taken, such as photos in matrix form, and then use some approaches to convert it to a lower-dimensional space. Any dimensional feature can be reduced to a 2D or 1D plane using this method. Furthermore, this technique was utilize to minimize features of the subcortical and cortical volumes that were retrieved using the Freesurfer toolbox. In this experiment, Principal component analysis (PCA) employed as dimensionality reduction technique [79].

Basically, PCA builds a sequence of starting features and used that $k < d$ to transform the information from d -dimension field to a k -dimension region. The attained variables k are referred as the principal components (PCs). Except for the variation that is already accounted for in all subsequent components, each PC is directed to the maximum variance. The first component, in comparison to the following components, covers the larger deviations. The following formula can be used to compute PCs:

$$PC_i = a_1X_1 + a_2X_2 + \dots + a_jX_j, \quad (3.2)$$

Where in equation (3.2), PC_i represent the i th principal component, X_j shows the original feature j , and a_j is the numerical coefficient for X_j . PCA is the most extensively utilized feature selection approach. Whereas, it just reduces the dimension of the features. Using a feature selection method, on the other hand, the model only runs on the basis of selected features and does not change. We first use PCA to minimize the dimension of a feature, then use the RFE feature selection approach to select the key features.

3.2.4.2 Recursive Feature Elimination (RFE)

RFE is a wrapper type feature selection strategy. This approach builds models from a subset of input characteristics and chooses the best ones based on performance parameter. The finest example of feature selection in a wrapper is RFE. The RFE algorithm is specified as follows in Algorithm 3.1.

3.2.5 Classification

To examine the classification accuracy based on subcortical and cortical characteristics, five different commonly used classifier algorithms were testified. A comparative studies of the used classifiers are described in table 3.2.

3.2.5.1 Random Forest (RF)

Random forest is a supervised learning method for classification and regression in ML. It's a classifier that takes the average of the outcomes of

Algorithm 3.1 Recursive feature elimination

- 1: Train a model with all of the predictors in the training sample
- 2: Model Performance should be calculated
- 3: Call BuildTree to calculate variable relevance or ranking
- 4: **for** *Each subset size $S_i, i = 1, \dots, S$* **do**
- 5: Keep just the most significant S_i properties
- 6: Train the model on the training set using S_i predictors
- 7: Determine the model's performance
- 8: Recalculate the predictor ranks
- 9: **end for**

BuildTree(N):

- 10: **if** only one class is represented by N instances **then**
 - 11: **return**
 - 12: **else**
 - 13: Select $x\%$ of the potential dividing features in N at random.
 - 14: To split on, choose the feature F with the biggest information gain
 - 15: Create f nodes of N, N_1, \dots, N_f , where F might have f different values
 - 16: **for** $i = 1$ to F **do**
 - 17: Set the contents of N_i to D_i , where D_i is all instances in N
 - 18: F_i
 - 19: Call BuildTree(N_i)
 - 20: **end for**
 - 21: **end if**
 - 22: Determine the performance profile over S_i
 - 23: Decide on the optimal amount of predictors
 - 24: Select the model that corresponds to the best S_i
- end**
-

multiple decision trees applied to different subsets of a dataset in order to increase the dataset's projected accuracy.

It is also called as a meta-estimator since it employs the aggregate to increase the model's prediction and minimize over-fitting by fitting a large number of decision trees to a wide range of data. Even though the sub-sample length remains the same as the original size of data, the samples are formed through substitution.

It forms a "forest" from a cluster of decision trees acquired by the "bagging" method. The bagging technique is based on the premise that merging many learning models boosts the end outcome. Rather than reliance on a single tree structure, the RF collects forecasts from each tree and predicts the final outcome based on the majority of votes.

3.2.5.2 Support Vector Machine (SVM)

Widely used supervised learning classifier for data classification and prediction. However, it is commonly used in machine learning to overcome classification difficulties.

The SVM approach's purpose is to find the optimal line or decision boundary for classifying n-dimensional space into classes so that following data points may be conveniently placed in the correct category. In that case hyperplane is the best border option.

SVM techniques categorize information and train models with extremely restricted degrees of polarity, yielding a three-dimensional classifier that extends beyond the X/Y predictive axes. As demonstrated in Figure 3.2, SVM is used to discover extreme points/vectors that help in hyperplane design.

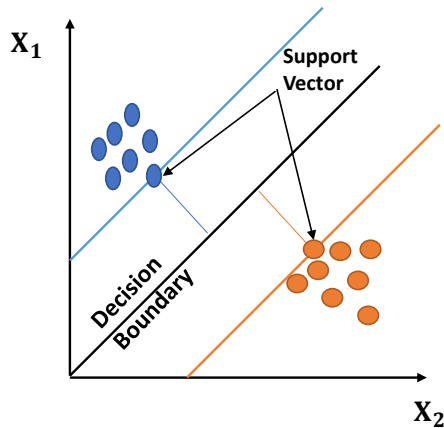


Figure 3.2: Support Vector Machine

3.2.5.3 K-Nearest Neighbors

K nearest neighbors (KNN) is indeed a simple approach which retains both existing samples and classifies new based on similarities (e.g., distance functions).

KNN has been utilized as a non-parametric approach in statistical estimates and pattern recognition since the early 1970s. It's a lazy learning technique since it doesn't strive to develop a generic internal model; instead, it only keeps instances of the training data. The categorization is determined by a majority vote of the each point's k nearest neighbours.

A case is assigned to the class with the most representation among its neighbors, as defined by a distance function, by a majority of votes cast by its neighbors. If $K = 1$, the case is simply assigned to the class of its nearest neighbor.

3.2.5.4 Naive Bayes

The Naive Bayes (NB) family of probabilistic algorithms determines how

likely each data point is to fall into one or more of a set of categories (or not). It is a supervised learning method to deal with classification issues based on the Bayes theorem. It's a probabilistic classifier, meaning it makes predictions based on the likelihood of an object.

3.2.5.5 Softmax Classifier

Softmax classifier is a simplified representation of logistic regression models that can be used to solve problems involving multiple mutually inimitable classes. Softmax regression is a sort of logistic regression that converts a data input into a vector of features that represent a probabilities and accumulates to 1. Softmax classifier has known applications in discriminative models like Cross-Entropy and Noise Contrastive Estimation.

3.2.6 Methodology

The technique may be described as follows:

- 1 Collect the pre-processed dataset consist of AD, eMCI, IMCI and NC groups.
- 2 Extract subcortical and cortical features using FreeSurfer toolbox.
- 3 Features normalization.
- 4 Dimension reduction.
- 5 Feature selection method performed.
- 6 Classification of AD in four groups using ML classifiers.

Table 3.2: A Comparative Study of Machine Learning based Classification techniques

ALGORITHM	FEATURES	LIMITATIONS
Random Forest	- Handle both categorical and continuous variables.	- A small change in the data can cause the algorithm to change significantly.
	- Noise has a lower impact on Random Forest.	- Compared to other algorithms, algorithm computations can be much more complex.
	- Outliers are usually tolerated by Random Forest and handled automatically.	- The technique may become too sluggish and useless for real-time predictions if there are a huge number of trees.
SVM	- In high-dimensional spaces, it is effective.	- More speed and size are required in training and testing.
	- Appropriate method when the sample size is less than the number of dimensions.	- In many cases, classification necessitates high complexity and extensive memory requirements.
	- Even if the information is not linearly separable in the base feature map, it performs well.	- Nonlinear problems are not suitable.
K-Nearest neighbor	- Classes are not required to be linearly separable.	- Finding the closest Neighbors in a big training data set might take a long time.
	- The learning process has no cost.	- It is responsive to unnecessary or disruptive attributes.
	- Ideal for multimodal classes.	- The number of dimensions impacts algorithm performance.
Naive Bayes	- Implementation is simple.	- The algorithm's precision decreases as the amount of data decreases.
	- Exceptionally high computational efficiency and classification rate.	- A large number of data are needed to acquire good results.
	- For the majority of classification and prediction problems, it predicts accurate results.	- Based on the assumption that all predictors (or features) are independent.
Softmax Classifier	- Makes no assumptions about class distributions in feature space.	- Not good for complex model, as it cannot handle a large number of categorical variables.
	- Model coefficients can be perceived as indicators of information gain.	- Prone to overfitting.
	- A natural probabilistic perspective on class predictions.	- This technique does not work if the independent variables are not associated with the target variable.

3.2.6.1 Architecture

3.2.6.2 Implementation Details

The datasets (ADNI) as whole is massive, and the huge in dimension. Therefore, other adjustments were necessary to make the learning feasible. The extracted feature values in the ML process usually are in high dimensional space, for that a large computational power is required to train a high dimensional data. So far, the gained output performance is not impressive due to the long time required to converge the data. In this perspective dimensionality reduction techniques in conjunction to transform high-dimensional feature values into low-dimensional feature values and select the best ones based on their performance can be useful to obtain the desired output. To accomplish that, PCA as a dimensional reduction method and RFE as a best feature selector in this experiment were acquired. All categorization problems were executed on Python 3.8.5 installed on Ubuntu 20.04 LTS. There were four categories of data in this thesis (AD, eMCI, IMCIs, and CN) and two types of characteristics (sMRI imaging modalities) (from where cortical and subcortical volumes were extracted from each subjects). Thus, the proposed method was tested on five different types of classification problems, all of which are binary class problems. Initially, around 873 features from each sMRI scan were extracted using Freesurfer v.6.0 automated toolkit. Later, those attributes were passed through the dimensionality reduction and feature selection method simultaneously to lower the dimension and to choose the most relevant attributes from all 873 features which were submitted to the classifier model. The block diagram for ML approaches is shown in Figure 3.3. As classifiers in this experiment, random forest, a support

vector machine (SVM), naive bayes, KNN, and softmax were testified. To get impartial performance assessments, the participants were randomly assigned to two groups, one is for training and another for evaluation, with a 70:30 split. In the training set, the cross-validation approach was utilized to discover the optimum hyperparameter values for the objective. The optimal hyperparameter values for the training set were identified using a 5-fold cross-validation (5f-CV) approach. For the aforementioned technique, the appropriate parameter value was utilized to build the 5 classifier using the training data, and the classifier's performance was thus assessed by utilizing the remaining 30% of data in the test samples. In this approach, the study was able to obtain unbiased performance estimates for each categorization problem and looked at four different options like accuracy, sensitivity, specificity, and Youden Index performance measure values for each group.

3.3 Deep Learning based Alzheimer's Disease Classification

3.3.1 Dataset

T1-weighted MRI data from ADN1/ GO of 163 AD patients, 163 MCI patients, and 163 normal controls (NC) were utilized in this investigation throughout a 24-month period. Table 3.3 shows the demographics of each category.

T1-weighted MR images were acquired sagittally in this experiment adopting a volumetric 3D MPRAGE with an image resolution of $1.25 \times 1.25\text{mm}^2$ in-plane and 1.2mm thick sagittal slices. The bulk of these scans were acquired with 1.5 T scanners. Furthermore, the structural MRI images used in the study had already been assessed for quality and had underwent gradwarp, B1 non-uniformity correction, and N3 processing steps.

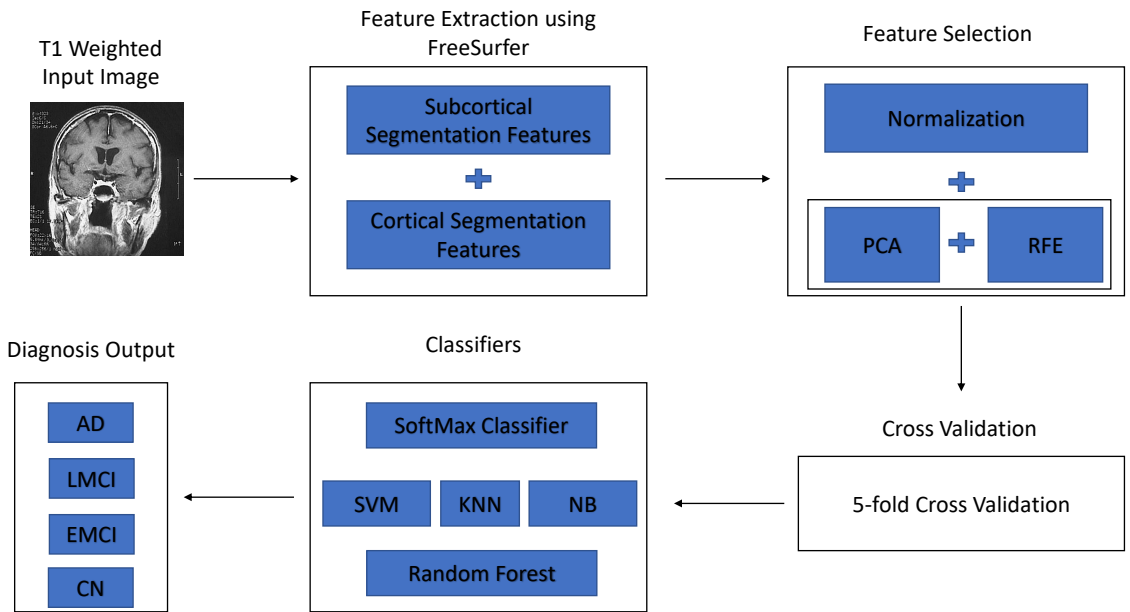


Figure 3.3: Machine learning based Proposed Method Architecture.

3.3.2 Image Patch Generation

Since MRI images are volumetric. As a result, a 3D Convolutional Neural (3D CNN) Network [80] is the DL model. In comparison 3D CNN models are more computationally expensive and time-consuming to train than 2D CNN models due to their high-dimensional nature. Another difficulty is that most of today's medical databases are rather tiny. Due to a shortage of data, it is difficult to train a deep network that can generalize to high degrees of complexity.

As a consequence, 489 3D MRI images with $192 \times 192 \times 160$ dimensions were employed in the experiment, which could not be input directly into a 2D CNN model. The 3D MRI images were initially scaled down to $96 \times 96 \times 96$. The

Table 3.3: Representation of ADNI subject’s demographics.

Group	No.of Subjects	Age Group	Gender M/F
AD	163	78.04 ± 6.40	82/81
CN	163	78.02 ± 5.33	78/85
MCI	163	77.14 ± 7.02	99/64

AD: Alzheimer’s disease; CN: Normal control;

MCI: Mild Cognitive Impairment.

images were then split into axial, coronal, and sagittal slices. Then, at the start and finish, certain slices were discarded because they lacked relevant information. The slices were similarly standardized, with a mean of zero and a standard deviation of one. The 2-D CNN model was then trained using patch slices chosen at random from the axial, coronal, and sagittal planes.

3.3.3 Deep Learning Architectures

Concept of CNN employed in this work was inspired by the human visual cortex. The human eye’s receptive field receives information, identical to the convolution procedure, which compresses the input picture and generates the subset of features by working with the input’s receptive field. Convolution employs several layers, including ReLU activation properties, max-pooling layers, and fully connected layers.

These procedures are applied to each input in order to generate a final result in the form of a binary or multiclass labels. A collection of neurons, shared hyperparameters, local connection, and shift invariance are utilized to link the convolution operation to enhance the network performance. We developed

an end-to-end convolution design for multi-label AD biomarker detection that employs the entire imaging volume as an input. In addition, we evaluate the proposed approach with the prominent CNN designs used for classification tasks, such as VGG-Net and ResNet. These models have been used effectively in a variety of applications to categorize, identify, label, and detect positions [81].

3.3.3.1 The VGG-Net Model

The VGG network is the name given to the pre-trained CNN model introduced by Simonyan and Zisserman [82] at Oxford University in early 2014. The Image Net ILSVRC dataset, which comprised images from 1000 distinct classes, was used to train VGG (Visual Geometry Group). For training, 1.3 million photos were used, and 50,000 were used for validation. VGG-19, a VGG architectural variant with 19 deeply connected layers, has consistently beaten other cutting-edge models. The model is composed of totally connected and strongly linked convolutional layers, which allows for improved feature extraction including the use of max-pooling (rather than averagepooling) for down-sampling prior to classification with the SoftMax activation function. The VGG-19 model is used as a baseline method in this study, with the ADNI dataset assisting in the classification of various stages of Alzheimer. Figure 3.4 depicts the construction of VGG-19.

3.3.3.2 ResNet Model

At ILSVRC-2015, Residual Network [83] took first place in the classification, object tracking, and identification tasks. The researchers were curious if boosting

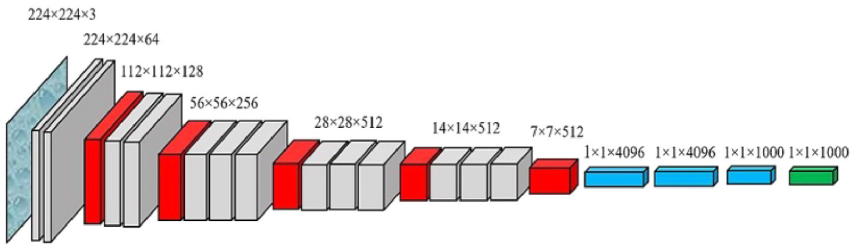


Figure 3.4: Basic Architecture of VGG-Net [78].

learning simply required stacking additional network layers on top of one another. They found the degradation problem, in which earlier models' performance, such as VGG, worsened rather than improved after a specific number of layers. To address this issue, they introduced the residual function, which serves as the foundation of a residual network (ResNet). ResNet was simply adopted from the non-bottleneck 50-layer design, where links with broadening dimension appear to be either (A) identity links, i.e., zero padding, or (B) projection links, i.e., convolutions with 1×1 (kernel) size, and has been used as the basic model to characterize AD using the ADNI dataset. ResNet-50's fundamental design is illustrated in Figure 3.5.

3.3.3.3 Proposed Model

Convolution layer are the fundamental building blocks of any cnn architecture that use advanced nonlinear function to achieve the best results. To identify AD, the suggested approach leverages a DL model to automatically collect knowledge from whole brain MRI scans. The suggested pipeline is presented in Figure 3.6 and consists of three basic steps: brain size scaling, 3D volume slicing, and CNN analysis. We envisioned a simple yet effective cnn model that is influenced by

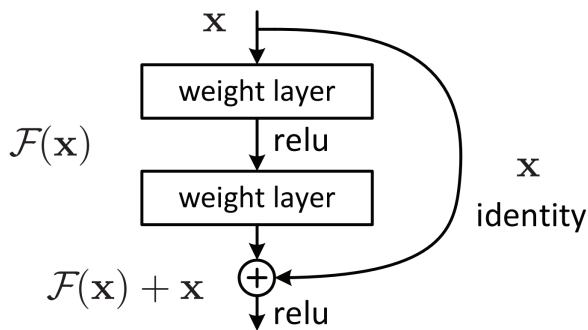


Figure 3.5: The fundamental representation of residual network [84].

the architecture design of ResNet and ConvMixer [85]. Where the proposed model performs conventional convolution, depth-wise convolution, and point-wise convolution concurrently, preceded by a skip conv layer to learn multi-level attributes from MRI data.

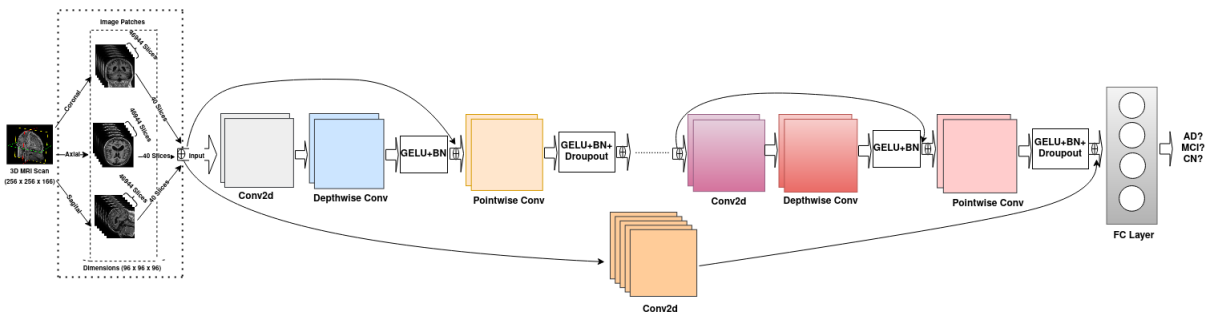


Figure 3.6: The layout of the proposed model.

A. Standard Convolution: The standard folding procedure is depicted in Figure 3.7. A conventional convolutional layer receives a $D_F \times D_F \times M$ feature

map I as input and provides a $D_G \times D_G \times N$ feature output O , where D_F signifies the spatially width and height of the square input matrix, M defines the number of input feature map channels, and N defines the number of output feature channels. The feature is taken from the typical convolutional layer's $k \times D_k$ sized convolution kernel. D_k is the convolution kernel's spatial breadth and height. The conventional convolution calculation procedure from feature map I to feature map O is described by:

$$G_{k,l,n} = \sum_{i,j,m} K_{i,j,m,n} \cdot I_{k+i-1,l+j-1,m}, \quad (3.3)$$

In Equation (3.3), I stands for input feature maps, G stands for output feature maps, and k stands for convolution kernels. i and j determine the position of the convolutional kernel elements. The positions of the element in the input and output feature maps are determined by k and l , respectively, whereby m is the channel of the input feature map and n is the channel of the output feature map.

The following are the standard convolution parameters:

$$F = M \times N \times D_k^2. \quad (3.4)$$

The cost of computing standard convolution is shown by:

$$G = M \times N \times D_F^2 \times D_k^2, \quad (3.5)$$

Where in equation (3.4) and (3.5) F represents number of model parameters, G represents the computation complexity, M represents the input feature channels, N represents the output feature channels, D_F represents the spatial width and height squared of the object map's input features, and D_k represents the convolution kernel's spatial width and height.

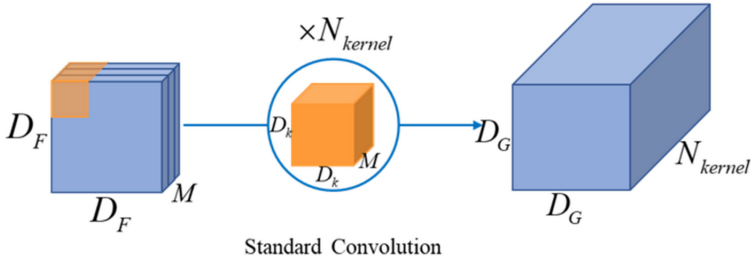


Figure 3.7: Demonstration of the standard convolution architecture [86].

B. Depth-wise and Point-wise Convolution Operation: Convolution in depth is a sort of convolution that processes each channels of the input image independently. It is used to retrieve spatial information in each dimension. A common convolution approach for the output feature map is point-by-point convolution. The Figure 3.8 depicts a depthwise and point-wise structure, where $D_f \times D_f \times M$ represent the size of the input picture, D_f represents the height and width of the input image, M represents the number of channels in the map, and $D_g \times D_g \times M$ reflects the output feature map's size created by convolution. The height and width of the output picture are represented by D_g , which is equivalent to the number of channels in the input data. It is utilized as the following convolution's input. For point-by-point convolution, the dimension of the convolution kernels is 1×1 , and the number of channels on each convolution filter must be equal to the number of channels in the input feature map. When the convolution kernels numbers are N , the resulting feature map is $D_g \times D_g \times N$.

The depthwise convolution output feature map is expressed as:

$$\bar{G}_{k,l,m} = \sum_{i,j} K_{i,j,m} \cdot I_{k+i-1,l+j-1,m}, \quad (3.6)$$

In equation (3.6), I indicates the input feature patterns, \bar{G} signifies the output image features, and K represents the kernels I and j define the convolution

kernel's element location. The variables k and l define the location of the given input feature and the extracted feature map, respectively, while m denotes the input feature map channels.

The cost function (G_2) and the convolution parameters (F_2) in depth are calculated as follows:

$$F_2 = M \times D_k^2 \quad (3.7)$$

and

$$G_2 = M \times D_i^2 \times D_k^2 \quad (3.8)$$

The parameter is simply proportionate to the amount of feature mapping channel and convolution kernels provided. The computation complexity depends on number of given input mapping providers, convolutional kernel, and quadratic input feature mapping function. The output feature mapping N does not need to be considered when determining the convolution depth parameter and computing cost. When compared to relations (3.4) and (3.5), equations (3.7) and (3.8) clearly demonstrate the simplicity of depthwise convolution. However, unlike a standard convolutional layer, depth-wise convolution just filters the input channels rather than combining them to produce new features. As a result, an extra layer of 1×1 convolution is required to create new features [87]. The new feature pattern is created by combining the depth-wise and 1×1 pointwise convolutions. As a result, the parameters (F_3) and cost function (G_3) may be determined as illustrated in equations (3.9) and (3.10):

$$F_3 = M \times D_k^2 + M \times N \quad (3.9)$$

and

$$G_3 = M \times D_i^2 \times D_k^2 + M \times N \times D_i^2 \quad (3.10)$$

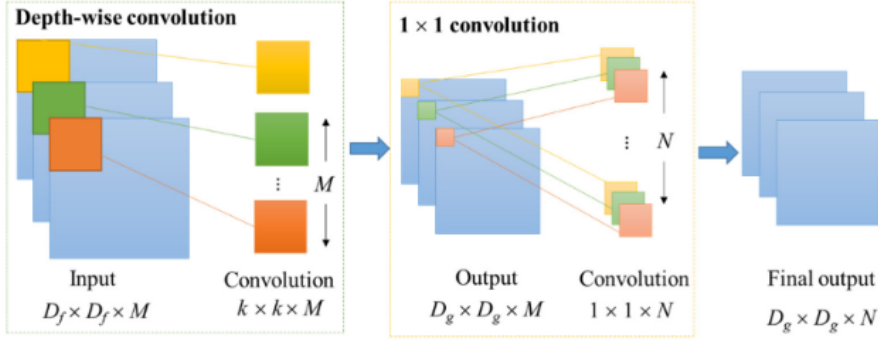


Figure 3.8: Demonstration of the depthwise convolution [86] architecture .

C. Proposed Model Implementation Details: This study analyzed 3D structural image data of 489 patients to generate a 2D model (163 AD, 163 MCI, and 163 CN). The original scans were rescaled to a level of $96 \times 96 \times 96$. 40 brain fields were produced from each area of the brain (axial, coronal, and sagittal) for training and validation, in total obtaining 58680 characteristic areas, of which 19560 corresponded to the AD group, 19560 to the CN, and 19560 to the MCI group.

The network, as depicted in Figure 3.6, was composed of three distinct types of layers. The input layer took N gray level picture ($X_n, n \in [1, N]$) patches, sized them to 96×96 , standardized them to the range $[0, 1]$, and sent them into the system. The second kind was a convolution layer, which was utilized

in this study as the proposed body design. The presented CNN model was built by combining four basic convolution layers and three convolution blocks, each one comprises of a skip connection-based depth-wise convolution coupled by point-wise (i.e., kernel size 1×1) convolution. The GELU activation function, Batch Normalization, and dropout layer follow each convolution in the block. A residual convolution is also used in this model, which was influenced by the skip connection type. The convolution filters were set at 5×5 in size and 256 in number. This was preserved all across the model to guarantee that same weights are distributed across different groupings of pixels in a picture. The fully connected layer, the third type of layer, was composed of a set of input and output neurons that formed the learnt linear combination of all neuron from the preceding layer after flowing through a non-linearity. Inputs and outputs of the fully linked layer were no longer spatially ordered, but rather represented a 1D vector.

During the studies, each subject was arbitrarily split into three main sections: training (60%), validation (10%), and test (30%). The presented technique was programmed in Python 3.8.10 and evaluated in the Ubuntu 20.04-x64 framework on a machine equipped with an NVIDIA RTX3090 GPU. Initially, the network weights were randomly set, and the Adaptive Moment Estimation (Adam) optimizer was employed with an initial rate of 0.001 and a decay rate of 0.9. The batch size was adjusted to 32 to avoid overfitting, and a dropout layer was included to avoid overfitting. We applied the proposed method in multiclass classification to distinguish among the patients with AD, MCI, and CN categories.

4 Results

4.1 Performance Metrics

The anticipated results in the diagnostic tasks are labeled as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), as shown in Table. 4.1. A positive sample that was correctly expected to be positive is known as TP. A TN sample is the one that was expected to be negative with certainty. The sign FP denotes a negative sample that was mistakenly identified as a positive sample. The symbol FN denotes a positive sample that was misidentified as a negative sample.

Table 4.1: Confusion Matrix

True Class	Predicted Class	
	G2	G3
G1	TP	FN
G1	FP	TN

Performance metrics such as accuracy, specificity, sensitivity, and the Youden Index employed to assess the efficacy of machine learning methods. On the other hand, to analyze deep learning-based diagnostic model following commonly used indicators were used: accuracy, sensitivity, specificity, precision, and receiver operating characteristic curve (ROC curve), F1 score,.

Accuracy (4.1) denotes the proportion of correctly identified samples among test samples.

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

As shown in the equation (4.2), specificity reflects the portion of samples correctly diagnosed, and sensitivity in equation (4.3) represents a model's ability to recognize AD patients in all positive samples.

$$Specificity = \frac{TN}{TN + FP} \quad (4.2)$$

$$Sensitivity(Recall) = \frac{TP}{TP + FN} \quad (4.3)$$

Equation (4.4), shows Youden Index (YI), indicates the effectiveness of biomarker.

$$YI = SEN + SPEC - 1 \quad (4.4)$$

Furthermore, in the equation (4.5), precision is defined as the proportion of accurately predicted positive occurrences out of all anticipated positive observations.

$$Precision = \frac{TP}{TP + FP} \quad (4.5)$$

The F1 score is determined by averaging Precision and Recall. As illustrated in the equation (4.6):

$$F1 - score = 2 \left(\frac{Precision \times Sensitivity}{Precision + Sensitivity} \right) \quad (4.6)$$

Finally, a ROC curve (receiver operating characteristic curve) is a plot that illustrates a classification model's performance of over all classification criteria. This graph depicts two parameters:

The first parameter is the True Positive Rate (TPR), sometimes known as a recall (4.7).

$$TPR = \frac{TP}{TP + FN} \quad (4.7)$$

The second parameter is the False Positive Rate (FPR), which is specified in the equation (4.8).

$$FPR = \frac{FP}{FP + TN} \quad (4.8)$$

4.2 Machine Learning Results

In the scenario of machine learning-based sMRI image processing, classification performance was evaluated using subcortical and cortical characteristics, and the findings are shown in Table 4.3 and Table 4.4. Moreover, classification reports are shown in Figure 4.1, 4.2 and 4.3 which helps to compare the performance visually.

Table 4.2: AD vs eMCI classification results

Algorithms	ACC	SEN	SPEC	YI
Softmax Classifier	94.47	97.39	91.79	89.18
SVM	97.87	100	95.65	95.65
Naive Bayes	93.62	92.0	95.45	87.45
Random Forest	91.91	100	86.08	86.08
KNN	91.49	88.46	95.24	83.7

Table 4.3: AD vs IMCI classification results

Algorithms	ACC	SEN	SPEC	YI
Softmax Classifier	91.66	89.07	93.83	82.9
SVM	95.83	100.0	93.33	93.33
Naive Bayes	85.42	94.44	80.0	74.44
Random Forest	92.49	98.22	87.62	85.87
KNN	89.58	86.36	92.31	78.67

Table 4.4: AD vs HC classification results

Algorithms	ACC	SEN	SPEC	YI
Softmax Classifier	92.20	100	87.86	87.86
SVM	97.83	95.0	100.0	95.0
Naive Bayes	92.17	96.46	89.79	86.25
Random Forest	93.33	94.74	92.31	87.05
KNN	89.13	95.24	84.0	79.0

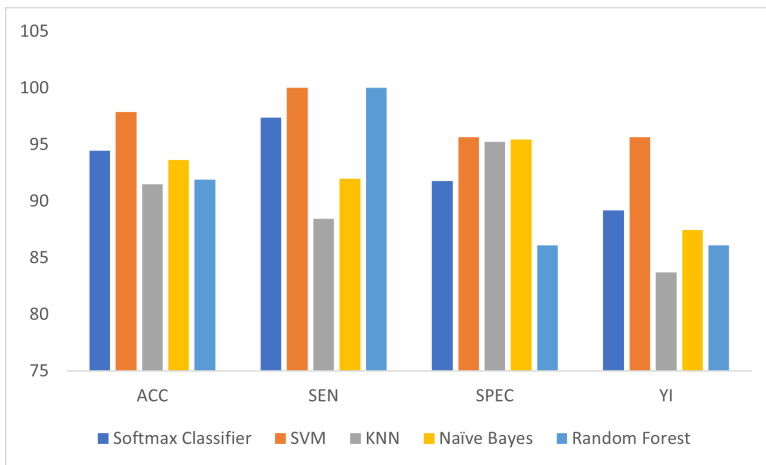


Figure 4.1: Performance comparison for AD vs eMCI between classifiers.

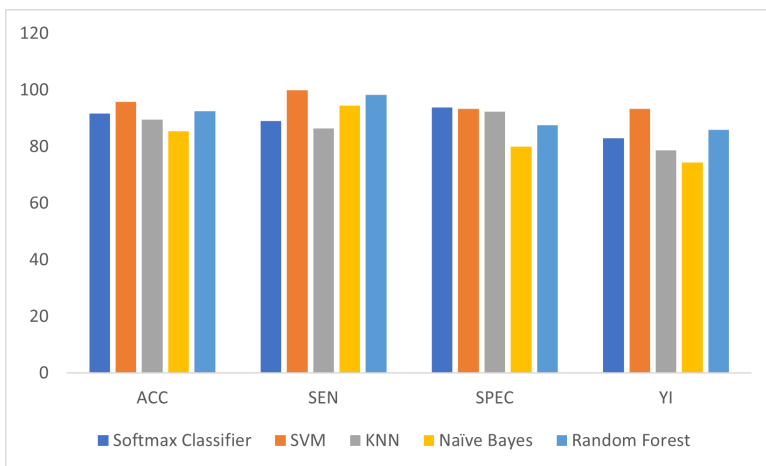


Figure 4.2: Performance comparison for AD vs lMCI between classifiers.

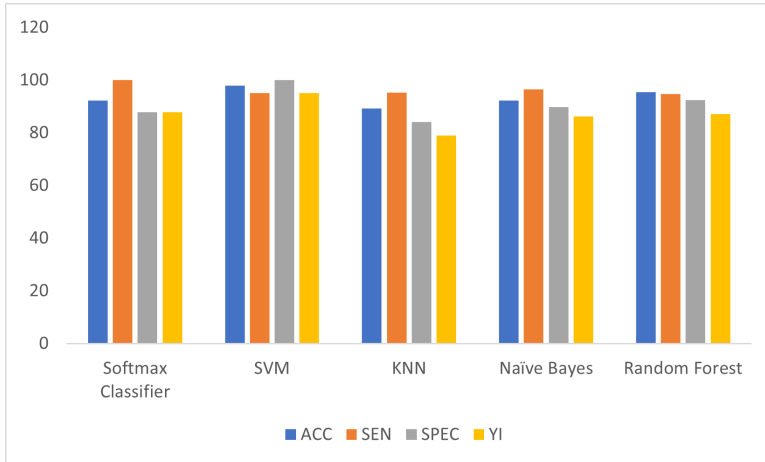


Figure 4.3: Performance comparison for AD vs HC between classifiers.

4.2.1 Results Summary

Table 4.2 illustrates the categorization findings for AD vs eMCI. Generally, all of the strategies outlined performed better. However, SVM performed substantially superior in terms of AD diagnosis, with an accuracy of 97.87% in this binary classification. In terms of the distinction between AD and IMCI. Table 4.3 displays the performance evaluation for AD vs IMCI, which demonstrates that SVM efficiency is 95.83%, which is higher than other classification algorithms. Similarly, in the AD vs HC classification, as shown in Table 4.4, SVM surpasses the other machine learning models with an effectiveness of 97.83%.

4.3 Deep Learning Results

The ADNI data is utilized to evaluate the presented method's performance. With a maximum accuracy of 96.41%, the suggested technique performed well in a multi-class classification challenge (CN vs MCI vs AD). The table 4.5 compares the proposed work to other model that can be trained on the same information. In

terms of accuracy, the comparison demonstrates that, even though having a fewer parameters, the suggested model beats the benchmark models.

Table 4.5: Comparison of multiclass classification model performances

Methods	# Parameters (Millions)	#FLOPs (Billions)	ACC	SEN	SPE	Precision	F1-score
ResNet50	23.587	0.65	95.27	95.31	97.63	95.39	95.32
VGG19	38.940	3.59	92.77	93.15	96.49	92.77	92.75
Proposed Model	3.514	1.42	96.12	94.99	97.73	95.50	95.23

FLOPs: Floating points operations; Maximum value of each column is denoted by bold figures;

ACC: Accuracy; SEN: Sensitivity; SPEC: Specificity.

ResNet and VGG Net were used as comparative models. ResNet obtained an accuracy of about 95.34% using commonly applied performance measures such as sensitivity (SEN), specificity (SPE), precision, F1 score, and accuracy (ACC), which is 2.08% higher than the VGG (93.26%) due to feature propagation improvements and skip connection, but lower than the suggested technique. The proposed model has the greatest accuracy of 96.41%, that is 1.07% better than ResNet and 3.15% higher than the VGG network.

Other performance measures showed that this work surpassed the baseline models. It achieved 97.73% specificity and 95.50% accuracy, which are both greater than baseline models. ResNet outperformed, with sensitivity and F1 scores of 95.31 percent and 95.32%, respectively. The relevant ROC for the

ADNI dataset is shown in 4.4, and so is the related confusion matrix in 4.5. The efficiency and degradation curves for the training and validation databases are shown in 4.6 and 4.7, respectively. In this work, I proved the efficiency of our CNN model for binary and multiclass classifications. Binary classification training is done in three scenarios between AD, MCI and CN.

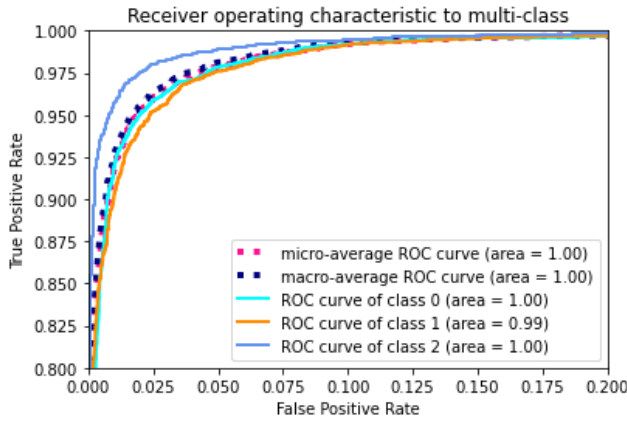


Figure 4.4: Class 2 (AD), Class 1 (MCI), and Class 0 (CN) Receiver operating characteristics (ROC) curve

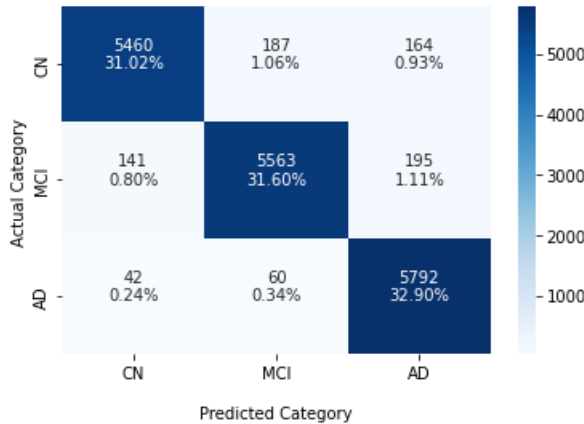


Figure 4.5: Testing Data Confusion matrix

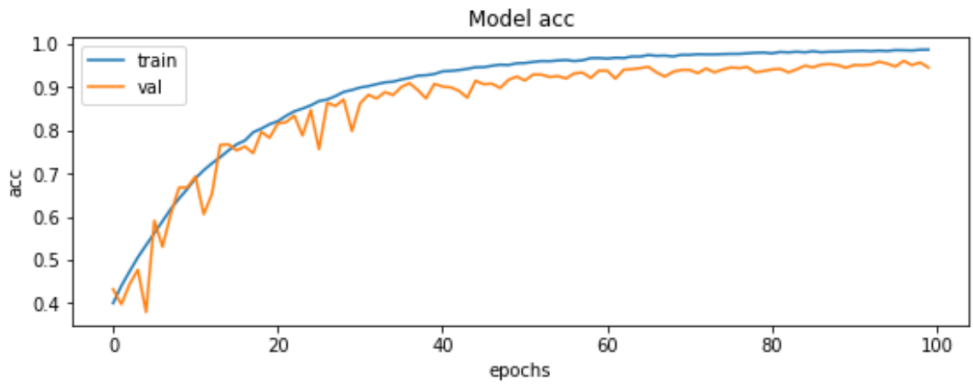


Figure 4.6: Training and validation accuracy curve

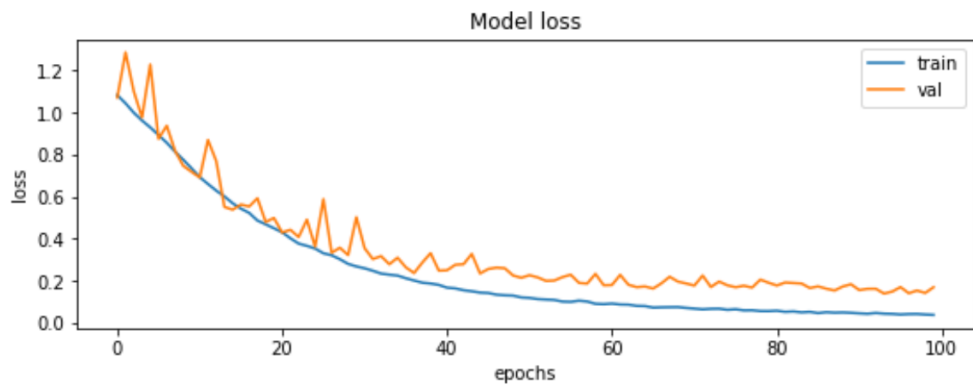


Figure 4.7: Loss curve for training and validation data

5 Discussion

There are remedies for Alzheimer’s disease symptoms, and various research are being undertaken to help in the disease’s rapid recognition. Several classification algorithms depending on T1-weighted images have been presented in this respect to distinguish among AD, MCI, and Healthy Control (HC) subjects.

Patients with AD were most typically classified using structural and functional parameters. Liu et al. [88], for example, provided ROI-based approaches for extracting brain characteristics, which were subsequently categorised using ML approaches such as the multiple kernel boosting (MKBoost) approach. It obtained 94.65% accuracy for AD versus CN, 89.63% accuracy for AD vs MCI, and 85.79% accuracy for MCI vs CN classification using a single structural MRI modality data. Sun et al. [89] achieved similar findings. In their paper, they introduced a novel SVM-based learning approach for extracting spatial-anatomical information, as well as a group lasso penalty to impose structural sparsity. Their proposed approach had an accuracy of 95.1% for AD versus CN classification, 70.8% for MCI vs CN classification, and 65.7% for AD vs MCI classification. S. Kadoury [90] also suggested a technique for semantically annotated PET image feature group classification. Their strategy produced 91.2% efficiency for AD versus HC categorization. Khajehnejad et al. [91] use a manifold-based semi-supervised learning technique to diagnose AD. Their approach was 93.86% accurate.

In contrast to previous work, the goal of this study is to increase the accuracy of a model created by integrating the feature selection approach. This work explore the effectiveness of multiple ML classifiers (SVM, Random Forest,

Nave Bayes, K-nearest neighbor, and Softmax classification method) for the identification of AD utilizing an efficient combined feature selection technique in AD, eMCI, IMCI, and HC volumes.

However, in order to employ such ML techniques for disease identification, pre-processing procedures, which are often time consuming and computationally costly, must be applied. As a result, scientists have concentrated their efforts on building computer-based systems which are designed to detect Alzheimer at an early stage. CNN-based image recognition is now widely employed in medical diseases diagnosis, according to [21]. But, creating an efficient CNN model capable of providing decent outcomes is neither practical nor realistic. As a result, in this study, we suggested a method for categorizing MRI images employing CNN features that has higher accuracy and fewer parameters.

To enhance classification performance, previous models just increased the depth and complexity of the network. Which frequently experienced vanishing gradient problem. In this context, this study aimed to ameliorate the vanishing gradient problem, encourage feature reuse, and considerably reduce the number of parameters by presenting a modified convolution network. A CNN network structure includes a regular convolutional layer as well as a depthwise-pointwise convolutional layer that was integrated to extract additional characteristics that are sensitive to brain activity or anatomical changes in a broad region. The proposed design addresses the aforementioned concerns by employing the notion of feature mixing, as defined in [92], [93]. Study primarily use depth-wise convolution to mix spatial information, point-wise convolution to blend channel areas, and the skip layer to mix global data attributes while lowering model complexity.

Furthermore, the suggested model was compared to existing deep learning

(DL) approaches. Hosseini et al. [6] suggested a 3D convolutional auto-encoder-based approach in their paper. They employed a pre-trained algorithm to detect anatomical form modifications in sMRI images. They next tested their algorithm on a CAD Dementia MRI dataset without first removing the skull and obtained 89.1% accuracy in multi-class classification. Also, their model obtained 97.6% for AD versus NC, 95% for AD vs MCI, and 90.8% for MCI vs NC on binary class tasks. Basaia et al. [68] introduce a DL framework based on the structural cross-sectional MRI scans for predicting AD diagnosis. Their suggested model was 99.2% correct for AD against CN, 87.1% correct for MCI vs CN, and 75.4% correct for AD vs MCI. Using the OASIS data set, Liu.J et al. [94] built a CNN-based framework that accomplished 78.02% validity for multiclass classification, 84.65% accuracy for MCI versus CN, 72.96% accuracy for AD versus MCI, and 75.2% accuracy for MCI versus CN classification. Later, in the same study, using ADNI data they enhanced their work to lower the number of parameters by utilizing a deep separable convolution model and reached 77.79% accuracy by decreasing the CNN model parameters by 87.94%. Liu.M et al. [94] created an architecture integrating 3D Densely connected deep cnn (3D Dense Net) and multi-task CNN to learn attributes from the segmented hippocampus. They got 88.9% accuracy for AD vs NC and 76.2% efficiency for MCI vs NC. Furthermore, in their publication, Xu et al. [95] suggested a customized Tresnet design for extracting information from gray matter MRI images. They obtained 86.9% accuracy in discriminating AD versus CN and 63.2% efficiency in classifying several classes. Finally, Aderghal et al. [96] developed a transfer learning approach based on DL to automatically identify brain images, depending on minimal ROI (few slices of the hippocampal region). Their analysis reveals that AD against CN is 91.86%, AD vs MCI is 69.95%, and MCI vs CN is 68.52%.

Table 5.1 examines numerous DL techniques in terms of parameters, while Table 5.2 highlights the comparison of various research.

Table 5.1: Number of Parameters compared between proposed and published methods.

References	Architecture	Parameters	Reduction (%)
Hosseini et al., 2016 [6]	3D Convolutional Auto Encoder	76.22	95.39%
Liu. J et al., 2021 [73]	Multi Layer Neural Network	29.80	88.22%
Liu. J et al., 2021 [73]	Depth separable convolution	11.15	68.52%
Aderghal et al., 2020 [96]	Le-Net	62.50	94.38%
Xu et al., 2021 [95]	Tresnet_L + SK_module	52.0	93.25%
Base Line Model	VGG19	38.94	90.98%
Base Line Model	Resnet50	23.58	85.11%
This Study	Proposed Method	3.51	-

*Number of Parameters are measured in millions

Moreover, from a methodological perspective, hardware throughput must be recognized as a vital role. To do this, floating point operations (FLOPs) are used as a sign of power use, also the number of parameters [97]. Generally, a network having fewer parameters and FLOPs demanded less memory to preserve the architecture, needed less hardware memory, and therefore it is more favorable to the embedded system. Figure 5.1 displays an assessment of the accuracy and computing costs between the benchmark and recommended models. The proposed approach has the benefit of having a lower number of FLOPs opposed to the VGG19 model, and also a lower number of parameters by

Table 5.2: Classification results of experiment and published methods.

References	Subjects	No. of Samples	AD/MCI/CN			AD vs CN			MCI vs CN			AD vs MCI		
			ACC (%)	SEN (%)	SPEC (%)	ACC (%)	SEN (%)	SPEC (%)	ACC (%)	SEN (%)	SPEC (%)	ACC (%)	SEN (%)	SPEC (%)
Liu et al., 2018 [88]	MRI	AD-200, MCI-280, CN-230	-	-	-	94.65	95.03	91.76	84.79	88.91	80.34	88.63	91.55	86.25
Sun et al., 2018 [89]	MRI	AD-137, sMCI-134, pMCI-76, CN-162	-	-	-	95.10	93.80	83.80	70.80	72.10	69.10	65.70	63.20	67.30
Hosseini et al., 2016 [6]	MRI	AD-70, MCI-70, CN-70	89.10	-	-	97.60	-	-	90.80	-	-	95.00	-	-
Basaia et al., 2019 [68]	MRI	AD-294, MCI-253, CN-352	-	-	-	99.20	98.90	99.50	87.10	87.80	86.50	75.40	74.50	76.40
Aderghal et al., 2020 [96]	MRI	AD-294, MCI-672, CN-627	-	-	-	92.30	93.95	90.65	78.48	77.72	81.44	79.16	82.72	78.36
Liu, M et al., 2020 [94]	MRI	AD-97, MCI-233, CN-119	88.90	86.60	90.80	-	-	-	76.20	79.30	69.80	-	-	-
Liu, J et al., 2021 [73]	MRI (OASIS)	AD-90, MCI-136, CN-266	78.02	83.21	75.32	82.50	-	-	84.65	82.35	79.50	72.96	78.34	82.15
Liu, J et al., 2021 [73]	MRI (ADNI)	AD-99, MCI-353	-	-	-	-	-	-	-	-	-	75.32	80.13	65.32
Xu et al., 2021 [95]	MRI	AD-85, MCI-244, CN-133	63.20	84.00	85.40	86.90	82.10	88.30	-	-	-	-	-	-
Proposed Method (Machine Learning)	MRI	AD-78, LMCI-79, EMCI-78, HC-78	96.12	94.99	97.73	97.00	94.78	98.40	96.00	96.29	96.22	90.00	88.00	91.71
Proposed Method (Deep Learning)	MRI	AD-163, MCI-163, CN-163	96.12	94.99	97.73	97.00	94.78	98.40	96.00	96.29	96.22	90.00	88.00	91.71

Abbreviations: AD = Alzheimer's disease; MCI = Mild Cognitive Impairment; CN = Cognitive Normal; Maximum value of each column is denoted by bold figures;

90.98%. Despite having the same amount of FLOPs as ResNet50, the proposed scheme has 85.11% less parameters. Experiments demonstrate that the proposed methodology produces better outcomes with fewer number of parameters and a lower computing cost.

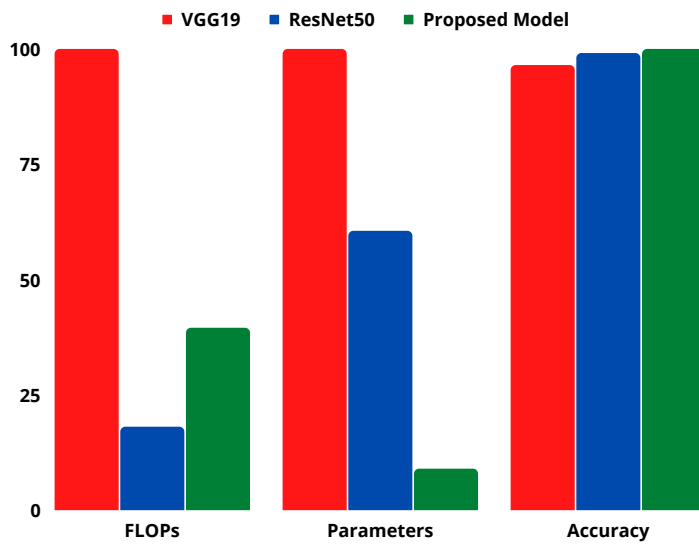


Figure 5.1: Model comparison for FLOPs, parameters and accuracy

6 Summary and Conclusion

This thesis comprises of machine learning (ML) frameworks to classify AD and other diagnostic groups using ADNI dataset. For ML case, this thesis proposed a combination of dimension reduction and feature selection methods. The objective is to integrate feature selection approaches to adequately identify AD from MCI (early or late AD), and a normal group acquired from the ADNI dataset. This study used a mixer of subcortical and cortical attributes retrieved using a computerized framework. To execute classification tasks, five classifiers were utilized (Softmax, KNN, SVM, Nave Bayes, and Random Forest). Following the recommended methodologies, the experimental results demonstrated satisfactory accuracy in three scenarios (AD vs HC, AD vs eMCI, and AD vs lMCI).

Moreover, in the instance of deep learning (DL), this work designed a CNN for 3D whole brain scans, and the optimal accuracy was found using an isotropically repeated convolution block network configuration. The presented technique outperforms current state-of-the-art technologies. Furthermore, the process works completely automated (no further knowledge or human involvement is required) and incredibly fast. The proposed methodology might be used to discover key trends in information, confirm previous expert conclusions, aid in diagnostic situations, and even reveal patterns for chronic conditions other than Alzheimer's.

Future study might explore towards attaining equivalent or better outcomes for pre-processed data for skull alignment and removal. Finally, it would be intriguing to include patient information to supplement the data provided by MRIs, influence decision-making, and link it to the patient's medical history.

REFERENCES

- [1] S. Gauthier, P. Rosa-Neto, J. Morais, and C. Webster, “World alzheimer report 2021: Journey through the diagnosis of dementia,” *Alzheimer’s Disease International: London, UK*, 2021.
- [2] J. Zhang, B. Zheng, A. Gao, X. Feng, D. Liang, and X. Long, “A 3d densely connected convolution neural network with connection-wise attention mechanism for alzheimer’s disease classification,” *Magnetic Resonance Imaging*, vol. 78, pp. 119–126, 2021.
- [3] P. Lewczuk, M. Łukaszewicz-Zajac, P. Mroczko, and J. Kornhuber, “Clinical significance of fluid biomarkers in alzheimer’s disease,” *Pharmacological Reports*, vol. 72, no. 3, pp. 528–542, 2020.
- [4] M. A. DeTure and D. W. Dickson, “The neuropathological diagnosis of alzheimer’s disease,” *Molecular neurodegeneration*, vol. 14, no. 1, pp. 1–18, 2019.
- [5] S. Lahmiri and A. Shmuel, “Performance of machine learning methods applied to structural mri and adas cognitive scores in diagnosing alzheimer’s disease,” *Biomedical Signal Processing and Control*, vol. 52, pp. 414–419, 2019.
- [6] E. Hosseini-Asl, R. Keynton, and A. El-Baz, “Alzheimer’s disease diagnostics by adaptation of 3d convolutional network,” in *2016 IEEE international conference on image processing (ICIP)*, IEEE, 2016, pp. 126–130.

- [7] J. Islam and Y. Zhang, “Brain mri analysis for alzheimer’s disease diagnosis using an ensemble system of deep convolutional neural networks,” *Brain informatics*, vol. 5, no. 2, pp. 1–14, 2018.
- [8] M. Liu, D. Cheng, K. Wang, and Y. Wang, “Multi-modality cascaded convolutional neural networks for alzheimer’s disease diagnosis,” *Neuroinformatics*, vol. 16, no. 3, pp. 295–308, 2018.
- [9] D. Lu, K. Popuri, G. W. Ding, R. Balachandar, and M. F. Beg, “Multimodal and multiscale deep neural networks for the early diagnosis of alzheimer’s disease using structural mr and fdg-pet images,” *Scientific reports*, vol. 8, no. 1, pp. 1–13, 2018.
- [10] T. Zhou, K.-H. Thung, X. Zhu, and D. Shen, “Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis,” *Human brain mapping*, vol. 40, no. 3, pp. 1001–1016, 2019.
- [11] S. Sarraf, D. D. DeSouza, J. Anderson, G. Tofighi, *et al.*, “Deepad: Alzheimer’s disease classification via deep convolutional neural networks using mri and fmri,” *BioRxiv*, pp. 430–441, 2017.
- [12] K. Brueggen, M. Dyrba, A. Cardenas-Blanco, A. Schneider, K. Fliessbach, K. Buerger, D. Janowitz, O. Peters, F. Menne, J. Priller, *et al.*, “Structural integrity in subjective cognitive decline, mild cognitive impairment and alzheimer’s disease based on multicenter diffusion tensor imaging,” *Journal of neurology*, vol. 266, no. 10, pp. 2465–2474, 2019.
- [13] C. Lian, M. Liu, J. Zhang, and D. Shen, “Hierarchical fully convolutional network for joint atrophy localization and alzheimer’s disease diagnosis

- using structural mri,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 4, pp. 880–893, 2018.
- [14] C.-Y. Wee, P.-T. Yap, W. Li, K. Denny, J. N. Browndyke, G. G. Potter, K. A. Welsh-Bohmer, L. Wang, and D. Shen, “Enriched white matter connectivity networks for accurate identification of mci patients,” *Neuroimage*, vol. 54, no. 3, pp. 1812–1822, 2011.
- [15] J. Islam and Y. Zhang, “A novel deep learning based multi-class classification method for alzheimer’s disease detection using brain mri data,” in *International conference on brain informatics*, Springer, 2017, pp. 213–222.
- [16] F. Al-Turjman, M. H. Nawaz, and U. D. Ulusar, “Intelligence in the internet of medical things era: A systematic review of current and future trends,” *Computer Communications*, vol. 150, pp. 644–660, 2020.
- [17] J. A. Sidey-Gibbons and C. J. Sidey-Gibbons, “Machine learning in medicine: A practical introduction,” *BMC medical research methodology*, vol. 19, no. 1, pp. 1–18, 2019.
- [18] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [19] R. Jain, N. Jain, A. Aggarwal, and D. J. Hemanth, “Convolutional neural network based alzheimer’s disease classification from magnetic resonance brain images,” *Cognitive Systems Research*, vol. 57, pp. 147–159, 2019.
- [20] H. Taheri Gorji and N. Kaabouch, “A deep learning approach for diagnosis of mild cognitive impairment based on mri images,” *Brain sciences*, vol. 9, no. 9, pp. 217–226, 2019.

- [21] R. Mishra, S. Urolagin, J. Jothi, A. Neogi, and N. Nawaz, “Deep learning-based sentiment analysis and topic modeling on tourism during covid-19 pandemic,” *Frontiers in Computer Science*, vol. 3, no. 3389–3399, 2021.
- [22] H. Braak, E. Braak, J. Bohl, and H. Bratzke, “Evolution of alzheimer’s disease related cortical lesions,” *Alzheimer’s Disease—from Basic Research to Clinical Applications*, pp. 97–106, 1998.
- [23] C. Wattmo, E. Londos, and L. Minthon, “Risk factors that affect life expectancy in alzheimer’s disease: A 15-year follow-up,” *Dementia and geriatric cognitive disorders*, vol. 38, no. 5-6, pp. 286–299, 2014.
- [24] A. Wimo and M. J. Prince, *World Alzheimer Report 2010: the global economic impact of dementia*. Alzheimer’s Disease International, 2010.
- [25] N. Villain and B. Dubois, “Alzheimer’s disease including focal presentations,” in *Seminars in neurology*, Thieme Medical Publishers, vol. 39, 2019, pp. 213–226.
- [26] L. J. Bain, K. Jedrzejewski, M. Morrison-Bogorad, M. Albert, C. Cotman, H. Hendrie, and J. Q. Trojanowski, “Healthy brain aging: A meeting report from the sylvan m. cohen annual retreat of the university of pennsylvania institute on aging,” *Alzheimer’s & dementia: the journal of the Alzheimer’s Association*, vol. 4, no. 6, pp. 443–455, 2008.
- [27] C. R. Jack Jr, D. A. Bennett, K. Blennow, M. C. Carrillo, B. Dunn, S. B. Haeberlein, D. M. Holtzman, W. Jagust, F. Jessen, J. Karlawish, *et al.*, “Nia-aa research framework: Toward a biological definition of alzheimer’s disease,” *Alzheimer’s & Dementia*, vol. 14, no. 4, pp. 535–562, 2018.

- [28] C. Hutton, B. Draganski, J. Ashburner, and N. Weiskopf, “A comparison between voxel-based cortical thickness and voxel-based morphometry in normal aging,” *Neuroimage*, vol. 48, no. 2, pp. 371–380, 2009.
- [29] P. Tiraboschi, L. Hansen, L. Thal, and J. Corey-Bloom, “The importance of neuritic plaques and tangles to the development and evolution of ad,” *Neurology*, vol. 62, no. 11, pp. 1984–1989, 2004.
- [30] H. Li, C.-C. Liu, H. Zheng, and T. Y. Huang, “Amyloid, tau, pathogen infection and antimicrobial protection in alzheimer’s disease—conformist, nonconformist, and realistic prospects for ad pathogenesis,” *Translational neurodegeneration*, vol. 7, no. 1, pp. 1–16, 2018.
- [31] G. M. McKhann, D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack Jr, C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux, *et al.*, “The diagnosis of dementia due to alzheimer’s disease: Recommendations from the national institute on aging-alzheimer’s association workgroups on diagnostic guidelines for alzheimer’s disease,” *Alzheimer’s & dementia*, vol. 7, no. 3, pp. 263–269, 2011.
- [32] M. S. Albert, S. T. DeKosky, D. Dickson, B. Dubois, H. H. Feldman, N. C. Fox, A. Gamst, D. M. Holtzman, W. J. Jagust, R. C. Petersen, *et al.*, “The diagnosis of mild cognitive impairment due to alzheimer’s disease: Recommendations from the national institute on aging-alzheimer’s association workgroups on diagnostic guidelines for alzheimer’s disease,” *Focus*, vol. 11, no. 1, pp. 96–106, 2013.
- [33] S. L. Bartels, R. J. van Knippenberg, C. Malinowsky, F. R. Verhey, and M. E. de Vugt, “Smartphone-based experience sampling in people with

- mild cognitive impairment: Feasibility and usability study,” *JMIR aging*, vol. 3, no. 2, e19852, 2020.
- [34] G. B. Frisoni, N. C. Fox, C. R. Jack, P. Scheltens, and P. M. Thompson, “The clinical use of structural mri in alzheimer disease,” *Nature Reviews Neurology*, vol. 6, no. 2, pp. 67–77, 2010.
- [35] A. Nordberg, J. O. Rinne, A. Kadir, and B. Långström, “The use of pet in alzheimer disease,” *Nature Reviews Neurology*, vol. 6, no. 2, pp. 78–87, 2010.
- [36] S. Mankhong, S. Kim, S. Lee, H.-B. Kwak, D.-H. Park, K.-L. Joa, and J.-H. Kang, “Development of alzheimer’s disease biomarkers: From csf-to blood-based biomarkers,” *Biomedicines*, vol. 10, no. 4, pp. 850–862, 2022.
- [37] L. Berg, D. W. McKeel, J. P. Miller, J. Baty, and J. C. Morris, “Neuropathological indexes of alzheimer’s disease in demented and nondemented persons aged 80 years and older,” *Archives of neurology*, vol. 50, no. 4, pp. 349–358, 1993.
- [38] J. C. Morris, M. Storandt, J. P. Miller, D. W. McKeel, J. L. Price, E. H. Rubin, and L. Berg, “Mild cognitive impairment represents early-stage alzheimer disease,” *Archives of neurology*, vol. 58, no. 3, pp. 397–405, 2001.
- [39] E. Bagyinszky, Y. C. Youn, S. S. A. An, and S. Kim, “The genetics of alzheimer’s disease,” *Clinical interventions in aging*, vol. 9, pp. 79–100, 2014.
- [40] W. Hollingworth, C. J. Todd, M. I. Bell, Q. Arafat, S. Girling, K. R. Karia, and A. K. Dixon, “The diagnostic and therapeutic impact of mri:

- An observational multi-centre study,” *Clinical radiology*, vol. 55, no. 11, pp. 825–831, 2000.
- [41] M. I. Jordan and T. M. Mitchell, “Machine learning: Trends, perspectives, and prospects,” *Science*, vol. 349, no. 6245, pp. 255–260, 2015.
- [42] M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain, and A. J. Aljaaf, “A systematic review on supervised and unsupervised machine learning algorithms for data science,” *Supervised and unsupervised learning for data science*, pp. 3–21, 2020.
- [43] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, “A comparative study of fine-tuning deep learning models for plant disease identification,” *Computers and Electronics in Agriculture*, vol. 161, pp. 272–279, 2019.
- [44] C. Käding, E. Rodner, A. Freytag, and J. Denzler, “Fine-tuning deep neural networks in continuous learning scenarios,” in *Asian Conference on Computer Vision*, Springer, 2016, pp. 588–605.
- [45] Y.-H. Chen, T. Krishna, J. S. Emer, and V. Sze, “Eyeriss: An energy-efficient reconfigurable accelerator for deep convolutional neural networks,” *IEEE journal of solid-state circuits*, vol. 52, no. 1, pp. 127–138, 2016.
- [46] L.-C. Hsu, C.-T. Chiu, K.-T. Lin, H.-H. Chou, and Y.-Y. Pu, “Essa: An energy-aware bit-serial streaming deep convolutional neural network accelerator,” *Journal of Systems Architecture*, vol. 111, pp. 4609–4613, 2020.

- [47] Y.-J. Lin and T. S. Chang, “Data and hardware efficient design for convolutional neural network,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 5, pp. 1642–1651, 2017.
- [48] Prabhu, *Understanding of Convolutional Neural Network (CNN) — Deep Learning — RaghavPrabhu*, <https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148>, [Accessed 08-Jun-2022], 2019.
- [49] B. C. Education, *What are convolutional neural networks?* [Online]. Available: <https://www.ibm.com/cloud/learn/convolutional-neural-networks>.
- [50] J. P. Kim, J. Kim, Y. H. Park, S. B. Park, J. San Lee, S. Yoo, E.-J. Kim, H. J. Kim, D. L. Na, J. A. Brown, *et al.*, “Machine learning based hierarchical classification of frontotemporal dementia and alzheimer’s disease,” *NeuroImage: Clinical*, vol. 23, pp. 101 811–101 823, 2019.
- [51] X. Long, L. Chen, C. Jiang, L. Zhang, and A. D. N. Initiative, “Prediction and classification of alzheimer disease based on quantification of mri deformation,” *PloS one*, vol. 12, no. 3, e0173372, 2017.
- [52] H. Guo, F. Zhang, J. Chen, Y. Xu, and J. Xiang, “Machine learning classification combining multiple features of a hyper-network of fmri data in alzheimer’s disease,” *Frontiers in neuroscience*, vol. 11, pp. 615–627, 2017.
- [53] L. Khedher, J. Ramirez, J. M. Górriz, A. Brahim, and I. Illán, “Independent component analysis-based classification of alzheimer’s disease from segmented mri data,” in *International Work-Conference on the Interplay between Natural and Artificial Computation*, Springer, 2015, pp. 78–87.

- [54] T. Tong, R. Wolz, Q. Gao, R. Guerrero, J. V. Hajnal, D. Rueckert, A. D. N. Initiative, *et al.*, “Multiple instance learning for classification of dementia in brain mri,” *Medical image analysis*, vol. 18, no. 5, pp. 808–818, 2014.
- [55] D. Zhang, Y. Wang, L. Zhou, H. Yuan, D. Shen, A. D. N. Initiative, *et al.*, “Multimodal classification of alzheimer’s disease and mild cognitive impairment,” *Neuroimage*, vol. 55, no. 3, pp. 856–867, 2011.
- [56] C. Salvatore, A. Cerasa, P. Battista, M. C. Gilardi, A. Quattrone, and I. Castiglioni, “Magnetic resonance imaging biomarkers for the early diagnosis of alzheimer’s disease: A machine learning approach,” *Frontiers in neuroscience*, vol. 9, pp. 307–313, 2015.
- [57] J. Baron, G. Chetelat, B. Desgranges, G. Perchey, B. Landeau, V. De La Sayette, and F. Eustache, “In vivo mapping of gray matter loss with voxel-based morphometry in mild alzheimer’s disease,” *Neuroimage*, vol. 14, no. 2, pp. 298–309, 2001.
- [58] Y. Gupta, K. H. Lee, K. Y. Choi, J. J. Lee, B. C. Kim, G. R. Kwon, N. R. C. for Dementia, and A. D. N. Initiative, “Early diagnosis of alzheimer’s disease using combined features from voxel-based morphometry and cortical, subcortical, and hippocampus regions of mri t1 brain images,” *PLoS One*, vol. 14, no. 10, e0222446, 2019.
- [59] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” *nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [60] S. Vieira, W. H. Pinaya, and A. Mechelli, “Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological

- disorders: Methods and applications,” *Neuroscience & Biobehavioral Reviews*, vol. 74, pp. 58–75, 2017.
- [61] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, “A guide to deep learning in healthcare,” *Nature medicine*, vol. 25, no. 1, pp. 24–29, 2019.
- [62] I. R. Silva, G. S. Silva, R. G. de Souza, W. P. dos Santos, and A. d. A. Roberta, “Model based on deep feature extraction for diagnosis of alzheimer’s disease,” in *2019 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2019, pp. 1–7.
- [63] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, “Early diagnosis of alzheimer’s disease with deep learning,” in *2014 IEEE 11th international symposium on biomedical imaging (ISBI)*, IEEE, 2014, pp. 1015–1018.
- [64] S. Alinsaif, J. Lang, A. D. N. Initiative, *et al.*, “3d shearlet-based descriptors combined with deep features for the classification of alzheimer’s disease based on mri data,” *Computers in Biology and Medicine*, vol. 138, p. 104 879, 2021.
- [65] S.-H. Wang, P. Phillips, Y. Sui, B. Liu, M. Yang, and H. Cheng, “Classification of alzheimer’s disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling,” *Journal of medical systems*, vol. 42, no. 5, pp. 1–11, 2018.
- [66] S. Ahmed, K. Y. Choi, J. J. Lee, B. C. Kim, G.-R. Kwon, K. H. Lee, and H. Y. Jung, “Ensembles of patch-based classifiers for diagnosis of alzheimer diseases,” *IEEE Access*, vol. 7, pp. 73 373–73 383, 2019.

- [67] C. Feng, A. Elazab, P. Yang, T. Wang, F. Zhou, H. Hu, X. Xiao, and B. Lei, “Deep learning framework for alzheimer’s disease diagnosis via 3d-cnn and fsbi-lstm,” *IEEE Access*, vol. 7, pp. 63 605–63 618, 2019.
- [68] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, R. Santangelo, M. Filippi, A. D. N. Initiative, *et al.*, “Automated classification of alzheimer’s disease and mild cognitive impairment using a single mri and deep neural networks,” *NeuroImage: Clinical*, vol. 21, pp. 101 645–101 655, 2019.
- [69] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, “Residual and plain convolutional neural networks for 3d brain mri classification,” in *2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)*, IEEE, 2017, pp. 835–838.
- [70] G. Folego, M. Weiler, R. F. Casseb, R. Pires, and A. Rocha, “Alzheimer’s disease detection through whole-brain 3d-cnn mri,” *Frontiers in bioengineering and biotechnology*, vol. 8, 2020.
- [71] A. Abrol, M. Bhattarai, A. Fedorov, Y. Du, S. Plis, V. Calhoun, A. D. N. Initiative, *et al.*, “Deep residual learning for neuroimaging: An application to predict progression to alzheimer’s disease,” *Journal of neuroscience methods*, vol. 339, pp. 108 701–108 713, 2020.
- [72] A. Basher, B. C. Kim, K. H. Lee, and H. Y. Jung, “Volumetric feature-based alzheimer’s disease diagnosis from smri data using a convolutional neural network and a deep neural network,” *IEEE Access*, vol. 9, pp. 29 870–29 882, 2021.
- [73] J. Liu, M. Li, Y. Luo, S. Yang, W. Li, and Y. Bi, “Alzheimer’s disease detection using depthwise separable convolutional neural networks,”

- Computer Methods and Programs in Biomedicine*, vol. 203, pp. 106 032–106 043, 2021.
- [74] L. M. Besser, W. A. Kukull, M. A. Teylan, E. H. Bigio, N. J. Cairns, J. K. Kofler, T. J. Montine, J. A. Schneider, and P. T. Nelson, “The revised national alzheimer’s coordinating center’s neuropathology form—available data and new analyses,” *Journal of Neuropathology & Experimental Neurology*, vol. 77, no. 8, pp. 717–726, 2018.
- [75] B. Fischl, M. I. Sereno, R. B. Tootell, and A. M. Dale, “High-resolution intersubject averaging and a coordinate system for the cortical surface,” *Human Brain Mapping*, vol. 8, no. 4, pp. 272–284, 1999.
- [76] J. S. Lee, C. Kim, J.-H. Shin, H. Cho, D.-s. Shin, N. Kim, H. J. Kim, Y. Kim, S. N. Lockhart, D. L. Na, *et al.*, “Machine learning-based individual assessment of cortical atrophy pattern in alzheimer’s disease spectrum: Development of the classifier and longitudinal evaluation,” *Scientific reports*, vol. 8, no. 1, pp. 1–10, 2018.
- [77] Y. Gupta, K. H. Lee, K. Y. Choi, J. J. Lee, B. C. Kim, and G.-R. Kwon, “Alzheimer’s disease diagnosis based on cortical and subcortical features,” *Journal of healthcare engineering*, vol. 2019, 2019.
- [78] M. y kamil, “A deep learning framework to detect covid-19 disease via chest x-ray and ct scan images,” *International Journal of Electrical and Computer Engineering*, vol. 11, pp. 844–850, Feb. 2021. DOI: 10.11591/ijece.v11i1.pp844-850.
- [79] I. T. Jolliffe and J. Cadima, “Principal component analysis: A review and recent developments,” *Philosophical Transactions of the Royal Society A:*

- Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20 150 202, 2016.
- [80] H. Wang, Y. Shen, S. Wang, T. Xiao, L. Deng, X. Wang, and X. Zhao, “Ensemble of 3d densely connected convolutional network for diagnosis of mild cognitive impairment and alzheimer’s disease,” *Neurocomputing*, vol. 333, pp. 145–156, 2019.
- [81] X. Liu, Q. Xu, and N. Wang, “A survey on deep neural network-based image captioning,” *The Visual Computer*, vol. 35, no. 3, pp. 445–470, 2019.
- [82] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [83] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [84] R. Atmudi, A. Rao, and P. Mohapatra, “Covid-19 face mask prediction using machine learning techniques,” Nov. 2020.
- [85] A. Trockman and J. Z. Kolter, “Patches are all you need?” *arXiv preprint arXiv:2201.09792*, 2022.
- [86] C.-F. Wang, *A basic introduction to separable convolutions*, <https://towardsdatascience.com/a-basic-introduction-to-separable-convolutions-b99ec3102728?source>, Last accessed on April 12, 2022, 2018.

- [87] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” *arXiv preprint arXiv:1704.04861*, 2017.
- [88] J. Liu, M. Li, W. Lan, F.-X. Wu, Y. Pan, and J. Wang, “Classification of alzheimer’s disease using whole brain hierarchical network,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 15, no. 2, pp. 624–632, 2018. DOI: 10.1109/TCBB.2016.2635144.
- [89] Z. Sun, Y. Qiao, B. P. Lelieveldt, M. Staring, A. D. N. Initiative, *et al.*, “Integrating spatial-anatomical regularization and structure sparsity into svm: Improving interpretation of alzheimer’s disease classification,” *NeuroImage*, vol. 178, pp. 445–460, 2018.
- [90] S. H. Nozadi, S. Kadoury, A. D. N. Initiative, *et al.*, “Classification of alzheimer’s and mci patients from semantically parcelled pet images: A comparison between av45 and fdg-pet,” *International journal of biomedical imaging*, vol. 2018, 2018.
- [91] M. Khajehnejad, F. H. Saatlou, and H. Mohammadzade, “Alzheimer’s disease early diagnosis using manifold-based semi-supervised learning,” *Brain sciences*, vol. 7, no. 8, pp. 109–123, 2017.
- [92] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, “Large kernel matters—improve semantic segmentation by global convolutional network,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4353–4361.

- [93] I. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, D. Keysers, J. Uszkoreit, M. Lucic, *et al.*, “Mlp-mixer: An all-mlp architecture for vision,” *arXiv preprint arXiv:2105.01601*, 2021.
- [94] M. Liu, F. Li, H. Yan, K. Wang, Y. Ma, L. Shen, M. Xu, A. D. N. Initiative, *et al.*, “A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in alzheimer’s disease,” *Neuroimage*, vol. 208, pp. 116 459–16 662, 2020.
- [95] Z. Xu, H. Deng, J. Liu, and Y. Yang, “Diagnosis of alzheimer’s disease based on the modified tresnet,” *Electronics*, vol. 10, no. 16, pp. 1908–1921, 2021.
- [96] K. Aderghal, K. Afdel, J. Benois-Pineau, G. Catheline, A. D. N. Initiative, *et al.*, “Improving alzheimer’s stage categorization with convolutional neural network using transfer learning and different magnetic resonance imaging modalities,” *Heliyon*, vol. 6, no. 12, e05652, 2020.
- [97] R. Tang, W. Wang, Z. Tu, and J. Lin, “An experimental analysis of the power consumption of convolutional neural networks for keyword spotting,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2018, pp. 5479–5483.

ACKNOWLEDGEMENTS

All praise is due to the Lord of the Worlds alone.

I would like to sincerely acknowledge my advisor Professor Kwon, Goo-Rak for all his technical and personal support during my tenure at Chosun University. He has been tremendously supporting me and patiently guiding me throughout these two years. I am grateful to him for trusting me to conduct this research when I did not have enough background of medical image processing field. I appreciate him for always being humble to answer all my trivial queries and sharing insightful ideas to lay the foundation of my dissertation.

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

Moreover, some of my friends including Muhammad Usman, M. Sohail, M.Kashif Mobeen and Faheem have helped me in many ways during my studies. Lastly, I would acknowledge to the Chosun University, and National Institute for International Education (NIIED) for the financial assistance. And also, I would like to thank ADNI and its collaborators for their great efforts, large amounts of research work and willingness to share their dataset, without which this thesis and the original work described herein would not be possible.