American International University – Bangladesh

Thesis

Early Detection of Alzheimer's Disease Using Deep Learning

Submitted By

18-37258-1	Mohammad Fahim Shahriar
18-37266-1	Bazlul Bari Mozharul Haq
18-37503-1	Tamzid Ahmed
18-37535-1	Md. Atiqur Rahman

Department of Computer Science

Faculty of Science & Information Technology

American International University, Bangladesh

11 October, 2021

Declaration

We certify that this thesis is our original work, and that we have not submitted it for another degree or diploma at any university or other tertiary education institution in any form. This book acknowledges information obtained from other's published and unpublished work, and a list of references is provided.

Mohammad Fahim Shahriar
18-37258-1
Department of Computer Science

M. Haq

Bazlul Bari Mozharul Haq
18-37266-1
Department of Computer Science

Tamzid

Tamzid Ahmed
18-37503-1
Department of Computer Science

Atiqur

Md. Atiqur Rahman
18-37535-1
Department of Computer Science

Approval

The thesis titled "Early Detection of Alzheimer's Disease Using Deep Learning" has been submitted to the following respected members of the board of examiners of the department of computer science in partial fulfilment of the requirements for the degree of Bachelor of Science in Computer Science on October 22,2021 and has been accepted as satisfactory.

Dr. S. M. Hasan Mahmud
Assistant Professor & Supervisor
Department of Computer Science
American International University-Bangladesh

Dr. M. M. MANJURUL ISLAM
Assistant Professor & External
Department of Computer Science
American International University-Bangladesh

Dr. Md. Mahbub Chowdhury Mishu
Assistant Professor & Head (Undergraduate)
Department of Computer Science
American International University-Bangladesh

Professor Dr. Tafazzal Hossain
Dean
Faculty of Science & Information Technology
American International University-Bangladesh

Dr. Carmen Z. Lamagna
Vice Chancellor
American International University-Bangladesh

Acknowledgement

At first, all praise to almighty Allah for giving us the strength and courage to continue working throughout the last four years and especially the last two semesters. With His blessing, we have finally completed the thesis even through a rough pandemic and economic hardships.

We also want to take this opportunity to thank our parents for their relentless support that saved us from outside hardships.

We especially like to thank our supervisor, Dr. S. M. Hasan Mahmud Sir for his relentless support in times when we had no direction to go. We are grateful to Hasan sir for always keeping time for us even in busier schedules and communicate with us to show us our flaws and aspects of us that we can improve. We honor the support Sir has provided us and dedicate this work to him.

We also like to thank google for freely providing a useful tool such as google colab without which this thesis would not be possible. The computation power google can provide through cloud is game changing and inspired us to be more so that one day we can influence many lives as google influenced ours.

Abstract

Alzheimer's Disease is a very prevalent form of dementia which is not only incurable, but also very costly to treat. It is important to detect Alzheimer's in early stages to start treatment early and lessen its eventual drastic effects. It is also very difficult to detect as patients often show symptoms that are similar to normal aging elderly people. This thesis tries to tackle the problem by proposing a convolutional neural network model. The model provides a testing accuracy of 79% in 5 class classification and 79.5% in 3 class classification. To authors knowledge, no work has been performed on five classes. While progress have been made on 3 classes, the accuracy ranges from 66-75% in literature reviews, noting that more work needs to be done in early detection of AD. In binary classification, the accuracy reached 88.2% for AD vs CN. Overall, CNN produced best accuracies for different classes when compared with KNN, ANN, SVM, RF.

This thesis also does a comparative study where it compares CNN's precision and recall with other classifiers such as KNN, ANN, SVM, RF to recommend models to specific class scenarios. This thesis tests the mentioned classifiers for 5-class, 3 class and 3 binary classifications to show which algorithm is better for which classes by comparing class wise precision and recall scores. CNN and KNN performed best in various classifications where they often shared the best precision or recall for a particular class. RF shined in some cases when it dealt with the CN class. We hope this thesis will provide an understanding to future researchers to create more robust models that can actually replace clinical diagnosis.

Key Words: Machine Learning, Deep Learning, Computer Vision, Alzheimer's Disease, Mild Cognitive Impairment, Dementia, Convolutional Neural Network, K-Nearest Neighbor, Support Vector Machine, Artificial Neural Network, Random Forest, Comparative Study.

Table of Contents

List of Figures

List of Tables

Abbreviations

AD	Alzheimer’s Disease
ADNI	Alzheimer’s Disease Neuroimaging Initative
BIMC	Blessed-Information-Memory-Concentration
BLSA	Baltimore Longitudinal Study of Aging's
BP	Back Propagation
BS	Bayes Statistics
CAD	Computer Aided diagnostics
CDR	Clinical Dementia Rating
CN	Cognitive Normal
CNN	Convolutional Neural Network
CSF	Cerebrospinal Fluid
CSVM	Clustered Support Vector Machine
CT	Computerized Tomography
DMF	Data Mining Framework
EEG	Electroencephalogram
ELM	Extreme Learning Machinesqq
GMM	Gaussian Mixture Models
LMOCV	Leave Multiple Out Cross Validation
LOOCV	Leave One Out Cross Validation
LSN	Large Scale Network
LVQ	Learning Vector Quantization Networks
MCI	Mild Cognitive Impairement
MCIc	MCI patients who eventually become AD patients
MCInc	MCI patients who did not become AD patients
MLDA	Maximum Uncertainty Linear Discriminant
MRI	Magnetic Resonance Imaging

NMF	Non-negative Matrix Factorization
NMSE	Normalized Mean Square Error
OPLS	Othogonal Partial Least Squares
PCA	Principle Component Analysis
PNN	Probabilistic Neural Networks
PLR	Penalized Logistic Regression
PLS	Partial Least Square
RBF	Radial Basis Networks
ReLu	Rectified Linear Unit
RF	Random Forest
ROI	Region of Interest
SLR	Sparse Logistic Regression
SPECT	Single Photo Emission Computed Tomography
SRC	Sparse Representation-Based Classifier
VAF	Voxel As Feature
VBM	Voxel-based Morphometry
VFI	Voting Feature Intervals
VOIs	Volumes of Interest

CHAPTER 1: INTRODUCTION

This chapter provides a general idea of the thesis. It explains the motivation behind this thesis, the objectives to be achieved through this thesis. as well as provides a basic idea of the overall structure of this thesis.

1.1 Motivation

Alzheimer’s disease is considered as a prominent kind of dementia with no effective form of cure as of yet. It is estimated in 2016 that there are around 47 million cases of Alzheimer’s dementia worldwide while by 2050, that number is projected to touch 131 million [1]. This can be attributed to the general increasing of age in the population. As being the most common form of dementia, the cost associated with AD is estimated to be 818 billon US dollars worldwide in the near future[1]. With this alarming progress, it becomes increasingly important to tackle this disease in different ways. While the disease is incurable, early detection show promise of better treatment outcomes as well as tackling potential side effects [2].

Diagnosing AD at early stages can better prepare the involved families to tackle eventual complications raised by the disease. Early detection also means more time for people to take preventive actions that can help them cope with side effects of the disease. Traditionally AD has been diagnosed using clinical observations and regular cognitive evaluation. However, Recent progress in neuroimaging suggests that photo analysis of various scans can be a more dependable as well as effective approach. As a result, an increased focus is given on figuring out biomarkers and performing various machine learning and deep learning techniques for early detection of AD.

To attain this goal, Structural Magnetic Resonance Imaging (MRI) is a very interesting neuroimaging technique as it is a widely used, non-invasive and also great for recognizing structural changes in brain structure, which is related to AD. Besides, The Alzheimer’s Disease Neuroimaging Initiative (ADNI), a leading research project in this field, has a significant quantity of MRI image data that has been used widely for classification of patients at various stages of AD. Recent research shows great promise when deep learning techniques are used in ADNI MRI images for quick and accurate detection. This introduces new opportunities to further develop the understanding of deep learning while introducing more fine-tuned algorithms to ensure early and accurate detection.

1.2 Thesis Objective

The main goals of this thesis are:

1.Propose A Deep Learning Convolutional Neural Network for classifying various stages of AD for early detection. This is done to evaluate how effective early detection through deep learning is compared to traditional clinical diagnosis. This process involves training different classifier models and improving their accuracies and make the CNN better in terms of accuracy.

2. Compare various other image classification techniques with the proposed technique by comparing various performance matrices to suggest different methods for different use cases. This is achieved by comparing the class-wise precision and recall of different models.

1.3 Thesis Outline

Chapter 2. Background

In this chapter, we will describe the contextual theories as well as information needed to implement the proposed algorithms.

Chapter 3. Literature Review

This chapter will systematically review the related research papers to describe latest findings.

Chapter 4. Methodology

This chapter will describe the data, the data processing methods, feature extraction methods and the classification algorithms used.

Chapter 5. Results

The results derived from the algorithms will be presented in this chapter.

Chapter 6. Discussion and Conclusion

This portion will discuss the results and the shortcomings of the procedure.

References

This portion references all the papers used to complete this work.

CHAPTER 2: BACKGROUND

This chapter delivers a small overview over the background theory, ideas and methods used in this thesis.

2.1 Alzheimer’s Disease

Alzheimer's disease (AD) is a neurological illness responsible for impairing a person's behavior, thinking, and memory to the point where they frequently forget things and have trouble doing daily chores. Alzheimer's disease is the most prominent form of dementia, responsible for roughly 60-80 percent of all dementia cases [3].

AD symptoms are comparable to general human aging, thus making detection harder in early stages. It is critical to recognize that Alzheimer's disease, like dementia, is not a normal aspect of aging and that both of these conditions imply brain deterioration. The most common symptom is memory loss to the extent that daily activities are hampered. In time, signs of depression and apathy is also prevalent and some may also suffer from verbal and physical agitation [4]. With further disease progression, delusions, hallucinations, and aggression become a more common sight [4]. Overtime, the symptoms of dementia only keep worsening. Since AD has no cure at the moment, the goal is to lessen the development of the disease so that long term treatments can be used to deaccelerate dementia symptoms. This can be achieved through early detection. While there is no specific diagnostic method to detect AD, several approaches can be conducted to ensure smooth diagnosis. These approaches are: examining individual medical history, AD presence in family, family feedback regarding cognitive decline such as acute memory loss and behavioral changes.

The precise cause of AD is still not known. Recent research however, points out that it has an association with neuritic plaques and neurofibrillary tangles in the brain [5]. Neuritic plaques are made of a special protein called Amyloid beta. While it is strongly implicated with AD, its role as a causative factor is still debated. It is however, regarded as a general marker of the disease. The current research goal seems to be identifying the biomarkers such as Amyloid Beta, Neurodegeneration detracted by rise of CSF (cerebrospinal fluid), brain atrophy on MRI that allow early identification of AD related symptoms [6]. There are two categories of biomarkers. Amyloid deposition in the brain is an early biomarker while neurodegeneration (MRI, FDG, PET) is a late-stage biomarker. These bio-markers often indicate progress early progression of dementia, neurodegeneration or Alzheimer’s Disease. However, these can also be seen in elderly patients who are facing normal neurodegeneration due to aging and age-related complications.

2.1.1 Mild Cognitive Impairment

Mild Cognitive Impairment (MCI) is often considered as an initial stage of AD where the patients cognitive changes can be noticed by family members. However, often the changes are not significant enough to be classified as dementia. It can be considered as the path towards dementia.

2.1.2 Risk Factors

Age

The effect of AD can be easily seen in the elderly population. The risk of AD doubles every five years for people aged 65 and more and after the age of 85, the risk is almost 50 percent [3].

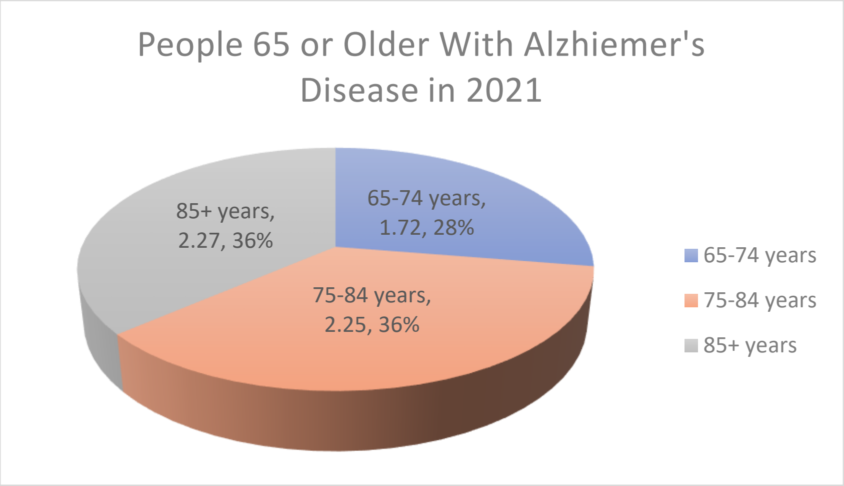


Figure 2.1. AD patients aged 65 or older in million [86].

Sex

In general, there are more women who are suffering from AD than men. While there is no conclusive evidence that women are more prone to AD than men, the reason there are more women AD patients than men are simply because on average cases, women live longer than man. And elderly people are at higher risk of AD. That is why in data, women are shown to be in greater risk of AD. Figure 2.2 shows the risk of AD in men and women of age 45 and 65 years.

Family History

Individuals whose close family members have suffered AD are more likely to follow the same fate. The risk increases with 2 or more family members with history of AD. Environmental as well as hereditary factors can play a part in carrying the disease in the family [3].

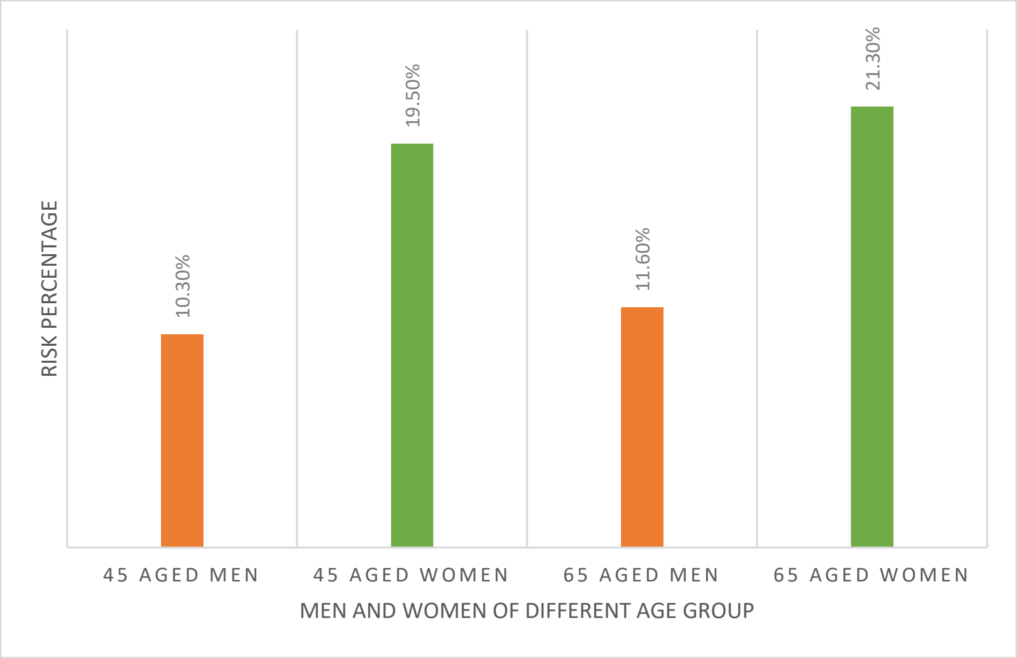


Figure 2.2. Risk of AD in men and women of 45 and 65 years [87].

Genetics

A potential cause of AD is genetical factors.AD is mainly influenced by two types of genes:

- Deterministic Genes
- Risk Genes

Deterministic genes directly cause AD so anyone inheriting the gene will eventually develop symptoms of AD. This is a rare case accounting only 1% of AD cases [7]. While this gene is rare, its discovery has paved the way for future research for understanding AD. Risk genes are genetical component that can increase the chances of diagnosed with AD but are not always the direct cause of AD.

2.1.3 Biomarkers

With the progress of AD, biomarker values keep increasing to reach unnatural values. Some biomarkers related to dementia are:

1. Amyloid beta (can be detected in CSF and PET amyloid imaging)

2.Loss of neuron measured detected by MRI.

3.Neurodegeneration that can be measured by calculating tau levels in CSF and synaptic dysfunction that can be calculated using FDG-PET.

4.Constant memory loss that can be evaluated through cognitive assessment.

5.Cognitive decline indicated by cognitive test.

The biomarkers 1,2 and 3 can be seen before the diagnosis of dementia while biomarkers 4,5 are classic dementia diagnosis indicators [8].

2.1.4 Biochemistry

AD is a protein misfolding disease. In protein misfolding, the polypeptide folds incorrectly, causing its final three-dimensional structure to be incorrect. As a result, the protein doesn't perform its intended work. Often the protein quality control (POC) systems break down these proteins to be used for future protein synthesis. However, in cells that are aging and cells with genetical diseases may not properly control this buildup of misfolded protein. As a result, these protein blocks may hamper basic cell level activities 11. In the case of AD, amyloid beta protein folds abnormally in brains of AD patients [3][5]. These amyloid protein fragments often form a sticky plaque like structure. These structures block the signaling between brain cells, triggering immune response and thus the eventual death of neurons through neurodegeneration.

Amyloid plaques are "a characteristic sign of a pathological diagnosis of AD," according to the researchers. [11]. The formation and accumulation of these plaques can be detected by biomarkers. The protein quaintly can be measured in plasma as well as in cerebrospinal fluid (CSF).

Alzheimer's Disease can also be caused by abnormal clustering of tau protein. Tau protein is a protein associated with microtubule that stabilizes microtubules in the cytoskeleton of the cell. It keeps the microtubules straight, causing molecules to pass freely between different cell components. However, when the protein tangles and creates twisted strands, the microtubules loss their structure and disintegrate, causing the obstruction of Ion and nutrient transport from cell to cell and leading to eventual death of neurons. This continuous accumulation of neurofibrillary tangles as well as formation of beta-amyloid plaques eventually cause death of neurons and breakage of synapses, causing various cognitive problems and memory decline [12]. These plaques and tangles, on the other hand, can also be noticed in the brains of older people who do not develop symptoms of Alzheimer's disease during their lifetime. [5] [13]. Since such scenarios are rare in the brains of young individuals, it is assumed that the AD related symptoms in older patients means that they represent "pathological aging" [15] or preclinical AD [13][14]. This implies the disease may exist, but there are no clinically noticeable changes in cognition. Currently AD and natural aging cannot be easily differentiated from one another.

Morris et al. (1996) investigated the links between “cognitively normal aging, very mild Alzheimer's disease, and the presence of neocortical senile plaques” in a study. The findings suggested that neuritic plaques may signify pre-symptomatic or undetected early symptomatic Alzheimer's disease, rather than being a natural component of aging. This is "consistent with the idea that amyloid beta accumulation is an early pathogenetic event in Alzheimer's disease progression" [13].

However, according to Dickson et al. (1992) [15], "cerebral amyloid buildup (plaques) is not necessarily connected to clinically obvious cognitive impairment." They argue that other factors such as loss of neurons and synapses, as well as significant cytoskeletal abnormalities (tangles), are necessary for dementia in Alzheimer's disease.”

Tiraboschi et al. (2004) published a research [5] that looked at the relationship between neuritic plaques and neurofibrillary tangles with the progression of AD in 102 patients with dementia and AD and 29 healthy people. After death, cases of Alzheimer's disease were classified based on their most recent Mini-Mental Situation Evaluation. Thioflavin-S preparations were used to target neuritic plaques and neurofibrillary tangles in the interior parietal, midfrontal, superior temporal, hippocampus, and entorhinal cortices.

The results showed that around 87 percent of the normal subjects had neurofibrillary tangles, while only 37 percent of the subjects had neocortical neuritic plaques. Among the normal subjects, 19 percent had hippocampal plaques. However, among AD patients, less than 10% had neocortical tangles and tangles were absent in half of the cases with subjects who died with mild AD. Plaques were common within all AD patients. While increase in density of neurofibrillary tangles and neuritic plaques resulted in increase of the harshness of dementia, noteworthy changes occurred for neurofibrillary tangles.

After the comparison of cognitive normal subjects with AD diagnosed individuals, they only significant difference was the amount of neuritic plaques. As a result, the authors declared neurofibrillary tangles sensitivity to be a lower marker than neuritic plaques sensitivity. They then come to the conclusion that deterioration in AD is accelerated by neuritic plaques and neurofibrillary tangles at various phases of AD. They also came to the conclusion that only neuritic plaques are linked to early indications of Alzheimer's disease. This finding has prompted researchers to further look for the formation of neuritic plaques.

2.1.5 Symptoms

The primary symptoms of AD are having difficulty remembering new information or events which eventually may lead to scenarios such as disorientation, behavioral changes, confusion about past events, being more suspicious of family, difficulty in performing basic cognitive tasks, severe memory loss as the disease progresses [9]. The primary symptoms may often be more apparent to family members than outsiders.

In initial phases of AD, in the hippocampus of the brain, destructive pairing of plaques as well tangles occur. Patients often feel short time memory loss, indicating earlier signs of Alzheimer's. As time progresses, proteins synthesized using the direct/risk genes enters other parts of the brain, gradually increasing complications related to cognitive activities. This causes damages to various parts of the brain, creating various problems in different stages of the disease. When the front brain, responsible for thoughts and logic, gets affected, the patient starts struggling with logical thoughts. This also causes a decline in emotional control, causing erratic behavior. As top of the brain gets affected, the patient feels more paranoid and hallucinated. As the disease progresses, the proteins infect the rear part of brain, erasing even the deepest memories. Eventually, the control centers that regulate heart beat and breathing gets severely affected and the patient faces eventual death.

2.1.6 Diagnosis

A patient is considered as an AD patient when the disease indicators are strongly visible. However, since many symptoms are related to aging, the differentiation can be difficult. Besides, AD is a long sluggish process without any specific event that informs its presence until early-stage dementia is noticed. So, it can be quite challenging to classify transition points for each person [9]. It is also difficult to pinpoint the transitional point from "symptomatic predementia phase to dementia onset" [41]. However, recent trend in using machine learning and deep learning algorithms for classification and prediction show promise and may become a certified way of early detection as further research is made.

2.1.7 Treatment

While progress has been made in detection and treatment of AD, it is still a hopeless disease that ensures eventual death. The current treatments mainly try to tackle the progression of the disease and lessen the impact of the symptoms. This becomes effective in case of early detection. These treatments deal with symptoms such as cognitive decline and the psychological problems resulting from it. These treatments are also used to lessen the effect of behavioral problems and ensure environmental adjustments to assist the patients to deal with basic daily activities with less burden.

2.2 Structural Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a non-invasive neuroimaging method capable of producing 3D images at any depth and orientation without using ionizing radiation that is found in x-ray and CT scans. This makes the procedure relatively safer. It is often chosen for diagnostics related to soft tissues as it can provide great contrast to soft tissues such as brain, body fluids and muscles [10]. What MRI does is mapping the water molecule positions in different tissues where they exist in different densities.

2.2.1 Physics

Human tissue consumes water. Water is composed of oxygen and hydrogen. The nucleus of hydrogen atom is also a proton, which is a positive charge and spins around its axis. This implies that it can produce its own magnetic field and, as a result, be impacted by another magnetic field. In MRI, strong magnetic fields are created by passing electric current through a coil of the MRI machine. These magnetic fields cause the protons to spin parallel or anti-parallel relative to the magnetic field direction. While the ratio of these two directions is almost 1 to 1, a very few electrons (9 out of 2 million) spins in the parallel direction instead of the anti-parallel direction. These extra electrons require more energy to be directed to anti-parallel spin and thus are mainly used in MRI. Let us call these protons “remaining” protons.

The axes of these spinning protons are redirected using radio waves with a frequency equivalent to the Larmor Frequency. It is proportional to the strength of the applied magnetic field [41].

$$\omega_0 = gB_0 \tag{2.01}$$

Here, B_0 is the applied magnetic field strength and g is the gyromagnetic ratio [41]. The value of g depends on the source matter of the nuclei in consideration. For hydrogen, this value is, g=42.6MHz/Tesla [16][41].

The few “remaining” protons will absorb the radio waves and with the absorbed energy, will flip their axes. Once the radio waves stop emitting, the spinning axis of the protons will reposition to equilibrium state and as a result, the protons will emit energy in the form of RF signals that is further processed to create gray scale 3D MRI images. While the image is processed, signal weighting is also a factor of consideration. It affects the contrast between tissue sections in the image. There are mainly two types of signal weighting:

T1 and T2. T1-weighted images are better for brain scans as they point out significant differences between white and grey matter.

2.2.2 Imaging

The RF signals returned from the protons are then transformed into electrical flow that is then digitized and further processed to create a series of images. These images show a thin slice of the designated area. In these grayscale images, different densities are shown using different shades of gray. While the produced images may vary depending on the signal weighting used by the machines, T1 weighting creates images where structures such as dense bone, air and sections containing few hydrogen protons will look black while concentrated proton areas and body-fat will show white and so on.

2.2.3 Slicing

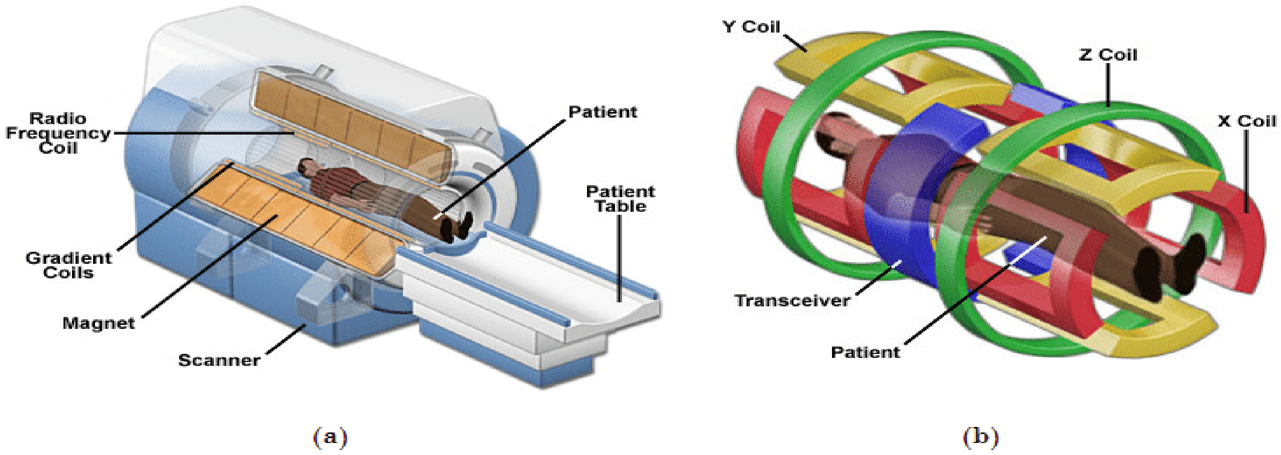


Figure 2.3. (a) MRI Scanner Cutaway (b) MRI Scanner Gradient Magnets (MRI: A Guided Tour, 2015) [88].

The MRI scanner contains three additional magnets (illustrated in figure 2.1) named after each axis called X, Y and Z. They are relatively weak compared to the primary magnet. These magnets are oriented to different planes with respect to the patient's body. These magnets allow the scanner to image the designated areas in slices by turning the magnets on and off repeatedly. The resulting image slices can be X-Y slice (transverse slice), Y-Z (sagittal plane) slice and X-Z (coronal plane) slice. These slices can be illustrated in figure 2.4.

Since MRI doesn't use ionizing radiation, the procedure relatively safer. It is often chosen for diagnostics related to soft tissues as it can provide great contrast to soft tissues such as brain, body fluids and muscles

[10]. What MRI does is mapping the water molecule positions in different tissues where they exist in different densities.

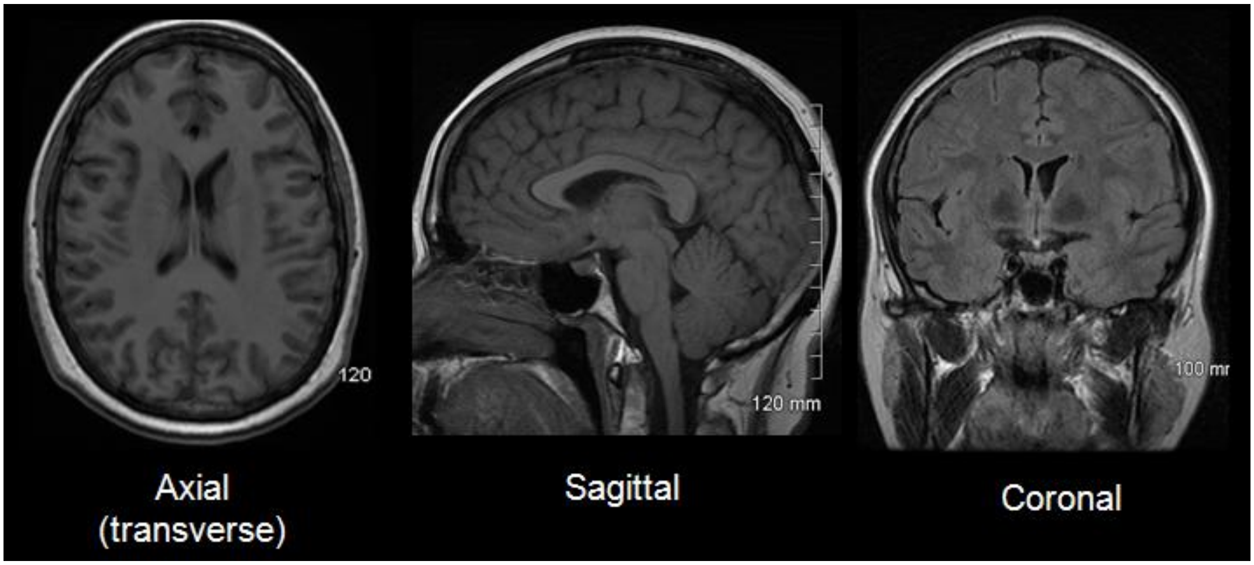


Figure 2.4. Brain Image Slicing Planes [89].

2.3 Organizations

This subsection tries to honor the organizations that work to provide data and resources to ensure an improved understanding of the key ideas and provide a framework to design and perform future AD research. This chapter also acknowledges the efforts they made to make all this vital clinical information and datasets public, without whom the progress made in tackling Alzheimer's Disease would be near impossible.

2.3.1 The Alzheimer's Disease Neuroimaging Initiative

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a multi-year project aiming to "creating clinical, imaging, genetic, and biochemical biomarkers for rapid diagnosis of AD" [17]. The ADNI1 trial started in 2004 and included 400 people with Mild Cognitive Impairment, 200 people with early Alzheimer's, and 200 older people as controls. From 2009 to 2011, the study was prolonged with ADNI GO, which included 200 people who were diagnosed with early MCI in order to look at biomarkers in the early stages of Alzheimer's disease. This act makes sure that researchers have access to reliable clinical data on AD and that they can properly understand the pathology of Alzheimer's disease using biomarkers, allowing for earlier diagnosis of the disease [41]. It will also supply clinical trial data to back new study approaches, disease stoppage, and handling, as well as continue and enhance the exchange of their data repository. The purpose of ADNI1 was to create more reliable diagnostic tools for early detection of

AD and to identify the pathology using biomarkers. The program was effective in generating a library of datasets storing brain scan data, as well as developing initial classification techniques for Alzheimer's disease and a standard diagnostic testing procedure. The Alzheimer's Disease Neuroimaging Initiative (ADNI) is an important step toward better diagnostic techniques and the development of viable medicines to delay the spreading of Alzheimer's disease and, eventually, cure and prevent it.

2.3.2 The Alzheimer’s Association

The Alzheimer's Association is a humanitarian American health organization dedicated to providing Alzheimer's disease care, support, and research resources. As part of their research focus, they hold the Alzheimer's Association International Conference on Alzheimer's Disease (ICAD) to review and update existing information on AD. ICAD is the world's largest association of Alzheimer's researchers. The association regularly publish the official periodical “Alzheimer's & Dementia” , which contains the latest clinical research breakthroughs of AD.

2.3.3 The Laboratory of Neuro Imaging (LONI)

LONI is a research facility of the University of Southern California's Institute for Neuroimaging and Informatics. It was and continues to be one of the most prestigious neurological research institutes in the United States. They conduct research and develop computerized brain maps that are population-based and disease-specific. They also help with study participant training and medical data transfer [18].

They've also built technical resources to support the comprehension of the links between structural and functional features of the brain. They're also working on "...the difficulty of comparing data among persons as well as across modalities..." [18], with a focus on functional data sets, as well as imagining software to deliver research outcomes effectively. LONI is significantly engaged in various kinds of national and international partnerships, some of which it serves as a hub for. Several of their brain mapping work has concentrated on AD, and LONI maintains the ADNI data in specific information storage systems.

2.4 Machine Learning & Deep Learning

Machine learning deals with automating computations without explicit programming so that machines can learn and improve from experience to complete specific tasks without human intervention [19]. It is a multi-disciplinary field drawing inspiration from various disciplines such as mathematics, programming, philosophy, artificial intelligence and statistics. It is also used in various fields ranging from neurobiology to finance to crime and more.

A simple machine learning system requires three key components: training data for training (learning) the model, a model that will use the data to train or learn, and a training algorithm which the model will follow. Generally, the learning means fitting and fine tuning the model's parameters in accordance with the training algorithm by using the supplied training data. A portion of data is separated before training for upcoming validation called "Validation Data". After that, to calculate the real-world performance of the model, it is tested on testing data and various accuracy metrics are evaluated for further optimization. The aim of this whole learning process is to ensure the machine can improve its functions from its own experience (training data) and perform well on unknown instances of data (testing data).

Machine Learning is divided into three separate groups: supervised, unsupervised and reinforcement learning. Supervised learning works on labelled data, where a set of input is associated with a specific output class. In unsupervised learning, the data is not labelled, so the machine have to figure out hidden patterns in the dataset to predict the outputs.in reinforcement learning, the machine performs a series of actions to reach a goal and based on the rewards or errors it receives, it chooses its next task and tries to reach the goal by maximizing rewards.

There are different machine learning models with different prior assumptions on input-output mapping or the distribution of data. These assumptions are made to find the target function and predict outputs for unknown instances of input that the model hasn't encountered yet. These assumptions make learning possible for machine learning models and these are called inductive biases. The importance of inductive bias is that it demonstrates how a learning algorithm generalizes outside the training instances. In case of choosing the proper algorithm for tackling a specific problem, many factors need to be considered such as the number of the instances (training examples), the completeness of the dataset, the type of data in question, insight about the data known to the learner. Whoever implements the algorithm should also have the confidence that the algorithm will build a model that will be competent enough to correctly classify unknown instances.

Deep learning is a part of machine learning that are basically neural networks that has three or more layers that is neural networks with hidden layers. This method tries to bring out hidden patterns or features from high dimensional data to learn as well as produce meaningful results that can be used for classification, detection, prediction etc. Generally, a deep learning model processes features as it goes to higher levels of hidden layers and in each layer, combination of features is computed from previous layers to calculate whether certain feature or feature combinations weight more in deciding the class of an instance. Then these values are sent as input for the next layer. In this process, the network is essentially learning the useful feature patterns that determine the class of an object. With each layer, the tasks become more complex as the model figures out more complex and concrete feature sets. For example: in a facial recognition deep network of four layers, the first layer nodes may identify the bright and dark pixels of the images. In second layer, the nodes may use these features to identify edges and simple shapes. The next layer nodes may learn to identify complex shapes using the input features and in the output layer, the nodes may enable the model to compute previous features to be able to learn face shapes and textures to determine whether an image is of a person or not.

Deep Learning is particularly inspired by neurological networks of the visual cortex where visual information is processed through various stages of transformation, creating a more complex processing stages by using information on previous state [21]. This idea is seen when training artificial neural networks where outputs from previous layers are used to create more sophisticated feature sets. Some of the prominent deep learning techniques are Recurrent Neural Network (RNN), Deep Neural Networks (DNN), Convolutional Neural Network (CNN) and Deep Auto Encoders (DAE).

While deep learning is an old idea, it has become very relevant recently with increased interest in machine learning fueled by increased performance gains in graphical hardware and software. Due to the increase in graphical performance, deep learning algorithms run many times faster than they were 10 years ago. This made training big and deep neural networks feasible and practical. This development has increased the use of deep learning tools in visuals and graphics. This is very noticeable with the progress in real time facial detection as well as self-driving autonomous systems. At present, Deep neural networks that rely on GPU-implementations has achieved remarkable accuracy results on many standardized image classification datasets, even exceeding human performance in some cases [22]. With more progress to GPU hardware, the usage of GPU-based implementations of deep learning is expected to grow with time.

While comparing with cognitive models, it becomes apparent that deep architectures may be the future of learning complicated-function that can be used for high-level problem solving such as problems related to computer vision [23]. In early days of deep learning, shallow networks were easier to train compared to deep networks as purely supervised model trainings often provided worse outcomes for deep networks [24]. This is due to issues like vanishing gradient problem, local optima problem and pathological curvature problem. The gradients calculated in backpropagation could become increasingly smaller as networks get deeper and deeper. As a result, the weights of the first layers can stop upgrading meaningfully after a time, thus complicating the training process. Besides, the weights may not be updated meaningfully for a time being because the gradient descent may be flatter for a time before finding a gradient descent that points to a local-optima or possibly the global optima. Due to such constraints, progress in deep networks was stale for a time. However, as newer methods and optimization techniques were introduced, deep learning has been implemented in many fields with remarkable success in many cases. Sutskever et al. (2013) described supervised techniques such as stochastic gradient descent can show good performance when applied in Deep Neural Networks [25], when better initialization techniques are applied. With increased GPU-hardware support and increased adoption of Deep Learning in real life smart-automation projects, it is expected that deep learning will dominate the technology industry and computer science research the upcoming years.

2.5 Machine Learning Classifiers

This part deals with the classifiers that have been used for preparing and comparing performance of the models.

2.5.1. Artificial Neural Networks

Artificial neural networks are derived from machine learning that produce models for processing data and build effective and robust machine learning models. It is structured in a way that mimics the complex structure of interconnecting neurons in brains. Each unit of the network, often referred to as artificial neurons or nodes, takes inputs in the form of a real valued number from other neurons to generate a real valued output that can also be used as inputs of other neurons. These nodes are often structured in different layers and nodes take input from a previous layer, processes the values and send the output to neurons of the next layer. While this process has similarities with information transmission of neurons, there are some significant differences between biological neuron structures and artificial neural networks. ANN doesn’t model all complex structures of brain networks and ANN introduced methods that brain networks don’t follow. ANN nodes often output a single value at a time while brain neurons output a time series of spikes [41].

Throughout this thesis, the phrase "Neural Networks" means feedforward neural networks unless otherwise mentioned. Feedforward networks are neural networks where the nodes do not create a directed cycle, thus making them multi-layered networks that feed from input layers and processes in hidden layers to provide desired output in output layers through a one directional data-flow. Fig 3 shows a simple feedforward network with two hidden layers and bias inputs. Here, the layers are fully connected. A fully connected layer means every single node is connected with every node of the previous layer and the next layer. Each node is an artificial neuron known as a perceptron. A perceptron receives a real-valued vector as input. This input can either be output of other neurons in a previous layer or raw input. It also takes a bias. Each input link has a modifiable weight attached to it. The perceptron calculates the weighted sum using the node input values and the associated weights and calculates the bias. This total result is used as an input for an activation function which classifies the result based on a specific threshold. If the result is bigger than the threshold, then the output is 1 and it is 0 otherwise. Different activation functions handle this differently and will be further discussed later.

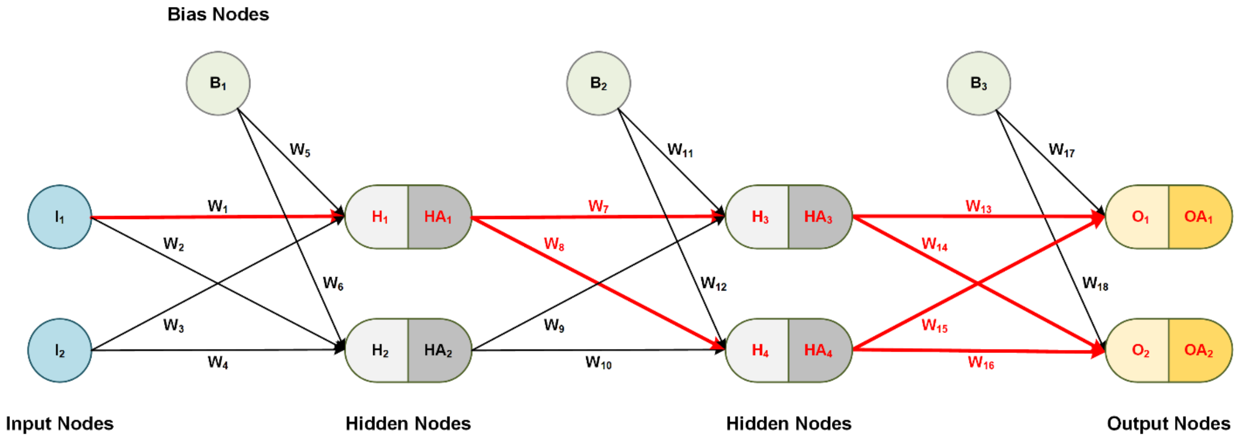


Figure 2.5. A simple Feedforward networks [90].

Since single layer perceptrons can only learn from data that show linearly separable patterns, they can be used to design primitive Boolean functions such as AND, OR, NAND, NOR [19] etc. However, multilayer perceptions can be used to classify more complex non-linearly separable patterns.

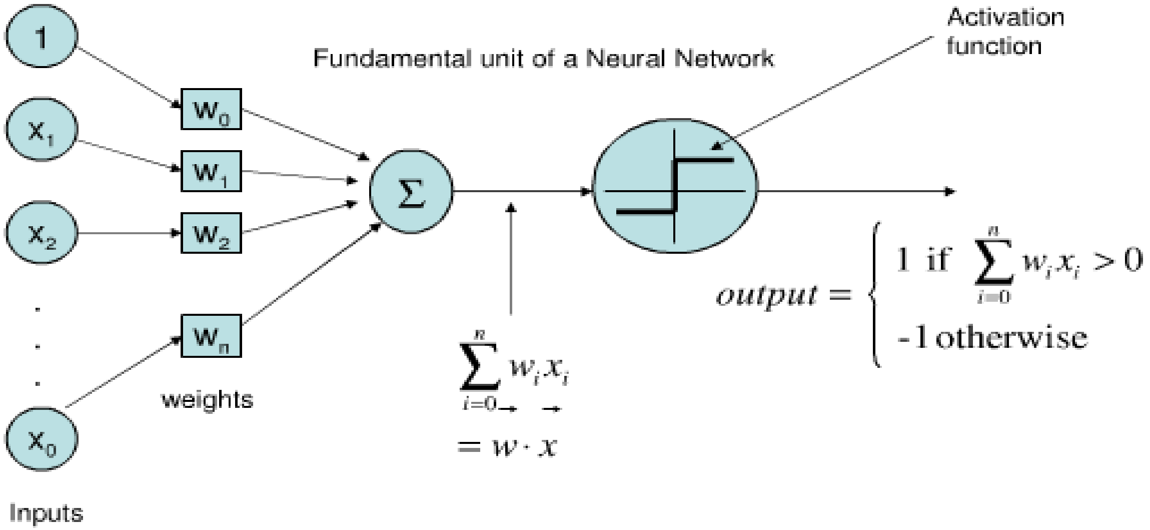

Activation function
Fundamental unit of a Neural Network
1
X_1
X_2
.
.
.
X_0
W_0
W_1
W_2
W_n
Σ
weights
Inputs
$\sum_{i=0}^{n} w_i x_i = \vec{w} \cdot \vec{x}$
$output = \begin{cases} 1 \text{ if } \sum_{i=0}^{n} w_i x_i > 0 \\ -1 \text{ otherwise} \end{cases}$

Figure 2.6. A perceptron [91].

Artificial neural networks are very robust when dealing with data with noise. So, they are frequently used for solving problems related to sensor data that are noisy such as images and audio. These networks can also be trained to deal with problems related with speech and audio, medical diagnostics, object and movement detection, facial recognition, autonomous driving, automatic trading systems etc. With advancements in deep learning, interest in artificial neural networks has increased steadily.

2.5.2. Deep Neural Networks

Deep neural networks (DNN) are artificial neural networks with many hidden layers. Generally, deep neural networks have more depth and larger dimensionality compared to ANN's. As a result, there are more nodes and layers and more weights for tuning. So, these models require relatively more powerful computer hardware and computation power. The advantage is more complex models can be trained. Deep architectures can be used to model highly complex functions by increasing the layers and the number of nodes in a single layer [26].

Deep networks are not a new idea, but due to high computational requirements, it was not a practical solution for complex problems that already relay on many variables. Since the initial weights are random in the network, this randomness can have some issues on deep networks. Lower initial weights mean the training will take a long time before any significant development. And if the weights are very large, then the network may get stuck on in a local optimum. These issues are tackled with the introduction of unsupervised pre-training. Here, one layer is pre-trained at a time with unsupervised greedy training [27]. Then the usual supervised training with backpropagation begins. This supervised training is known as fine tuning as it further tunes the weights of the network. While DNN is a usable solution for image classification, currently, CNN's are more widely used for tackling computer vision related problems [28].

2.5.3. Convolutional Neural Networks

A convolutional neural Network (CNN) is a type of neural network that works on the basis of a feedforward structure and used widely in applications of Computer Vision, especially in image classification and recognition. Its biological inspiration is the complex visual processing of mammals [23]. In a CNN, the idea is to filter the images to extract meaningful features and then the model works on the features to classify instances of images. These filters are just a set of multipliers, which can often be represented by a matrix. Visually, the matrix representation is often called as filter kernel. CNN require little data pre-processing as the algorithm operates on pixel level and automatically performs feature selection.

Individual neurons in the convolutional layer receive input from a small area of the image. A small neighborhood of input neurons will only be connected to a single neuron in the first convolution layer. This differs from artificial neural networks which are often fully connected. The neighborhood of input neurons is often called as a receptive field. Since each neuron of the convolution layer works with a specific receptive field at a time, it is more efficient compared to Ann's.

The weights in the filters are at first assigned randomly and the model fine tunes the weights as it is trained with more training instances. Each filter looks for a specific pattern or shape in images. These shapes are eventually refined into distinguishable features through which the model can accurately classify which class an instance belongs to. To control the convolutional layer, hyper parameters such as filter count, kernel shape, padding and stride value in each dimension are modified to better tune the model. Here, stride means how many pixels the filter will move in every convolution. Figure 6 shows an example of a convolution layer where the input size is 6x6, filter size is 3x3, filter quantity is 1, stride value is 1 and padding is 0.

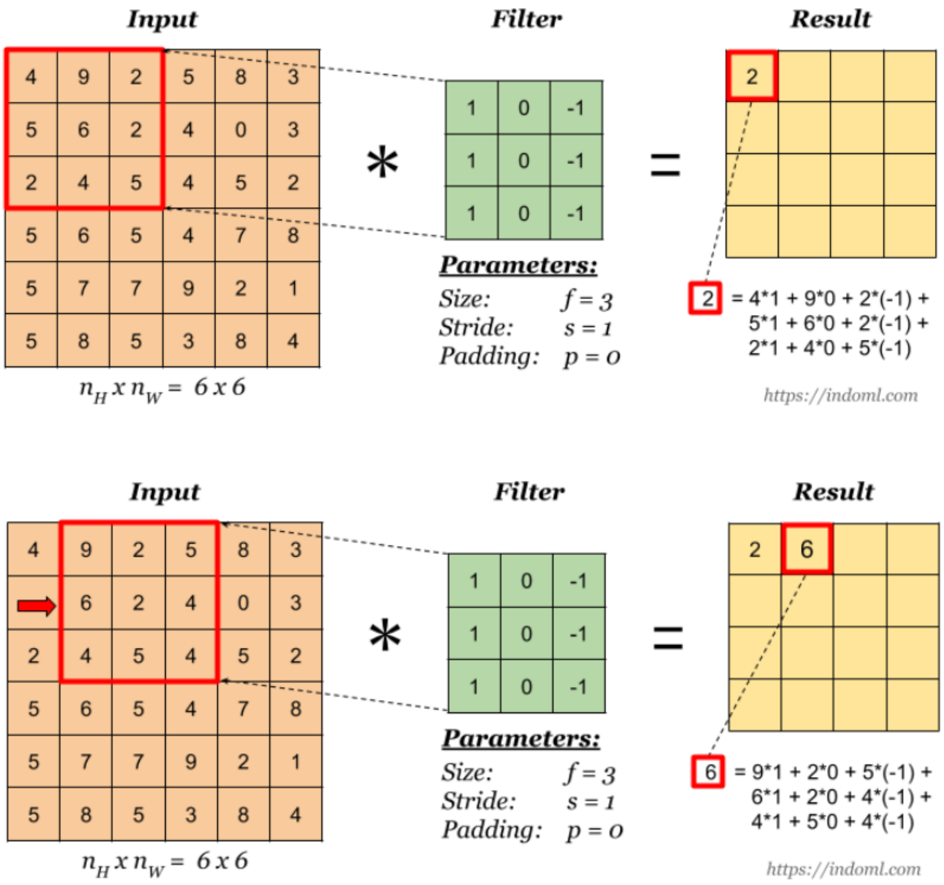


Figure 2.7. A convolution layer visualization with 6x6 input, 3x3 filter and output matrix [93].

The output size of the convolution can be measured using the equation:

$$\text{Output size} = \frac{\text{Input - Filter} + 2 \times \text{Padding}}{\text{Stride}} + 1 \qquad (2.04?)$$

In the pooling layers, the dimension of the images is reduced. The maximum or average values are taken from sections of the input matrix and a new matrix is created. Maxpool takes the maximum value from a region while avgpool takes the average value from the region. The region is the size of the pool. Often maxpooling is used as it takes the greater variant features which improve generalization performance, resulting faster convergence and better image recognition. Figure 7 shows applying max pooling and average pooling on an 4x4 input space with a 2x2 pool region.

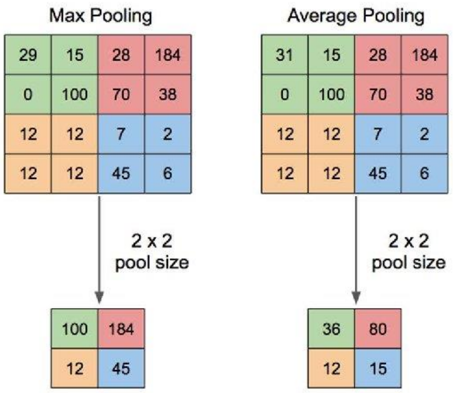


Figure 2.8. Illustration of max pooling and average pooling [94].

2.5.4. Random Forest

Random forest is a collective learning technique. It combines many decision trees to solve complex classification and regression problems. Random forest builds multiple decision trees to train separately on separate datasets. These datasets are generated from the original dataset using bootstrap aggregation or bagging. The created or “child datasets”, are created with randomized feature subset sampling and random oversampling of instances. This creates variety in datasets and the many different decision trees that works with different child datasets converge and predict more accurately compared to a single decision tree that is sensitive to data changes and can show high variance. The model classifies a test instance by finding the majority class predicted by the decision trees. Since a single instance can be repeated multiple times in a dataset, each dataset will not have all the instances of the parent dataset resulting in less correlation among the trees. These excluded instances are then used to calculate the error rate and identify important features to further improve model accuracy. “Out-of-bag samples” are a term used to describe the sample data that is used for testing [29] which are often one-third of the training dataset. The number of features for each child datasets as well as the number of trees to be trained are important hyper parameters to tune and better fit the model.

Random forest is a popular image classification for its performance. Each tree in the forest models multi-layered pixels from images and the forest creates a response variable which it then calculates by evaluating the responses from the decision trees. Every internal node has a testing logic that splits the data space [30]. The leaf nodes are labeled by the posterior probability of the object being in a particular class with respect to the existing testing logic. Then the forest averages the posterior probabilities of the classes and the class that provides the highest value is selected as the class of the input image [30].For example, if majority of the trees classifies an animal as a cat, then the forest will average that outcome and also classify the animal as a cat. Since it is less sensitive to input parameters, overfitting is not as prevalent as decision trees and pruning is not required.

2.5.5. Support Vector Machine

Support vector machine (SVM) is a machine learning algorithm that works on training instances of data to fit a generalized model which overall works well on testing instances. This is done through constructing hyperplanes to divide the dataset into distinct groups. While SVM works good on linearly separable data, to tackle with data that is not linearly separable, various kernels such as polynomial and radial basis function (RBF) kernel are used [31]. These kernels are used for mapping the data to higher dimensions and separate the instances using hyperplanes.

If the data are distributed in one dimension, then they can be separated using a dot. If the data is distributed in two dimensions, then a line can separate them assuming the data is linearly separable. If the data is three dimensional, then the distributions can be separated using a plane. While all of these separators are hyperplanes, the term hyperplane is more used when to separate data with three or more dimensions.

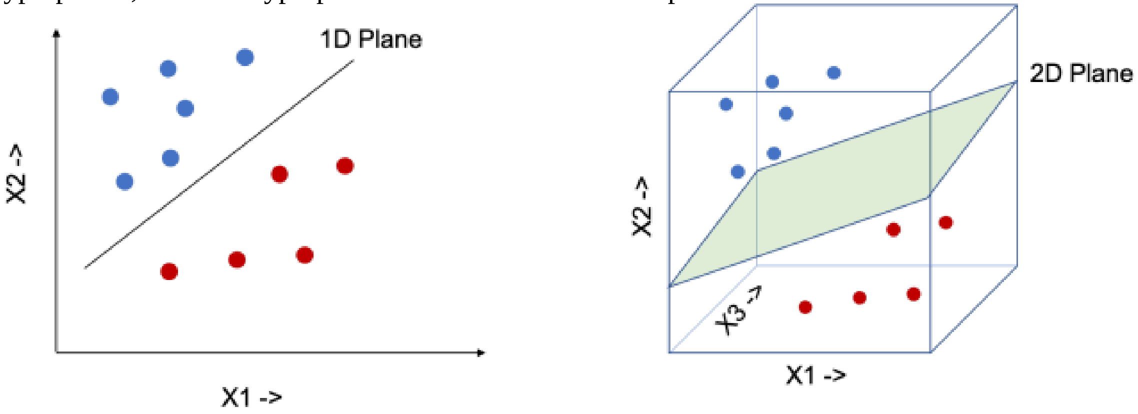


Figure 2.9. Support vector machine hyperplane in 1D and 2D [95].

While separating the data distribution, the hyperplane may separate the data in many ways. Consider a scenario where there are data of two classes. One way of drawing a hyperplane would be to take the highest margin from data groups, that is, staying at the middle of both classes. Because the hyperplane's margin is maximal from two close points of different classes, it's known as the maximum margin hyperplane [32]. This works well when the training and testing data have similar distribution. However, in case of outliers in training data, the model may overfit and misclassify testing instances. To prevent this, soft margin is used where very small number of training data is misclassified but the hyperplane is margined in a way that it can classify test instances more accurately.

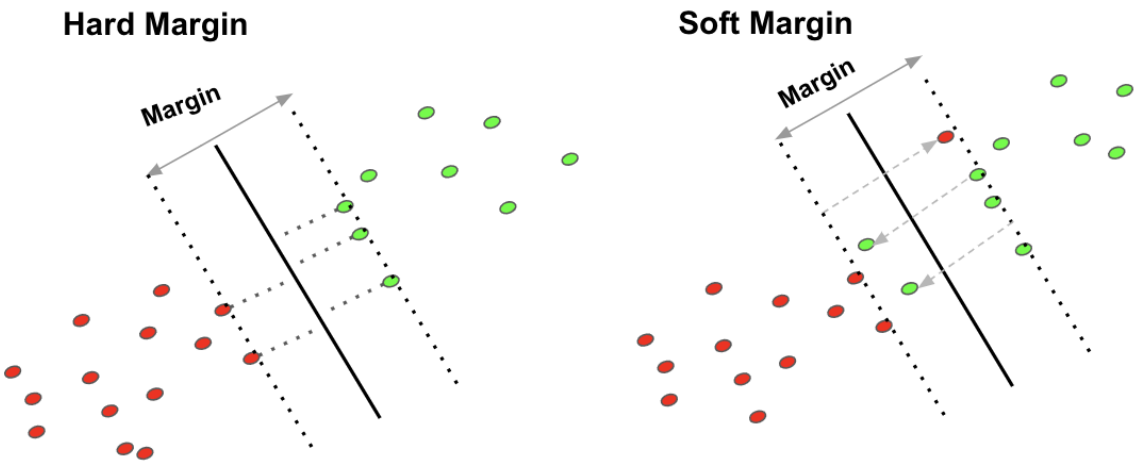


Figure 2.10. Hyperplane margin for different data distribution. For the first data, there are no outliers, so hard margin is used. For second data, using hard margin will overfit the data as the data has some outliers. So soft margin is used while misclassifying some training data for better testing performance [96].

So far, we have discussed the concepts of a linear support vector machine. However, when the data is not linearly separable, then we have to use nonlinear support vector machine [33]. In this case, the kernel function can be used to plot the data to higher dimensions. Then we can find ways to separate the data through hyperplanes. However, this also increases the curse of dimensionality, a scenario where with increase in new variables due to the increase in dimension, the solution space also increases exponentially. This problem is often solved by tuning the hyperparameters and testing the output results.

While SVM works mainly as a binary classification algorithm, it can also be modified to be used for multi class classification. This can be done by separating the dataset into datasets consisting of a pair of classes to fit a binary classification model on every dataset. In this way, a binary classifier such as SVM can be used for multiclass classification. Two approaches are used to achieve this are: one-vs-rest (OvR) and one-vs-one (OvO). In OvR, a class is compared with all other classes combined. This means the classification occurs between one class and the combination of other classes. As a result, the problem becomes a binary classification problem. One classifier is created for every class and the classification values are compared to measure the best result. This means this approach becomes increasingly difficult with the increase in class number in the initial dataset. Besides, since one class is the combination of many classes, the model may not be well trained due to data imbalance. In OvO, each class is compared with another class. As a result, for each pair of classes, a binary classifier is created. These classifier results are compared and the maximum value providing class is classified as the target class. While this approach creates more classifiers, it is less prone to imbalanced data compared to OvR.

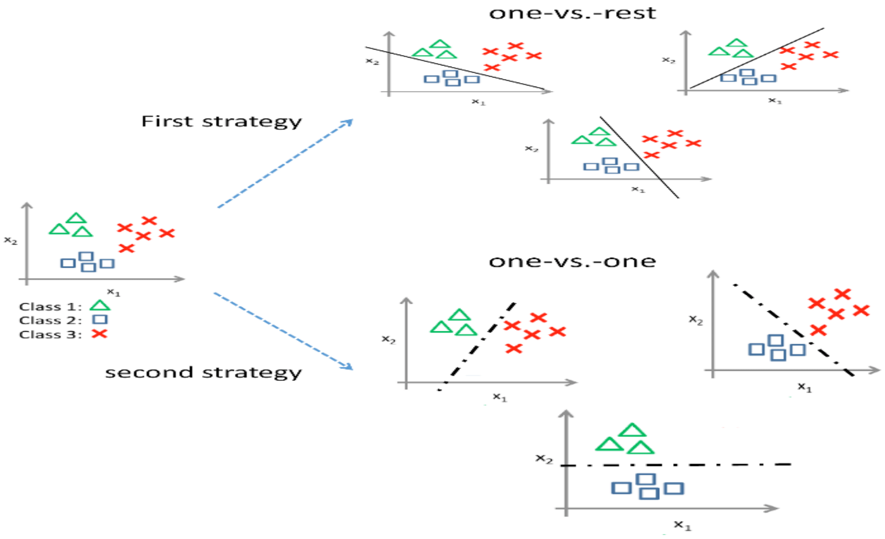


Figure 2.11. One-Vs-Rest and One-Vs-One classification [97].

2.5.6. K-Nearest Neighbor

K-Nearest Neighbor (KNN) is a very straight forward classification algorithm that classifies an unknown instance by identifying the class of its closest neighbor. KNN follows the principle that similar things stay together and classify instances based on this principle. The algorithm compares the features of an instance with features of previously labeled examples and calculates how close the features are. The class with least feature distance is then selected as the class of that instance. Since it often considers more than a single neighbor for classification, it is called K-NN, where k is the number of points it takes into account for classification. Sometimes, all the point of interests can be of a single class while sometimes there can be multiple classes (this occasionally happens with larger values of K). When k number of nearest instances contain multiple classes, the class that contains the highest number of nearest instances is classified as the target class.

KNN is also a lazy learning technique as there is no training in KNN, rather it memorizes the dataset and compares the features at runtime. Since the training dataset is loaded at runtime in the memory, it is called a Memory-Based Classification [34]. However, this means KNN is relatively computationally expensive at runtime. While choosing a value of K, it is optimal to choose an odd number which is not a multiple of the number of classes. This is done to prevent scenarios where equal number of class instances are near the target instance.

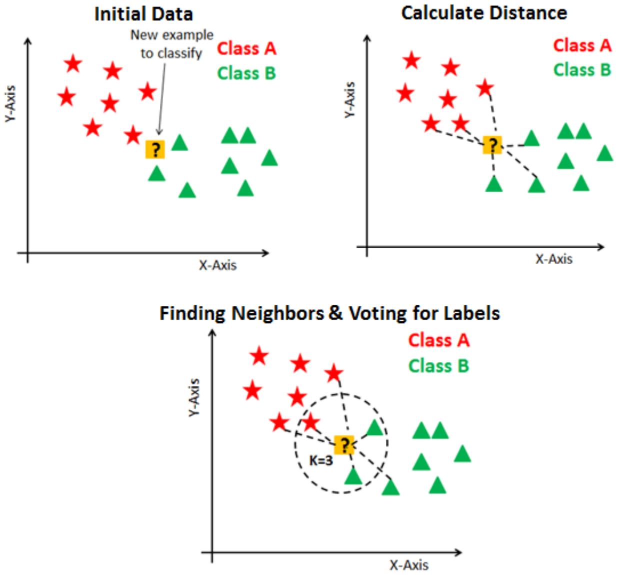


Figure 2.12. The figure shows KNN in action. At first the algorithm chooses a new example for classification. It then calculates the distance between the target instance and other instances to identify nearest neighbors. It then chooses k=3 as the number of nearest neighbors to be considered. Since there are more instances of class B in the circle, the instance is classified as part of class B [98].

For image classification, we convert the images into a fixed-length vector containing pixel values stored as real numbers. Then we compare the distance between the target instance values and the previous feature values [35]. The distance is often calculated using Euclidean distance, the linear distance between two points. Sometimes, Manhattan distance is also used for measuring the distance.

2.6 Training Algorithms

The initial weights in a neural network are often initialized using random numbers with in some cases the numbers follow normal distribution with standard deviation of 1 while the bias is chosen as a small constant. In the training process, initially some training data pattern is fed to the input layer which further processes the info and pushes it to deeper layers and calculates outputs in the output layer. The output values are compared with the target values and through backpropagation, the algorithm tries to minimize the difference in target and output value i.e., the cost function. In each iteration, the cost function value is used to modify the weights so that the cost is reduced. How much the weights can be modified depend on the adaptive learning rate.

Backpropagation is a popular learning algorithm which is used to adjust the weights of a model. Through backpropagation, the effect of weights on a particular node is evaluated for every node. This means backpropagation can see which nodes really impact a particular output. So, it tries to minimize cost function by tweaking the weights in each network layers to the previous layers and tweaks the weights that provide the greatest descent in cost function value. However, the most optimum combination of weighs is not a certainty as the model can get stuck in a local-optima. The overfitting problem is also there where the model may be very good at classifying training instances but perform poorly to generalize real life instances. This is often handled by using more data, ensembling techniques and regularization methods.

2.7 Regularization

Dropout

Dropout is a regularization technique where at each layer, certain percentages of neurons are deactivated, that is the associated activations are assigned to 0 [39]. This is done to reduce the dependency of the model on some particular features. When a model is trained, due to low amount of data or repetitive occurrence of particular features in the dataset, the model may depend heavily on those features alone. This may cause overfitting and make the model less useful in test cases. Dropout is used to solve this overfitting issue.

By using dropout, the nodes of a layer are forced to not rely heavily on a particular node in previous layer. Dropout can also be considered as an efficient model averaging technique [40]. This is because for each training sample, random nodes are inactivated. This essentially means different neural networks are trained with different samples and these networks are averaged to reduce overfitting and predict the best possible outcome. We can fix the dropout value for each layer to fix the percentage of nodes that will be inactive for the current training sample. This value can range between 0 and 1, ranging from making all nodes active to making all nodes inactive respectively. Figure 2.13 shows dropout in action where in the original network, certain percentage of nodes are inactivated in each network layer. By using different dropout values, the model is fine tuned to reduce dependency on singular characteristics as well as using it as method to average the model as different nodes will be deactivated with each iteration of model training.

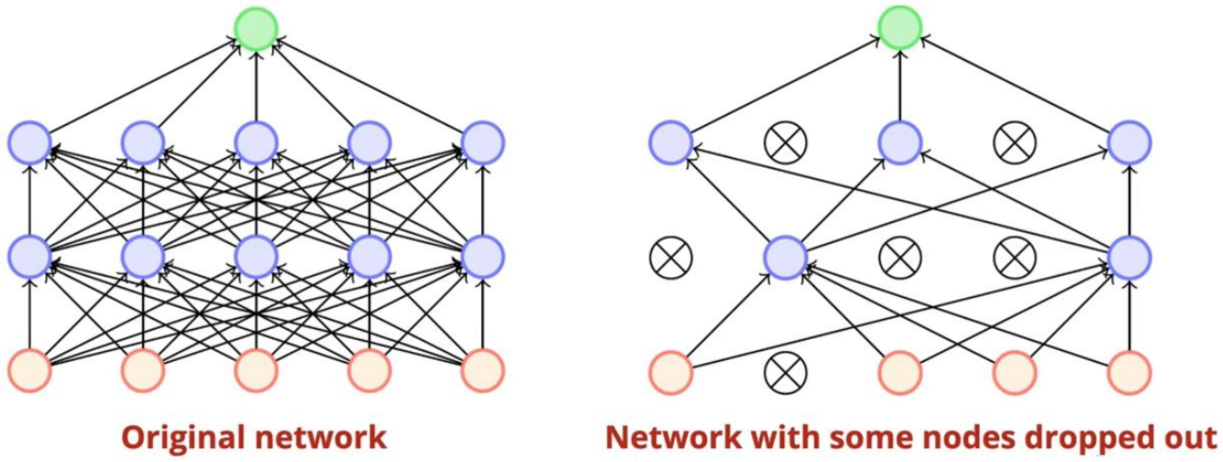


Figure 2.13. Dropout is used to inactivate certain percentage of nodes in each layer of the neural network. For the input layer (in orange), dropout = 0.2, for the first hidden layer (first purple layer from bottom), dropout = 0.6, and for the second hidden layer (second purple layer from bottom), dropout = 0.4 [99].

2.8 Activation Functions

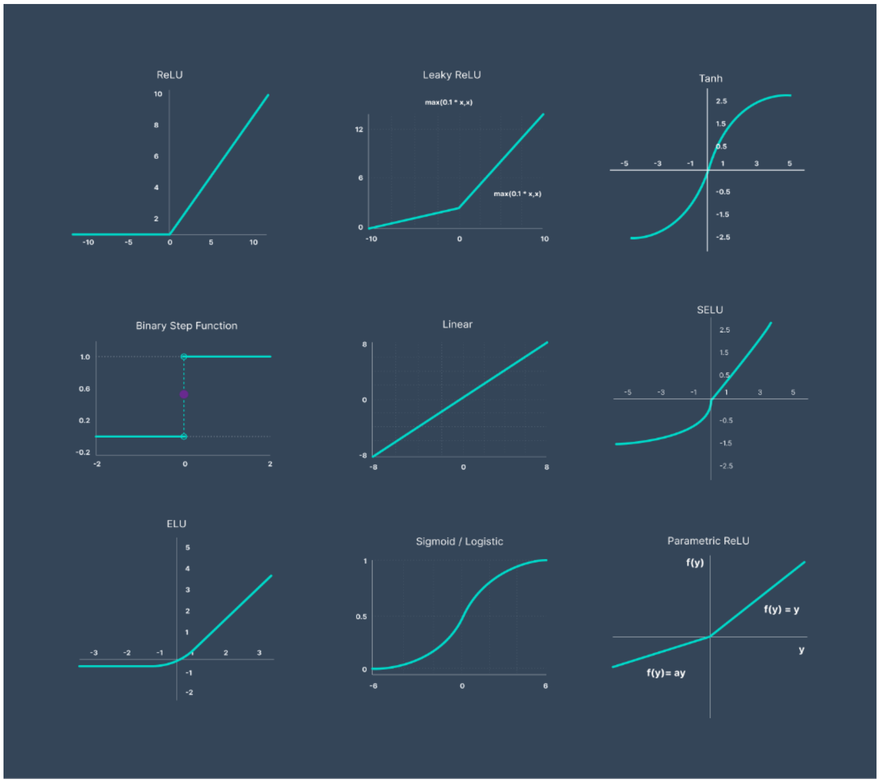


Figure 2.14. Various Activation Functions [92].

An activation function can be defined as an entity that dictates the final output of a perceptron. An activation function has a threshold value. When the function takes input from a perception’s calculation, it compares the calculated value (weighted sum) with threshold value and gives a binary output based on that. There are many activation functions available for machine learning such as ReLU, sigmoid, tanh, SoftMax etc. For classifying instances, often SoftMax is used in the final output layer. It labels the values of its input vectors between 0 and 1 and these values can later be interpreted as probabilistic values that predicts the corresponding class directed by the input vector. Activation functions can be linear and nonlinear. Linear activation functions have boundaries that are also linear so networks that need to adjust with linear changes of the inputs work well with linear activation functions. However, in practice, the inputs and errors are frequently nonlinear in properties and thus require nonlinear activation functions to better adjust the neural network [20]. As a result, non-linear activation functions are preferred and more widely used compared to linear activation functions.

CHAPTER 3: LITERATURE REVIEW

This chapter deals with the related literature reviewed for the purpose of this thesis.

3.1. Overall Description

The papers were primarily searched in google scholar. Prominent literature reviews dealing with AD and early classification were used to track papers that are in line with our thesis goals. A total of 40 papers were reviewed, 15 focused on SVM, 13 focused on neural networks and 12 focused on deep learning and ensemble methods.

It is important to identify which type of neuroimaging technique is used for solving a particular problem. In our review, majority of the studies or research used structural MRI scans to tackle their research questions. Some researchers used SPECT while others used a combination of sMRI and PET scans. While PET neuroimaging is considered as the better neuroimaging technique, PET is not frequently used in our reviewed papers. A probable reason is that PET scans are not as easily available as sMRI and may often require ethical permission from authorities and patients involved in the test. The figure below gives a numeric representation of the neuroimaging techniques used in the reviewed papers.

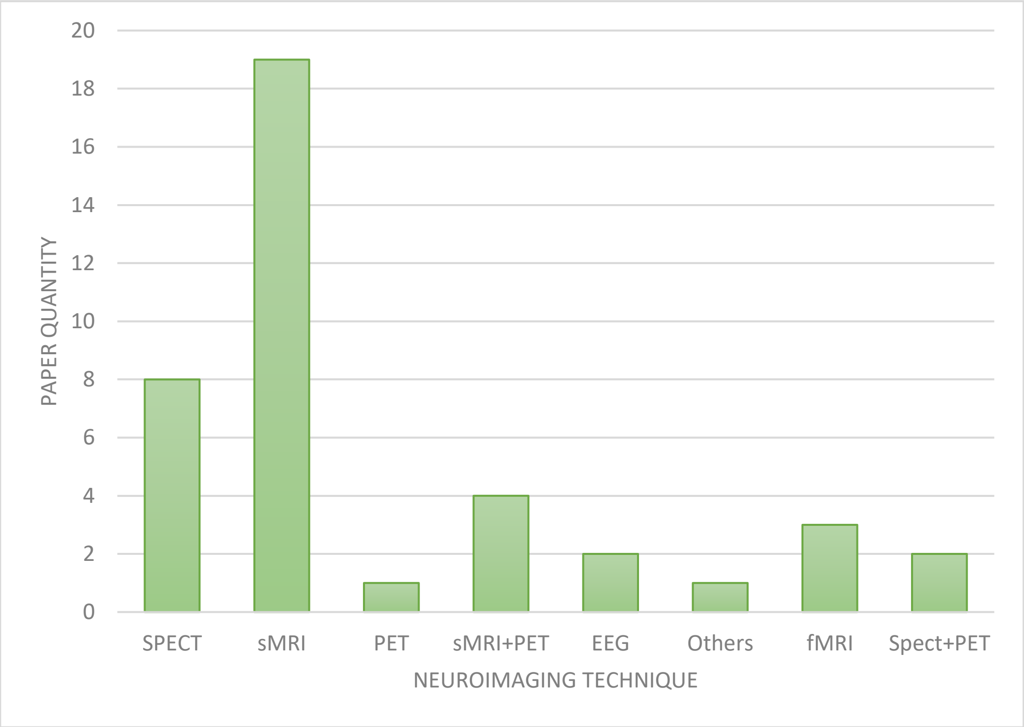


Figure 3.1. Various Neuroimaging Techniques found in the review.

While most papers try to work on early detection of AD, a lot of the papers we reviewed focus more on binary classifications. While binary classifications are a step in the right direction, it is important to understand that without classifying early stages of MCI reliably, it is not feasible to create an early detection AD model that can successfully detect initial stages of dementia. Classifying CN vs MCI can be more effective in attaining the goal as MCI is often an early symptom of AD. However, as we can see in the figure below, most of the focus is given on CN vs AD. This can be explained with the fact that CN and AD scans have more distinctions between them compared to CN and MCI. So, finding significant results can be tough. This problem is more eminent when the classification is not binary, as models struggle when there are more than two classes.

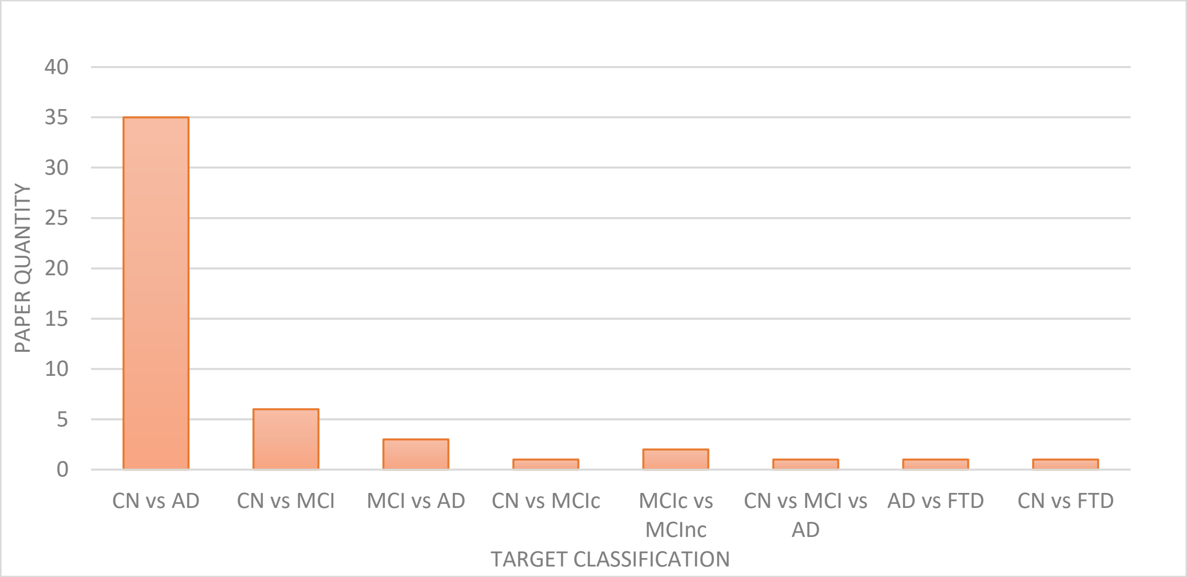


Figure 3.3. Various target classifications used in the reviewed papers and their quantity.

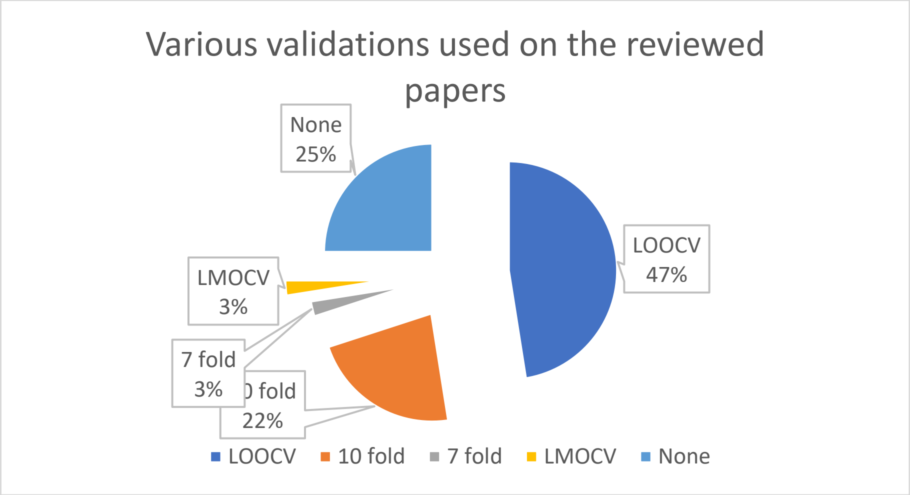


Figure 3.2. Percentage of papers using different validation techniques.

The papers have used various validation techniques for determining how reliable the model is at generalizing data. Around 47% of the papers used LOOCV (Leave One Out Cross Validation), which is the most popular validation technique in this paper. It is interesting to note that around 25% of the papers didn’t feel the need to use any validation technique. The next most popular validation technique is 10-fold cross validation with 22% of the papers using it. Some papers however, preferred to use LMOCV and 7-fold classification over the others (3% in both cases).

3.2. SVM

G. Fung and J. Stoeckel (2007) [46] think that when employing a classifier-based method, a comparable location inside the volume frame of reference across many volumes corresponds to a same anatomical position. This allows meaningful voxel-by-voxel comparisons of images. On the other hand, unprocessed pictures do not satisfy these criteria. They lacked a comprehensive knowledge of their patients' anatomy since their application contained just HMPAO-SPECT pictures of the themes. The term "functional imagery" refers to this. They are used to demonstrate just the subject's localized blood flow. Regional cerebral blood flow does provide some fundamental anatomical information, but only because blood flow and underlying anatomy are inherently linked. They added all volumes together and then calculated a mean volume to get a more accurate answer. Initially, the mean volume was projected onto the mid-sagittal plane using a reversed registration. The mean was then reversed to make it symmetrical. Following that, all volumes were brought into sync with the current volume.

The experiment produced data for 90 training instances and 33 testing cases selected at random from different universities. They achieved a 90.9% CSVM and an 88.7 % FLD Sensitivity Specificity throughout testing. After the second trial, the FLD Sensitivity Specificity rose to 100%, while the CSVM increased to 93.0 percent.

C. Sandeep and L. Patnaik (2006) [47] used wavelet-coded images as inputs to build SVMs in MATLAB 7.1. This is often a two-dimensional classification technique. They approach categorization of MRI brain photos in this paper as a problem of binary pattern classification. They employ a classifier to determine if a wavelet-coded magnetic resonance image is normal or abnormal.

The wavelet coefficients were computed for 52 MR images of the brain, each having a resolution of 256 256 pixels. A brain magnetic resonance image processed with the DAUB4 level-1 wavelet delivers 17161 wavelet approximation coefficients, whereas secondary and tertiary level wavelets provide 4761 and 1444 coefficients, respectively. The third stage of wavelet decomposition substantially decreases the size of the input vector while maintaining a high classification rate. The resulting vector (17,161) is too big to use as the input. They found that level-2 features are the simplest and most appropriate for self-organizing maps and SVMs in neural networks, while primary and tertiary level features provide weaker accuracy.

López et al. (2009) [48] demonstrates the use of a computer-aided diagnostic (CAD) tool. This was a multidimensional approach that overcomes the restriction of sample size by doing Principal Component Analysis (PCA) on the feature vector. As a result, they successfully reduce coaching sample sizes. While PCA has been used qualitatively throughout history, the coefficients have never been used as classification features. Following the PCA transformation, the resultant feature vectors are utilized for building a Bayesian classifier capable of classifying new instances of image using the posterior knowledge.This technique achieves an accuracy of 91.21% for SPECT images and 98.33% for PET images, improving upon their previous work.

Ramrez et al. (2009) [49] use SPECT images to aid the first diagnosis of AD related dementia in conjunction with a CAD system. The proposed approach, which combines SVM principles and advanced feature extraction algorithms, aims to reduce physicians' subjectivity while interpreting SPECT images to improve early AD diagnosis accuracy.

The proposed features outperform the recently developed VAF technique, which prefers linear SVM over quadratic, RBF, and polynomial kernels because to the input space's high dimension. With these and other changes, the proposed method obtained a 90.38% accuracy rate for the first diagnosis of ATD.

As part of the BLSA neuroimaging supporting study, Fan et al. (2008) [50] gathered annual MRI images from aging elderly adults (Resnick et al., 2000). The BLSA neuroimaging project, which began in 1994, is a prospective study dealing with the structure and workings of the brain in elderly population. It investigates the morphological, functional, and cognitive changes that people go through as they age, as well as when they suffer cognitive impairment. At the time of first participation, all participants had zero symptoms of dementia and were free from other neurodegenerative illnesses, as well as serious disorders and metastatic cancer (described in Resnick et al. 2000). A mental state assessment using the (BIMC) test, as well as a full neuropsychological screening, are included in each annual appointment. Professional examiners gave the CDR (Morris et al., 1989) scale to nearly half of the BLSA autopsy participants each year, as well as those who scored three or more BIMC mistakes.

The analytical technique presented in this article has shown the method of differentiating between disease-specific and natural aging atrophy during a T1 weighted structural MRI scan. Klöppel et al. (2008) [51] utilized linear SVMs to localize critical voxels for classifying scans into two groups. The para-hippocampal gyrus and parietal cortex included the voxels with the greatest confidence in a complete brain SVM classification of AD from controlled groups.

With both group data sets being joined, 95.6 % of trials correctly assigned patients to the acceptable group when the leave-one-out method and complete brain MRI were used (with 94.1% specificity and 97.1% sensitivity). After that, as the first group is utilized for instruction and second group for testing, the percentage of patients labeled to the approved group was 96.4% with full 100% sensitivity and 92.9% specificity. However, as the second group was used for teaching and first group for checking, the percentage of patients correctly labeled to the authorized group was 87.5% with 95.0% sensitivity and 80.0% specificity.

The technique described by Mesrob et al. (2008) [52] is as follows: "Individual magnetic resonance scans are divided into anatomical ROI using a labeled registration template. Along with the traditional parcellation given by Automated Anatomical Labeling (AAL), they provide a refined parcellation that results in a greater degree of anatomical information."

They devised an approach for feature selection that was compatible with the SVM-RFE technique. Though the regions were selected based on data rather than prior knowledge, they included bodies such as para-hippocampal gyrus, the hippocampus, precuneus, and therefore the temporal lobes, which are expected to undergo early changes in the illness. On Cohort 1, feature selection improved categorization accuracy (98.9% rather than 76.5%).

Magnin et al. (2009) [53] used SVM classification of complete brain MRI to create and assess a novel automated approach for differentiating between (AD) patients and normal individuals. To categorize the themes, they utilized SVM and to ensure the results were reliable, they used statistical approaches like bootstrap resampling. They were able to classify AD and healthy individuals with an accuracy of 94.5 % on average with 91.5% average sensitivity and 96.6% average specificity.

R. Chaves et al. (2009) 0 utilizes a CAD method for initial diagnosing of AD. The method employs a statistical learning theory classifier and feature selection from SPECT images. The NMSE features created from cubic blocks inside the temporoparietal region attain a top accuracy of 98.3% when an almost linear SVM is constructed over the twenty most prominent features retrieved. This new method outperforms established methods for early-stage AD diagnosis.

Plant et al. (2010) [55] utilized a novel DMF and combined it with SVM, BS, and VFI to construct a pattern matching quantitative index for predicting MCI to AD conversion. A total of 32 individuals with AD, 24 people with MCI and 18 normal healthy individuals had their brains scanned. Nine out of twenty-four MCI patients transitioned to AD after an average of 2.5 years of follow-up. Using feature selection techniques, the parts of the brain with the maximum accuracy for distinguishing between AD and healthy individuals were discovered, with up to 92% classification accuracy. The AD clusters found were utilized to look for regions of the brain linked to AD conversion in MCI patients.

Segovia et al. (2010) [56] use GMMs to provide a new method for automatically identifying ROIs in functional brain imaging. This method overcomes the limited sample size issue in categorizing functional brain pictures for AD diagnosis.

The LOOCV technique is implemented for validating the supervised CAD system's results on SPECT and PET image databases reaching an accuracy score of 96.67%.

Padill et al. (2010) [57] suggested a new CAD approach for early AD diagnosis enabled by the analysis of SPECT images using the NMF technique. The preprocessing methods have been performed to decrease input file dimensionality for tackle the "curse of dimensionality" issue.

The proposed NMF + SVM approach achieves a classification accuracy of up to 94% and a respectable sensitivity and specificity score (both above 90%), making it a suitable way for categorizing SPECT images. To be comprehensive, a comparison of the proposed approach to a lately released PCA + SVM methodology is provided. The NMF + SVM technique beats the standard PCA + SVM approach and the traditional VAF + SVM approach.

Abdulkadir et al. (2011) [58] examined the effect of categorizing information through hardware acquisition enhances performance of SVM. Around 518 MRI sessions were used from 226 normal individuals and 191 people with a probable AD from the ADNI study.

Using a training set of all 417 people, they achieved the best accuracy of 87%. Also, models trained with 95 patients in each diagnostic class and purchased utilizing a variety of scanner settings obtained an 84.22% detection accuracy when tested on a similar-sized independent set.

Illán et al. (2011) [59] use two distinct approaches for SVM ensemble aggregation. These approaches were chosen based on the use of feature selection. These models are implemented using the pasting votes technique, which classify people according to their vote summation. Each component voted once on a particular issue, resulting in a set of Boolean choices on that topic. They offer two approaches for combining these votes: (a) majority voting, which involves every part contributing to the ending choice, and (b) relevance voting, which involves only a subset of components. For SPECT pictures, this approach resulted in 96.91% accuracy.

Cuingnet et al. (2011) [60] tested ten different techniques on 509 ADNI patients to judge the reliability of the models. The classification tests were MCIc vs MCInc, CN vs AD and CN vs MCIc. For training and fine tuning, 81 CN, 67 MCInc, 39 MCIc, and 69 AD patients were employed. To give an neutral evaluation of the approaches' capability, the rest of the free samples of 67 MCInc, 37 MCIc, 68 AD and 81 CN samples were used. Whole-brain approaches yielded great accuracies when comparing AD to CN (with sensitivity and specificity reaching to 81% and 95% scores respectively).

3.3. Neural Networks

Huang et al. (2008) [61] treated 10 patients with probable AD and 12 normal individuals of comparable age. Six men and four women aged 65.46.6 years were included in the research, with an average of 11.24.1 years of education. The average score on the miniature mental state examination (MMSE) was 17.65.4[10]. Clinical diagnosis was made using NINCDS/ADRDA criteria. Six people with AD had severe dementia, whereas four had moderate dementia. Each subject had neurological and psychological examinations, as well as MRI scanning to rule out the possibility of other types of dementia. The twelve healthy controls were aged 62.17.2 years, had completed 10.74.7 years of education, and had an MMSE score of 28.11.8. This study shown that a combination of VBM and ANN may be used to establish the first diagnosis of AD. The new method is more objective, repeatable, and automated than the previous one. It is very helpful for medicinal purposes.

Savio et al. (2009) [62] utilized 4 distinct ANN models to classify individuals having mild AD against healthy individuals. These models were: BP, RBF, LVQ, and PNN. For gathering characteristics from volume information of the brain, VBM detection clusters are utilized.

They assessed the performance of classifiers built using a variety of various training and architecture approaches using a tenfold cross-validation methodology. After assessing several designs for the VBM's SPM, they discovered that the basic GLM structure with zero covariates is able to recognize modest differences between AD patients and healthy individuals, resulting to the development of ANN classifiers with 83% average exclusionary accuracy. Given the database's size, an accuracy rate of 83 percent is very remarkable.

Ahmadlou et. al. (2010) [63] presents a novel chaos–wavelet approach for an EEG based detection of AD. They do so by using a recently found concept called visibility graph (VG). This technique works on the principle that “nonlinear characteristics may not disclose major alterations between AD and healthy individuals across the band-limited EEG, but can reveal substantial differences in certain sub-bands,” Thus, the wavelet decomposed VGs and sub-bands of the EEGs are used to calculate the intricacy of the EEGs. Furthermore, a binary-stage classifier was used to achieve an accuracy of 97.7% utilizing ANOVA-selected features.

Lopes et. al. (2010) [64] presented a pattern recognition technique that proved to be successful in recognizing patterns of waves that had similarities with patterns recorded in the data-collection, facilitating for a qualitative and quantitative analysis of study of EEG data using PANN. The proposed method has an 82% sensitivity score, but a 61% specificity score.

X. Long and C. Wyatt (2010) [65] examined magnetic resonance images from the OASIS database. The OASIS data collection is the result of a cross-sectional study of 416 people aging between 18 and 96 years. The substantia alba distance correctly assigns 94.67 percent of people to the approved category. When the grey matter distance is utilized, an extra improvement of 97.33 percent is attained, which is superior than the MDS and SVM findings.

In this study, D. Zhang and D. Shen (2011) [66] explore the possibility of using MCI patients to assist in the categorization of AD from cognitive normal individuals using multiple classifier imaging and CSF biomarkers. Protein categorization, severe prostate cancer diagnosis, and skull stripping have been effectively accomplished using modeling methods that are partially supervised. MLapRLS and MRLs have AUC values of 98.5% and 94.6% respectively. These findings demonstrate the utility of mLapRLS in strengthening AD classification through the utilization of extra data.

Qunitana et al. (2012) [67] recruited a whole sample of 522 participants in the Neurinoma trial. They were divided into three groups: 346 healthy seniors, 79 MCI patients, and 97 AD patients. The proportion of people properly categorized was very high (66.67 %). ANNs were used to predict diagnosis in groups of healthy individuals, AD and MCI patients. The authors presented that their method outperformed LDA in terms of classification power and classifier sensitivity. These results support previous studies demonstrating that ANNs are a flexible tool capable of disentangling nonlinear relationships, such as those between cognition and aging.

Mahanand et al. (2012) [68] suggested a totally novel method for identifying brain areas associated with AD using MRI imaging. The method includes the newly established Self-adaptive Resource Allocation Network (SRAN) for classifying AD using voxel-based morphometric characteristics extracted from magnetic resonance images. The SRAN classifier is a sequential learning algorithm that utilizes self-adaptive thresholds to choose appropriate training samples and reject redundant data in order to avoid over-training. As a result, the SRAN classifier is used in ICGA to pick different sets of characteristics. The ICGA–SRAN classifier is used to choose 10, 20, 30, 45 characteristics from the total of 5788. The research demonstrates unequivocally that effective identification of AD may often be accomplished with just ten ICGA-selected features and the SRAN classifier.

Chyzhyk et. al. (2014) [69] presents a groundbreaking, ELM based, wrapper feature selection method for the basic classifier. The evolutionary wrapper feature selection method is composed of a Genetic Algorithm (GA) that iteratively explores possible feature combinations. The average accuracy of a cross-validation assessment of every feature selection is used to generate the GA fitness function. The saliency of a feature is determined by its classification accuracy marginal distribution. The results of 30 rounds of cross-validation utilizing these features for tenfold classification. The simplest outputs are generated using just five hidden units and six features, validating the method's selection of high-quality features.

R. Mahmood and B. Ghimire (2013) [70] implemented the suggested improved automated detection of AD technique completely in software. The program starts by reading the training data from the OASIS dataset. This data set includes a cross-sectional sample of 457 individuals aging between 18 and 96 years, some of whom had early-stage Alzheimer's disease (AD). Each participant receives three or four distinct T1-weighted MRI pictures attained during a single imaging session. Each theme is right-handed and features men and women. One hundred of the study's participants over the age of 60 were diagnosed with very moderate to mild AD.

Al-Naami et al. (2013) [71] devised a fusion method for differentiating between conventional and AD MRI imaging. This comprehensive method examines about 27 MRI samples obtained from Jordanian hospitals. Low pass morphological filters are used to route the recovered statistical outputs away from the intensity histogram and toward the descriptive box plot. Additionally, an artificial neural network (ANN) is utilized to assess this method's efficacy. Finally, the t-test results were compared to the ANN's classification accuracy (100%). The suggested technique's robustness is often cited as an effective method of diagnosing and identifying the kind of AD picture.

Jie et. al. (2014) [72] evaluates their method utilizing functional connection networks from 12 individuals with MCI and 25 individuals with NC. According to the experimental results, their suggested technique reaches 91.9% classification accuracy, 100.0% sensitivity, 94.0% balanced accuracy, and 0.94 neighborhood under the receiver operating characteristic function, indicating that it has significant potential for MCI classification and supported connectivity networks. Additional connection study shows that the linkage of specific areas of the brain differs between MCI patients' healthy individuals, suggesting that MCI individuals suffer from lower functional linkage than healthy subjects, which is consistent with earlier findings.

Ortiz et. al. (2014) [73] validated its proposed technique by segmenting 818 pictures from the ADNI using Statistical Parametric Mapping (SPM). The authors collaborated with ADNI and used the suggested technique on partitioning ROIs that are linked with AD. Additionally, since this technique compresses the discriminative information stored in the brain, it is often utilized to find out a smaller collection of distinct characteristics for classification. The suggested approach obtains a classification accuracy of up to 90% when categorizing CN and AD patients, and a classification accuracy of 84 percent when identifying MCI and AD patients.

3.4. Deep Learning and Ensemble Methods

Wang et al. (2007) [74] looked into the aberrant functional connectivity of the whole brains of early Alzheimer's patients. They then examined the universal ordination of these anomalies using the resting-state fMRI technique. In this study, 18 Alzheimer's patients and 26 cognitively normal people took part. The authors took a linear regression technique to their research. With an accuracy of 81 percent, their method was the most accurate.

Fan et al. (2008) [75] employed MRI and PET scanned pictures to demonstrate that the combined examination of MRI and PET images yield superior classification accuracy results than either method alone. They extracted features using voxel-based morphometry. Using LOOCV, they succeeded in achieving a classification accuracy rate of up to 100%. Using only MRI and the same information, they were able to reach a 93% accuracy rate. The accuracy obtained only from PET scans was extremely low.

Davatzikos et al. (2008) [76] used a multi-dimensional pattern classification of a dataset containing MRI images to diagnose individual AD and Frontotemporal Dementia (FTD) patients. They used MRI scans from 37 Alzheimer's patients and 12 Frontotemporal Dementia individuals (FTD). The MRI scans were analyzed using high-dimensional pattern classification and voxel-based analysis in this investigation. They were able to differentiate among AD from FTD scoring a mean accuracy rate of 84.3%.

Hinrichs et al. (2009) [77] proposed a new Alzheimer's disease predictive classification framework that blends Linear Program Boosting with novel spatial smoothness regularization. The algorithm was implemented in MATLAB. This study relied on ADNI data. This collection includes MR scans, FDG-PET images, and data containing cognitive biomarkers and neuropsychological biomarkers. They used grey matter probability maps (GMPS) and white matter probability maps (WMPs) for classification and training. When applying GMPs on MR image data, the classification accuracy was 82 percent, the sensitivity was 85 percent, and the specificity was 80 percent. When applied to FDG-PET images with 'Y' spatial augmentation, the proposed method achieved 84% classification accuracy, 82% specificity, and 84% sensitivity.

Using SPECT and PET images, López et al. (2009) [78] demonstrated a CAD method to detect AD in initial states based on Bayesian categorization. Instead of voxel-as-features (VAF), they employed the PCA method that helped minimize the feature space dimension. They acquired 88.6% accuracy for SPECT and 98.3% accuracy for PET pictures whereas VAF produced 78.5% and 96.7% accuracy. Their method outperformed the referenced VAF method.

Horn et al. (2009) [79] used perfusion SPECT images to test an automatic technique for determining the likelihood of a diagnosis among AD and FTD. They employed perfusion SPECT scans from 91 FDT patients and 82 AD patients in their study. To make diagnoses, they used a variety of linear and non-linear classification approaches. K-NN+PLS and SVM performed the best among these approaches. However, the learning outcome for the entire dataset was closer to K-NN+PLS than SVM. The K-NN classifier produced the maximum accuracy for K=42. K-NN + PLS had 88% accuracy score, with sensitivity and specificity of 93% and 85%, respectively.

Desikan et al. (2009) [80] explored whether automated magnetic resonance imaging-based assessments might accurately identify people with moderate cognitive impairment. There were 313 participants in this study. Ninety-seven people were chosen from the OASIS database for the training cohort, while 216 participants were selected from the ADNI database for the validation cohort. For this analysis, they used logistic regression. For both AD and MCI classifiers, they achieved a maximum accuracy of 95%.

Ramirez et al. (2010) [81] used SPECT images classification to demonstrate a CAD system to detect AD at initial stages. A random forest (RF) predictor and a PLS regression model are used in their proposed method. They created a baseline PCA method for comparison. They demonstrated that with increase in tree quantity in forest, the RF classifier's generalization error converges to a limit. PLS had the maximum accuracy of 96.9% (sensitivity = 100%, specificity = 92.7%). This result beat the PCA result, which had a peak accuracy of 88.7% (sensitivity = 94.6%) and a specificity of 80.5%. They demonstrated that the PLS-RF system they presented outperformed several previously available AD CAD systems.

Chen et al. (2011) 0 employed LSN analysis for classification of individuals suffering from AD, amnestic MCI, and CN individuals. The dataset's voxel-wise time series comprised 20 patients with AD, 15 individuals with aMCI, and 20 cognitively normal (CN) people were acquired using resting-state fMRI. Around 116 segments were selected as ROI in the brain. For error estimates, they employed the LOOCV method. They achieved a maximum accuracy of 87% when comparing CN to AD (sensitivity = 85%, specificity = 80%). The highest accuracy was 91% for CN vs. MCI, with 95% sensitivity and 93.90% specificity.

Westman et al. (2011) 0 examined and merged MRI data from the European Union AddNeuroMed initiative with the ADNI. The MRI scans of 295 people with Alzheimer's disease, 444 people with MCI, and 335 healthy people were used in this investigation. To differentiate between AD patients and normal individuals, they employed OPLS to latent structure techniques for the individual and the combined cohorts. They employed 7-fold cross-validation to validate their method. They concluded that multivariate data analysis paired with quantitative structural MRI might reliably distinguish Alzheimer's disease patients from older persons.

To identify AD using structural MRI, Rao et al. (2011) [84] utilized SLR with optional spatial regularization (sMRI). T1 weighted sMRI scans of 75 individuals with AD as well as 65 normal individuals were used in this study. The model strength was tested with the 10-fold cross-validation. The authors also used SLR, SRSLR, PLR, and MLDA. SLR and PLR performed better than the other two techniques. Overall, classification accuracy for SLR and SRSLR was 85.26% (sensitivity 90%, specificity 80%). SRSLR, on the other hand, produced spatially smoother classifiers than SLR.

Liu et al. (2012) [85] developed a local patch-based subspace ensemble technique for more reliable classifying that produces numerous single classifiers based on distinct subsets of local patches and merges them. This approach was put into test using a 652 MRI scan dataset from the ADNI database, which included 198 individuals with AD, 225 people with MCI, and 229 normal controls. The SRC approach was utilized to build each weak classifier. For the final decision, multiple weak classifiers were combined. Their method yielded a maximum accuracy of 90.8% for AD classification, with 86.32% sensitivity and 94.76% specificity. The accuracy for MCI categorization was 87.85%, with 85.12% sensitivity and 90.40% specificity.

CHAPTER 4: METHODOLOGY

This chapter deals with describing the dataset and tools used for this thesis and the algorithms and techniques used for pre-processing, feature extraction and the eventual classification.

4.1 Dataset Handling

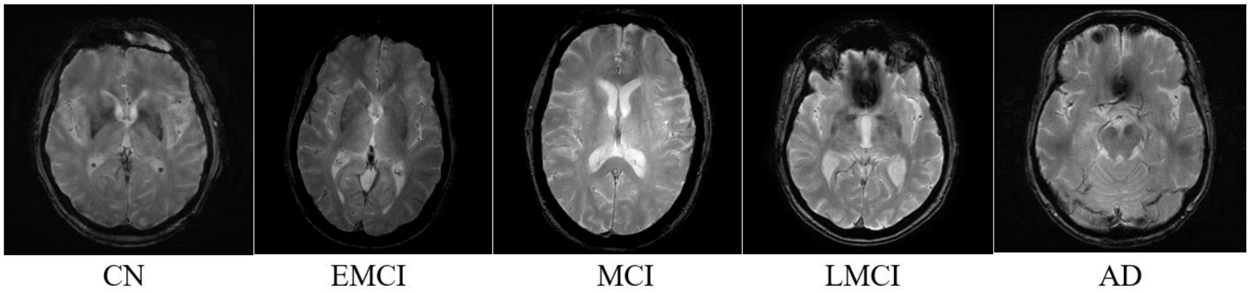


Figure 4.1. Brain MRI coronal plane scans of different phases of AD

The dataset used in this thesis is derived from T1-weighted MRI images from ADNI 1. ADNI 1 is a subset of MRI neuroimaging initiative of ADNI database (website: adni.loni.usc.edu). The original 3d images were converted to 2d images and preprocessed and formatted into JPG images. Then this derived dataset is uploaded to Kaggle for general purpose uses. The images are axial section of the brain of different patients in different stages of AD. The dataset consists of five classes: Cognitive Normal (CN), Early Mild Cognitive Impairment (EMCI), Mild Cognitive Impairment (MCI), Late Mild Cognitive Impairment (LMCI) and AD (Alzheimer's Disease). Each image is 256x256 pixel. In total, there are 1296 images with 171 AD images, 580 CN images, 240 EMCI images, 72 LMCI images and 233 MCI images.

The ADNI images are already preprocessed as working on images straight from MRI machines is very difficult. This is also done by ADNI to ensure less difference between different MRI scanning machines and less preprocessing needed by the researchers to actually work on them. This makes working with such data less stressful.

3D-MRI images are very high dimensional large sized data. Without custom pre-processing the data, working directly on them is impractical. While working on data with high dimensionality, the model will require many data to ensure better distribution of features. Otherwise, the model will struggle to generalize so many features and will perform very poorly. However, since the MRI images are already very large and high dimensional, feeding more data is very inefficient and computationally expensive and not feasible in real life applications. To tackle this issue, we used the derived version of the dataset from Kaggle where the 3d images are already converted to 2d to reduce dimensionality. MRI images of the brain also contains portions such as the skull, skin and space around a patient's skull which are information not needed for the classification purpose. So, the parts are cut down from the final image to decrease the dataset size and overall training time. The data is also downscaled to 128x128 to reduce computational expenses.

4.2 Methods

This chapter discusses the methods we applied to prepare the data and use it for model training, testing and validation. The following figure gives an overview of the entire process. The 2d MRI images go through a basic pre-processing. Then the images are flattened to be used by different algorithms (except CNN). The flattened image vectors are then sent to models for the training process. For CNN, the images aren't flattened. Instead, convolution is applied and the image is then flattened and sent to ANN. As the models get trained, they are used for various classifications.

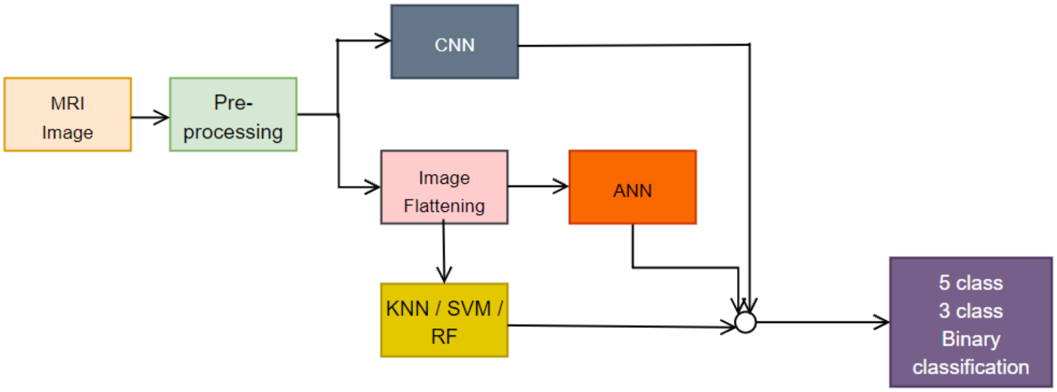


Figure 4.2. Process flow of the methodologies used.

4.2.1 Data Pre-processing

Normalization and Downscaling

The dataset used in this thesis is already normalized as every image is 256x256 pixels. However, to reduce computational expense, the images are converted from 3 channel RGB images to grayscale images. These are further downscaled for increasing model efficiency. Since the data is already normalized, the data is only needed to be smaller to decrease computational cost and increase model performance. Downscaling is performed to decrease dimension of the data and make it more feasible for the machine learning algorithm to train on. In this thesis, to attain this goal, the data was downscaled and resized from 256x256 to 128x128 using OpenCV.

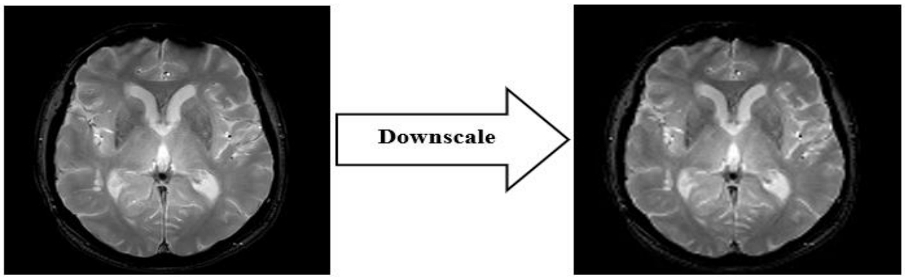


Figure 4.3. Downscaling data from 256x256 to 128x128.

Random Oversampling through Data Augmentation

When the dataset is not balanced i.e., there is a noticeable disparity in the number of instances for each class, the model may learn more about a particular class and ignore or fail to learn important features of another class due to lack of instances. This makes the model more biased towards the majority classes. To prevent this, random oversampling can be used by using random rotation as a data augmentation technique. In this thesis, since the dataset is not balanced, random oversampling has been used in the minority classes to balance the data. Here, the minority class images have been rotated and then oversampled to increase the number of instances of the minority classes and match with the majority class.

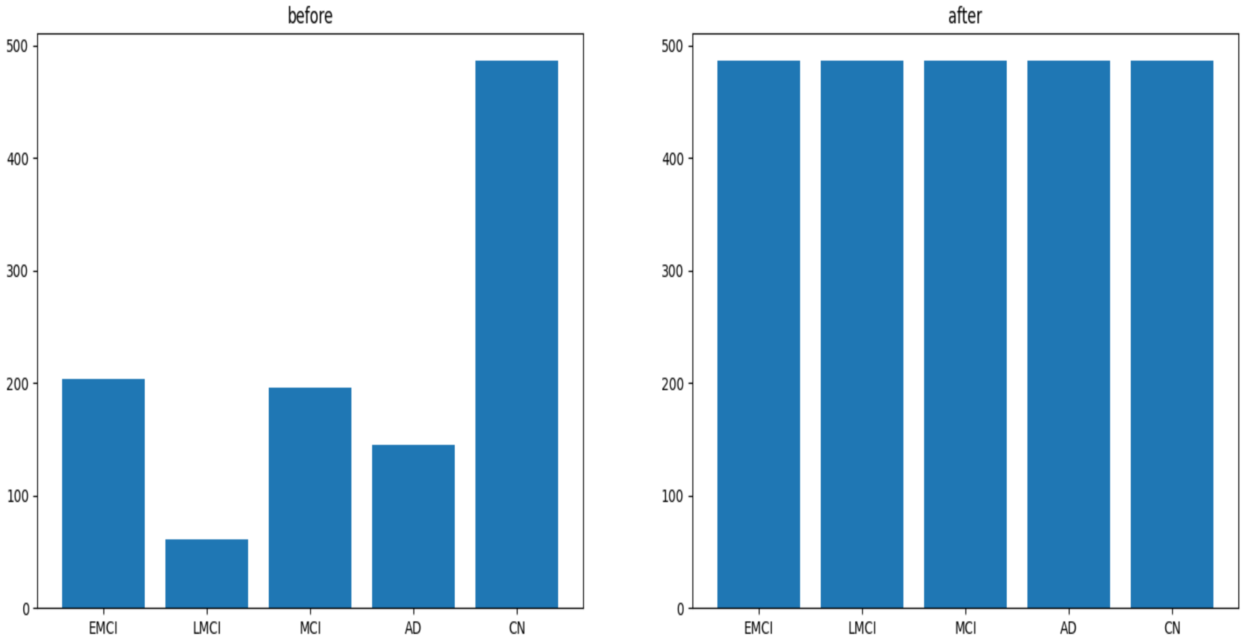


Figure 4.4. Data distribution before and after oversampling

Smoothening

Smoothening is used to reduce noise in image, making classification easier for training models. This noise reduction also often creates a blurry effect. For this thesis, gaussian blur has been used to reduce noise from images using a 3x3 sized Gaussian smoothing kernel.

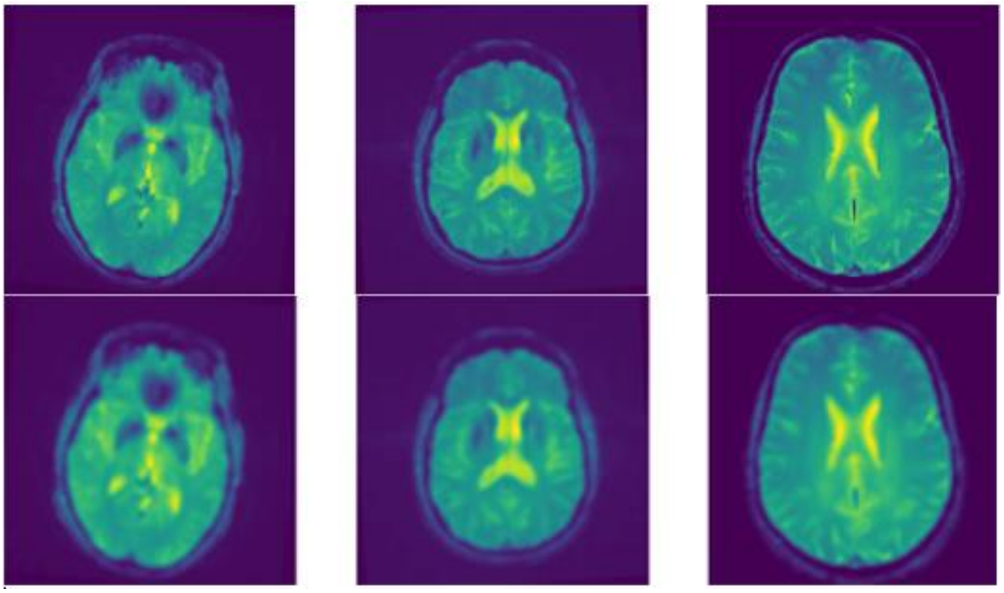


Figure 4.5.Gaussian blur in affect. The first row represents images before applying Gaussian blur and the second row represents images after applying the blur.

Data Splitting

For proper evaluation of the model, the dataset is split into a training and testing set with 80% images in training set, and 20% images in testing set. In the training set, 20% of the images were used in validation set.

4.2.2 Model Training and classification

The processed dataset is used to train 5 different models. These models try to classify the target class from 5 classes to compare the models. Then the classification problem for the CNN is split into 5-class, 3-class and binary classification problems. The 5-class classification classifies among EMCI, MCI, LMCI, AD and CN. The 3-class classification tries to classify whether a patient is in MCI, AD or CN stage. The binary classifications are CN vs AD, MCI vs AD and CN vs MCI.

The preprocessed images are directly loaded in CNN. For the rest of the architectures, the image has been flattened to convert the matrices into one dimensional vector.

Random Forest

The preprocessed and flattened images are used in random forest for training the model using the implementation in scikit-learn library in python 3.8. While for most parameters, the default values are used, the number of trees used is set to 150 while the depth of the tree is set to 25. This resulted in an overall satisfactory result.

Support Vector Machine

Here, the flattened images are trained using the linear SVM implementation in scikit-learn library. Some hyper parameters have their default values while others have been slightly tuned. Since we don't know whether the data distribution will fit into center, fit interval is set to true. It is assumed that the number of examples is more than the number of features and so the dual value is set to False. By setting C=0.5, moderate regularization has been used as high regularization needs very high number of iterations which significantly increases training time and decreases overall model speed for very negligible accuracy gain. The classification used is one vs rest. Since data balancing is performed in pre-processing, the class weights are set to balanced. The maximum number of iterations is set to 1600 which results model convergence.

KNN

The flattened images are loaded into the k-nearest neighbor classifier implemented in scikit-learn library. Again, some hyperparameter tuning have been used to better fit the model. The K value is set to 7 as higher values provide poor accuracy. For measuring the distances, Manhattan distance is used. The weights are evaluated with respect to their distance from the instance i.e., nearest points weight more in deciding the class. The leaf size for the resulting tree is set to 35.

Neural Network

Artificial Neural Networks and Convolutional Neural Networks have been used to solve the classification problem. While the preprocessed image is flattened for the ANN, such steps are not needed for CNN. The neural network models are implemented using implementations from Keras in TensorFlow.

The ANN model has 5 dense layers with 4 layers using ReLU activation and the final layer uses SoftMax activation. After the second and fourth dense layer, batch-normalization and dropout is used. In model compilation, adam optimizer is used and the categorical cross-entropy is used as the loss function. The flow of operation is shown in figure 4.6.

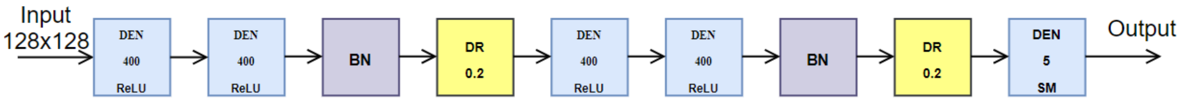


Figure 4.6. Flow diagram of the ANN used in this thesis. DEN=Dense Layer, BN=Batch Normalization, DR=Dropout, SM= SoftMax.

The CNN model has two normal convolutional layers followed by a maxpooling layer. The convolution layers have 16 kernels with kernel size of 5x5 with he_uniform initializer and ReLU activation. Then there are three sets of two separable convolutional layers with 3x3 filter size, same padding and ReLU activation and a batch normalization and maxpooling layer. Then there are two sets of the same layers with dropout layers in between. Then the outputs go to a flatten layer and then goes to three sets of dense layers with

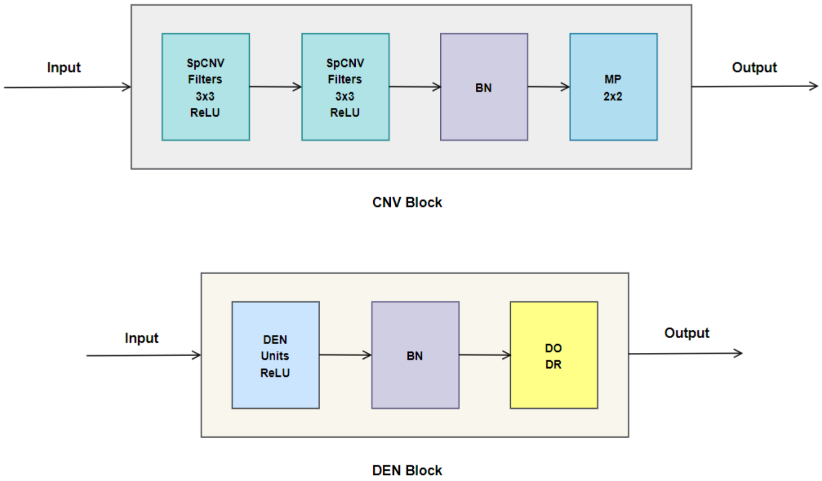


Figure 4.8. The convolution and Dense Block used in CNN. CNV=Convolution Layer, SpCNV=Spatial Sparse Convolution Layer, BN=Batch Normalization, MP=Maxpooling, DEN=Dense Layer, DO=Dropout, DR=Dropout Rate

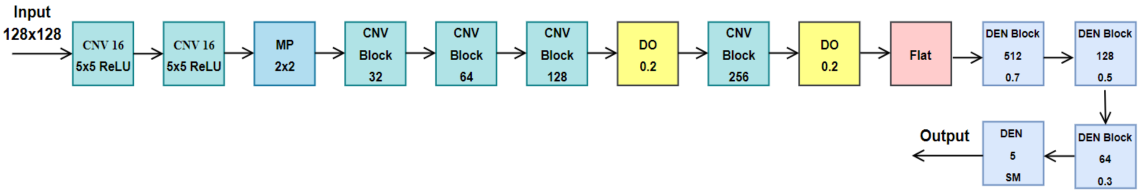


Figure 4.8. Flow Diagram of CNN used in this thesis. CNV=Convolution Layer, MP=Maxpooling, DO=Dropout, SM=SoftMax

ReLU optimizer and with normalization and dropout layers. Then there is a final dense layer with SoftMax activation. The model is then compiled with adam optimizer and categorical cross-entropy loss function.

4.3 Tools and Libraries Used

To implement the necessary solutions to solve the initial problem, we have used tools and libraries to assist us. Various tools and libraries have been used for specific use cases in the implementation section. These libraries and tools help assisting machine learning and deep learning research especially in the subfield of computer vision. These tools also help the development of different methods and increase overall functionality. The ones for this thesis are all written in python as the implementation written for thesis is in python. This chapter will describe those tools and libraries and the reasons for using them in our implementations.

4.3.1 TensorFlow

TensorFlow is a very popular library used for machine learning. It is an open-source system developed by the Google Brain team. It is mainly written in C++ but provides python APIs. This makes it both fast and convenient to use. TensorFlow is mainly used in computational calculations, machine learning and deep learning [42]. The name is given as the system inputs data as tensors or multi-dimensional matrices and flows the input throw various processes which is often associated with a dataflow graph where the nodes resemble a computational operation and the edges are associated with a multidimensional array (tensor).

TensorFlow is heavily used in machine learning applications. It has pre-built architectures of several machine learning and deep learning algorithms such as convolutional neural networks, recurrent neural networks for fields such as computer vison, natural language processing, label generation, classification and prediction etc.

Since TensorFlow offers its tools and APIs in python, it is very easy to learn and implement. And since there are multiple level of abstractions [42], the user can work on the level he feels comfortable with. TensorFlow applications work on multiple OSes as well as on different hardware architectures. It can be run on CPU, GPU, multi-GPU and even in dedicated TPUs. In this thesis, we are using TensorFlow in Jupiter notebook using the ipynb format. It has been used for model training, testing and evaluation, especially models related to neural networks. For example, for convolutional neural networks, function such as conv2d, maxpool2d, dropout etc. has been used to build and regularize the model. Other functions have also been used to validate the model and interpret the outcomes.

4.3.2 Scikit-learn

Scikit-learn is a Python package library used for machine learning. It includes a range of classification, regression, and clustering methods, as well as machine learning approaches for supervised and unsupervised settings, and is based on popular packages such as NumPy as well as SciPy. Because it is centers on delivering machine learning to non-specialists," the program is easier to use compared to some competitors [44]. Scikit-learn provides a good amount of machine learning methods, such as random forest, SVM, logistic regression, KNN, which can be used for solving problems related to classification, regression, clustering, preprocessing dimensionality reduction model selection. It can also be used to perform cross validation, feature selection and extraction.

In this thesis, scikit-learn is primary used to split the dataset into train and test set. It has also been used to call the confusion matrix which is a prime tool in this thesis for quantitative comparison.

4.3.3 Keras

Keras is a high-level API used by TensorFlow 2 which is written in python. It is heavily used in deep learning, particularly for neural networks. Keras is highly embedded with TensorFlow 2, so a lot of the functionalities of TensorFlow is attained through Keras. Its simplicity and modular nature make it popular among general users as well as experts. The goal of the team behind Keras is to make it as simple as possible to decrease the cognitive load of the user, which it did succeed in many contexts. This can be seen in modelbuilding where it provides the key building blocks of neural networks such as model and layer with ease. However, Keras doesn’t have the architecture to perform low level operations on its own and so it utilizes python as one of its key backends.

4.3.4 OpenCV

OpenCV is a cross platform open-source library that is used to tackle computer vison problems [45]. The library and its main interface are written in C++. However, it has bindings in Python, Java and with simply APIs, it is very easy to work on using python and incorporate with other tools such as TensorFlow and PyTorch. OpenCV provides an easy structure to work on for creating computer vision programs. These programs are often created to tackle problems related to image classification, object detection, feature recognition, gesture recognition, motion detection, medical imaging as well as robotics. Since OpenCV can provide a diverse functionality while also being fairly intuitive to use, it is well used by professionals from software developers to computer scientists as well as data scientists and machine learning engineers. While

OpenCV methods are sufficient for basic machine learning project requirements, there are options given to full tinker the methods to work on the hardware architecture level.

OpenCV was developed by Intel to be used as a CPU focused tool [45]. It then emerged as an easy-to-use machine learning tool where most of the complicated coding is already done. This makes it easier for individuals who just started understanding and working on computer vison problems. In this thesis, OpenCV has been heavily used for preprocessing of image. It has been used for reading and processing image, rotating the images at an angle from center, downscaling the images for better computation and smoothening them using Gaussian blur.

4.3.5 Comparison of libraries

	TensorFlow	OpenCV	Scikit-learn
CNN	YES	NO	NO
ANN	YES	YES	YES
KNN	NO	YES	YES
Random Forest	YES	YES	YES
Support Vector Machine	YES	YES	YES

Table 4.1. Existence of different models in various libraries.

	TensorFlow	OpenCV	Scikit-learn
Sigmoid	YES	YES	NO
ReLU	YES	YES	NO
SoftMax	YES	NO	NO
SoftPlus	YES	NO	NO
Linear	YES	NO	NO
tanh	YES	NO	NO

Table 4.2. Support for different activation functions in different libraries.

Libraries such as TensorFlow, OpenCV as well as scikit-learn provides support for various architectures such as CNN, ANN, random forest, SVM, KNN etc. However, every architecture is not implemented in every library. Table 4.1 points out which algorithm is implemented inside which library.

These libraries also support different activation functions. Table 4.2 points out which function is built in into which library.

The libraries also provide various training algorithms to train the model. Again, not every algorithm is supported by every library. Table 4.3 shows the support between various training algorithms and libraries.

	TensorFlow	OpenCV	Scikit-learn
Stochastic Gradient Descent	YES	NO	NO
Batch Gradient Descent	YES	NO	NO
Levenberg-Marquardt	YES	YES	NO
Adam	YES	NO	NO

Table 4.3. Support for training algorithms among different libraries.

Regularization techniques are often needed to deal with overfitting of the models. Certain libraries provide more support for regularization. This support relationship is given in the table 4.4 below”

	TensorFlow	OpenCV	Scikit-learn
L1 Regularization	YES	NO	NO
L2 Regularization	YES	NO	NO
Dropout	YES	NO	NO
Momentum	YES	NO	NO

Table 4.4. Support for regularization techniques among libraries.

CHAPTER 5: RESULTS

This section will describe the results found in the methodology section. The classifications are performed in 5 class data, 3 class data and a set of binary classifications. All results are shown and described in the following sections.

5.1 Confusion Matrix

In supervised learning, confusion matrix is a popular evaluation technique where the prediction classes and the true classes are organized in a matrix structure. In one axis, the predictive labels are placed and in the other axis, the actual labels are placed. Then the blocks are filled with numbers representing how many times the model classified instances correctly and incorrectly.

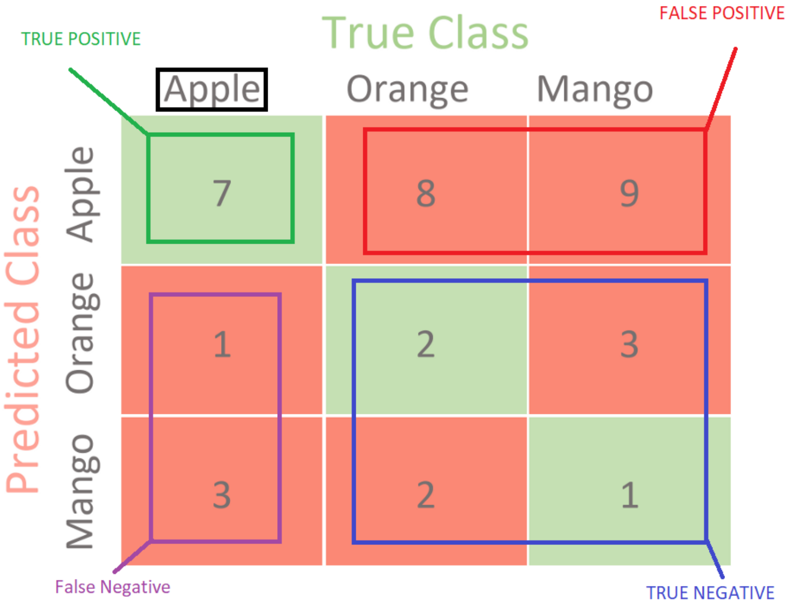


Figure 5.1. Confusion matrix for three classes [100].

Figure 13 represents a 3x3 confusion matrix. Consider a scenario where a classifier needs to classify whether an instance is an apple or an orange or a mango. The classifier has correctly classified 7 apples that is the predicted and the true class is the same. For this example, we are considering apple as the target class. So, for apple, these 7 instances are true positives (TP). The classifier also predicted 1 apple as orange and 3 apples as mango. These are false negatives (FN) as the classifier classified them as not an apple but the classification label is false. The classifier classified 8 oranges and 9 mangoes as apples. These are false positives (FP) because the classifier was positive that these instances were apples but the classification label was false. The remaining blocks hold counts of fruits that are not apples nor they are classified as an apple. So, they are true negatives (TN) as the algorithm classified them as non-apples which is also true.

Using these 4 parameters, we can calculate the accuracy, precision, recall and specificity of the model to better understand and measure its performance. Accuracy is the measure of how correctly the model can classify an instance.

$$Accuracy = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(2.3)}$$

Precision tries to measure how many positive classifications are actually positive. It indicates how precise the model is at making positive predictions. Precision is considered in cases where false positives play an important role.

$$Precision = \frac{\text{TP}}{\text{TP+FP}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(2.4)}$$

Recall measures how many instances of a particular class are correctly classified. That is, recall indicates how many instances of a true class are positively classified. It shows how good the model is at recognizing a particular class. Recall is considered in cases where false negatives pay a key role.

$$Recall = \frac{\text{TP}}{\text{TP+FN}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(2.5)}$$

5.2 Overall Performance

The below table shows the testing accuracy of various algorithms in different class distributions.

Architecture	5-class	3-class (AD vs MCI vs CN)	CN vs AD	MCI vs AD	MCI vs CN
KNN	0.750	0.742	0.868	0.745	0.761
CNN	0.790	0.795	0.882	0.810	0.800
ANN	0.633	0.740	0.823	0.776	0.761
Random Forest	0.747	0.792	0.864	0.765	0.757
SVM	0.617	0.667	0.784	0.686	0.736

Table 5.1. Testing accuracy of the used algorithms.

Here, CNN gives the best testing accuracy for 5 class while 75% accuracy while SVM struggles the most with 61.7% accuracy. This can be attributed to the fact that SVM linear kernel is struggling with multiclass data that may not be linearly separable in higher planes.

For the comparison of AD vs MCI vs CN, the models performed similarly with CNN and RF being the best two scoring 79.5% and 79.2% test accuracy respectively. SVM still struggled with 66.7% test accuracy. Every algorithm except KNN received a slight increase in accuracy scores.

For CN vs AD, we see a high increase in accuracy scores. This can be attributed to the high contrast between AD patients and cognitive normal healthy individuals. Here, the CNN scored the highest accuracy of 88.2% and SVM scored the lowest with 78.4 %. Since KNN and SVM works great in binary classification problems, they received the highest increase in accuracy from 3 class with 12.6% and 11.7% increase respectively.

Since early stages of Alzheimer’s can often be very similar with normal aging of elderly people, there are similarities between CN images and MCI images. For MCI vs CN, CNN scored the highest accuracy with 80% while SVM scored the lowest accuracy with 73.6%.

It is difficult to diagnose what stage of dementia a patient is in as the symptoms of MCI and AD are very similar, only in AD the symptoms are more prominent and other symptoms also occur. So, there are similarities between these images. For MCI vs AD, CNN scored best with 81% accuracy while SVM struggled with an 68.6% accuracy.

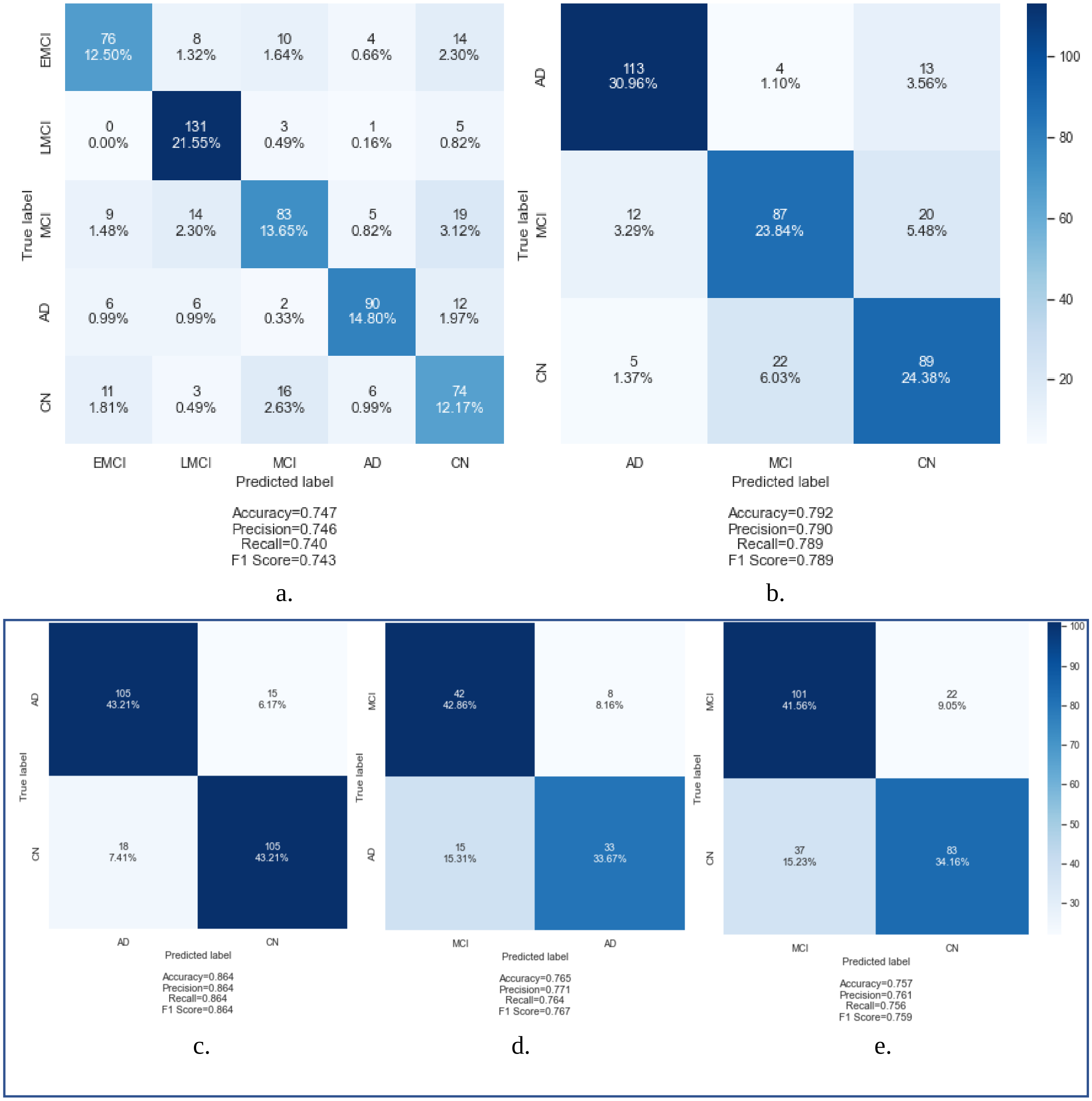
These accuracies of the algorithms are in line with expectations. Classifying the 5 classes are difficult, especially when it includes AD, MCI and CN. Three class classification is a bit easier as there are a smaller number of classes to worry about. AD vs CN will provide the best accuracy as the images are most distinct compared to other class comparisons. While MCI vs AD and MCI vs CN may provide similar accuracy as the differences are not as prominent as AD vs MCI.

If we do not consider the class number as a comparison matric, then just by comparing values, CNN at binary classification gives the best accuracy of 88.2% for AD vs MCI while SVM at 5-class classification gives the lowest accuracy of 61.7% percent.

5.3 Progress of Algorithm performance

This section will describe how each algorithm progressed as number of classes have been decreased.

5.3.1. Random Forest (RF)



a.

b.

c.

d.

e.

Figure 5.2. Confusion Matrices of RF classifier for a. 5-class classification b. AD vs MCI vs CN c. AD vs CN d. MCI vs AD e. MCI vs CN

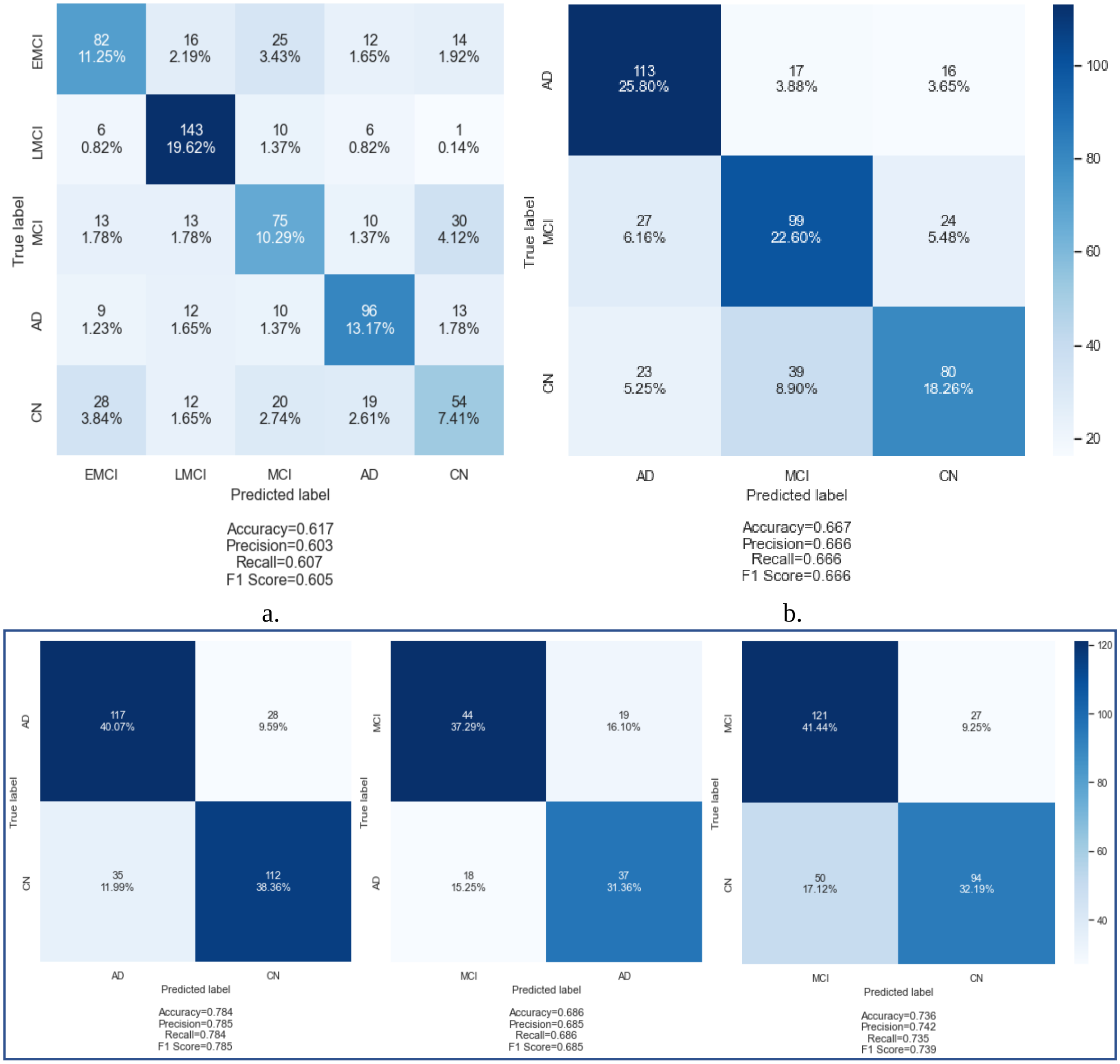
In 5 class, we can see the algorithm particularly struggle at the lower half of the portion, particularly, some instances of CN have been classified as EMCI or MCI. This is expected as CN and MCI types are similar. This is also prevalent in the 3-class classification where some CN instances are labeled as MCI and vice versa. In binary classifications, the algorithm has a tendency of labeling AD and CN labels as MCI. This shows some bias of the model for MCI. Since AD vs CN has no MCI involved, the class performs best fewer false positives and negatives ensuring an accuracy score of 86.4%.

RF	Classes	Accuracy	Precision	Precision Average	Recall	Recall Average
5-class	EMCI	0.747	0.75	0.746	0.68	0.740
	MCI		0.73		0.64	
	LMCI		0.81		0.94	
	AD		0.85		0.78	
	CN		0.60		0.67	
AD vs MCI vs CN	MCI	0.792	0.77	0.790	0.73	0.789
	AD		0.87		0.87	
	CN		0.73		0.77	
AD VS CN	AD	0.864	0.85	0.864	0.88	0.864
	CN		0.88		0.85	
AD VS MCI	AD	0.765	0.80	0.771	0.69	0.764
	MCI		0.74		0.84	
CN VS MCI	CN	0.757	0.79	0.761	0.69	0.756
	MCI		0.73		0.82	

Table 5.2. Classification-wise Matrix values of RF

As we look at the table, we can see the increase of accuracy as class number decreases and the similarity in accuracy when MCI is involved in binary classification. We can also see increase in average precision and recall in the same pattern. The five-class classification has the least precision and recall average while AD vs CN has the highest precision and recall average. However, individual class wise breakdown shows a better picture. In 3 class classification, RF holds the highest precision score of 0.77 for MCI. It also holds the highest recall score of 0.84 for MCI in AD vs MCI. RF holds highest precision and recall in 3 class classification (0.87 in both cases) for AD. It holds the highest precision and recall in AD vs CN (0.88,0.85) for CN. The lowest precision for MCI is 0.73 in 5 class classification and in CN vs MCI. The lowest recall for MCI comes from 5 class classification (0.64).

5.3.2. Support Vector Machine (SVM):



a.

b.

c.

d.

e.

Figure 5.3. Confusion Matrices of SVM classifier for a. 5-class classification b. AD vs MCI vs CN c. AD vs CN d. MCI vs AD e. MCI vs CN

The SVM classifier struggled throughout the classification problems. This can be attributed to the use of linear kernel for classification. In five class, SVM particularly performed worse than any other models. It is particularly noticeable in the CN row where the instances were classified to different classes in significant numbers. The model also struggled in distinguishing between MCI and CN in the same classification. In the three-class classification, while the accuracy did increase, it is still not good with a certain bias for MCI classifying other classes as MCI.As expected, the model performed best with AD vs CN, however the increase in accuracy while not as prominent as other algorithms, is significant.

SVM	Classes	Accuracy	Precision	Precision Average	Recall	Recall Average
5-class	EMCI	0.617	0.59	0.603	0.55	0.607
	MCI		0.54		0.53	
	LMCI		0.73		0.86	
	AD		0.67		0.69	
	CN		0.48		0.41	
AD vs MCI vs CN	MCI	0.667	0.64	0.666	0.66	0.666
	AD		0.69		0.77	
	CN		0.67		0.56	
AD VS CN	AD	0.784	0.77	0.785	0.81	0.784
	CN		0.80		0.76	
AD VS MCI	AD	0.686	0.66	0.685	0.67	0.686
	MCI		0.71		0.70	
CN VS MCI	CN	0.736	0.78	0.742	0.65	0.735
	MCI		0.71		0.82	

Table 5.3. Classification-wise Matrix values of SVM.

In the above table, the increase of accuracy, precision and recall is shown as class number decreases. The values follow the same pattern followed by rf, but the values are less in quantity. Among the binary classes, AD vs MCI shows the less accuracy, average precision and recall values. In 5 class, the precision and recall values are small. While LMCI show a great increase in values, LMCI was the smallest class and due to oversampling, this class received the highest number of sampled data and thus is prone to overfitting. Overall, SVM scored better precision and recall for MCI in CN vs MCI (0.71,0.82). For CN, the best precision and recall are recorded in AD vs CN (.80,76). For AD, best precision and recall are recorded in AD vs CN (0.77,0.81).

5.3.3. K-Nearest Neighbor (KNN):

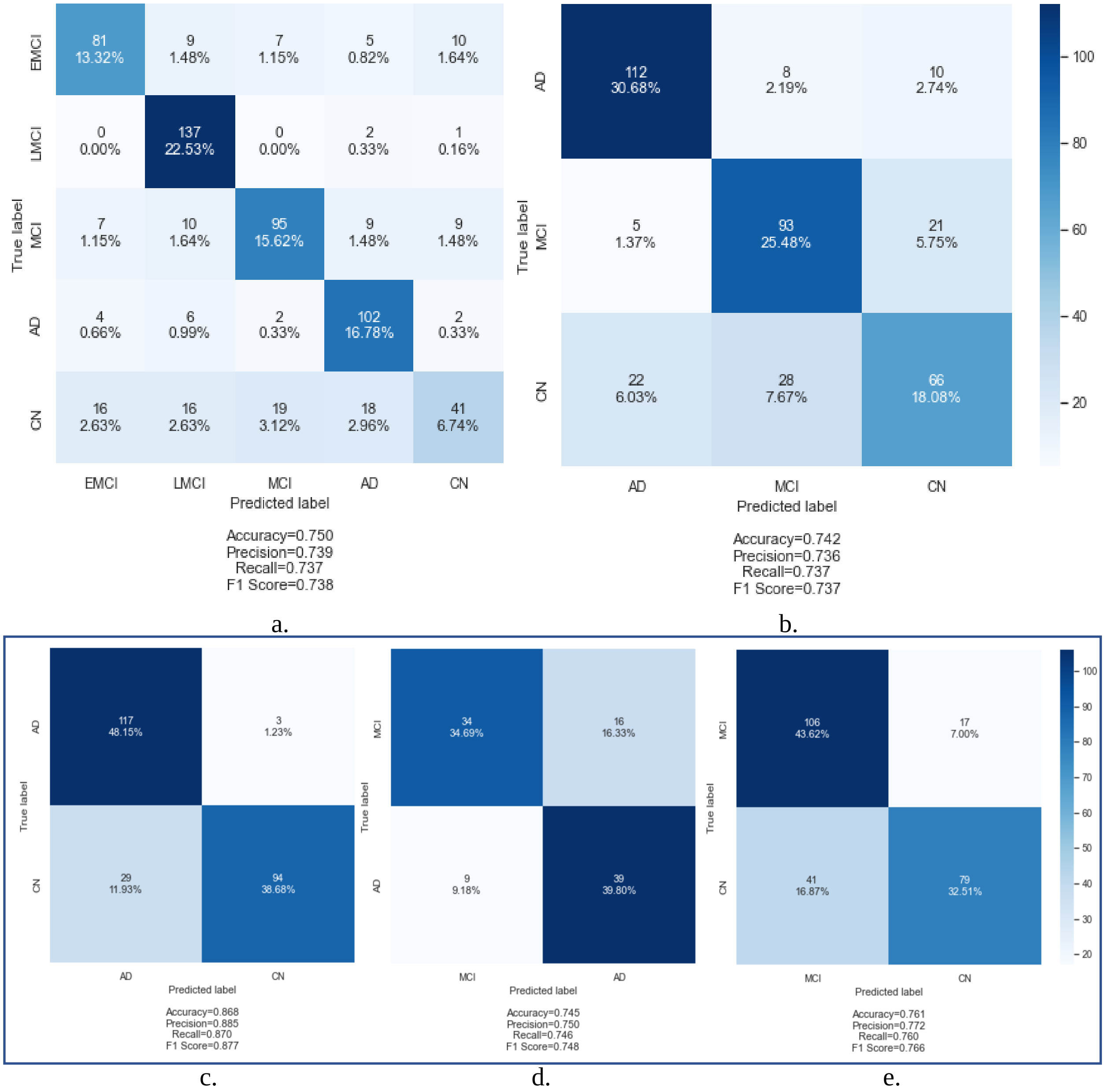


Figure 5.4. Confusion Matrices of KNN classifier for a. 5-class classification b. AD vs MCI vs CN c. AD vs CN d. MCI vs AD e. MCI vs CN

Here, from 5 class to 3 class, the accuracy actually decreased by 1 %. KNN particularly struggled classifying CNN correctly which mainly affected the overall performance of the classifier. In 5 class, the true levels of CN instances are classified into other classes in significant portion. The model also struggled a bit at distinguishing between AD and MCI. The model faced the same issue of distinguishing CN in 3 classes. In binary classification, the results improved, particularly in AD vs CN classification with 86.8% accuracy.

KNN	Classes	Accuracy	Precision	Precision Average	Recall	Recall Average
5-class	EMCI	0.750	0.75	0.739	0.72	0.737
	MCI		0.77		0.73	
	LMCI		0.77		0.98	
	AD		0.75		0.88	
	CN		0.65		0.37	
AD vs MCI vs CN	MCI	0.742	0.72	0.736	0.78	0.737
	AD		0.81		0.86	
	CN		0.68		0.57	
AD VS CN	AD	0.868	0.80	0.885	0.97	0.870
	CN		0.97		0.76	
AD VS MCI	AD	0.745	0.71	0.750	0.81	0.746
	MCI		0.79		0.68	
CN VS MCI	CN	0.761	0.82	0.772	0.66	0.760
	MCI		0.72		0.86	

Table 5.4. Classification-wise Matrix values of KNN.

KNN is the first and only model where the accuracy did not increase from 5 class to 3 class. However, the overall values are in line with other good performing algorithms. It is discussed before that LMCI is highly affected by oversampling but this affect is not so prevalent in KNN data. MCI has the highest precision value in CN vs MCI (0.82) and highest recall value in CN vs MCI (0.86). While AD has the highest precision value in AD vs CN vs MCI (0.81) and highest recall value in AD vs CN (0.97). CN has the highest precision and recall value in AD vs CN (0.97,0.76). KNN’s struggle with CN in 5 class is more apparent here with 0.65 precision value and only 0.37 recall value which are also the lowest values for CN in KNN.

5.3.4. Artificial Neural Network (ANN):

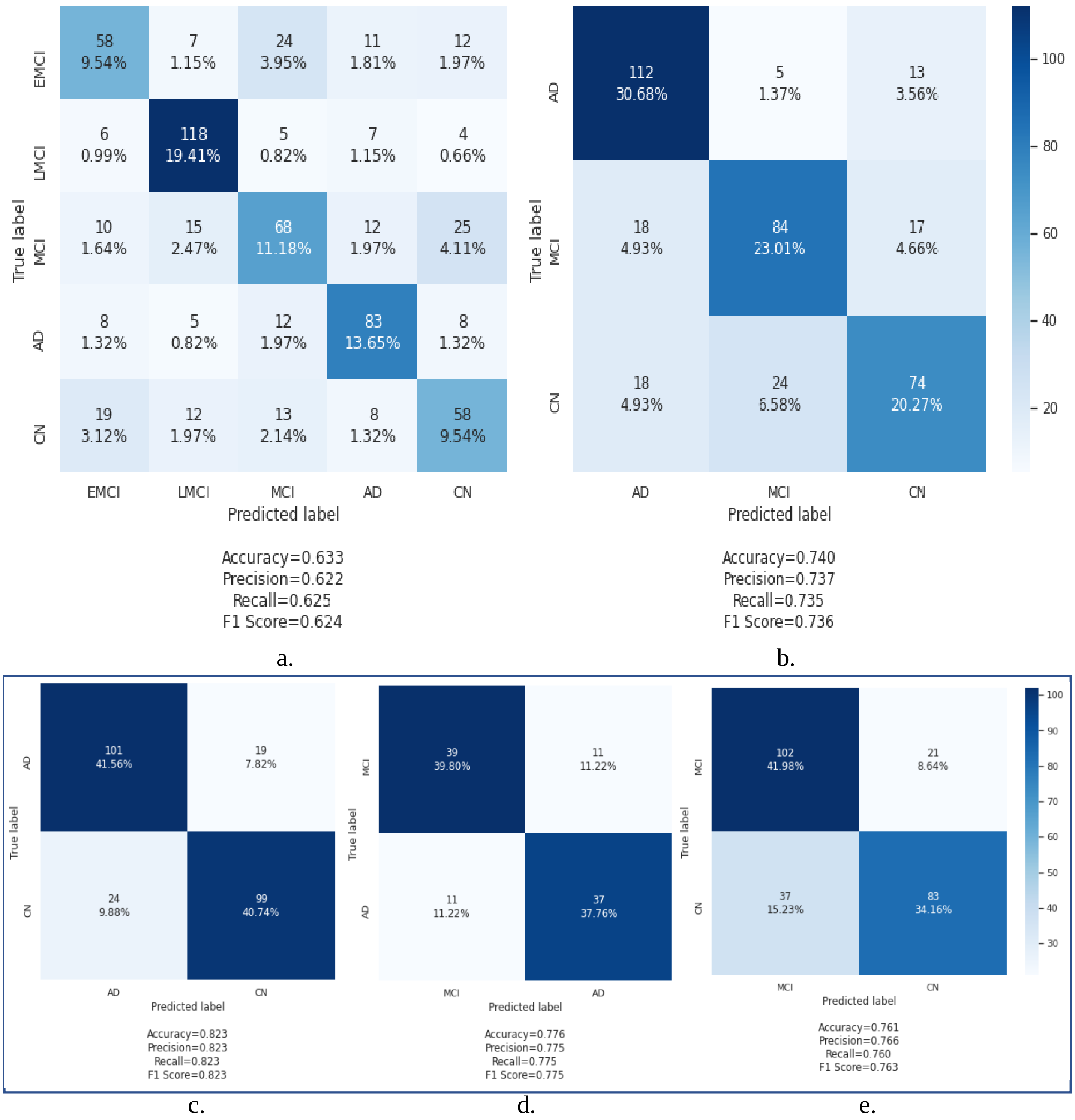


Figure 5.5. Confusion Matrices of ANN classifier for a. 5-class classification b. AD vs MCI vs CN c. AD vs CN d. MCI vs AD e. MCI vs CN

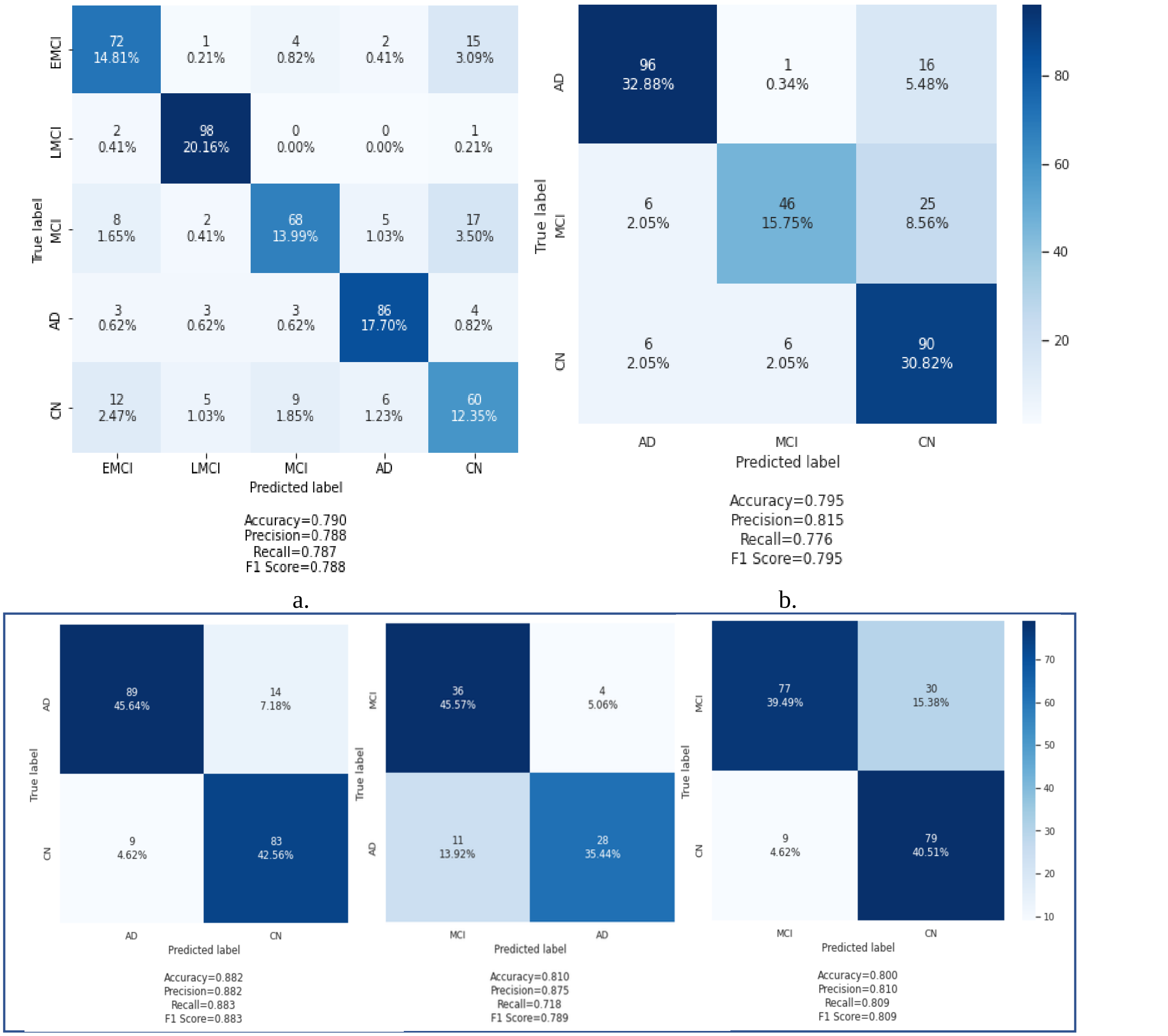
Here, ANN struggled particularly in classifying CN and MCI. The accuracy increased and the misclassification problem lessened when the classifications are binary. In both 5-class and 3-class, ANN classified some instances of CN and MCI as other classes. The accuracy again improved significantly in AD vs CN.ANN did improve a significant number in accuracy from 5 class to 3 class (11%).

ANN	Classes	Accuracy	Precision	Precision Average	Recall	Recall Average
5-class	EMCI	0.633	0.57	0.622	0.52	0.625
	MCI		0.56		0.52	
	LMCI		0.75		0.84	
	AD		0.69		0.72	
	CN		0.54		0.53	
AD vs MCI vs CN	MCI	0.740	0.74	0.737	0.71	0.735
	AD		0.76		0.86	
	CN		0.71		0.64	
AD VS CN	AD	0.823	0.81	0.823	0.84	0.823
	CN		0.84		0.80	
AD VS MCI	AD	0.776	0.77	0.775	0.77	0.775
	MCI		0.78		0.78	
CN VS MCI	CN	0.761	0.80	0.766	0.69	0.760
	MCI		0.73		0.83	

Table 5.5. Classification-wise Matrix values of ANN.

In the table, it is apparent the jump of accuracy, precision and recall from 5 class to 3 class. While LMCI is showing relatively good value compared to other classes in 5 class, it is already discussed that it can be argued that this is do the overfitting and bias due to oversampling of the set which was also the smallest class before oversampling. Here, MCI has the highest precision value in AD vs MCI (0.73) and highest recall value in CN vs MCI (0.83). While AD has the highest precision value in AD vs CN (0.81) and highest recall value in AD vs MCI vs CN (0.86). CN has the highest precision and recall value in AD vs CN (0.84,0.80). The precision value of CN is also the highest precision value of ANN. The highest recall value for ANN is 0.86 for AD in AD vs MCI vs CN.

5.3.5. Convolutional Neural Network (CNN):



a.

b.

c.

d.

e.

Figure 5.6. Confusion Matrices of CNN classifier for a. 5-class classification b. AD vs MCI vs CN c. AD vs CN d. MCI vs AD e. MCI vs CN

The CNN performed the best across all other algorithms. While it did struggle a bit with classifying CN instances, it was not as bad performing as the others. The accuracy in 5 class falls because the algorithm was struggling with CN and MCI which as we discussed before, are similar. In 3 class, CNN did misclassify some instances of MCI and AD as CN, so the increase in accuracy was very small. In binary classifications however, CNN performed better in all classes compared to all other algorithms.

CNN	Classes	Accuracy	Precision	Precision Average	Recall	Recall Average
5-class	EMCI	0.790	0.74	0.788	0.77	0.787
	MCI		0.81		0.68	
	LMCI		0.90		0.97	
	AD		0.87		0.87	
	CN		0.62		0.65	
AD vs MCI vs CN	MCI	0.795	0.87	0.815	0.60	0.776
	AD		0.89		0.85	
	CN		0.69		0.88	
AD VS CN	AD	0.882	0.91	0.882	0.86	0.883
	CN		0.86		0.90	
AD VS MCI	AD	0.810	0.88	0.825	0.72	0.813
	MCI		0.77		0.90	
CN VS MCI	CN	0.800	0.72	0.810	0.90	0.809
	MCI		0.90		0.72	

Table 5.6. Classification-wise Matrix values of CNN.

Here, compared to the other algorithms, the accuracy, precision and recall of every class improved significantly as class number decreased. Here, MCI has the highest precision value in CN vs MCI (0.90) and highest recall value in AD vs MCI (0.90). While AD has the highest precision value in AD vs CN (0.91) and highest recall value in 5-class (0.87).CN has the highest precision and recall value in AD vs CN (0.86,0.90). These values are constantly improving and seems better than the previous algorithms. However, to get a more in depth understanding, a more class-wise comparison is needed to determine which model performed best in which classification problem and had better precision or recall or both values for which particular class. This analysis will be performed in the next section.

5.4 Classification-wise Comparison of Different Algorithms:

5.4.1. 5-class:

Classifier for 5-class	Accuracy	Precision					Recall				
		EMCI	MCI	LMCI	AD	CN	EMCI	MCI	LMCI	AD	CN
RF	0.747	0.75	0.73	0.81	0.85	0.60	0.68	0.64	0.94	0.78	0.67
SVM	0.617	0.59	0.54	0.73	0.67	0.48	0.55	0.53	0.86	0.69	0.41
KNN	0.750	0.75	0.77	0.77	0.75	0.65	0.72	0.73	0.98	0.88	0.37
ANN	0.633	0.57	0.56	0.75	0.69	0.54	0.52	0.52	0.84	0.72	0.53
CNN	0.790	0.74	0.81	0.90	0.87	0.62	0.77	0.68	0.97	0.87	0.65

Table 5.7. Matrix values of 5-class classifiers.

Here, we can see that CNN has the highest accuracy. However, for classifying instances as EMCI, KNN and RF has the best precision value of 0.75 and CNN has the best recall value of 0.77. For classifying MCI, CNN has the highest precision value of 0.81 and KNN has the best recall value of 0.73. For classifying LMCI, CNN has the best precision value of 0.90 while KNN has the best recall value of 0.98. While classifying AD, CNN has the best precision value of 0.87 while KNN has the best recall value of 0.88. To classify CN, KNN has the best precision value of 0.65 while RF has the best recall value of 0.67.

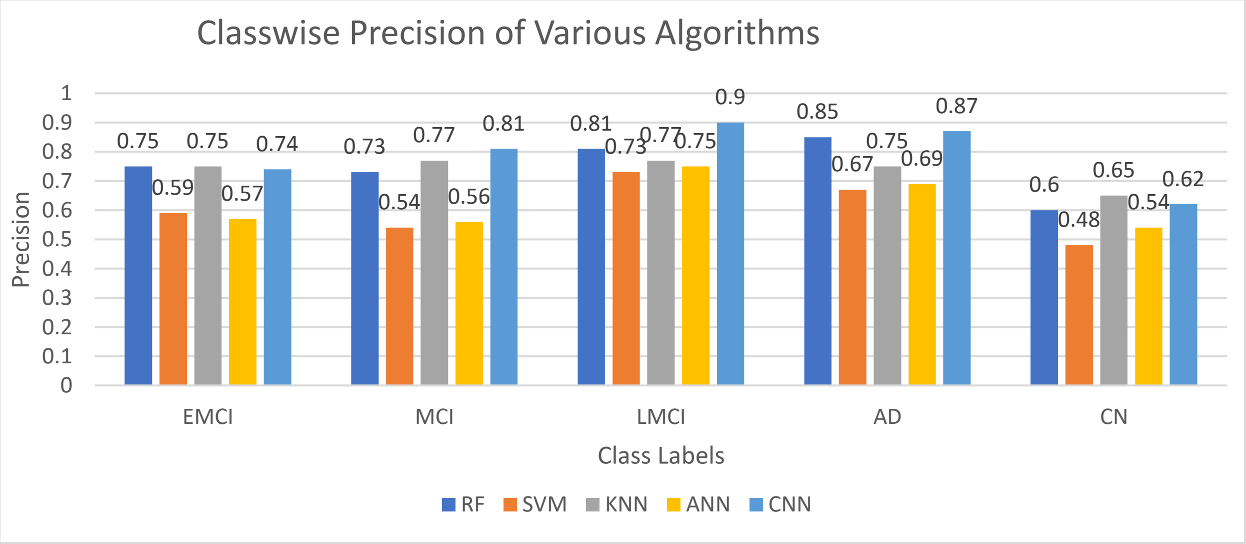


Figure 5.7. Class-wise precision of different algorithms for 5 classes.

So, when the goal is to choose the best algorithm for early detection, CNN is the go-to algorithm as it shows the best accuracy. However, if the goal is to have really small number of false positive for a particular class, then the algorithm with highest precision for that class should be chosen. For example: if a hospital needs patients that are only AD patients who need immediate treatment, then they would look for a model that has high precision i.e., low false positive rate. So, he would choose the CNN model for this job. On the other hand, if some individual is checking whether he is cognitively normal, then he would need a model that has good recall for CN (In our case, its RF). Otherwise, the false negative rate may be high, so he may misdiagnose himself as not CN, and may take medicines that are not at all suitable for his current state. So, it's important to choose the right model based on the need.

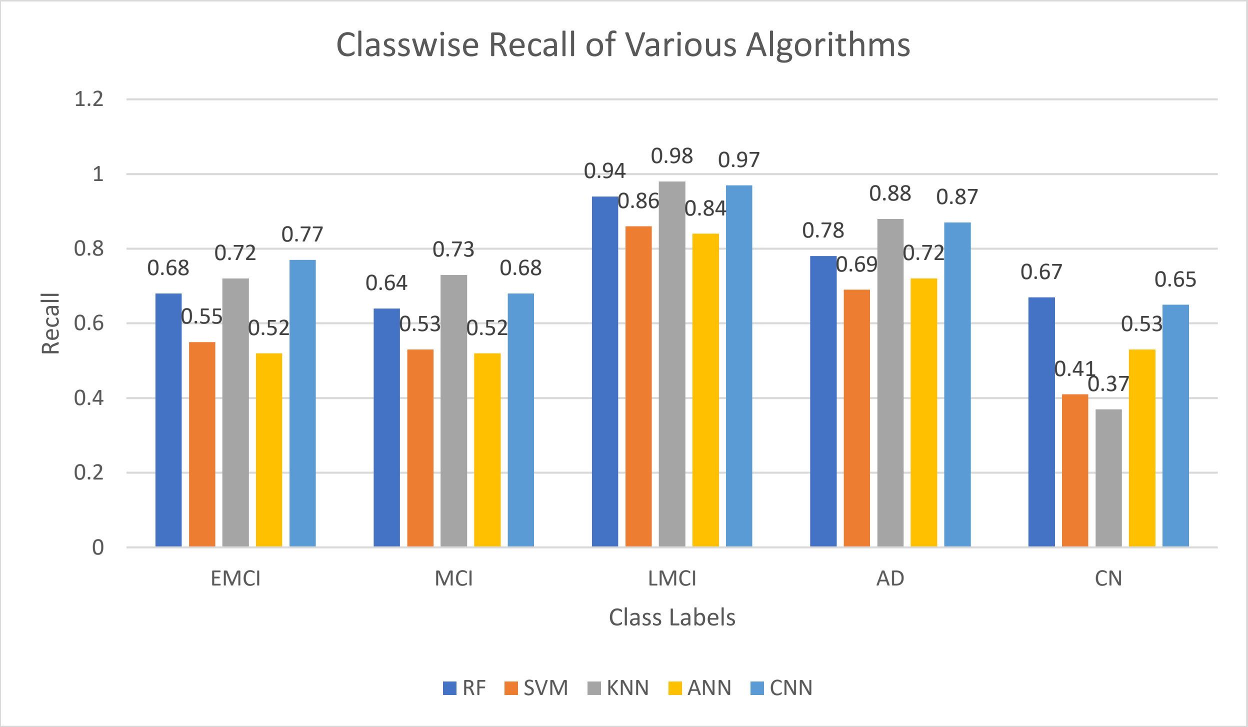


Figure 5.8. Class-wise recall of different algorithms for 5 classes.

5.4.2. 3-class (AD vs MCI vs CN):

Here, again CNN has the highest accuracy. For AD, CNN has the highest precision (0.89) while RF has the highest recall (0.87). So, CNN can be used for scenarios where for AD, the false positive rate is more important than the false negative rate. Similarly, RF can be used in scenarios where the false negative rate is more important than false positive rate of AD. For MCI, CNN has the highest precision (0.87) and KNN has the highest recall (0.78). For CN, RF has the highest precision (0.87) while CNN has the highest recall (0.88).

Classifier for AD vs MCI vs CN	Accuracy	Precision			Recall		
		AD	MCI	CN	AD	MCI	CN
RF	0.792	0.87	0.77	0.73	0.87	0.73	0.77
SVM	0.667	0.69	0.64	0.67	0.77	0.66	0.56
KNN	0.742	0.81	0.72	0.68	0.86	0.78	0.57
ANN	0.740	0.76	0.74	0.71	0.86	0.71	0.64
CNN	0.795	0.89	0.87	0.69	0.85	0.60	0.88

Table 5.8. Matrix values of 3-class (AD vs MCI vs CN) classifiers

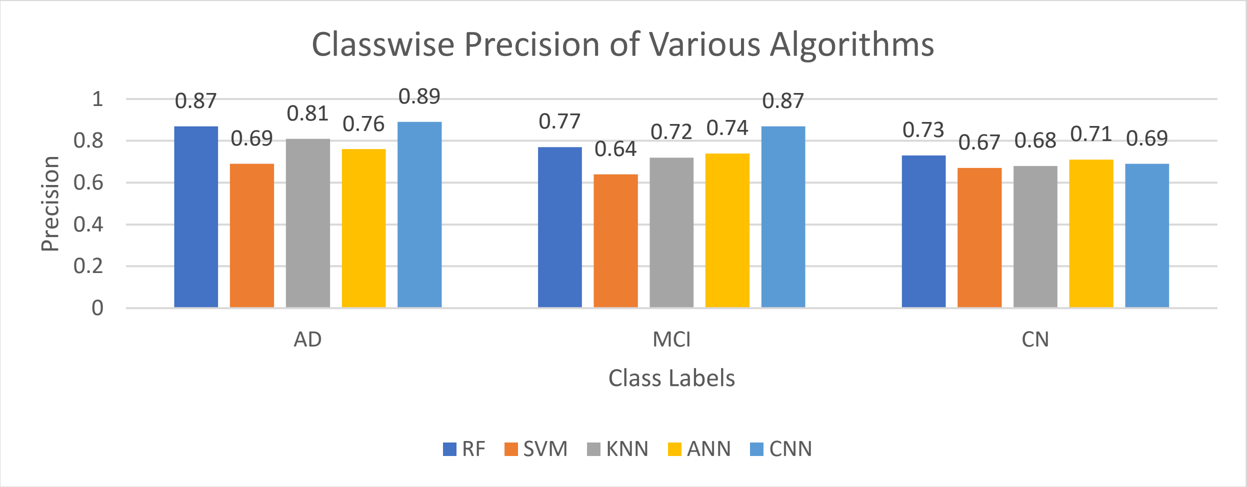


Figure 5.10. Class-wise precision of different algorithms for AD vs MCI vs CN

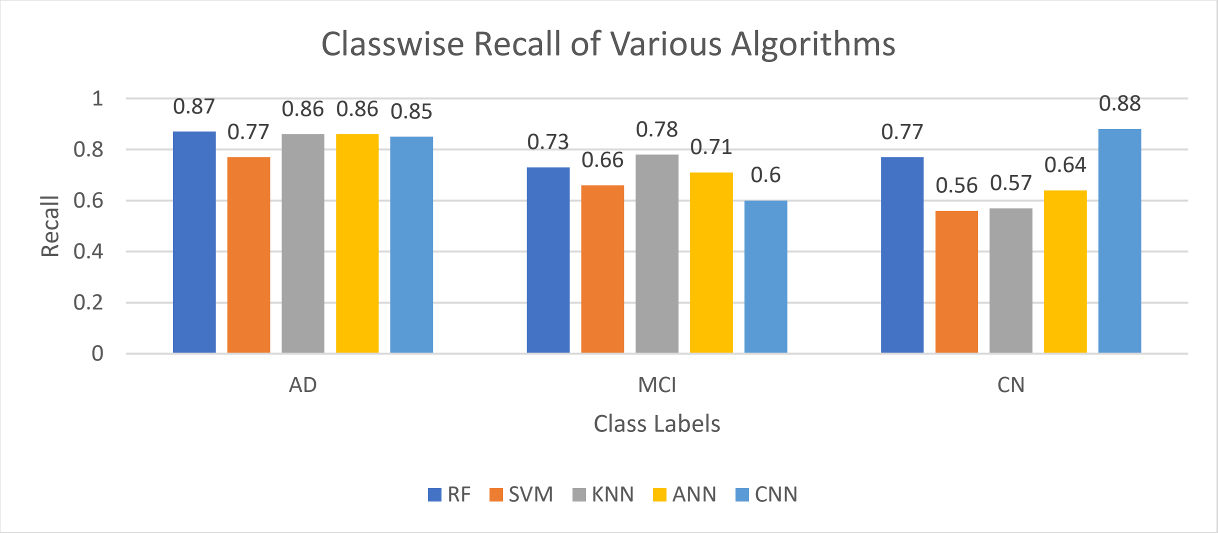


Figure 5.9. Class-wise recall of different algorithms for AD vs MCI vs CN

5.4.3. Binary classification (AD vs CN):

Classifier for AD vs CN	Accuracy	Precision		Recall	
		AD	CN	AD	CN
RF	0.864	0.85	0.88	0.88	0.85
SVM	0.784	0.77	0.80	0.81	0.76
KNN	0.868	0.80	0.97	0.97	0.76
ANN	0.823	0.81	0.84	0.84	0.80
CNN	0.882	0.91	0.86	0.86	0.90

Table 5.9. Matrix values of binary classification (AD vs CN)

In this classification, CNN has the best classification accuracy of 88.2%. For AD, CNN has the highest precision value (0.91) while KNN has the highest recall value (0.97). For CN, KNN has the highest precision value (0.97) while CNN has the highest recall value (90). This means that for cases that deals with low false positive rate and high false negative rate as well as cases that deals with high false positive rate and low false negative rate, both KNN and CNN can be used. Which method to use depends on the focus of the problem and on the target class for which the problem exists (AD/CN).

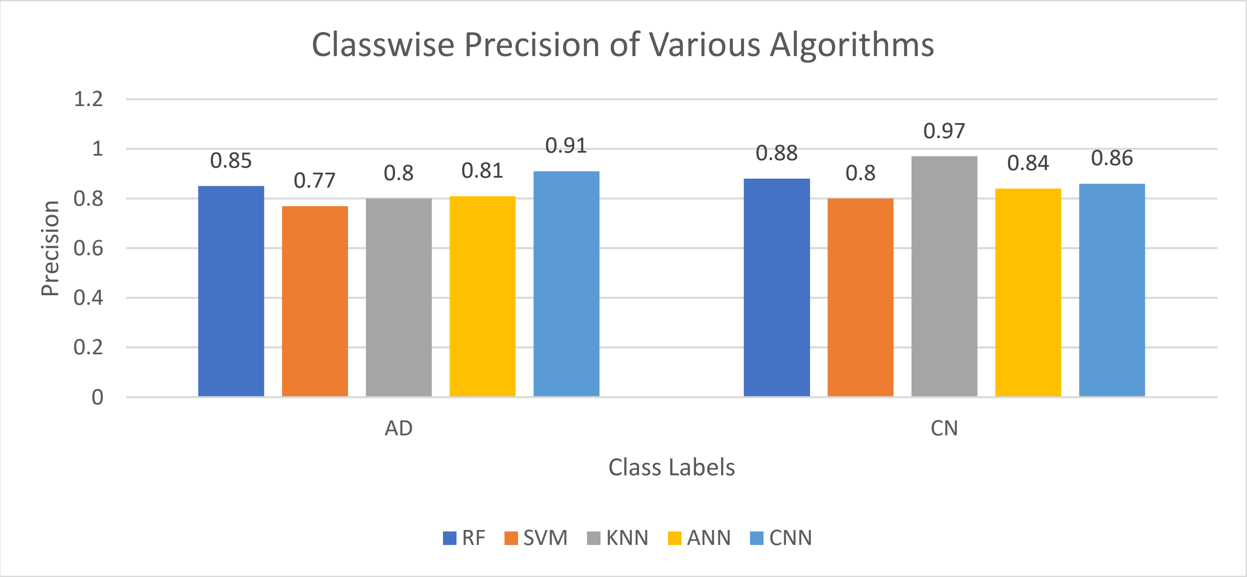


Figure 5.11. Class-wise precision of different algorithms for AD vs CN

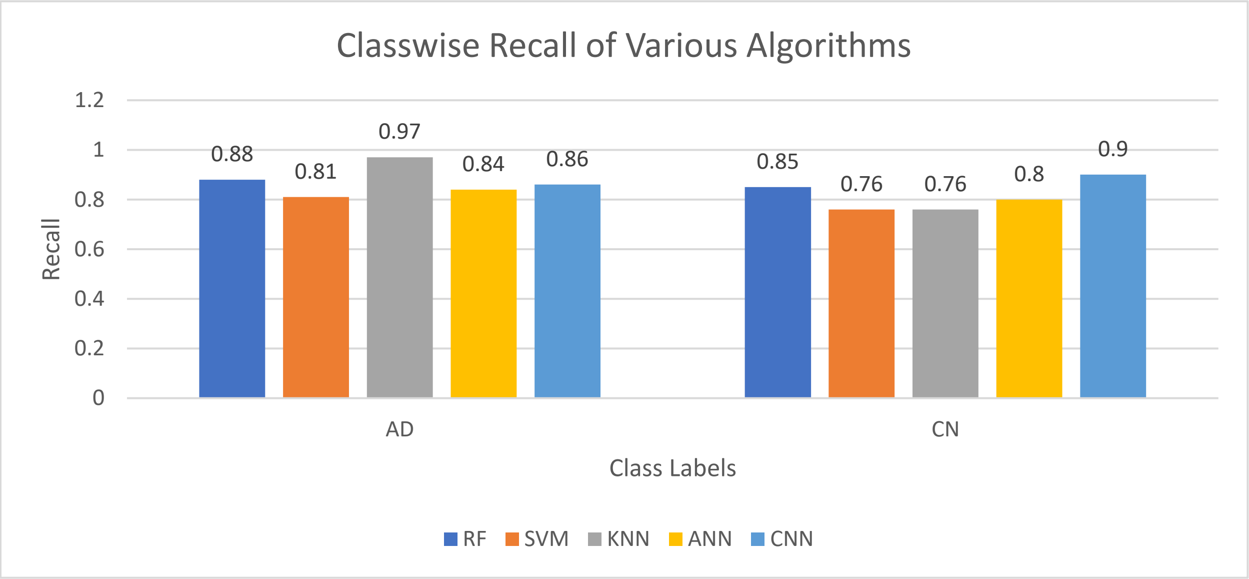


Figure 5.12. Class-wise recall of different algorithms for AD vs CN

5.4.4. Binary classification (MCI vs AD):

Classifier for MCI vs AD	Accuracy	Precision		Recall	
		MCI	AD	MCI	AD
RF	0.765	0.74	0.80	0.84	0.69
SVM	0.686	0.71	0.66	0.70	0.67
KNN	0.745	0.79	0.71	0.68	0.81
ANN	0.776	0.78	0.77	0.78	0.77
CNN	0.810	0.77	0.88	0.90	0.72

Table 5.10 Matrix values of binary classification (MCI vs AD)

Here, CNN continues its lead in accuracy (0.810). For MCI, KNN provides the highest precision (0.79) and CNN provides the highest recall (0.90). For AD, CNN scores the highest in precision (0.88) and KNN scores the highest recall (0.81). This result is similar to the previous one (AD vs CN), however, the previous values have higher magnitude value.

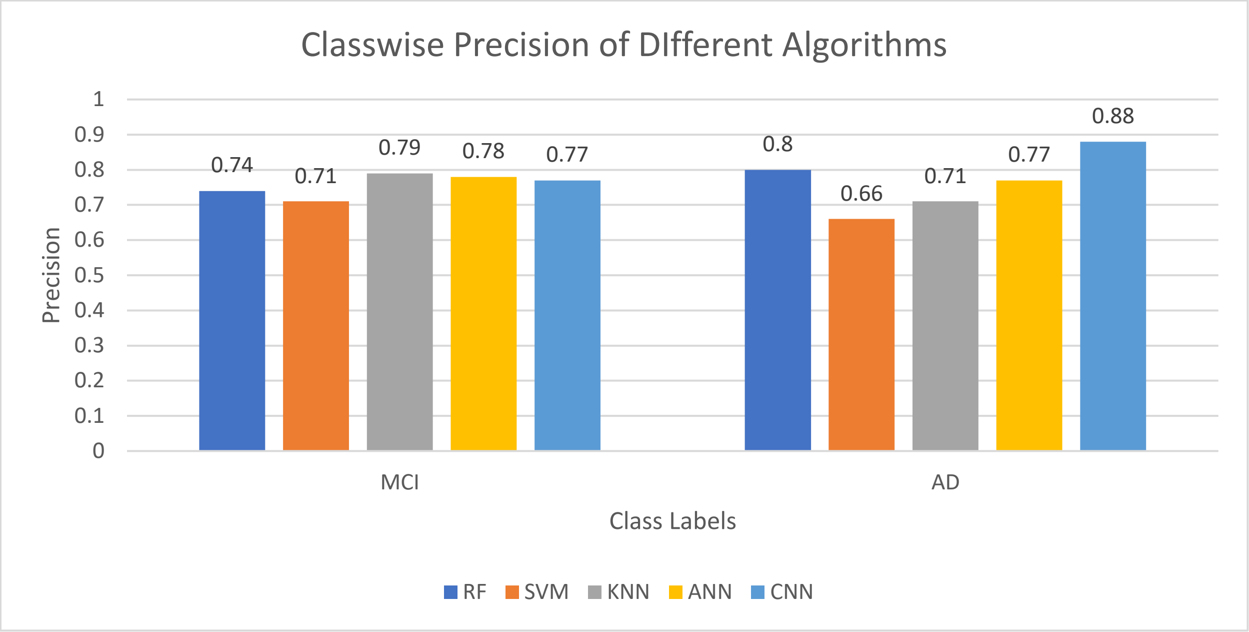


Figure 5.13. Class-wise precision of different algorithms for MCI vs AD

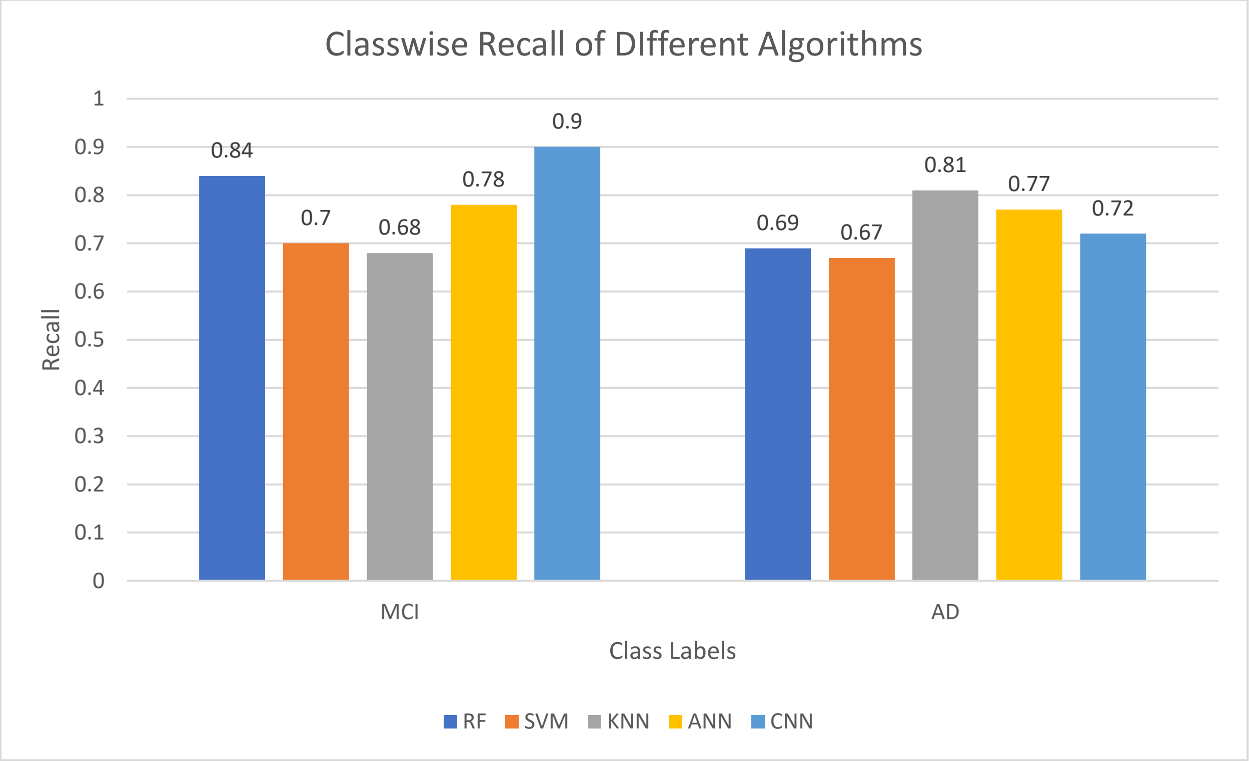


Figure 5.14. Class-wise recall of different algorithms for MCI vs AD

5.4.5. Binary classification (MCI vs CN):

Classifier for MCI vs CN	Accuracy	Precision		Recall	
		MCI	CN	MCI	CN
RF	0.757	0.73	0.79	0.82	0.69
SVM	0.736	0.71	0.78	0.82	0.65
KNN	0.761	0.72	0.82	0.86	0.66
ANN	0.761	0.73	0.80	0.83	0.69
CNN	0.800	0.90	0.72	0.72	0.90

Table 5.11. Matrix values of binary classification (MCI vs CN)

Here, CNN again keeps its lead in accuracy (0.800) For MCI, CNN scored the highest precision value (0.90) while KNN scored the highest recall value (0.86). For CN, KNN scored the highest precision value (0.82) and CNN scored the highest recall value (0.90). This result continues the pattern of the binary classifications that if CNN is good in one matric, then KNN will be better at the other matric. This means for the three classes AD, MCI and CN, for each binary classification combination, KNN and CNN can be used to deal with needs that either require good precision or recall.

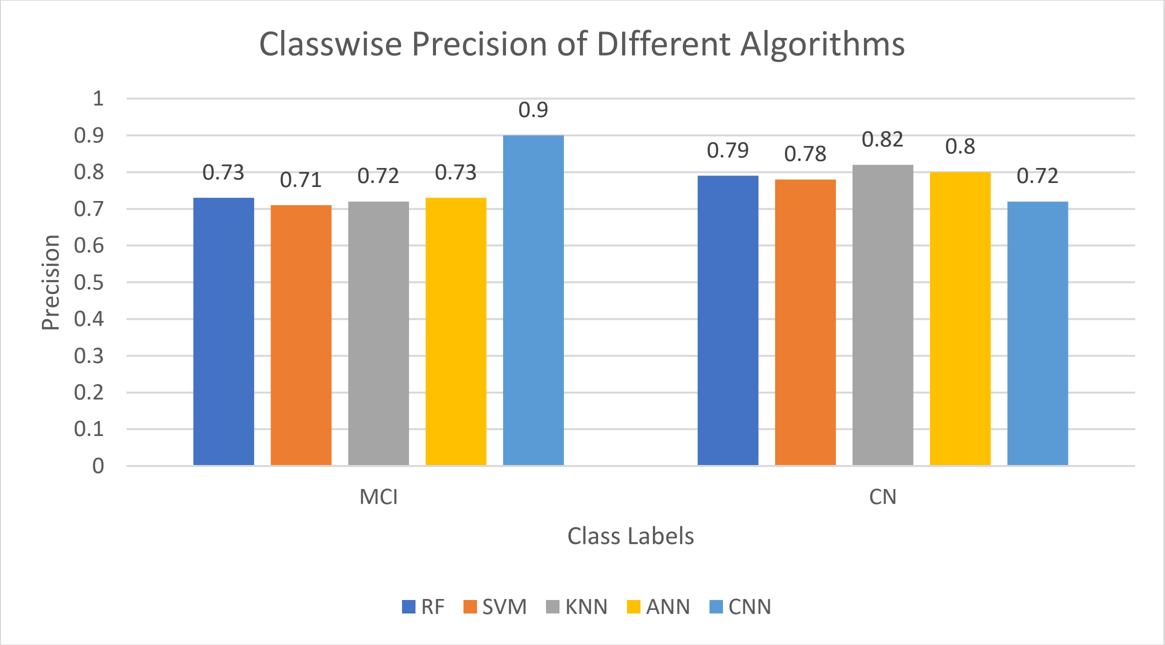


Figure 5.15. Class-wise precision of different algorithms for MCI vs CN

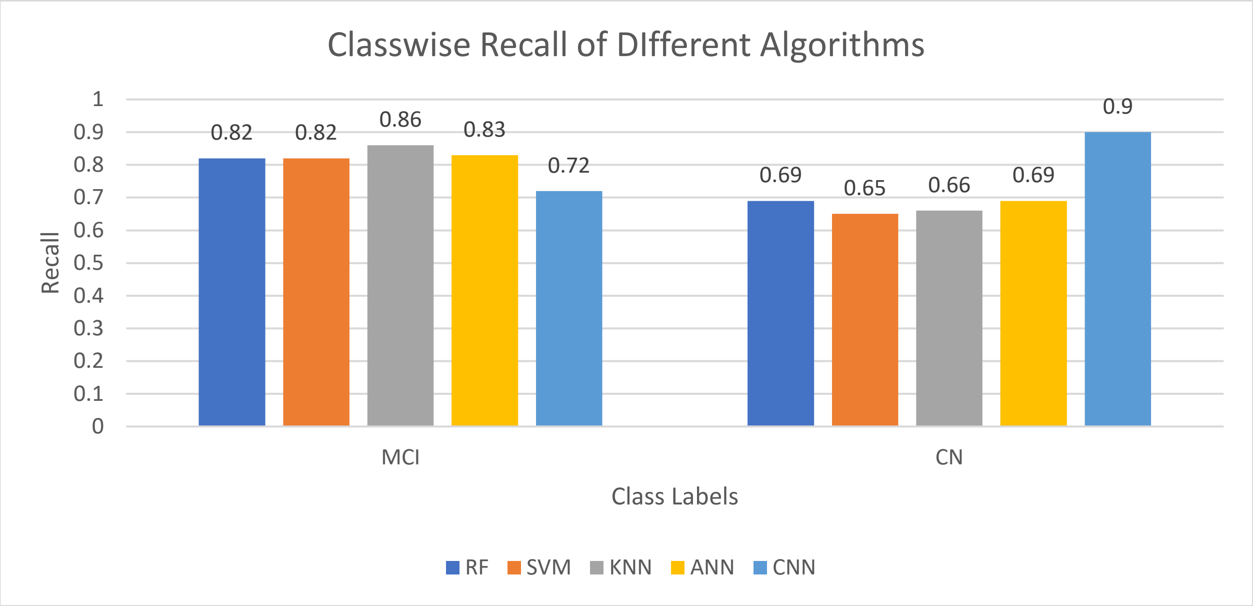


Figure 5.16. Class-wise Recall of different algorithms for MCI vs CN

CHAPTER 6: DISCUSSION AND CONCLUSION

This chapter will conclude the thesis by describing the findings and the usages of the findings, the limitations faced while doing the task as well as future works of interest.

6.1 Significance of Result

This comparative study showed that each algorithm has their own strength and work better in specific cases. So, the user has to figure out the exact problem they want to tackle and how many types of datasets they want to handle. If they want early detection of various stages of AD, then CNN is the overall algorithm whether it's a 5-class or binary class classification since it has the highest accuracy value in all cases. However, if the problem deals with identifying specific classes while considering the effect of their false positive or false negative rates, then the user must determine the number of classes he may deal with and the class he will prioritize and whether his system is sensitive to false positive rate or false negative rate. If the problem has more weights for false positives, then he can use the model that provided best precision for the target class. If the problem has more weights for true negatives, then he can use the model that provided best recall for the target class. One thing to note is that in binary classifications, for a particular class, if CNN had the best precision value, then KNN had the best recall value and vice versa. So, using both algorithms can serve getting the best precision and recall for both classes. In three class classification, in some cases RF provided the better precision and recall value for different classes. So, for multiclass classification, RF can also be used along with CNN and KNN depending on the target class and the problem at hand. In, 5 class, RF did perform on par with CNN and KNN at calculating the precision of EMCI while doing better in calculating recall of CN. In 3 class, RF scored higher in recall for AD and in precision for CN. So, in certain specific cases, RF is the go-to choose, while in other cases it's a choice between CNN and KNN depending on the context. But in terms of overall testing accuracy, CNN is the clear choice.

The results found in this study is comparable to the already performed literature reviews. While the accuracy values were higher in some cases, in most cases, they were similar with the expected results. No accuracy stood particularly high compared to the accuracy values found in the literature review. However, to the authors knowledge, little work is done in early detection especially when the number of classes are 4 or more. We are not aware of a study that worked with 5 classes in early detection. This means the 5-class classification creates new knowledge or shows a direction to how early detection of Alzheimer's disease can be handled when the class number is 5. As for the other classifications, there was no breakthrough accuracy that stood higher than the already reviewed papers. However, this thesis did provide another insight on using precision and recall on such multiclass dataset which is rare for even 3 class datasets. So, these values can be used to get further insight about the inner workings of the algorithm in distinguishing instances that has so much similarities. In binary classifications, existing research show promise of greater accuracy, precision and recall. So, further tuning, training and testing is needed to fully realize the potential of the models.

6.2 Limitations

This section describes the limitations faced throughout thesis.

Data set

A key issue faced while working on early detection is finding quality dataset in 2d image format. While ADNI has an ample amount of image data, they are all MRI 3D images and working on them directly was not feasible with our existing hardware and online tools such as google Colab. This means we were dependent on 2d MRI image data that was provided by third party users. In Kaggle, there were very few unique 2d MRI image datasets as many existing datasets were the same dataset that was separated into different class combinations. We found our dataset which had more classes than the existing dataset. The need for many classes is there because our goal was the early prediction of Alzheimer's disease. But similar to the other datasets, the dataset lacked a strong number of instances. Since MRI images of various states were very similar, we needed a lot of images to better train our models. However, since the number of images in the dataset was low, we had to oversample the minority images. This means the diversity of the image was reduced which in some cases make the classification process more difficult. Lack of strong 2d dataset with a good number of instances for different classes can be considered as a barrier for early AD research. Since the dataset is small, splitting it made the splits smaller. As a result, the model may not receive enough data to learn about the diverse kind of features that are actually present. And since the test dataset is small and MRI images of the dataset have similarities, this may have had an effect in increasing the testing accuracy. So, there can be bias and overfitting in the models trained using this dataset.

Pre-processing

Since the images were not 3d, in-depth pre-processing couldn't be performed as the images were already pre-processed at a level. While basic preprocessing is performed, if the data was 3d, then 2d slices could have been made based on the need of view. Brain splitting could be performed to get better brain slices.

More elaborate feature selection and extraction techniques could be used. No feature selection has been used in this thesis.

Algorithms

SVM and ANN particularly performed poorly in testing. If RBF kernel of SVM were used or more in-depth hyper parameter tuning was performed, then better results could have been extracted. While hyperparameter tuning have been used, it has not been used extensively. Often around 2-5 parameters are tuned for the algorithms. If more parameters are tuned, then better result could be ensured. Since the test class is small, that can be a reasoning behind increase in the value of accuracy. Since Training data sample is small, the models can be prune to oversampling. Training with more extensive data may have increased the realizabilities of the results.

6.3 Future Work

In future, this modeling can be performed to a more robust dataset to better train the models. And with better computational hardware, we may finally work on 3D images to handle the dataset based on our needs. While existing algorithms have been used and, in some cases, good accuracy and precision/recall have been obtained, other techniques or architectures of Neural networks can be used to increase model performance. This thesis focused on the supervised algorithms, in future, unsupervised algorithms such as deep belief network as well as auto encoders can be used. Besides, other architectures of convolutional neural networks such as AlexNet, VGG16, ResNet, Inception, GAN etc. can also be used to determine whether better accuracy can be obtained. In terms of neuroimaging type, pet scan images can be used to get a functional map of the glucose flow in the brain and detect stage of AD based on that data.

6.4 Conclusion

This thesis tried to solve early detection problem of AD which is a difficult problem by design. This is because the similarities between the different classes is not really apparent. Moreover, MRI scans healthy individuals who only show general signs of aging can be similar with people with mild cognitive disorder, making distinguishing between cognitive normal and mild cognitive disorder harder. Several researchers have attempted this problem before with various level of success. While binary classifications have shown prominent results in existing research and our work, things become increasingly difficult with multiclass classification, especially in cases where there are 4 or more classes. To our knowledge, no work is done on five class classifications while 4 class classifications have been performed, accuracy has not been the best in terms of acceptable accuracy in medical and health related fields. While our work is a step in the right direction, much more extensive work needs to be performed in more robust datasets to produce meaningful outcomes. Since clinical diagnosis is crucial in insuring proper treatment, it is highly advised that the models of this thesis be used as a guidance tool and not a replacement of traditional clinical diagnosis. In future, when a model can reliably detect at least 97% of the instances correctly, then we may have a discussion on whether detection through models can be a viable alternative to clinical diagnosis. But till then, clinical diagnosis must be the go-to procedure for early detection of Alzheimer's Disease.

REFERENCES

[1]. Prince, M., Comas-Herrera, A., Knapp, M., Guerchet, M., & Karagiannidou, M. (2016). World Alzheimer report 2016: improving healthcare for people living with dementia: coverage, quality and costs now and in the future.

[2]. Alzheimer's Association. (2019). 2019 Alzheimer's disease facts and figures. Alzheimer's & dementia, 15(3), 321-387.

[3]. Earlier Diagnosis. (2021). Retrieved 2 July 2021, from https://www.alz.org/alzheimers-dementia/research_progress/earlier-diagnosis

[4]. Lyketsos, C., Carrillo, M., Ryan, J., Khachaturian, A., Trzepacz, P., & Amatniek, J. et al. (2011). Neuropsychiatric symptoms in Alzheimer's disease. Alzheimer's & Dementia, 7(5), 532-539. doi: 10.1016/j.jalz.2011.05.2410

[5]. Tiraboschi, P., Hansen, L. A., Thal, L. J., & Corey-Bloom, J. (2004). The importance of neuritic plaques and tangles to the development and evolution of AD. Neurology, 62(11), 1984-1989.

[6]. Jack Jr, Clifford R., et al. "Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers." The Lancet Neurology 12.2 (2013): 207-216.

[7]. Is Alzheimer's Genetic? (2021). Retrieved 4 July 2021, from https://www.alz.org/alzheimers-dementia/what-is-alzheimers/causes-and-risk-factors/genetics

[8]. ADNI | Study Design. (2021). Retrieved 4 July 2021, from http://adni.loni.usc.edu/study-design/#background-container

[9]. Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., ... & Phelps, C. H. (2011). 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. Alzheimer's & dementia, 7(3), 270-279.

[10]. Radue, E. W., Weigel, M., Wiest, R., & Urbach, H. (2016). Introduction to magnetic resonance imaging for neurologists. Continuum: Lifelong Learning in Neurology, 22(5), 1379-1398.

[11]. Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., ... & Phelps, C. H. (2011). 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. Alzheimer's & dementia, 7(3), 270-279.

[12]. Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., ... & Trojanowski, J. Q. (2013). The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. Alzheimer's & Dementia, 9(5), e111-e194.

[13]. Morris, J. C., Storandt, M., McKeel, D. W., Rubin, E. H., Price, J. L., Grant, E. A., & Berg, L. (1996). Cerebral amyloid deposition and diffuse plaques in “normal" aging: Evidence for presymptomatic and very mild Alzheimer's disease. Neurology, 46(3), 707-719.

[14]. Berg, L., McKeel, D. W., Miller, J. P., Baty, J., & Morris, J. C. (1993). Neuropathological indexes of Alzheimer's disease in demented and nondemented persons aged 80 years and older. Archives of neurology, 50(4), 349-358.

[15]. Dickson, D. W., Crystal, H. A., Mattiace, L. A., Masur, D. M., Blau, A. D., Davies, P., ... & Aronson, M. K. (1992). Identification of normal and pathological aging in prospectively studied nondemented elderly humans. Neurobiology of aging, 13(1), 179-189.

[16]. Mackiewich, B. (1995). Basic Principles of MRI. Simon Fraser University, Vancouver.

[17]. ADNI | About. (2021). Retrieved 16 July 2021, from http://adni.loni.usc.edu/about/

[18]. Laboratory of Neuroimaging. (2021). Retrieved 16 July 2021, from https://www.loni.usc.edu/about_loni

[19]. Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260.

[20]. Sharma, S., & Sharma, S. (2017). Activation functions in neural networks. Towards Data Science, 6(12), 310-316.

[21]. Bengio, Y. (2009). Learning deep architectures for AI. Now Publishers Inc.

[22]. Ciregan, D., Meier, U., & Schmidhuber, J. (2012, June). Multi-column deep neural networks for image classification. In 2012 IEEE conference on computer vision and pattern recognition (pp. 3642-3649). IEEE.

[23]. Bengio, Y. (2007). Learning deep architectures for Al. Foundations and Trends in Machine Learning.-2009.—2 (1).-pp, 1-127.

[24]. Erhan, D., Courville, A., Bengio, Y., & Vincent, P. (2010, March). Why does unsupervised pre-training help deep learning?. In Proceedings of the thirteenth international conference on artificial intelligence and statistics (pp. 201-208). JMLR Workshop and Conference Proceedings.

[25]. Sutskever, I., Martens, J., Dahl, G., & Hinton, G. (2013, May). On the importance of initialization and momentum in deep learning. In International conference on machine learning (pp. 1139-1147). PMLR.

[26]. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. Neurocomputing, 234, 11-26.

[27]. Bengio, Y., Lamblin, P., Popovici, D., & Larochelle, H. (2007). Greedy layer-wise training of deep networks. In Advances in neural information processing systems (pp. 153-160).

[28]. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 1097-1105.

[29]. Horning, N. (2010, December). Random Forests: An algorithm for image classification and generation of continuous fields data sets. In Proceedings of the International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences, Osaka, Japan (Vol. 911).

[30]. Bosch, A., Zisserman, A., & Munoz, X. (2007, October). Image classification using random forests and ferns. In 2007 IEEE 11th international conference on computer vision (pp. 1-8). Ieee.

[31]. Tanveer, M., Richhariya, B., Khan, R. U., Rashid, A. H., Khanna, P., Prasad, M., & Lin, C. T. (2020). Machine learning techniques for the diagnosis of Alzheimer's disease: A review. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 16(1s), 1-35.

[32]. Noble, W. S. (2006). What is a support vector machine?. Nature biotechnology, 24(12), 1565-1567.

[33]. Suthaharan, S. (2016). Support vector machine. In Machine learning models and algorithms for big data classification (pp. 207-235). Springer, Boston, MA.

[34]. Cunningham, P., & Delany, S. J. (2020). k-Nearest neighbour classifiers: (with Python examples). arXiv preprint arXiv:2004.04523.

[35]. Kim¹, J. I. N. H. O., Kim, B. S., & Savarese, S. (2012). Comparing image classification methods: K-nearest-neighbor and support-vector-machines. In Proceedings of the 6th WSEAS international conference on Computer Engineering and Applications, and Proceedings of the 2012 American conference on Applied Mathematics (Vol. 1001, pp. 48109-2122).

[36]. Perez, A., Larranaga, P., & Inza, I. (2006). Supervised classification with conditional Gaussian networks: Increasing the structure complexity from naive Bayes. International Journal of Approximate Reasoning, 43(1), 1-25.

[37]. Griffis, J. C., Allendorfer, J. B., & Szaflarski, J. P. (2016). Voxel-based Gaussian naïve Bayes classification of ischemic stroke lesions in individual T1-weighted MRI scans. Journal of neuroscience methods, 257, 97-108.

[38]. Kamel, H., Abdulah, D., & Al-Tuwaijari, J. M. (2019, June). Cancer classification using gaussian naive bayes algorithm. In 2019 International Engineering Conference (IEC) (pp. 165-170). IEEE.

[39]. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), 1929-1958.

[40]. Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580.

[41]. Arvesen, E. (2015). Automatic classification of Alzheimer’s disease from structural MRI (Master's thesis).

[42]. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). Tensorflow: A system for large-scale machine learning. In 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16) (pp. 265-283).

[43]. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825-2830.

[44]. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825-2830.

[45]. Bradski, G., & Kaehler, A. (2008). Learning OpenCV: Computer vision with the OpenCV library. " O'Reilly Media, Inc.".

[46]. Fung, G., & Stoeckel, J. (2007). SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information. Knowledge and Information Systems, 11(2), 243-258.

[47]. Sandeep, C., & Patnaik, L. Jaganathan (2006) Classification of MR brain images using wavelets as input to SVM and neural network. Biomedical signal processing and control, 1, 86-92.

[48]. López, M., Ramírez, J., Górriz, J. M., Álvarez, I., Salas-Gonzalez, D., Segovia, F., & Puntonet, C. G. (2009, June). Automatic system for Alzheimer’s disease diagnosis using eigenbrains and bayesianclassification rules. InInternational Work-Conference on Artificial Neural Networks (pp. 949-956). Springer, Berlin, Heidelberg.

[49]. Ramírez, J., Górriz, J. M., López, M., Salas-Gonzalez, D., Álvarez, I., Segovia, F., & Puntonet, C. G. (2008, November). Early detection of the alzheimer disease combining feature selection and kernelmachines. InInternational Conference on Neural Information Processing (pp. 410-417). Springer, Berlin,Heidelberg.

[50]. Fan, Y., Resnick, S. M., Wu, X., & Davatzikos, C. (2008). Structural and functional biomarkers ofprodromal Alzheimer's disease: a high-dimensional pattern classification study. Neuroimage,41(2), 277-285.

[51]. Klöppel, S., Stonnington, C. M., Chu, C., Draganski, B., Scahill, R. I., Rohrer, J. D., ... & Frackowiak, R. S. (2008). Automatic classification of MR scans in Alzheimer's disease. Brain,131(3), 681-689.

[52]. Mesrob, L., Magnin, B., Colliot, O., Sarazin, M., Hahn-Barma, V., Dubois, B., ... & Benali, H. (2008, August). Identification of atrophy patterns in Alzheimer’s disease based on SVM feature selection andanatomical parcellation. InInternational Workshop on Medical Imaging and Virtual Reality (pp. 124-132). Springer, Berlin, Heidelberg.

[53]. Magnin, B., Mesrob, L., Kinkingnéhun, S., Pélégrini-Issac, M., Colliot, O., Sarazin, M., ... &Benali, H. (2009). Support vector machine-based classification of Alzheimer’s disease from whole-brain anatomicalMRI.Neuroradiology,51(2), 73-83.

[54]. Chaves, R., Ramírez, J., Górriz, J. M., López, M., Salas-Gonzalez, D., Alvarez, I., & Segovia, F. (2009). SVM-based computer-aided diagnosis of the Alzheimer's disease using t-test NMSE feature

selection with feature correlation weighting. Neuroscience letters,461(3), 293-297.

[55]. Plant, C., Teipel, S. J., Oswald, A., Böhm, C., Meindl, T., Mourao-Miranda, J., ... & Ewers, M. (2010). Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer'sdisease.Neuroimage,50(1), 162-174.

[56]. Segovia, F., Górriz, J. M., Ramírez, J., Salas-González, D., Álvarez, I., López, M., ... & Padilla, P. (2010). Classification of functional brain images using a GMM-based multi-variate approach. NeuroscienceLetters,474(1), 58-62.

[57]. Padilla, P., Górriz, J. M., Ramírez, J., Lang, E. W., Chaves, R., Segovia, F., ... & Álvarez, I. (2010). Analysis of SPECT brain images for the diagnosis of Alzheimer's disease based on NMF for featureextraction.Neuroscience letters,479(3), 192-196.

[58]. Abdulkadir, A., Mortamet, B., Vemuri, P., Jack Jr, C. R., Krueger, G., Klöppel,S., & Alzheimer's DiseaseNeuroimaging Initiative. (2011). Effects of hardware heterogeneity on the performance of SVMAlzheimer's disease classifier. Neuroimage,58(3), 785-792.

[59]. Illán, I. A., Górriz, J. M., López, M. M., Ramírez, J., Salas-Gonzalez, D., Segovia, F., ... & Puntonet, C. G.(2011). Computer aided diagnosis of Alzheimer’s disease using component based SVM.Applied SoftComputing,11(2), 2376-2382.

[60]. Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M. O., ... & Alzheimer'sDisease Neuroimaging Initiative.(2011). Automatic classification of patients with Alzheimer's diseasefrom structural MRI: a comparison of ten methods using the ADNI database.neuroimage,56(2), 766-781.

[61]. Huang, C., Yan, B., Jiang, H., & Wang, D. (2008, May). Combining voxel-based morphometry with artificalneural network theory in theapplication research of diagnosing alzheimer's disease. In2008 InternationalConference on BioMedical Engineering and Informatics(Vol. 1, pp. 250-254). IEEE.

[62]. Savio, A., García-Sebastián, M., Hernández, C., Graña, M., & Villanúa, J. (2009, September).Classification results of artificial neural networks for alzheimer’s disease detection. InInternationalConference on Intelligent Data Engineering and Automated Learning(pp. 641-648). Springer, Berlin,Heidelberg.

[63]. Ahmadlou, M., Adeli, H., & Adeli, A. (2010). New diagnostic EEG markers of the Alzheimer’s diseaseusing visibility graph.Journal of neural transmission,117(9), 1099-1109.

[64]. da Silva Lopes, H. F., Abe, J. M., & Anghinah, R. (2010). Application of paraconsistent artificial neuralnetworks as a method of aid in the diagnosis of Alzheimer disease.Journal of medical systems,34(6),1073-1081.

[65]. Long, X., &Wyatt, C. (2010, June). An automatic unsupervised classification of MR images in Alzheimer'sdisease. In2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition(pp.2910-2917). IEEE.

[66]. Zhang, D., & Shen, D. (2011, March). Semi-supervised multimodal classification of Alzheimer's disease.In2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro(pp. 1628-1631).IEEE.

[67]. Quintana, M., Guàrdia, J., Sánchez-Benavides, G., Aguilar, M., Molinuevo, J. L., Robles, A., ... & Neuronorma Study Team. (2012). Using artificial neural networks in clinical neuropsychology: High performance in mild cognitive impairment and Alzheimer's disease. Journal of Clinical and Experimental Neuropsychology, 34(2), 195-208.

[68]. Mahanand, B. S., Suresh, S., Sundararajan, N., & Kumar, M. A. (2012). Identification of brain regions responsible for Alzheimer’s disease using a Self-adaptive Resource Allocation Network. Neural Networks, 32, 313-322.

[69]. Chyzhyk, D., Savio, A., & Graña, M. (2014). Evolutionary ELM wrapper feature selection for Alzheimer's disease CAD on anatomical brain MRI. Neurocomputing, 128, 73-80.

[70]. Mahmood, R., & Ghimire, B. (2013, July). Automatic detection and classification of Alzheimer's Disease from MRI scans using principal component analysis and artificial neural networks. In 2013 20th International Conference on Systems, Signals and Image Processing (IWSSIP) (pp. 133-137). IEEE.

[71]. Al-Naami, B., Gharaibeh, N., & Kheshman, A. A. (2013). Automated detection of Alzheimer disease using region growing technique and artificial neural network. World Acad. Sci. Eng. Technol. Int. J. Biomed. Biol. Eng, 7(5).

[72]. Jie, B., Zhang, D., Wee, C. Y., & Shen, D. (2014). Topological graph kernel on multiple thresholder functional connectivity networks for mild cognitive impairment classification. Human brain mapping, 35(7), 2876-2897.

[73]. Ortiz, A., Górriz, J. M., Ramírez, J., Martinez-Murcia, F. J., & Alzheimer's Disease Neuroimaging Initiative. (2014). Automatic ROI selection in structural brain MRI using SOM 3D projection. PloS one, 9(4), e93851.

[74]. Wang, K., Liang, M., Wang, L., Tian, L., Zhang, X., Li, K., & Jiang, T. (2007). Altered functional connectivity in early Alzheimer's disease: A resting-state fMRI study. Human brain mapping, 28(10), 967-978.

[75]. Fan, Y., Resnick, S. M., Wu, X., & Davatzikos, C. (2008). Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. Neuroimage, 41(2), 277-285.

[76]. Davatzikos, C., Resnick, S. M., Wu, X., Parmpi, P., & Clark, C. M. (2008). Individual patient diagnosis of AD and FTD via high-dimensional pattern classification of MRI. Neuroimage, 41(4), 1220-1227.

[77]. Hinrichs, C., Singh, V., Mukherjee, L., Xu, G., Chung, M. K., Johnson, S. C., & Alzheimer's Disease Neuroimaging Initiative. (2009). Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. Neuroimage, 48(1), 138-149.

[78]. López, M., Ramírez, J., Górriz, J. M., Salas-Gonzalez, D., Alvarez, I., Segovia, F., & Puntonet, C. G. (2009). Automatic tool for Alzheimer's disease diagnosis using PCA and Bayesian classification rules. Electronics letters, 45(8), 389-391.

[79]. Horn, J. F., Habert, M. O., Kas, A., Malek, Z., Maksud, P., Lacomblez, L., ... & Fertil, B. (2009). Differential automatic diagnosis between Alzheimer's disease and frontotemporal dementia based on perfusion SPECT images. Artificial intelligence in medicine, 47(2), 147-158.

[80]. Desikan, R. S., Cabral, H. J., Hess, C. P., Dillon, W. P., Glastonbury, C. M., Weiner, M. W., ... & Alzheimer's Disease Neuroimaging Initiative. (2009). Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. Brain, 132(8), 2048-2057.

[81]. Ramírez, J., Górriz, J. M., Segovia, F., Chaves, R., Salas-Gonzalez, D., López, M., ... & Padilla, P. (2010). Computer aided diagnosis system for the Alzheimer's disease based on partial least squares and random forest SPECT image classification. Neuroscience letters, 472(2), 99-103.

[82]. Chen, G., Ward, B. D., Xie, C., Li, W., Wu, Z., Jones, J. L., ... & Li, S. J. (2011). Classification of Alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network

analysis based on resting-state functional MR imaging. Radiology, 259(1), 213-221.

[83]. Westman, E., Simmons, A., Muehlboeck, J. S., Mecocci, P., Vellas, B., Tsolaki, M., ... & Alzheimer's Disease Neuroimaging Initiative. (2011). AddNeuroMed and ADNI: similar patterns of Alzheimer's atrophy and automated MRI classification accuracy in Europe and North America. Neuroimage, 58(3), 818-828.

[84]. Rao, A., Lee, Y., Gass, A., & Monsch, A. (2011, August). Classification of Alzheimer's Disease from structural MRI using sparse logistic regression with optional spatial regularization. In 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 4499-4502). IEEE.

[85]. Liu, M., Zhang, D., Shen, D., & Alzheimer's Disease Neuroimaging Initiative. (2012). Ensemble sparse classification of Alzheimer's disease. NeuroImage, 60(2), 1106-1116.

[86]. Rajan, K. B., Weuve, J., Barnes, L. L., McAninch, E. A., Wilson, R. S., & Evans, D. A. (2021). Population estimate of people with clinical Alzheimer's disease and mild cognitive impairment in the United States (2020–2060). Alzheimer's & Dementia.

[87]. Chêne, G., Beiser, A., Au, R., Preis, S. R., Wolf, P. A., Dufouil, C., & Seshadri, S. (2015). Gender and incidence of dementia in the Framingham Heart Study from mid-adult life. Alzheimer's & Dementia, 11(3), 310-320.

[88]. Conti, D. (2021). Magnetic Resonance Imaging. Retrieved 11 October 2021, from https://www.researchgate.net/figure/a-MRI-Scanner-Cutaway-b-MRI-Scanner-Gradient-Magnets-MRI-A-Guided-Tour-2015_fig2_299512554.

[89]. Imaging Techniques. (2021). Retrieved 11 October 2021, from https://sites.google.com/site/postgraduatetraining/image-acquisition/the-basics?tmpl=%2Fsystem%2Fapp%2Ftemplates%2Fprint%2F&showPrintDialog=1.

[90]. Beqari, E. (2021). A Very Basic Introduction to Feed-Forward Neural Networks - DZone AI. Retrieved 11 October 2021, from https://dzone.com/articles/the-very-basic-introduction-to-feed-forward-neural.

[91]. Perceptron Function - GM-RKB. (2021). Retrieved 11 October 2021, from http://www.gabormelli.com/RKB/Perceptron_Function.

[92]. Baheti, P. (2021). 12 Types of Neural Networks Activation Functions: How to Choose?. Retrieved 11 October 2021, from https://www.v7labs.com/blog/neural-networks-activation-functions.

[93]. Manmohan, M. (2021). 57 CNN Basics. Retrieved 11 October 2021, from https://www.kaggle.com/manmohan291/57-cnn-basics.

[94]. Yani, M. (2019, May). Application of transfer learning using convolutional neural network method for early detection of terry's nail. In Journal of Physics: Conference Series (Vol. 1201, No. 1, p. 012052). IOP Publishing.

[95]. Jana, A. (2021). Support Vector Machines for Beginners - Linear SVM - A Developer Diary. Retrieved 11 October 2021, from http://www.adeveloperdiary.com/data-science/machine-learning/support-vector-machines-for-beginners-linear-svm/.

[96]. Liu, C. (2021). A Top Machine Learning Algorithm Explained: Support Vector Machines (SVMs) - Velocity Business Solutions Limited. Retrieved 11 October 2021, from https://www.vebuso.com/2020/02/a-top-machine-learning-algorithm-explained-support-vector-machines-svms/.

[97]. Armi, L., & Fekri-Ershad, S. (2019). Texture image analysis and texture classification methods-A review. arXiv preprint arXiv:1904.06554.

[98]. Navlani, A. (2021). KNN Classification using Scikit-learn. Retrieved 11 October 2021, from https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn.

[99]. Kumar, N. (2021). Batch Normalization and Dropout in Neural Networks Explained with Pytorch. Retrieved 11 October 2021, from https://towardsdatascience.com/batch-normalization-and-dropout-in-neural-networks-explained-with-pytorch-47d7a8459bcd.

[100]. Mohajon, J. (2021). Confusion Matrix for Your Multi-Class Machine Learning Model. Retrieved 11 October 2021, from https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826.