Recall-Net: A CNN-based Model for Four-class Classification of Alzheimer's Disease

by

M. M. Kamrul Hasan
17201049
Farhan Faisal Angan
17201153
Tasmim Rashid
17301232

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
September 2021

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

M. M. Kamrul Hasan
17201049

Farhan Faisal Angan
17201153

Tasmim Rashid
17301232

Approval

The thesis/project titled "Recall-Net: A CNN-based Model for Four-class Classification of Alzheimer's Disease" submitted by

1. M. M. Kamrul Hasan (17201049)
2. Farhan Faisal Angan (17201153)
3. Tasmim Rashid (17301232)

Of Summer, 2021 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on September 26, 2021.

Examining Committee:

Supervisor:
(Member)

Faisal Bin Ashraf
Lecturer
Department of Computer Science and Engineering
BRAC University

Co-Supervisor:
(Member)

Dr. Mohammad Zavid Parvez
Assistant Professor
Department of Computer Science and Engineering
BRAC University

Program Coordinator:
(Member)

Md. Golam Rabiul Alam, PhD
Associate Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

Sadia Hamid Kazi
Associate Professor
Department of Computer Science and Engineering
BRAC University

Abstract

Deep learning, a cutting-edge machine learning technique, has outperformed classical machine learning at detecting detailed structures in complex multi-dimensional data, particularly in the field of computer vision. As recent advances in neuroimaging techniques have created massive multimodal neuroimaging data, the application of deep learning to early diagnosis and automated categorization of AD has recently gotten a lot of interest. It also supports biomedical researchers in the identification of many diseases such as cancer, Alzheimer's, Malaria, and blood cell detection, among others. Deep learning is a subclass of machine learning techniques for extracting features and applying them to classification, image processing, and other tasks. A thorough Google Scholar search was conducted before and during our research to find deep learning publications on AD published between January 2010 and July 2020. After reading and evaluating, these articles were categorized and summarized according to their used algorithms and neuroimaging techniques. In our research, we have used CNN also known as ConvNet which is one of the most efficient deep learning-based neural networks to classify Alzheimer's patients from healthy individuals with the help of MRI data. We have collected our sMRI data from ADNI. Our dataset includes a total of 6400 patients who were categorized as non-demented or having mild to severe Alzheimer's disease. Our deep learning approach automatically distinguishes different stages of AD subjects according to their severity. Our proposed model was designed to assist with accurate classification of four classes e.g. NC, EMCI, MCI and AD. Though classification is very crucial for modeling a prediction model to find out the existence and intensity of the disease, it has always been quite difficult. Separating the distinctive features from the ROI is the most challenging part. Our proposed model has a classification accuracy of 79.77%. The performance of our work was compared to several other existing approaches for multi-class classification.

Keywords: Machine Learning, Deep Learning, Alzheimer, MRI, sMRI, MCI, ADNI, CNN.

Acknowledgement

First of all, all the gratitude is towards the Almighty Allah who has given us the strength to continue our work without any major hindrance and finish it successfully. Our team has given their best effort to finish our thesis on time. Moreover, we would like to show our wholehearted gratitude towards our supervisor and co-supervisor, Faisal Bin Ashraf and Dr. Mohammad Zavid Parvez Sir respectively for their assistance and motivation. In every step of our research, they have motivated us in every way to pull off our work successfully. Due to their knowledge and experience, our path towards achieving our goal has been so smooth. Furthermore, we are also very thankful to our family without whom we could not have done it. Lastly, we very much appreciate BRAC University for giving us the opportunity to do our research and provide us with all the necessary information and resources to achieve our objective.

Table of Contents

Declaration **i**

Approval **ii**

Abstract **iv**

Acknowledgment **v**

Table of Contents **vi**

List of Figures **viii**

List of Tables **ix**

Nomenclature **xi**

1 Introduction **1**

1.1 Motivation . . . 2
1.2 Aims and Objectives . . . 3
1.3 Thesis Overview . . . 3

2 Literature Review **4**

2.1 Literature Review . . . 4
2.2 Convolutional Neural Network . . . 8
2.3 How Different Layers of CNN works: . . . 10
2.3.1 Input Layer . . . 10
2.3.2 Convolution Layer . . . 10
2.3.3 Pooling Layer . . . 12
2.3.4 Fully Connected Layer . . . 13
2.4 VGG . . . 13
2.4.1 The Architecture of VGG16 . . . 14
2.5 Inception . . . 14
2.5.1 Inception v3 Concept . . . 15
2.5.2 The Architecture of Inception v3 . . . 16
2.6 ResNet . . . 18
2.6.1 ResNet50 Concept . . . 18

3 Methodology **22**

3.1 Research Framework . . . 22
3.2 Research Objectives . . . 22

3.3 Data Collection 23
3.4 Data Pre-processing 24
3.4.1 Data Labeling 25
3.5 Recall-Net: Our Proposed Model 25
3.5.1 Recall-Net Architecture 27

4 Experimental Result 29
4.1 Applying Recall-Net CNN Architecture 29
4.1.1 Accuracy Calculation Equation 29
4.1.2 Precision Calculation Equation 29
4.1.3 Recall Calculation Equation 30
4.1.4 F1-Score Calculation Equation 30
4.1.5 AUC Calculation Equation 30
4.2 Recall-Net Based on VGG16 30
4.2.1 Result of Recall-Net Architecture Based on VGG16 30
4.2.2 Result Discussion of Recall-Net Architecture Based on VGG16 31
4.3 Recall-Net Based on InceptionV3 34
4.3.1 Result of Recall-Net Architecture Based on Inception v3 . . . 34
4.3.2 Result Discussion of Recall-Net Architecture Based on Inception v3 35
4.4 Our Recall-Net Based on ResNet50 38
4.4.1 Result Discussion of Recall-Net Architecture Based on ResNet50 38
4.5 Comparison between Different Variation of Recall-Net 41
4.6 Compare the Performance of Recall-Net with Other Models 43

5 Summary, Conclusion, Discussion and Recommendations 44
5.1 Overview 44
5.2 Key Insights 45
5.3 Discussion 45
5.4 Recommendation For Practice 46
5.5 Recommendations For Future Work 46
5.6 Conclusion 47

Bibliography 52

List of Figures

2.1 Layers of Convolutional Neural Network 9
2.2 Steps of Convolution Layer 11
2.3 Stride 2 Utilization in Pooling Process 13
2.4 VGG16 Architecture. 14
2.5 Naive version of Inception and Dimension reductions of Inception . . 15
2.6 A 3X3 convolution 16
2.7 Auxiliary Classifier in Inception v3 17
2.8 Grid Reduction Technique 17
2.9 Inception v3 18
2.10 Residual learning: Skip Connection 19
2.11 Residual learning: Building Block 19
2.12 ResNet50 Architecture 20

3.1 Block Diagram of Our Proposed Model 23
3.2 Three Different Views of Subject Skull. 24
3.3 MRI of Skull Before and After Skull Stripping. 25
3.4 Skull Condition at Different Classes of AD. 26

4.1 ROC graph of a model 31
4.2 Accuracy of VGG16 32
4.3 Precision of VGG16 32
4.4 F1-Score of VGG16 33
4.5 AUC of VGG16 33
4.6 Confusion Matrix of VGG16 34
4.7 Accuracy of Inception v3 35
4.8 Precision of Inception v3 36
4.9 F1-Score of Inception v3 36
4.10 AUC of Inception v3 37
4.11 Confusion Matrix of Inception v3 37
4.12 Accuracy of Recall-Net Based on ResNet50 39
4.13 Precision of of Recall-Net Based on ResNet50 39
4.14 F1-Score of of Recall-Net Based on ResNet50 40
4.15 AUC of of Recall-Net Based on ResNet50 40
4.16 Confusion Matrix of ResNet50 41

List of Tables

2.1 ResNet-50 Layers and Associated Parameters 20

3.1 Size of Data Set in Four Different Classes 25
3.2 Summary of Our Proposed CNN Model, Recall-Net 27

4.1 Precision, Recall and F1 Score in VGG16 model 31
4.2 Precision, Recall and F1 Score in Inception V3 model 34
4.3 Precision, Recall and F1 Score in Recall-Net Based on ResNet model 38
4.4 Performance Comparison of Recall-Net with & without Pre-processing 42
4.5 Performance Comparison of Recall-Net with Different CNN Architectures as Base Model 42
4.6 Performance Comparison of Recall-Net with Existing Models 42

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

AAL Automated Anatomical Labelling

AD Alzheimer's disease

$ADNI$ Alzheimer's Disease Neuroimaging Initiative

ANN Artificial Neural Network

AUC Area Under Curve

AUC Area Under Curve

$cMCI$ Converter Mild Cognitive Impairment

CNN Convolutional Neural Network

CRD Clinical Dementia Rating

CSF Cerebrospinal Fluid

DBM Deep Boltzmann Machine

$DeepESrnet$ Deep Ensemble Sparse Regression Network

DLB Dementia with Lewy bodies

DNN Deep Neural Network

DT Decision Tree

DTI Diffusion Tensor Imaging

$EMCI$ Early Mild Cognitive Impairment

FC Functional Connectivity

$FDG-PET$ Fluorodeoxyglucose Positron Emission Tomography

$fMRI$ Function Magnetic Resonance Imaging

GLM Generalized Linear Model

KNN K-nearest Neighbors

$LMCI$ Late Mild Cognitive Impairment

LRN Local Response Normalization

MCI Mild Cognitive Impairment

$MMSE$ Mini-Mental State Examination

MRI Magnetic Resonance Imaging

$MTLA$ Medial Temporal Lobe Atrophy

NC Normal Control

$ncMCI$ Non-Converter Mild Cognitive Impairment

PET Positron Emission Tomography

$ReLU$ Rectified Linear Unit

$ResNet$ Residual Network

$ResNet$ Residual Network

RNN Recurrent Neural Network

ROC Receiver Operating Characteristics

ROC Receiver Operating Characteristics

ROI Region of Interest

SAE Small Area Estimation Methodology

SMC Substantial Memory Concern

$SMRI$ Structural Magnetic Resonance Imaging

SVM Support Vector Machine

VCI Vascular Cognitive Impairment

Chapter 1

Introduction

This chapter provides an overview of our thesis as well as the the problem definition, reason, goal of it and ultimately significance of our work.
AD is the most frequent reason behind dementia among elderly people. It is also the most prevalent neurodegenerative disease in the world, with over fifty millions of cases. Although this disease is presently the sixth greatest cause of mortality in the United States, new estimates suggest that it may rank third, after cancer and heart disease, as a cause of death for the elderly [52]. Although post-mortem examination of brain tissue is considered necessary for a conclusive diagnosis of AD, PET and biomarkers in the CSF in combination with a number of recent clinical criteria can assist in diagnosis of living patients [54]. There are several stages of this disease. Those stages are titled as: CN, SMC, EMCI, MCI, LMCI and AD [50]. The severe and terminal stage of Alzheimer's disease leads to death due to the complications arising from the high decrease in brain function. Although there is no cure for Alzheimer's disease, the correct medication and care can help control symptoms. As a result, if the right prognosis of the cognitive decline is recognized at an early stage of the disease, it can be delayed and managed. Therefore, right classification of AD can be extremely helpful in order to increase the patient's quality of lives and to smooth the development process of drug trials. Though conventional machine learning techniques can be useful, deep learning models are currently preferred due to their efficiency and better performance in the computer vision area.
Deep learning is a machine learning subclass that uses lots of layer of a neural network. In terms of the kind of data it utilizes and the kind of learning methods it employs, deep learning varies from standard machine learning [26]. DL has revolutionized computer vision area of machine learning research. Complex features and higher-level represention are derived from lower-level ones and such hierarchy of features is called deep architecture [50] which culminates in a detailed feature map of the data. There are several architectures of DL suited for several kinds of data. The neural networks in deep learning are collections of layers that identify correlations in certain datasets using a method that is roughly identical to how the human brain operates. Among those CNN, DNN and RNN have been most successful in fields like feature learning, computer vision, image processing, customer experience, industrial automation, medical research, text generation and many more [23]. In our research, we have used CNN best known for object detection and pattern recognition which includes image processing to voice recognition [26]. Generally, images are taken as input in CNN architecture and after going through layers of convolution,

it recognizes certain patterns through feature extraction. Taking this pattern as a reference, CNN finds similarities or dissimilarities among other inputs accurately. We have used CNN because ANN has more complex structure and has more training parameters than CNN which makes it computationally infeasible to train and use in computer vision area than CNN[13]. When compared to other classification models, the amount of pre-processing required by a ConvNet is significantly less [1]. We have used MRI data of the subject brain from the ADNI dataset for classifying the subjects affected with Alzheimer's from healthy subjects by building a model using several CNN architectures which include InceptionV3, ResNet50, VGG16 and etc.

1.1 Motivation

For the past two decades, AD has grown consistently to become quite a common disease in the world. This disease is currently the sixth leading cause of death in the United States. It's a brain disease that slowly erodes memory and cognitive abilities, as well as the ability to do even the most fundamental tasks [52]. The symptoms of AD generally appear in individuals in their mid-60s. Overproduction of hyperphosphorylation of τ protein [34][39] and amyloid-β (Aβ) [52] [43] are thought to trigger AD pathogenesis. This results in the accumulation of Aβ and neurofibrillary tangles, disrupting the nucleocytoplasmic transport between neurons leading to cell death. Initially, the hippocampus region [13] is affected by the disease. Since anatomically the hippocampus is responsible memory and neuroplasticity i.e. learning, therefore memory loss is one of the early symptoms of AD. It is the most common form of dementia in those over the age of 65. The percentage of people with AD increases with age as follows: 3% of people of age 65–74, 17% people of age 75–84 and 32% people of age 85 or older have Alzheimer's disease [26]. Dementia has become so widespread as a disease that someone develops dementia every three seconds somewhere in the world. Alzheimer's Association has stated that without new prevention approaches, the number of people with Alzheimer's disease in the United States may rise from around 5 million currently to between 11 million and 16 million by 2050 and from roughly 26 million to more than 100 million globally [46]. It is clear that if advances in this field are not made, the world will be affected in a variety of ways. Although there is no cure for Alzheimer's disease at this time, early identification and proper classification can enhance a patient's quality of life through appropriate therapy and adaption. The progression of Alzheimer's disease can be reduced significantly if this illness is diagnosed properly [30]. That is the reason behind our work on Alzheimer's disease classification. We have considered deep learning for our work since over the last several decades it has shown exceptional results in areas such as image processing, facial recognition, voice and audio recognition and many others [13]. Among a variety of deep learning designs we choose to use the CNN which has already proven to be more efficient and accurate than previous deep learning networks for image and pattern recognition. The application of CNN in the classification and detection of Alzheimer's disease has grown considerably in the recent years. That is why we have used CNN and its different variations of architecture to classify the four different classes of dementia in our research.

1.2 Aims and Objectives

The main goals of our research are:

- Pre-process MRI data of ADNI dataset subjects, which consists of adults of various age groups, was used to build train and test models.
- Applying different variations of Recall-Net with InceptionV3, ResNet50 and VGG16 as backbone to classify 4 classes of dementia.
- Analyzing the outcomes of several architectures of CNN and determining the one with the highest accuracy in order to determine the best CNN architecture for 4-class classification.
- Comparing the best performance of Recall-Net with previously applied models.

1.3 Thesis Overview

- **Chapter 1:** This chapter introduces our entire research. This chapter also gives an overview of Alzheimer's disease itself. Here, the aims and objectives behind our work have been discussed basically.
- **Chapter 2:** The second chapter is about the background study of our research. The theoretical knowledge about the base topic of our work has been discussed here. Also, the literature review of numerous previous works has also been given.
- **Chapter 3:** In chapter three, a brief discussion about our workflow has been done alongside a detailed discussion about the entire data pre-processing method. Moreover we outlined the architecture details of Recall-Net in this chapter.
- **Chapter 4:** Our chapter 4 consists of the result of our proposed model and a discussion about it. This chapter is contained by different results we obtained from our 23-layers custom model using various pre-trained base models.
- **Chapter 5:** This chapter concludes our work with reference to future work.

Chapter 2

Literature Review

2.1 Literature Review

In [25], the authors have classified the different stages of Alzheimer's disease with the help of a Deep Convolutional Neural Network. Here, the structural MRI and functional MRI data of several subjects were used as the dataset. Here basically, the authors worked to differentiate the MRI of normal subjects from the MRI of a subject having Alzheimer's disease. Here, a unique deep learning-based pipeline has been used to distinguish. This unique pipeline was formed based on three essential modules. Those are pre-processing, data conversion and classification. This study has presented two pipelines using MRI and fMRI data, deep learning-based classifiers and extensive pre-processing modules Later, they have processed the dataset very carefully with the pipelines alongside a GPU-based computing CN has been used to attain shift-invariant from low to high feature. Here fMRI data has been used to predict Alzheimer's Disease through medical imaging. Here, the authors have claimed that the pipeline they implemented has demonstrated a higher accuracy than other studies. Like, 99.9% for MRI pipelines and 98.84% for MRI pipelines.

In [3], the authors have mainly focused on determining the accuracy of MTA scores on different types of dementia including AD, DLB and VCI. This MTA has a 5-point scale in which 0 means absent and 4 means severe. This score is presented based on the width of CSF spaces and the height of the hippocampal formation. With the help of the correlation between MTA and measure of Spearman's rho pathology and linear regression, the validity of MTA's local pathological changes can be calculated. The study also mentions that to distinguish among AD, DLB and VCI, Receiver Operator Curve analysis is used. This diagnosis is also done with the help of MTA. Here, the authors have claimed that MTA on MRI has demonstrated more accurate results in terms of AD than other forms of dementia like DLB or VCI. The authors have also mentioned that further studies are required to distinguish AD from other forms of dementia as some forms of dementia like frontotemporal dementia have similar MTA.

In [9], the authors have mainly worked on differentiating the regional FC between AD and MCI. The data set used here was the combination of two sets of fMRI data which were scanned using this [2] protocol. The subject's medical history and extensive neuropsychological test were also analyzed as well. In this study, data was mainly analyzed using a software, MELODIC. As a part of pre-processing, removing non-brain tissue, motion correction and spatial smoothing was done. Using the dual-regression method, a voxel-wise comparison of functional connectivity has been done. Nonparametric permutation tests have been carried out for finding differences between groups of 4-dimensional files obtained with the help of MELODIC. The authors have also claimed that the value of regional functional connectivity is numerically in the middle of patients of AD and controls. It is also demonstrated that subjects having AD have shown a lower value of FC than subjects having sMCI. Although no difference in regional FC has been shown in the case between AD patients and cMCI patients. Further studies may require to differentiate AD patients from cMCI patients.

Liu et al., in their paper [14], worked on the problem of early detection of Alzheimer's disease for preventive care and preparedness of patients before serious damage. This approach implements stacked auto-encoders for learning and softmax regression for prediction. But the difficulty in this approach lies in the fact that it uses sparse autoencoders as its ability to successfully learn the complex patterns is hugely handicapped. Their classification classes include Non-Alzheimer's subject or NC, the possibility of progression in MCI as ncMCI and cMCI and AD With 77 NC subjects and 169 MCI subjects, their overall result 47.42% inaccuracy, 65.71% insensitivity and 83.75% in specificity.

Ortiz et al., in their paper [24], discussed an ensemble of deep learning techniques for early detection of Alzheimer's disease. Their suggested method first segments the anatomical parts or patches of the brain called AAL by using SVM. Then each part of the brain is trained by a different deep belief network. In the end, a voting scheme chooses the final result. It is a variation of multitask learning where each deep belief network is tasked with learning a specific part of the brain. Their method can only classify between NC and AD from MRI and PET images. Between different categories of choices in different stages of AD from the ADNI dataset, the accuracy, sensitivity and specific scores are respectively 90%, 86% and 94%.

Suk et al, in their paper [15], worked on the diagnosis of AD in the MCI stage in a patient. This paper acknowledged the limitation of hand-crafted feature engineering of the segments of the brain. Thus, it takes a different approach using DBM as a way of determining the underlying hierarchical structure within the features. Learning the hierarchical feature representation, classifies the image using SVM. Even though DBM is a much better approach than handcrafted feature engineering, hierarchical representations do not capture the complexity within the data. This method classifies MRI and PET Scans of patients into two categories i.e. AD and MCI. From the 93 AD patients, 204 MCI patients of different stages and 101 NC patients, the proposed method helps to achieve an accuracy of 95.35% in AD/NC classification and 85.67% in MCI/NC classification.

Ding et al., in their paper [37]. used 18-fluorodeoxyglucose PET exclusively to image the brain or predicting Alzheimer's Disease as they suspected, it is the most effective imaging technique for early diagnosis. The Study used the 1002 patients of ADNI datasets and 40 independent studies. The data was split into 90 and 10 for training and testing respectively. The basic architecture of the deep learning model used here is Inception V3. With image data pre-processed into a 100 X 100 X 90-pixel grid and some other basic pre-processing, the image data was trained. The model proposed here can differentiate between AD, MCI and non-AD/MCI patients. Due to the very nature of the 18-fluorodeoxyglucose PET image, it gives an edge for the patient for early diagnosis specifically. This method had an 82% specificity score and 100% sensitivity score and 75.8 months earlier than a final diagnosis.

The authors Shen et al. [11] have approached Multi-Task, Stacked Auto Encoder and Multi kernel SVM methods Using the ADNI data set. Previously, authors used Mean Signal from PET scans and Grey matter Tissue Volume from MRI. But the results were poor. The author Suk and Shen's idea of using the method of Stacked autoencoder which is an Artificial neural network that consists of several layers of sparse autoencoder along with the hidden layer which is connected to the input of the successive hidden layer of a neural network which proved to be given much more improved accuracy in deep learning with noisy autoencoder from MRI images. In their paper they have used a deep architecture to find highly nonlinear and complex patterns in data [37] They have also used the Logistic sigmoid function to find the accuracy. This SAE method They have used has a unique characteristic which is to learn the pattern of nonlinear relationship of input. As SAE is a supervised Model, they used another output layer on top of the SAE model to initialize input data from the data set. SAE uses a pre-training step known as fine-tuning . As they used the SAE Method it helped the deep learning neural network model different from CNN. They had followed Zhang and Shen's work [32] using multi-tasks and Multi kernel SVM learning which is a, supported vector learning which helps to go through feature vectors parameters and kernel induced mapping to predict the accuracy from the class label, Mini-mental state examination which used for diagnosing dementia but they tried in their paper to train there with this same method to diagnose AD. In recent years the MMSE has proven tests for identifying AD stages and classifying them. From their paper, we can use some of their methods and combine them with our collected data and methods to get a more accurate result.

In 2013 [10] authors Shen et al. tried to get the best accuracy to classifying AD using the Stacked Auto Encoder, Multi-Task and Multi kernel SVM method by creating a new model for deep learning. However, in 2017 they Came up with another model which they called the sparse regression model mixed-effects and compared this model with the result of their previous research [32]. In their newest paper with another, they did not only analyze their previous result but also used a new approach mainly focusing on MRI, PET and DTI data from the ADNI data set. In this paper, they used Logistic regression from the DTI data set. Even though logistic regression is used mainly in machine learning to predict curiosity by giving dimensional brain imaging features some target values. They have used sparse controlled parameters and weight coefficient matrix [10]. Then They Used deep ensemble sparse regression which is multiple neural network models to reduce the variety of prediction and general errors that cause failure to gain the most accuracy, mainly by assigning random weights to neural network layers on genetic data. Even though they used mainly these ML and deep learning methods to train their model they also explored some experimental methods such as the MOLE baseline regression model, Which is a combined method like DeepESrnet to achieve mean classification in AD. We think their previous paper's approach was more satisfying and easier to classify Alzheimer's, Rather than their experimental approaches.

In this paper [7], a whole new multimodal data fusion alongside a classification method was proposed. This method is based on the kernel combination for Alzheimar's Disease and Mild Cognitive Impairment. This particular method provides a completely unique way to combine heterogeneous data rather than the direct feature concentration method. The proposed method has used different weights for different data modalities. A total record of 202 subjects were taken into consideration here. There are 51 AD patients, 99 MCI patients and 52 other healthy subjects that were taken into consideration for the development and validation of the proposed multimodal classification method. To evaluate the performance of the method here, the researchers have used a 10-fold cross-validation strategy to calculate the accuracy of classification. One advantage of the proposed method is it can be easily and efficiently integrated into conventional SVM. The method achieves a relatively high accuracy for AD classification and MCI to AD converters classification which is 93.2% and 91.5% but is very impressive in terms of MCI classification which is 81.8%. This paper only considers the baseline data of ADNI patients. Further improvements can be made by considering both baseline and longitudinal data of the patients.

In this paper [5], an automated analysis method of the whole brain is proposed. It is claimed that this method is able to identify patients with AD from healthy elderly subjects. A total of 16 patients with AD were taken into consideration for preparing this paper. The 3D T1-weighted MR images of each patient were considered and given into ROI. This method is based on a structural MRI and it works automatically using certain regions of interest. A feature parameter is used which shows the discrimination between AD patients and healty ones. Here it is basically focused on the distribution of gray matter in the brain for the feature parameter which is estimated by the Gaussian mixture model. Here only a2 is used because it is more invariant with respect to other acquisition parameters and using it alone would make comparison of datasets acquired with different sequences. Here the SVM algorithm was used to classify the subjects. This classifier maps the input parameter to the feature space non-linearly. This non-linear mapping in this case was a radial basis function. Bootstrap resampling methods were also used to select test and training data to evaluate the accuracy of classification. In this paper, two other alternative approaches for cross-validation methods was proposed namely leave-one-out and n-fold cross-validation. Here, the researchers have come up with a 94.5% correct classification for AD. But this method does not really work for some exceptional cases. The SVM has not been able to classify at least 4 subjects correctly. Further improvement can be made to resolve these aspects.

Farooq et al. in their paper [27], implemented a 4 class AD classifier ranging from CN, LMCI, MCI to AD on MRI scan from ADNI dataset patients using GoogleLeNet, ResNet-18 and ResNet-152. The dataset they included used 33 AD, 22 LMCI, 49 MCI and 45 healthy patient's scan periodically over one, two and three years. Then image augmentation was applied along the horizontal axis of the brain because of the symmetry of the human brain. Thus 9506 images for each type of classification were acquired and a total of 38024 images combining all the types. After

65% train data, 25% test data and 10% validation data split, the model was Xaviar Initialized and was feeded through three different models GoogleLeNet, ResNet-18 and ResNet-152 to compare results. Performance of each model was evaluated in terms of accuracy, sensitivity, specificity and precision. The acquired results by this method was for GoogleLeNet 98.88% accuracy, for ResNet-18 98.01% accuracy and for ResNet-152 98.14% accuracy. Even though their models were very fundamental, their technique of consistent data selection among all the patients and augmentation techniques to increase the volume of data is very effective.

The paper [28] authored by Hon et al. proposed transfer learning of deep CNN model VGG and Inception to solve the problem of quality and consistent neuroimaging data with 10 times smaller dataset resulting in high accuracy. The authors used OASIS cross-sectional MRI dataset of 416 subjects where 200 were chosen randomly from the CN group and AD group. The grouping was done by CRD ranging from 0 to 2 where 0 means CN and greater than 0 value implies AD. Further image processing was done to make the data compatible with VGG16 and Inception V4 image size to 150X150 and 299X299 respectively. After an 80-20 percent train test split the data were trained using two different pretrained weight initialized architecture VGG-16 and Inception V4 for 100 epochs. The result shows Inception V4 based transfer learning has the highest accuracy of 96.25% where VGG16 based transfer learning has 92.3% accuracy. In our opinion, the insight of using transfer learning to solve the problem of small datasets in neuroimaging and leveraging the pretrained state-of-the-art models like VGG16 and Inception V4 is the reason for the excellent result. More improvement could have been made by testing the transfer learning approach with other pretrained models and more comprehensive data.

In their paper [29], Islam et al. presents a novel deep learning architecture inspired by Inception V4 to classify multiple stages of AD ranging from Nondemented, very mild AD, mild AD to moderate AD. They chose OASIS dataset's T1-weighted MRI scans of 461 subjects. Each of the subjects has 3 to 4 scans whose stages vary from nondemented to moderate AD. After applying some data augmentation techniques like scaling and reflection, the images were resized to 299X299X3 shape. Then the dataset was split into 70% training data, 20% test data and 10% validation data. As mentioned earlier they have modified the Inception V4 architecture, they actually changed the filter size of Inception module B and C from 1154 to 1152 and 2048 to 2144 respectively. After training with this modified architecture, they have achieved accuracy of 73.75% after 10 epochs of training. They also tried their modification of GoogleNet. In their proposed method, they changed the filter size in different parts of the module to fit the MRI data size. However, in our opinion it was redundant to change the filter size so slightly to achieve these results as the reasoning was not clear and accuracy increase due to the modification is not reproducible.

2.2 Convolutional Neural Network

A CNN which is also known as ConvNet, (see Fig 2.1), is actually a DL algorithm that takes images as input, learnable weights and biases to various aspects in the picture and then identifies them [54]. A ConvNet requires much less pre-processing than other classification techniques [13]. The input to a CNN is generally an order 3 tensor. For example, an image with H rows, W columns and 3 color channels (R, G, B) [33]. After one input layer, one or more hidden layers and one output layer make up the structure of a neural network [51]. CNN's input passes through a variety of processing steps in order. A layer can be a normalizing layer, convolution layer, pooling layer, loss layer, fully connected layer and a variety of other processing stages [33]. The hidden layers of the CNN are typically made up of a series of convolutional layers. ReLU is one of the most popular activation functions and it's generally followed by pooling layers, normalizing layers and fully connected layers. More layers and bigger networks are required for the effective usage of convolutional neural networks [44]. Unlike traditional neural networks, each layer's neuron is only linked to nodes in a specific region known as the local receptive field here, rather than all of the nodes in the preceding layer [54]. After convolutional layers, pooling layers are frequently employed. The convolutional layer applies a set number of different filters to obtain the complex features of input images. The pooling layer is the down-sampling layer to reduce the feature dimensions of the input [44]. These layers have been created in order to decrease the scale of feature maps and simplify the information. After several alternating convolutional and pooling layers, the fully connected layers are followed to compute the class scores, which can be one or more layers [44]. This section of CNN combines and aggregates the most important data from CNN's various procedures. As a result, these layers provide the input data's feature vector, which can be used for machine learning tasks like classification and prediction [48]. The softmax classifier is the final fully connected layer and it determines the likelihood of each class label over N classes. The width, height and depth of these layers are all three dimensions [54]. It's mostly used for classification, image processing, segmentation and certain other auto-correlated data. Text analysis, video recording, natural language processing, text categorization, audio recognition and other AI systems are some of the additional applications [49].

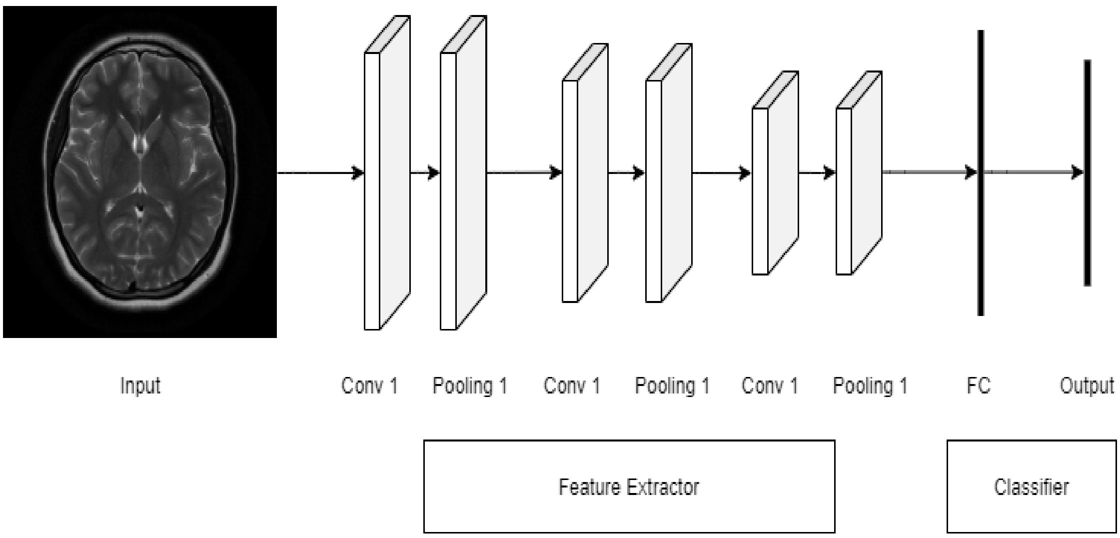

Figure 2.1: Layers of Convolutional Neural Network

2.3 How Different Layers of CNN works:

2.3.1 Input Layer

The leftmost layer of a neural network is the input layer (See Fig 2.1). Initially, the images we have collected from our dataset have to be fed as inputs to this layer of CNN. For greyscale images, the input layer has the same single map three times and for color images, the number of maps is three [41]. For every image, we get three dimensions (width x height x depth). These dimensions together are called pixel values. The height and width denote the size of the image taken as an input. The depth actually means RGB channels [42]. For example, in our research, we have used images of 224 x 224 x 3-pixel values. As the input layer cannot function in three-dimensional pixel values of input images, to make the layer function properly, the width and height of the image should be multiplied with each other and converted into a single column dimension. For example, an image having 224 x 224 dimensions should be converted to 224x224=50176. The dimension of input in the input layer will look like this (50176, n), where n is the number of training examples.

2.3.2 Convolution Layer

The convolutional layer is the most basic component of a Convolutional Neural Network. These layer parameters are constructed up of K learnable filters (also known as kernels) each of which has a nearly square width and height [33]. These filters are small in terms of spatial dimensions yet they cover the whole depth of the volume. The depth of the picture in CNN inputs refers to the number of channels in the image. RGB pictures have three levels of depth each of which corresponds to a channel [13]. This layer moved each of the K kernels over the input region, followed by element-wise multiplication and summation. The outcome of the computation is stored as a two-dimensional activation map. The output volume is obtained by stacking the activation maps with the depth dimension array (See Fig 2.2) [33]. As a result, each item in the output volume is the result of a neuron that only looks at a little fraction of the input. When a given sort of feature is observed at a specific spatial point, the network learns specific filters that activate. In CNN, each neuron is only connected to a small part of the input volume. The receptive field of a neuron is the size of the immediate region [8]. In case of our dataset, the input volume has a size of 224x224x3. Each image has a width of 224 pixels, height of 224 pixels and depth of 3. If we consider a receptive field size of 3x3 with depth 3, there will be 3x3x3 = 27 connections to input volume. On the contrary, keeping the receptive field size same and increasing the dept to 16, we will get 3x3x16 = 144 connection to input volume.

Depth: In our previous section of the convolution layer, we have come to know about the depth of the input image of CNN. The depth depends on the number of channels the input images have. In the case of the depth of the output volume, it is produced by the convolutional layers which can be set

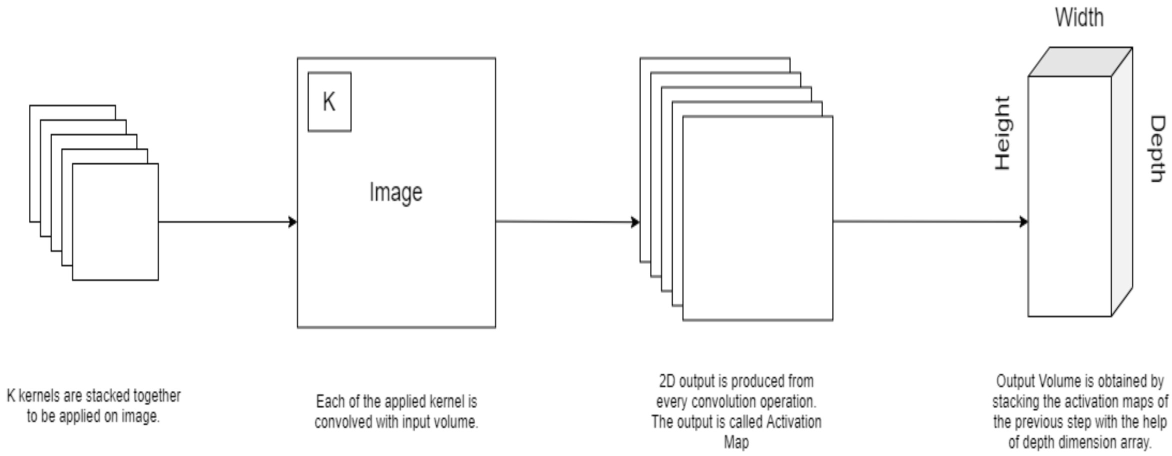

Figure 2.2: Steps of Convolution Layer

by hand through the number of neurons within the layer to the same area of the input [40]. This is also prevalent in other forms of ANNs, where all of the neurons in the hidden layer are directly connected to every single neuron already. But it reduces the model's capabilities of pattern recognition [17].

Stride: In CNN, parameters can be decreased. Hence, the number of side effects to the network also decreases. Stride helps CNN reduce these side effects through reducing parameters. Overlapping in any region can also be manipulated by controlling stride [13]. Stride can also be defined where the depth around the spatial dimension of the input for placing the receptive field. Lesser value of stride results in a heavily overlapped receptive field which produces large activations. On the contrary, a greater value of stride will reduce the overlapping amount and produce a lower spatial dimensional output [17]. For example, in Fig 2.3, we have taken a 7×7 image. If the filter is moved to one node each time, the output 5×5 image will have overlapping at three left matrices. On the other hand, if the filter is moved and made each stride 2, both the number of overlapping and output size will be decreased.

The Following Equation 2.1 gives the output size O for the given image of N×N dimension and F×F filter size.

$$O = 1 + \frac{N - F}{S} \tag{2.1}$$

Here, N = Input size, F = Filter size, S = Stride size

N = Input size

F = Filter size

S = Stride size

Padding: The information that resides at the border of input images often gets lost during the convolution layer. Zero-padding is quite useful for resolving this issue [13]. It is a simple process of padding the border of the input and is an effective method to give further control as to the dimensionality of the output volumes [17]. Hence, zero-padding has another benefit of managing the output size. For example, if we take N=7, F=3 and stride=1, we will get an output image of size 5 × 5 which actually a shrunk size of 7 × 7 input. Therefore, by adding zero-padding, the output we will get will be 7 × 7. Equation 2.1 is without zero-padding. The following Equation 2.2 is a modified equation including zero-padding.

$$O = 1 + \frac{N + 2p - F}{S} \tag{2.2}$$

Here, P = Number of zero-padding layer.

Padding helps to prevent network output size from shrinking with depth. Therefore, it is possible to have any number of deep convolutional networks [13].

2.3.3 Pooling Layer

A pooling function substitutes a net's output at a particular point with a summary statistic of adjacent outputs. This layer seeks to gradually lower the representation's dimensionality, reducing the number of parameters and the model's calculation complexity [21]. In the context of image processing, it's compared to lowering the resolution. The number of filters is unaffected by pooling. One of the most popular forms of pooling techniques is max-pooling [13]. They usually take the shape of max-pooling layers with kernels with a dimensionality of 2 × 2 and a stride of 2 along the spatial dimensions of the input in most CNNs. The activation map is reduced to 25% of its original size while the depth volume is maintained at its original size [17]. In max-pooling, the number 2 × 2 is one of the most commonly used sizes. When pooling is conducted in the top-left 2 × 2 blocks, it advances 2 and focuses on the top-right portion, as seen in Fig ***. This indicates that stride 2 is utilized in the pooling process. Because it does not retain the location of the information, down-sampling should only be used when the existence of information is critical [13]. The pooling layer takes the followings as input:

Volume of input images = Input image width (W_i) × Input image Height (H_i) × Input image depth (D_i).

Two hyperparameters = Spatial extent (F), Stride (S) as inputs.

Then, produces a volume of size Output image (W_o) × Output image (Ho) × Output image depth (D_o)$where$:

$$W_o = (Wi - F)/S + 1 \tag{2.3}$$

$$H_o = (Hi - F)/S + 1 \quad (2.4)$$

$$D_o = D_i \quad (2.5)$$

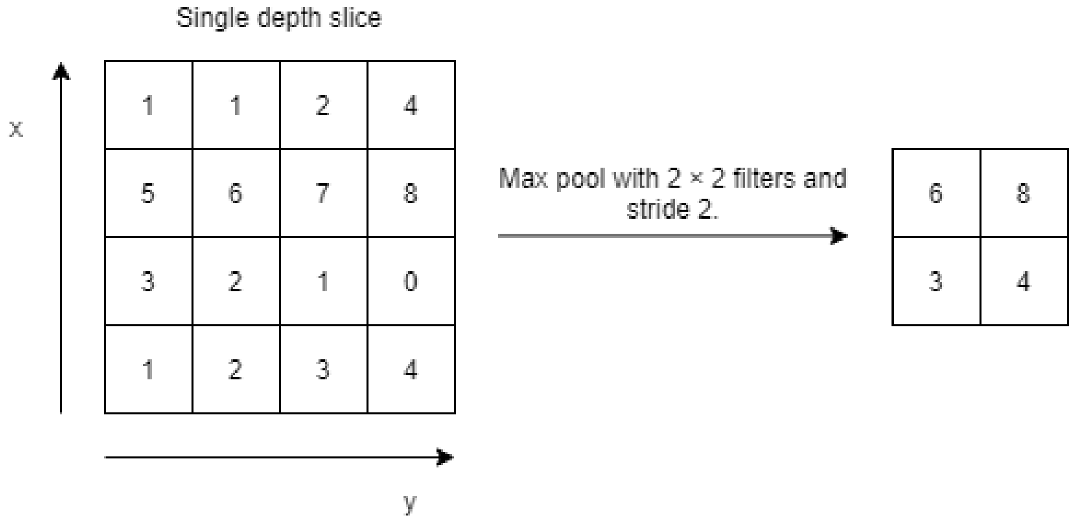

Figure 2.3: Stride 2 Utilization in Pooling Process

2.3.4 Fully Connected Layer

The fully connected layer is comparable to how neurons in a conventional neural network are organized. As a result, each node in a fully connected layer is connected to every node in the preceding and subsequent layers [13]. Each of the nodes in the final frames of the pooling layer is linked as a vector to the first layer from the fully connected layer, as shown in this diagram. One of the most significant disadvantages of the fully linked layer is that it contains a large number of parameters, which results in a time-consuming and complex calculations [38]. To reduce the complexity, some nodes and edges can be eliminated. The dropout technique can satisfy the absence of these nodes and edges while keeping the calculation complexity constant [18].

2.4 VGG

One of the most well-known models for image categorization and object identification is the VGG neural network. It's also recognized for its unwavering quality. While Inception CNN is concentrated on lengths and smaller window sizes, VGG focused on depth, which is another important characteristic of CNNs. It is made up of 19 layers. It stacks three 3x3 convolutional layers on top of each other to improve depth. The decrease of volume size is accomplished through max pooling.

2.4.1 The Architecture of VGG16

The input to the conv1 layer in this architecture is a 224* 224 RGB image. A succession of convolutional layers is applied to the image, each with a very small receptive field, where 3*3 is the smallest size that can capture the concepts of left/right, up/down and center. It also utilizes 11 convolution filters, which are a linear adjustment of the input channels followed by non-linearity in one of the configurations. For 3*3 convolution layers, the convolution stride is set to 1 pixel and the spatial padding of convolution layer input is set to 1 pixel to maintain spatial resolution after convolution. After that, there are five max-pooling layers that follow part of the Conv. layers and do spatial pooling. Note that not all Conv. layers are followed by max-pooling, but in VGG16 we can see that Conv. layers are. As a result, Max-pooling is done across a 2* 2-pixel frame using stride 2. Then, following a recent stack of convolutional layers of variable depth in various designs, three Fully-Connected layers are added. The first two each have 4096 channels, but the third uses 1000-way where each one is for each class of ImageNet. The last layer here is the soft-max layer, which comes after that. The fully linked levels are configured in the same way in all networks. All hidden layers use the ReLU for non-linearity. It's also worth noting that, with the exception of one, none of the networks utilize LRN, which improves performance on the imageNet dataset but increases memory usage and computation time. Fig. 2.4 illustrates the VGG16 network.

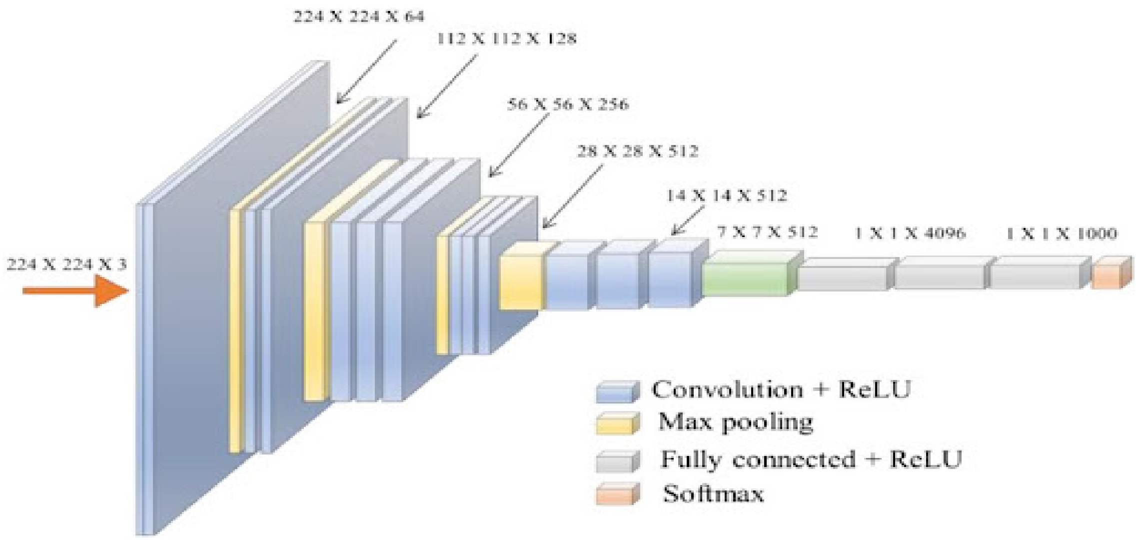

Figure 2.4: VGG16 Architecture.

2.5 Inception

In the evolution of CNN classifiers, the inception network was a key innovation. It is a convolutional neural network design with an extremely deep level of complexity. Following its development, the convolutional layer of the most popular convolutional neural networks was layered deeper to increase performance. It's a complicated design since it utilizes a variety of techniques to increase accuracy and speed. The network features one-to-one convolution. It also utilizes global average pooling [55] rather than completely linked layers. Convolutions of various sorts and sizes are used for the same input. The visible portion of an image might vary significantly in size. It's tough to pick the right kernel size because of the wide range. Bigger kernel size is preferred for globally distributed information, whereas a lower kernel size is preferred for locally distributed information. Furthermore, for highly deep networks, over-fitting is a major issue. The network has a hard time passing gradient updates. In these situations, inception is useful. Multiple-sized filters work on the same level in the beginning. Convolution on input is performed using three different sizes of filters (1x1, 3x3 and 5x5) [19]. The number of input channels is limited by adding an additional 1x1 convolution. It decreases the model size, which reduces the over-fitting problem [55]. Prior to the 1x1 convolution, max pooling is implemented. After being concatenated, the outputs are passed into the next module. Inception-v1, Inception-v2, Inception-v3, Inception-v4 and Inception-v5 have all been implemented due to the rapid advancement of the inception model.

Figure 2.5: Naive version of Inception and Dimension reductions of Inception

2.5.1 Inception v3 Concept

One of the well-recognized models that have achieved greater accuracy on the image-net datasets is inception v3. This model represents the result of numerous concepts developed over time by a range of researchers. This model was introduced in 2015 [20]. Inception v3 is a modified architecture of the previous inception of CNN architectures. Inception v3 architecture mainly focuses on the less expenditure of computational power. Inception Networks (GoogLeNet/Inception v1) have been shown to be more highly scalable than VGGNet, as well as in terms of the number of parameters created by the network and the economical incurred (memory and other resources). When making modifications to an Inception Network, it's important to keep the computational advantages in mind. As a result, because of the uncertainty about the new network's efficiency, adapting an Inception network for diverse use cases becomes difficult. Several network optimization approaches have been proposed in an Inception v3 model to remove the restrictions and make model adaption simpler. Here in this model, the approaches that are being used are Factorized convolutions, regularization, dimension reduction and parallel computation.

2.5.2 The Architecture of Inception v3

There are 5 steps that are used to build an Inception v3 architecture. This architecture was built by following these steps gradually. The first step is to use Factorized Convolutions, which is to lower the number of parameters in a network and therefore improves computing performance. It also monitors the network's efficiency.

Then there are Smaller Convolutions, in which they replace bigger convolutions with smaller convolutions, resulting in faster training. A 5 *5 filter, for example, contains 25 parameters; replacing it with two 3 3 filters reduces the number of parameters to 18 (3*3 + 3*3) Fig 2.6.

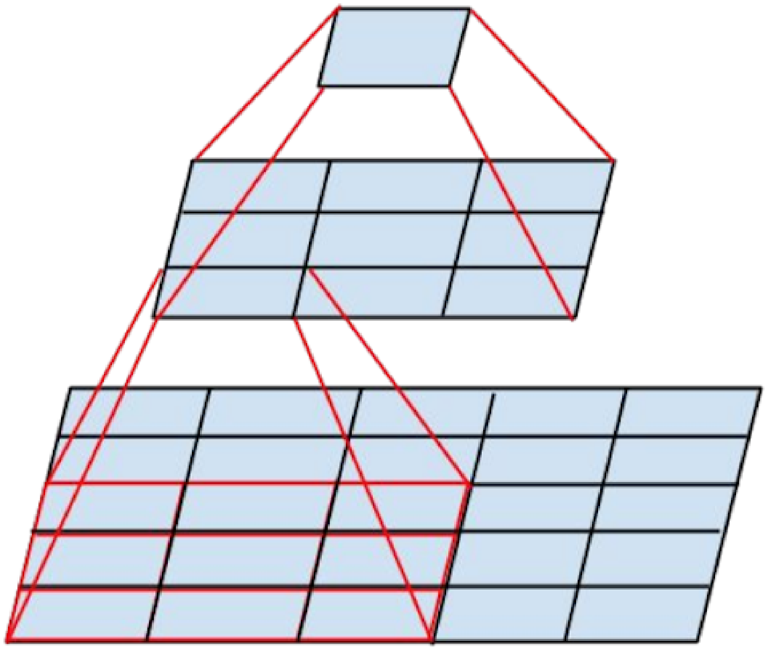

Figure 2.6: A 3X3 convolution

The third stage is Asymmetric convolutions. A* 3 convolution is sometimes substituted by a 1*3 convolution, preceded by a 3*1 convolution. The number of parameters would be slightly higher than the asymmetric convolution which is proposed if a 3*3 convolution was substituted with a 2*2 convolution.

In the fourth step during training, an auxiliary classifier which is a small CNN placed between layers, with the loss contributed to the main network loss figure[6]. Even though auxiliary classifiers used to be implemented in GoogLeNet for a deeper network, but here an auxiliary classifier is used as a regularizer in Inception v3 figure 2.7.

Grid size reduction is the final phase, which is generally accomplished by pooling operations.Figure 2.8.

After following all the steps, we can get the final architecture of inception v3 Figure

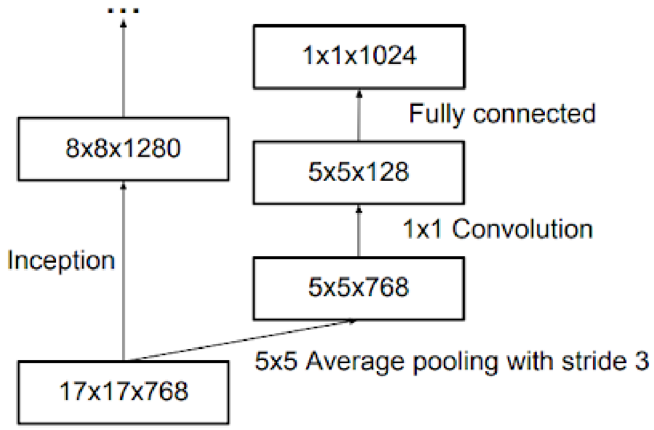

Figure 2.7: Auxiliary Classifier in Inception v3

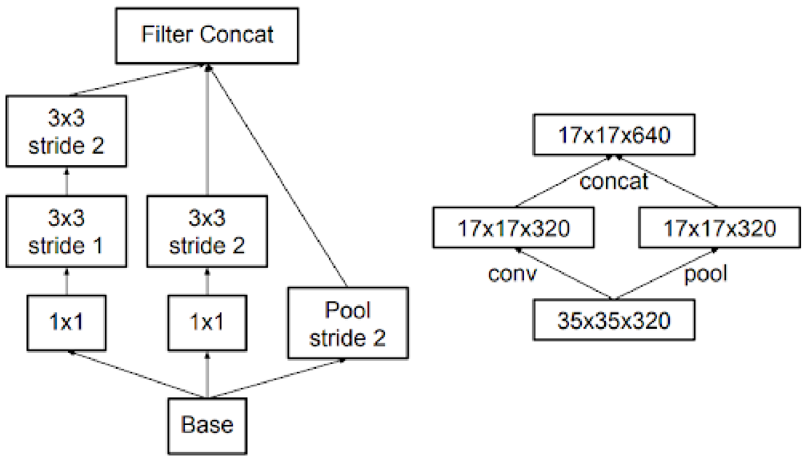

Figure 2.8: Grid Reduction Technique

2.9. This final architect simply shows that after using all the necessary steps one by one we can make the inception v3 model which is now one of the most used convolutional neural network architecture in classification cases.

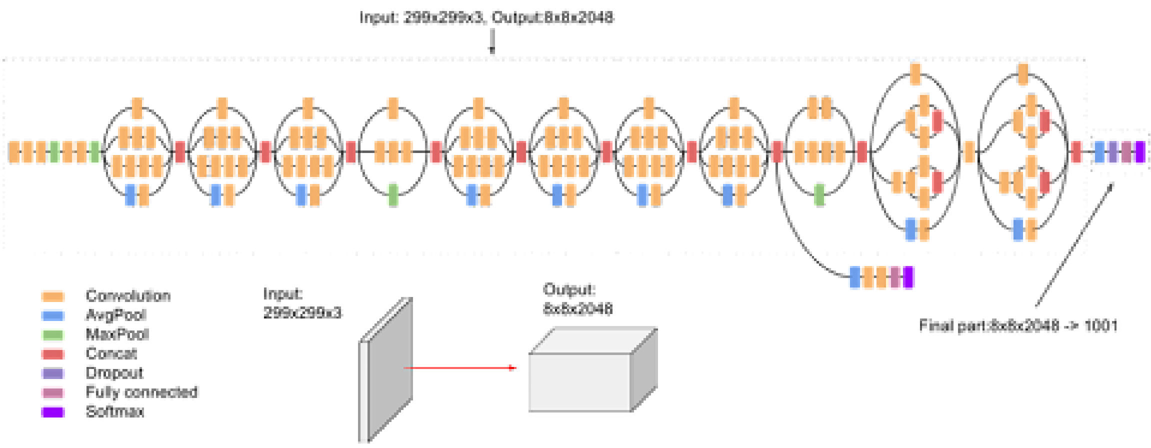

Figure 2.9: Inception v3

2.6 ResNet

ResNet is one of most recent state-of-art CNN architecture for image classification. It has been extremely competent in solving complex CNN problems. As a rule of thumb in deep learning, we have come to understand that the more layers the better. However, previous to this architecture, increasing layers after a point would cause decreased model performance. This decrease is caused mostly because of vanishing gradient problem. This has been a major concern for training very deep networks. To tackle this problem, in the paper[22] residual block was introduced. It is showen in Fig. 2.11[36]. The ResNet is made with several of these residual blocks. At its core, residual blocks use 'skip connection' shown in Fig. 2.10[22]. This skip connection helps the network to skip some connections thus allowing gradients to flow through alternative shortcuts. This allows the ResNet to add more layers to recognize more and more complex patterns in the image.

2.6.1 ResNet50 Concept

ResNet-50 is a 50-layer deep pre-trained network [33]. There are 48 convolutional layers illustrated in Fig. 2.12 [45], one max-pooling layer and one average pooling layer in this architecture. Shortcut connections have been added to its design, skipping three layers with non-linearities like ReLU and one convolution layer. This aids in the prevention of gradient vanishing in deep networks. ResNet50 was chosen above other models on the medical imaging dataset [53] because of its higher performance. All the layers and associated parameters are shown in Table 2.1

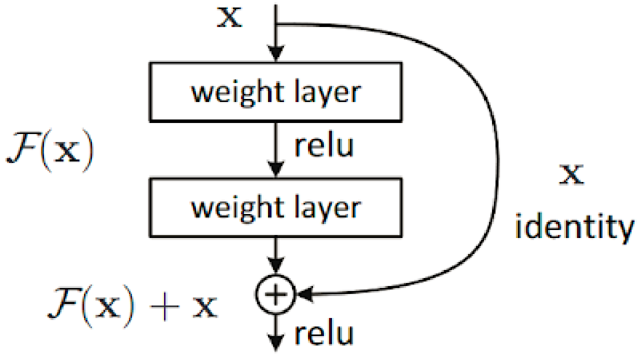

Figure 2.10: Residual learning: Skip Connection

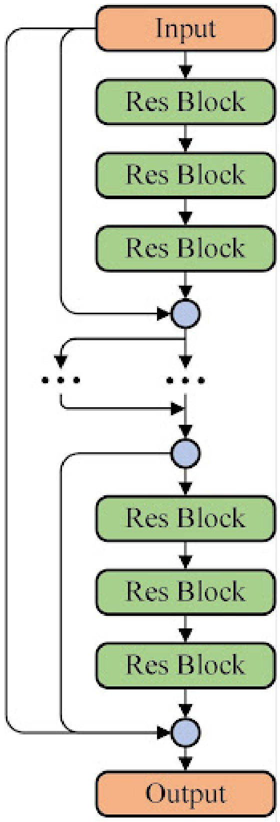

Figure 2.11: Residual learning: Building Block

Table 2.1: ResNet-50 Layers and Associated Parameters

Type	Filter Size	Channel number	Input Size	Times
Conv1	7 X 7	64	156 $\times$15$\times$	1
Max pool	3 $\times$3	-	77 $\times$7 $\times$ 4	1
	1 $\times$1	64	38 $\times$3 $\times$ 4	
Conv2	3 $\times$3	64	38 $\times$3 $\times$ 4	3
	1 $\times$1	256	38 $\times$3 $\times$ 56	
	1 $\times$1	128	19 $\times$1 $\times$ 28	
Conv3	3 $\times$3	128	19 $\times$1 $\times$ 28	4
	1 $\times$1	512	19 $\times$1 $\times$ 12	
	1 $\times$1	256	10 $\times$1 $\times$ 56	
Conv4	3 $\times$3	256	10 $\times$1 $\times$ 56	6
	1 $\times$1	1024	10 $\times$1 $\times$ 024	
	1 $\times$1	512	5 $\times$ $\times$ 12	
Conv5	3 $\times$3	512	5 $\times$ $\times$ 12	3
	1 $\times$1	2048	5 $\times$ $\times$ 048	
Avg pool	7 $\times$7	-	5 $\times$ $\times$ 048	1
Flatten	-	1	5 $\times$ $\times$ 048	1
Output	-	1	51200	1

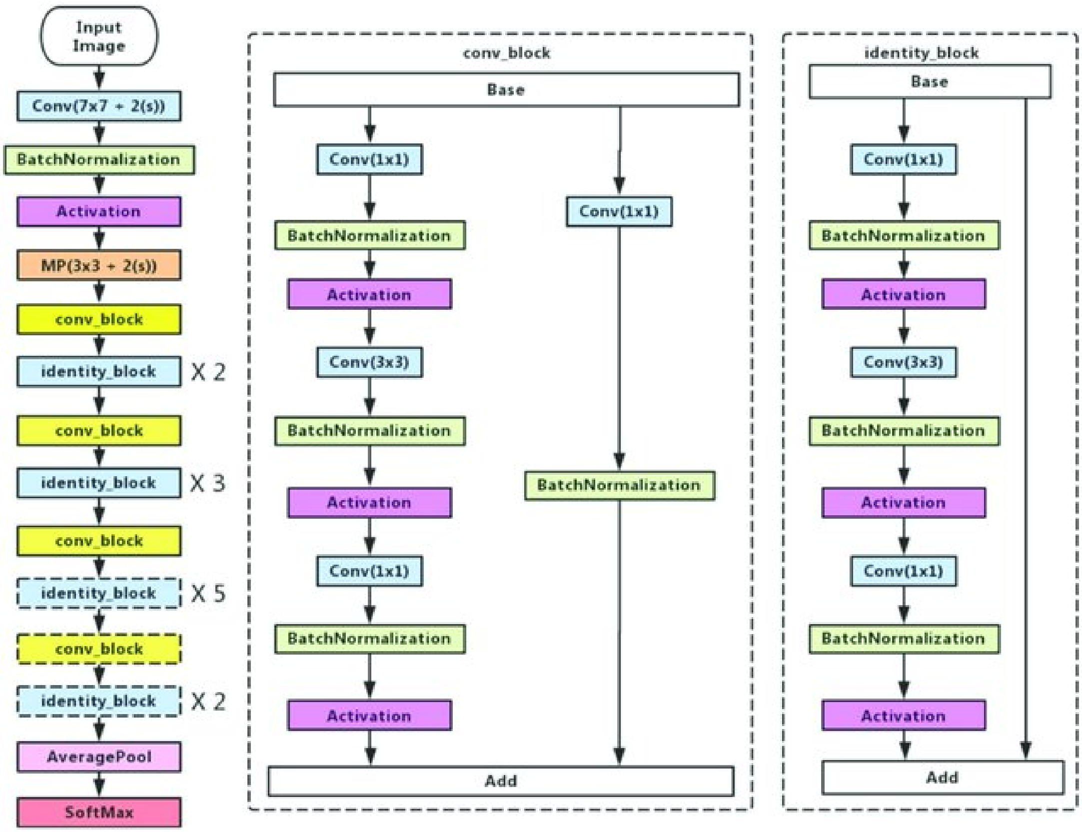

Figure 2.12: ResNet50 Architecture

Overall, this chapter contains the theoretical knowledge of our work along with our overviews about various previous works in related fields. Firstly, we have discussed about CNN and its architecture. Brief description about various layers of CNN has been given. This shows how various layers work inside the convolutional neural network. Moreover, we have discussed about various architectures of CNN and given a basic concepts about those layers. Lastly, we have provided our overview about numerous number of previous done researches in related fields.

Chapter 3

Methodology

This chapter consists of a summary of our data set acquisition process. Here, we have provided description of the process of data collection, pre-processing and data labeling and a brief analysis of the data set.

3.1 Research Framework

In the flowchart (see Fig 3.1), we have outlined our working approach from start to finish. We first obtained our data from the official ADNI website [6]. Afterward, we did some research on the data set. From there, we gathered cross-sectional data on Alzheimer's disease and adjusted the data as per our requirements. This is where we present a new image pre-processing approach in our data pre-processing section. In the data pre-processing section, we have done image resizing, denoising and ROI based slicing and skull stripping so that it can facilitate our new pre-processing method. After that, we split our data into two groups and labeled them as train data and test data respectively. We moved on to the next stage, which was to use our dataset to train our proposed model as well as other CNN models. Then we compare the results of each CNN model's accuracy, precision-recall and confusion matrix to our designed model. At last, we compare the best performance of any of our implemented CNN models with the models implemented in previous research.

3.2 Research Objectives

Many approaches and research have been done using various algorithms and models on Alzheimer's disease, as it is a non-curable disease. Techniques like machine learning and data mining algorithms including K-NN, DT, rule induction, Naive Bayes, GLM and deep learning algorithm are applied on ADNI dataset to classify the five different stages of the AD. Therefore, our objective here is to use deep learning-based neural network approaches and trying out different architectures of CNN like VGG, ResNet and Inception alongside our proposed model to do the 4-class

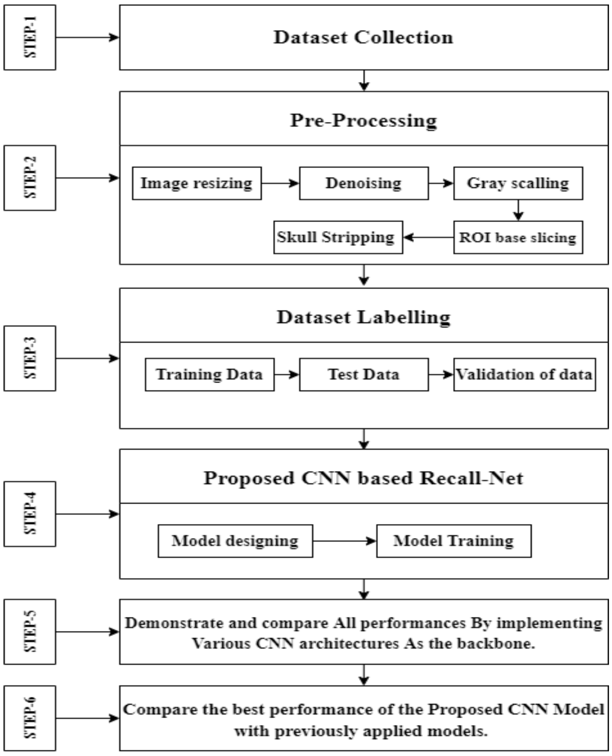

Figure 3.1: Block Diagram of Our Proposed Model

classification and achieve the highest success rate in the 4-class classification and detection of Alzheimer's disease. We want to classify MRI of different age groups with NC, EMCI, MCI and AD to see how this disease affects differently among these 4 classes. Our fundamental contributions in this paper is to pre-process the MRI data of ADNI dataset, applying different architectures, analyzing the outcomes and determining the one with the highest accuracy for 4-class classification.

3.3 Data Collection

For our research, The data set has been collected from an open-source medical data set called ADNI [6]. This data set contains 6400 subject's T1 weighted raw MRI scans which were in Dicom image format. This data set consists of selected axial views of MRI Scans. The age of the patients are between 40 to 97 years. The dataset consists of 53% female and 47% male subjects. In this data set, each subject has four T1 weighted MRI scans based on their visits and their MRI scanning session. In addition to that, 3200 subjects were non-demented from the same age group of 40 to 80 years. Lastly, 2100 subjects who are over the age of 70 have been diagnosed with AD traversing from very mild to moderate levels of AD.

3.4 Data Pre-processing

For our research purpose, as we mentioned before a new technique of image visualization method of pre-processing has been done on the MRI images. This pre-processing method has been used to get more tuned and accurate brain images so that when the customized CNN model gets implied on the data-set, it would get better accuracy. This image visualization method of pre-processing is also a self-design method of image pre-processing for our benefit. Here we tried to manually pre-process the raw images depending on the ROI. In Fig 3.2, three different views Sagittal view, Coronal view and Axial view of a subject's MRI scans are shown. Though the data-set consisted of 3 views of brain scans, our research and proposed method mainly need the 2D axial view of the brain. Therefore, the axial view images were selected only. First, The scan images were divided based on ROI. Then on those images, we implemented gray scaling and masking to get the best-tuned image. After a point, a threshold-based classification has been done to the images using the Selkow algorithm. Where pixel values are compared to a pre-determined threshold value[33]. During the thresholding process, each pixel value is compared to a preset threshold value. We set the value of the pixel to the maximum, which was 254 since it was not less than the threshold. To get only the inner portion of a brain without a skull, we have done image conversion based on 4 connected components to do the binary conversion. In Fig 3.3, we have shown the difference in skull images before and after skull stripping. We have done skull stripping because for research purposes the inner portion of the brain is most important [34]. Here we can see with our normal visual capability the images are very well tuned and clearer than the previous scans of the ADNI data-set. There you can see the main regions where AD occurs are very clear as well as the T1 white fluid, which is primarily responsible for AD, are also visible. Here in Fig 3.4, shows the skull condition of 4 different classes of AD when the data pre-processing method has been implied on the data set. At the end of pre-processing, we have labeled our whole processed data into train and test data and validation data.

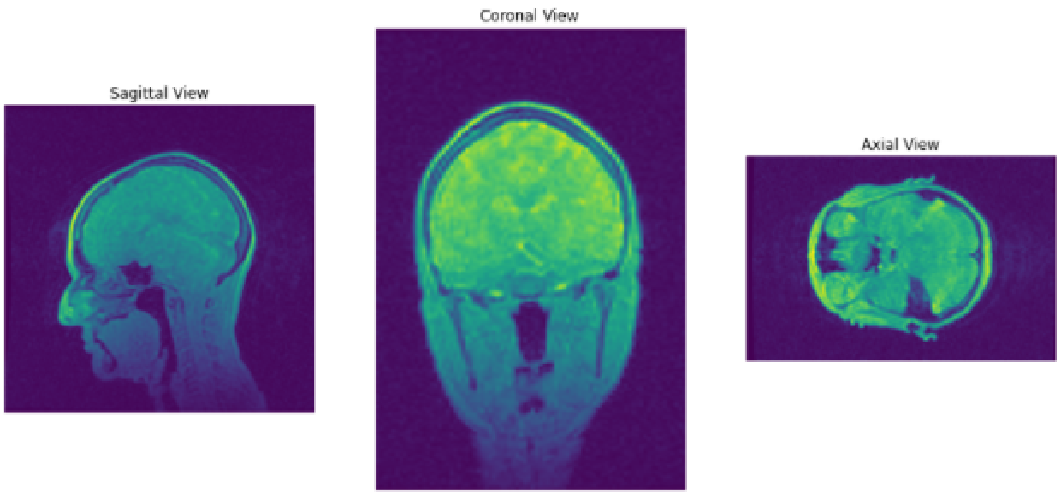

Figure 3.2: Three Different Views of Subject Skull.

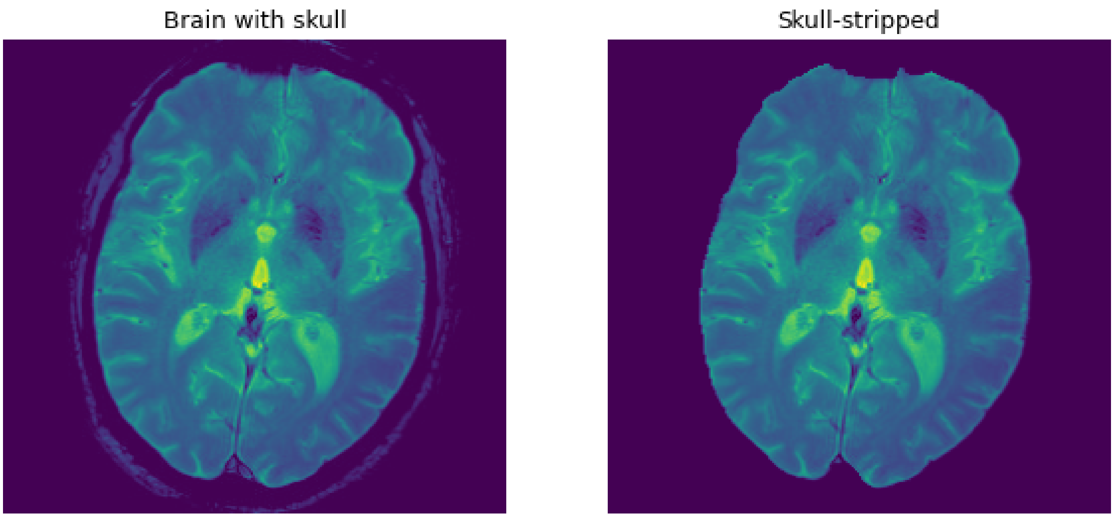

Figure 3.3: MRI of Skull Before and After Skull Stripping.

3.4.1 Data Labeling

As we have previously said, there are a total of 6400 MRI image scans in our ADNI dataset. Also as previously mentioned, we have implemented binary conversion for our proposed neural network model. To accomplish so, we first processed the data using our customized image-processing method and then classified it into four categories of AD phase which are NC, EMCI, MCI and AD. It's important to remember that for higher accuracy in binary conversion, the quantity of data in both the test and train parts of the conversion should be equal. After that, we split the data 80% for train data and 20% for test data to produce a four-class categorization of the cases (See Table. 3.1). Furthermore, we have 5120 MRI scans in the test dataset and 1280 MRI scans in the training dataset, all of these are split into four classes.

Table 3.1: Size of Data Set in Four Different Classes

	Non-demented	**Very Mild Demented**	**Mild Demented**	**Moderate Demented**
Train	2560	1791	717	52
Test	640	449	179	12

3.5 Recall-Net: Our Proposed Model

The name "Recall-Net" originates from the main symptom of Alzheimer's Disease itself. Patients suffering from AD can't recall their memories which initially start from trivial things like key and money to more serious things like the face of family members, how to eat, how to take bath or go to the bathroom. We hope that our proposed model "Recall-Net" will help patients to recall their precious memories.

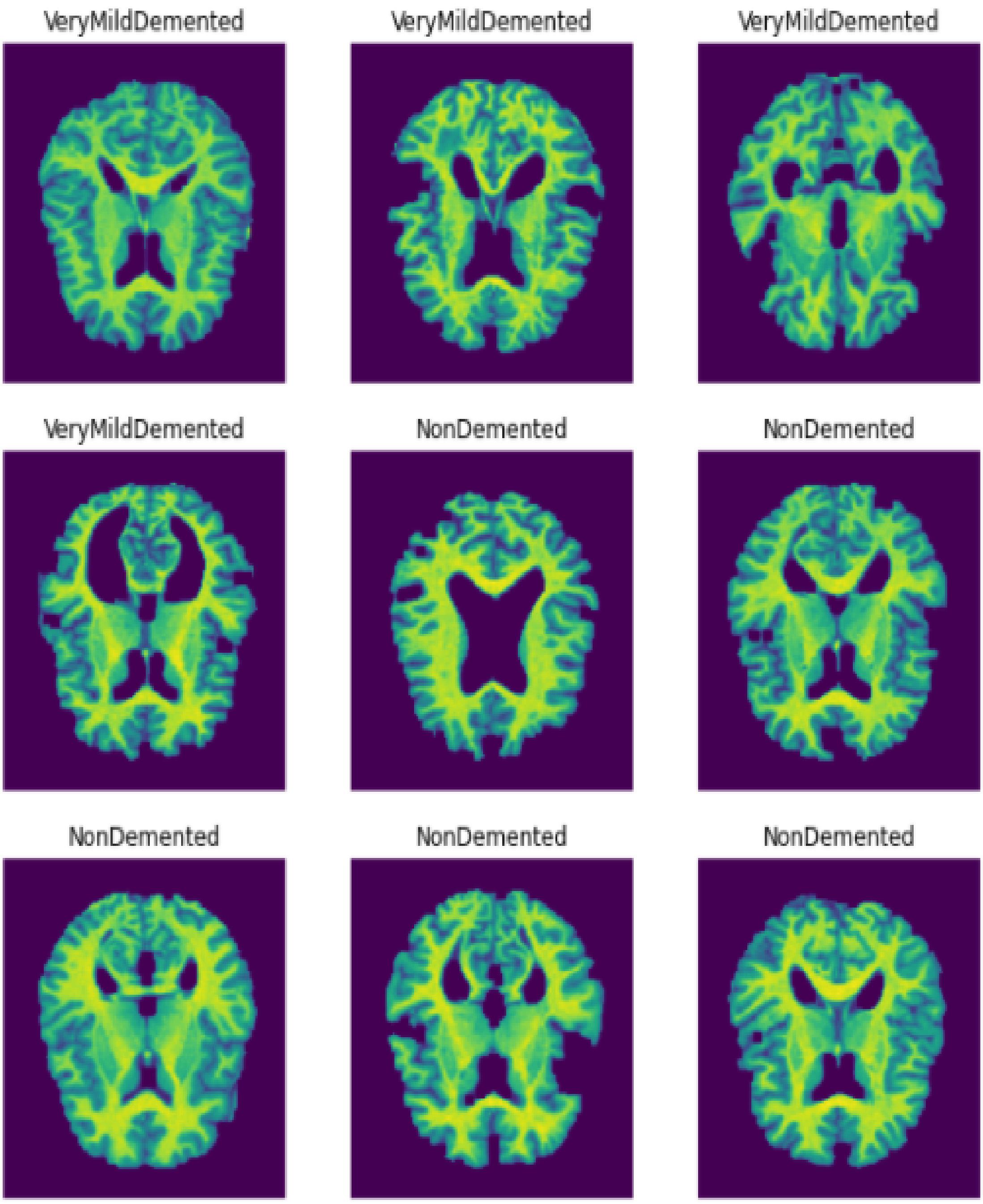

Figure 3.4: Skull Condition at Different Classes of AD.

Table 3.2: Summary of Our Proposed CNN Model, Recall-Net

Layer (Type)	Output Shape	Number of Parameter
resNet50 (Functional)	(None, 7, 7, 2048)	23587712
dropout (Dropout)	(None, 5, 5, 2048)	0
flatten (Flatten)	(None, 51200)	0
batch_Normalization$_9$4	(None, 51200)	204800
dense (Dense)	(None, 64)	3276864
batch_normalization_95 (Batch Normalization)	(None, 64)	256
activation_94 (Activation)	(None, 64)	0
dropout_1 (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 64)	4160
batch_normalization_96 ((Batch Normalization)	(None, 64)	256
activation_95 (Activation)	(None, 64)	0
dropout_2 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 64)	4160
batch_normalization_97 (Batch Normalization)	(None, 64)	256
activation_96 (Activation)	(None, 64)	0
dropout_3 (Dropout)	(None, 64)	0
dense_3 (Dense)	(None, 32)	2080
batch_normalization_98 (Batch Normalization)	(None, 32)	128
activation_97 (Activation)	(None, 32)	0
dropout_4 (Dropout)	(None, 32)	0
dense_4 (Dense)	(None, 32)	1056
batch_normalization_99 (Batch Normalization)	(None, 32)	128
activation_98 (Activation)	(None, 32)	0
dense_5 (Dense)	(None, 4)	132

3.5.1 Recall-Net Architecture

We are using ResNet50 as our base architecture and after layers of the base model, we have added our own 23 layers. Our model has the following steps: At first, we added a dropout layer and then flattened the output of the last layer and then normalized the result. After that, we added five Neural Network Blocks. In our proposed CNN model, each Neural Network Block includes a dense layer initialized with He uniform variance scaling initializer, batch normalization layer, ReLU activation layer and then dropout layer at 0.5 rates. Here are the layers following that.

Dropout Layer: Dropout layer solves the the problem of overfitting. Another five dropout layers have been added at 0.5 rates at the very end of the five neural network blocks.

Flatten Layer: A flatten layer is used after the dropout layer. This flatten layer will flat the output of the last layer.

Batch Normalization Layer: After the flatten layer, a normalization layer will avoid slowing down the process and at the same time and normalize the output layer through uniform distribution of input layers. Other five batch normalization layers have been added in five different neural network blocks.

Dense Layer: 5 dense layers have been used at the beginning of 5 neural network blocks. This layer is also known as the fully connected layer. As a fully connected layer computes the products between the weights and the preceding layer, these five layers will help us in obtaining the correct probabilities for various classes.

Activation Function: The Sigmoid function as given in Eq.3.1 is combined with another dense layer and the ReLU function as given in Eq.3.2 for our model. We also employed the Leaky ReLU activation function, as given in Eq.3.4, which gave the greatest results with Maxpooling 2D. The following is a list of the three activation functions:

$$f(x) = \frac{1}{1 + \mathrm{e}^{-x}} \tag{3.1}$$

$$ReLU = max(0, x) \tag{3.2}$$

$$f(x) = 1(x < 0)(x) + 1(x >= 0)(x), \tag{3.3}$$

$$\tag{3.4}$$

In the preceding equations, $x \in R$ and e is a small constant. The range of values for the Sigmoid function is 0 to 1. The Sigmoid function is defined as an S-shaped curve. It has a vanishing gradient problem, despite its simplicity. The gradient updates diverge too much in opposing directions since the output oscillates around zero, which makes optimization more difficult. ReLU is a simple and efficient activation function because of its ability to avoid vanishing gradient problem and nature of equation. Leaky ReLU was used to overcome the problem of "dying ReLU." A Leaky ReLU will have a small negative slope instead of becoming 0 when x is 0 .

At the last layer of the last Neural Network Block, instead of a dropout layer, we added a softmax dense block outputting the prediction of 4 classes.Table 3.2 shows the summary of our proposed modified CNN model.

Chapter 3 has been about the overall approach of our data set acquisition and the proposed model for our research. Chapter 4 contains the findings of our research and Chapter 5 consists of conclusions, discussion and some recommendations for future research.

Chapter 4

Experimental Result

4.1 Applying Recall-Net CNN Architecture

After the pre-processing we explained in chapter 3.3, we trained our proposed CNN architecture along with some other variation of Recall-Net architectures. Then for each model, we tried to predict the result of the test dataset. From the prediction of the test dataset and it's actual label, we calculated the accuracy, precision, recall, AUC and F1 score and generated a confusion matrix to get an idea about how each model is doing. Additionally, for each model we generated epoch vs loss, epoch vs accuracy, epoch vs precision etc.

4.1.1 Accuracy Calculation Equation

$$ACC = \frac{TP + TN}{P + N} \tag{4.1}$$

$$= \frac{TP + TN}{TP + TN + FP + FN} \tag{4.2}$$

Where, ACC = Accuray, TP =True Positive, TN = True Negative, P = Condition Positive, N = Condition Negative.[47]

4.1.2 Precision Calculation Equation

$$PPV = \frac{TP}{TP + FP} \tag{4.3}$$

$$= 1 - FDR \tag{4.4}$$

Where, PPV = Positive Predictive Value/ Precision, TP =True Positive, FP = False Positive, FDR = False Discovery Rate [47]

4.1.3 Recall Calculation Equation

$$\begin{align} PPV &= \frac{TP}{TP + FN} \tag{4.5} \\ &= 1 - FDR \tag{4.6} \end{align}$$

Where, PPV = Positive Predictive Value/Recall, TP =True Positive, FN = False Negative, FDR = False Discovery Rate [47]

4.1.4 F1-Score Calculation Equation

$$\begin{align} F1 &= 2 \times \frac{PPV \times TPR}{PPV + TPR} \tag{4.7} \\ &= \frac{2TP}{TP + FP + FN} \tag{4.8} \end{align}$$

Where, PPV = Positive Predictive Value, TPR =True Positive Rate, TP =True Positive, FP = False Positive, FN = False Negative. [47]

4.1.5 AUC Calculation Equation

AUC is the calculated area under ROC curve is a useful way to compare performance between different models. Its usage in machine learning models regarding medical imaging is quite prevalent nowadays. Generally the more area under the ROC curve i.e. AUC score , the better the model is. So to calculate, we have to first plot the ROC curve of that model. Then using the trapezoidal rule, AUC score can be computed. [47] shown in 4.1

4.2 Recall-Net Based on VGG16

As a part of experiment we implemented Recall-Net based on VGG16 to see on which kind of architecture perferms better with Recall-Net.

4.2.1 Result of Recall-Net Architecture Based on VGG16

Accuracy Among all the CNN architectures that we applied VGG16 also did not give us any good results in the case of accuracy, precision and recall, confusion matrix, ROC curve, model accuracy and model loss. With 10 epochs and a batch size

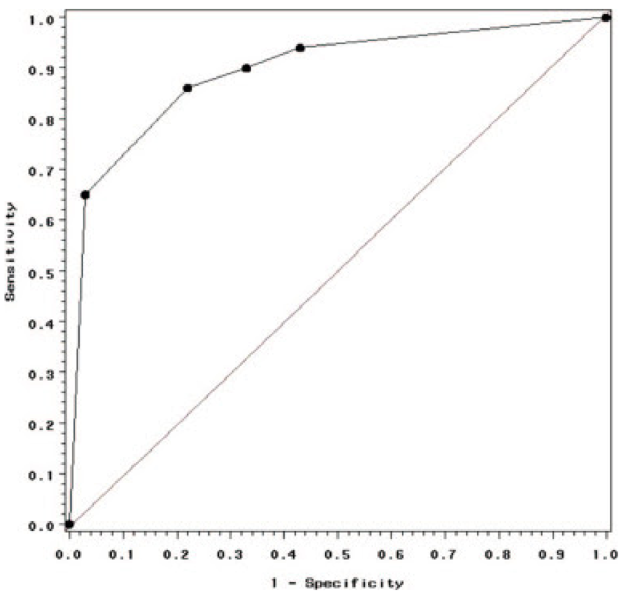

Figure 4.1: ROC graph of a model

of 32 both gave us the accuracy of only 78.00% Precision, Recall and f1 interpretation: Precision. We have got precision in overall four classes is 0.61% data which represents the poor performance of the model. Recall got a recall of 0.37% on our data which also is very poor for this model. Then the F1 score In our case 52% for demented data which is also poor for this model.

Table 4.1: Precision, Recall and F1 Score in VGG16 model

Accuracy	**Precision**	**Recall**	**AUC**	**F1-Score**
0.78557	0.49257	0.37842	0.82117	0.52469

4.2.2 Result Discussion of Recall-Net Architecture Based on VGG16

After the model was trained on the training data set for 10 epochs, we used our model to predict the labels of the test data set. And from the predicted labels and actual labels, we calculated several metrics for our model.

1. **Accuracy, Precision, Recall, F1 Score and AUC Score:** The model is also underfitting when it comes to our data-set, as seen by the Model accuracy, Model loss, Model precision and F1 score graphs. This implies that the model is biased. The confusion matrix can also help us figure this out. Moreover, we can conclude that the model's performance could be improved based on accuracy, confusion matrix, AUC, model accuracy and model loss figure 4.2, 4.3, 4.4 and 4.5.

2. **Confusion Matrix Interpretation:** As far as the Confusion matrix result, we can see that VGG16 does not perform well also figure 4.6

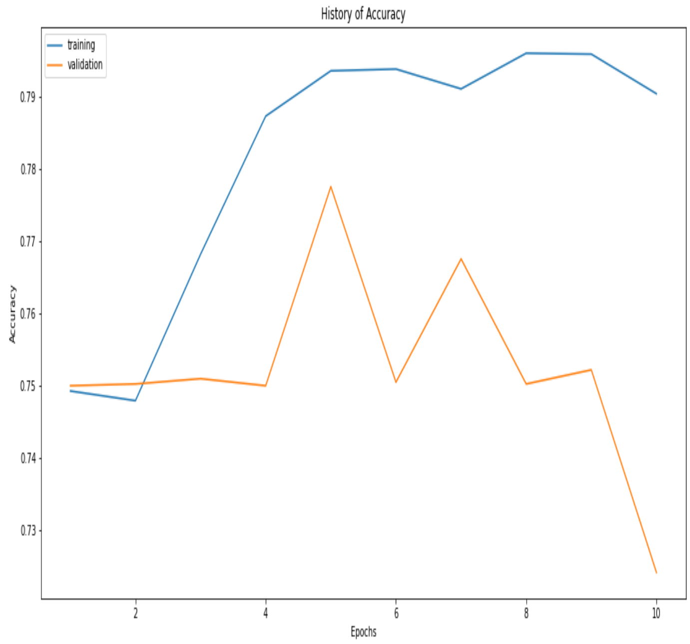

Figure 4.2: Accuracy of VGG16

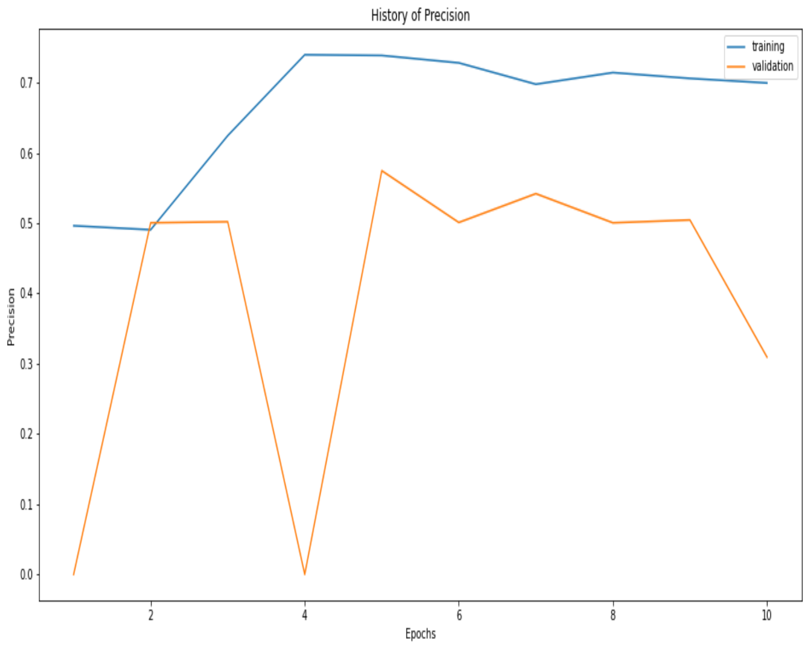

Figure 4.3: Precision of VGG16

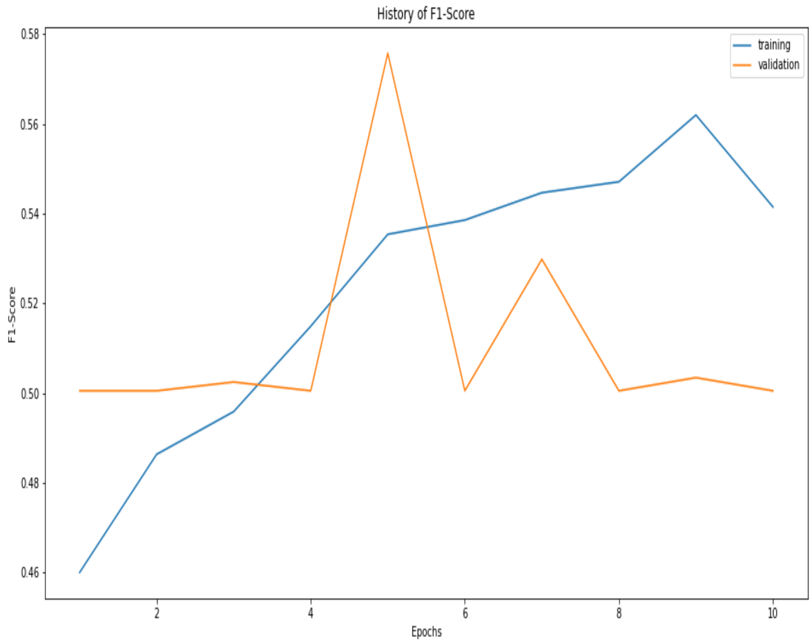

Figure 4.4: F1-Score of VGG16

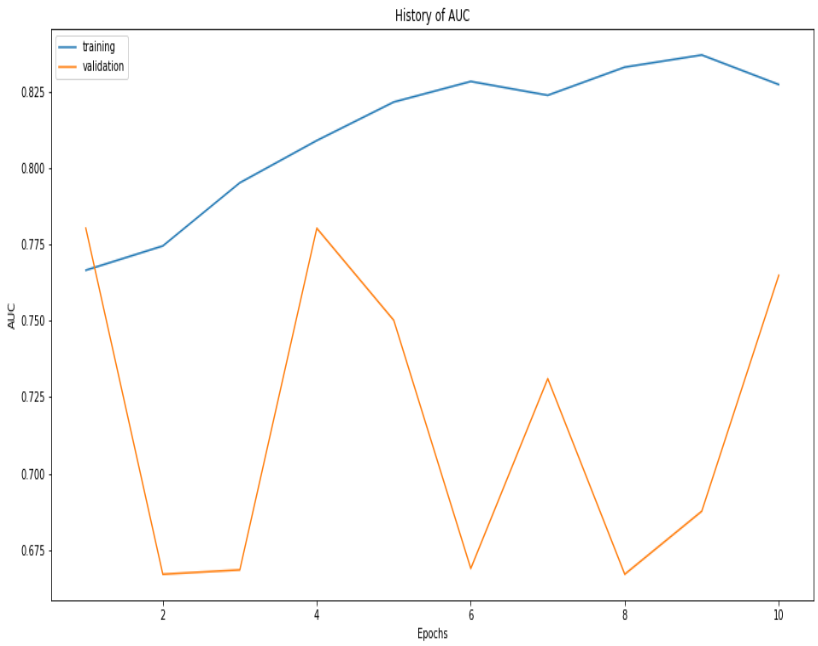

Figure 4.5: AUC of VGG16

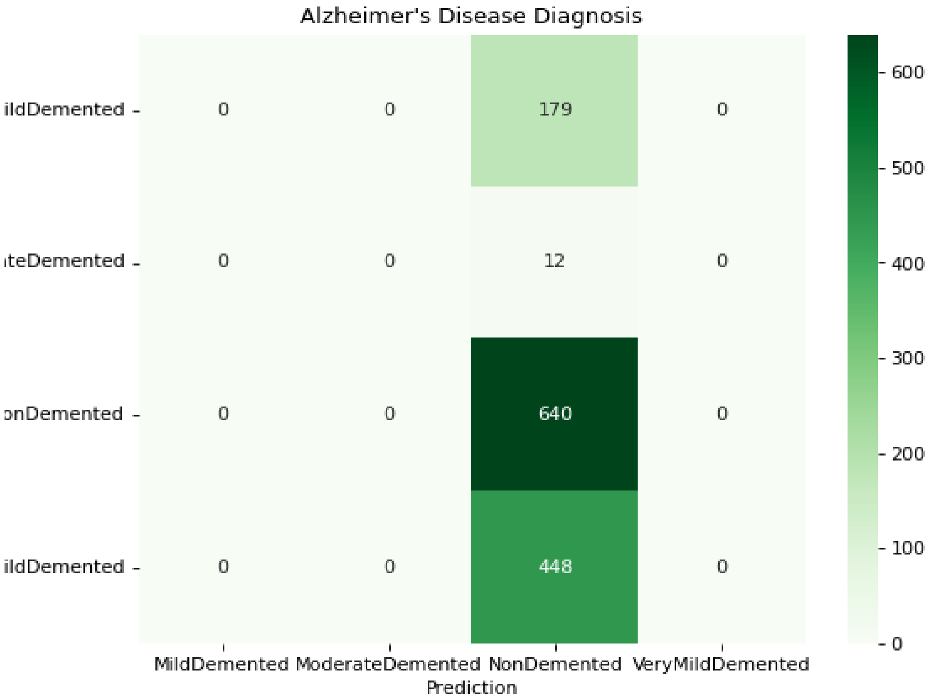

Figure 4.6: Confusion Matrix of VGG16

4.3 Recall-Net Based on InceptionV3

To try out different architecture to see on which kind of architecture performs better with Recall-Net, we implemented Recall-Net based on InceptionV3.

4.3.1 Result of Recall-Net Architecture Based on Inception v3

According to Precision, AUC, Recall and f1 interpretation of the Inception v3 model with 10 epochs and batch size 32, these results are one of the worst results that we got among all the CNN models we have used on our dataset, it had an accuracy of 0.75%. Also when we plot the confusion matrix, we get a significantly different result. The data that reflects the model's performance has a precision of 0.49% which is extremely poor for this model. As well as, we have a recall of 0.50% and AUC is 0.78% which are also poor .likewise.the F1 score for our data of this model is 0.50% which is not good Table 4.2.

Table 4.2: Precision, Recall and F1 Score in Inception V3 model

Accuracy	Precision	Recall	AUC	F1-Score
0.7623	0.5293	0.4449	0.8095	0.5097

4.3.2 Result Discussion of Recall-Net Architecture Based on Inception v3

After the model was trained on the training data set for 10 epochs, we used our model to predict the labels of the test data set. And from the predicted labels and actual labels, we calculated several metrics for our model.

1. **Accuracy, Precision, Recall, F1 Score and AUC Score:** From the Model accuracy, model loss, model precision and F1 score graph we can see that the model is underfitting when it comes to our data set. This means the model has biases. Which we can also determine from the confusion matrix. Overall from accuracy, confusion matrix, AUC, model accuracy and model loss figure we can say that performance of the model is not good at all figure 4.7, 4.8, 4.9 and 4.10.

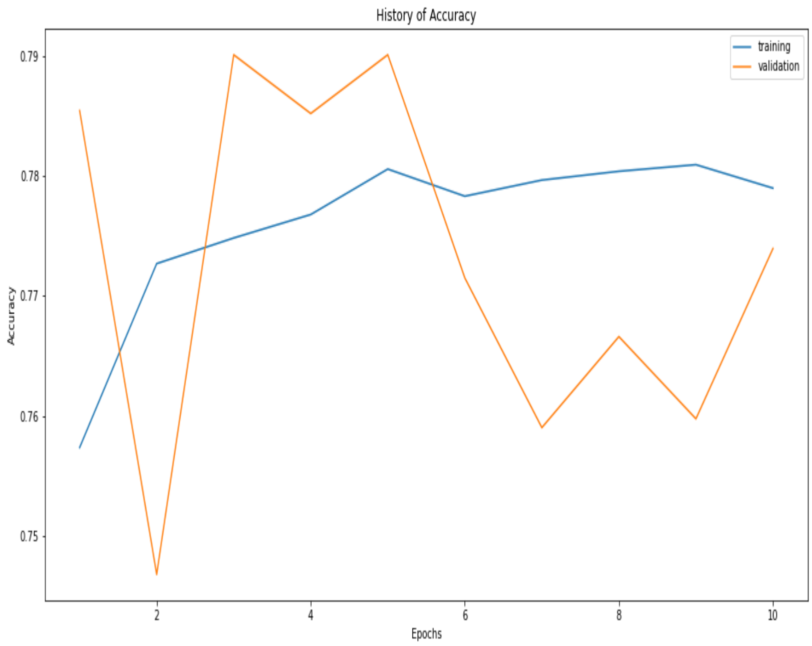

Figure 4.7: Accuracy of Inception v3

2. **Confusion Matrix Interpretation:** When we look at the confusion matrix (Fig. 4.11), we can see that it tried to predict non-demented, moderate demented, mild demented and very mild demented pictures, but the result is slightly disappointing. Also if we compare the accuracy with our self-designed architecture, the performance analysis differed significantly. Precision and recall, AUC, model accuracy and model loss all yielded varied results.

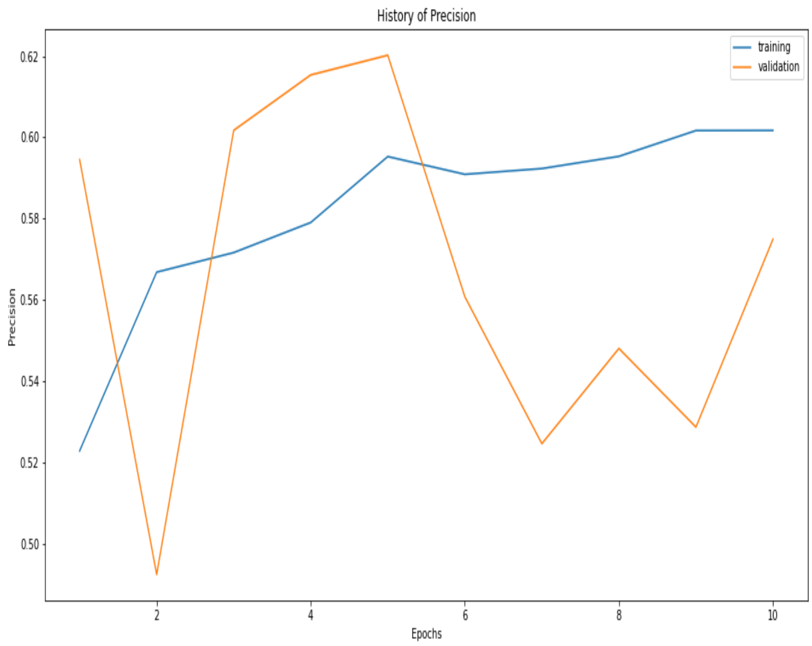

Figure 4.8: Precision of Inception v3

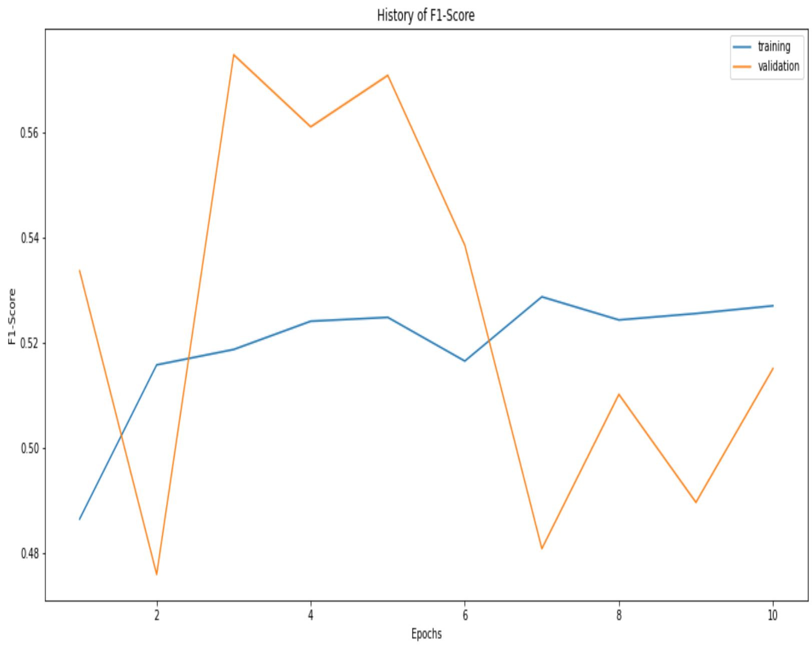

Figure 4.9: F1-Score of Inception v3

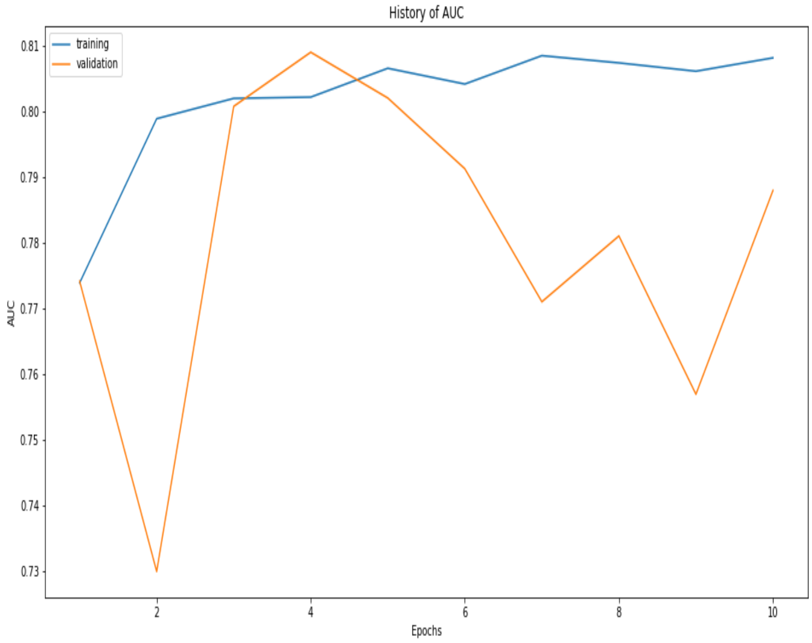

Figure 4.10: AUC of Inception v3

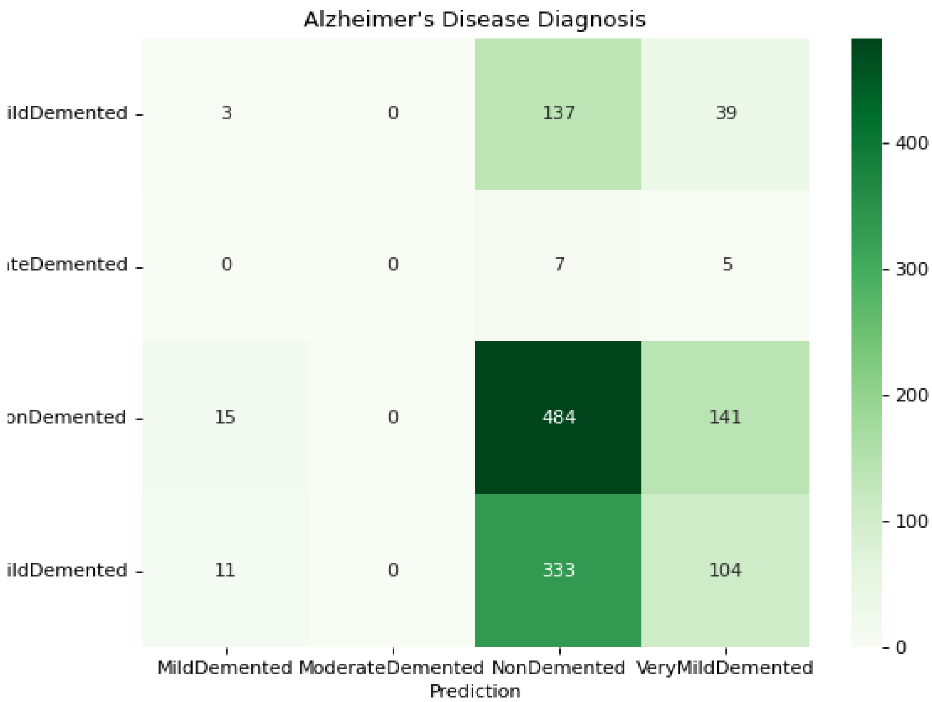

Figure 4.11: Confusion Matrix of Inception v3

4.4 Our Recall-Net Based on ResNet50

Various CNN architectures have been experimented to optimize them for classifying MRI images of Alzheimer's patients. Among those architectures, our proposed CNN model, Recall-Net, built on top of ResNet-50 that outperformed all other backbones.

4.4.1 Result Discussion of Recall-Net Architecture Based on ResNet50

After our model was trained on the training data set for 10 epochs, we used our model to predict the labels of the test data set. And from the predicted labels and actual labels, we calculated several metrics for our model. We can the result in Table 4.3

Table 4.3: Precision, Recall and F1 Score in Recall-Net Based on ResNet model

Accuracy	Precision	Recall	AUC	F1-Score
0.7977	0.6949	0.3401	0.8412	0.5512

1. **Accuracy:** After 10 epochs and with batch size of 50, have an accuracy of 79.77% which indicates our model was able to classify around 80% of the images into the right category.

2. **Precision:** As our model classified data into four classes, the precision was 69.49%. Even though it is quite high, it could have been improved with additional training time.

3. **Recall:** The recall value of our model based on ResNet50 was the lowest out of the other three base models we experimented with. It was around 34.01%.

4. **F1 Score:** The F1-Score was around 55.12%. It showed there is further scope of improvement in the training time.

5. **AUC Score:** The AUC score of our model after training was 84.12%.

6. **Confusion Matrix Interpretation:** The confusion matrix in fig. 4.16 shows us that the model accurately predicted the non-demented patients. And it predicted a very high number of patient of very mild demented category correctly. It also shows that it is misclassified some very mild demented patient as non demented. As a very mild demented patient is still at the beginning stage of Alzheimer, the model confused it sometimes with non demented patient. It also misclassified some non demented as very mild demented for the above mentioned reason.

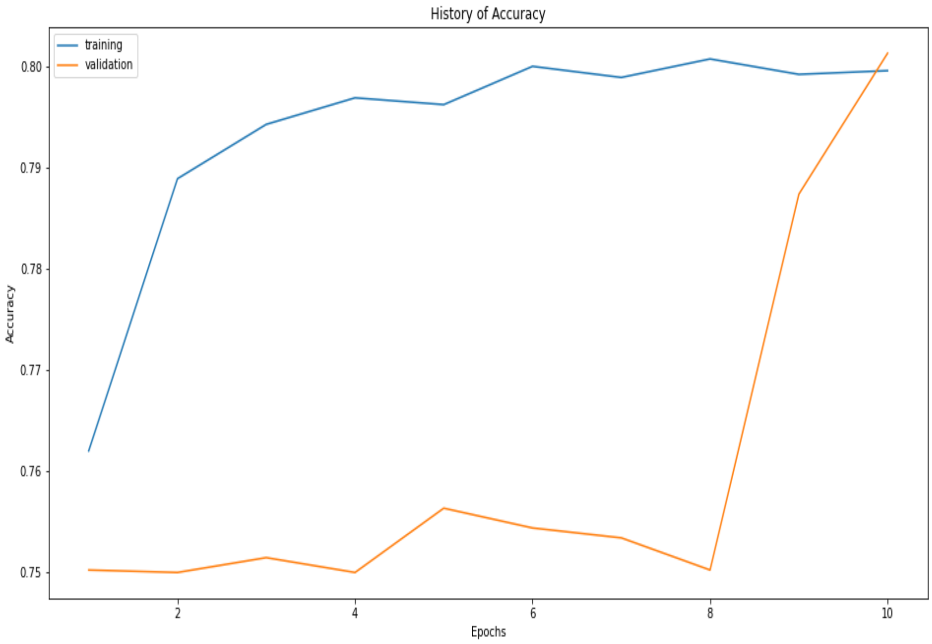

Figure 4.12: Accuracy of Recall-Net Based on ResNet50

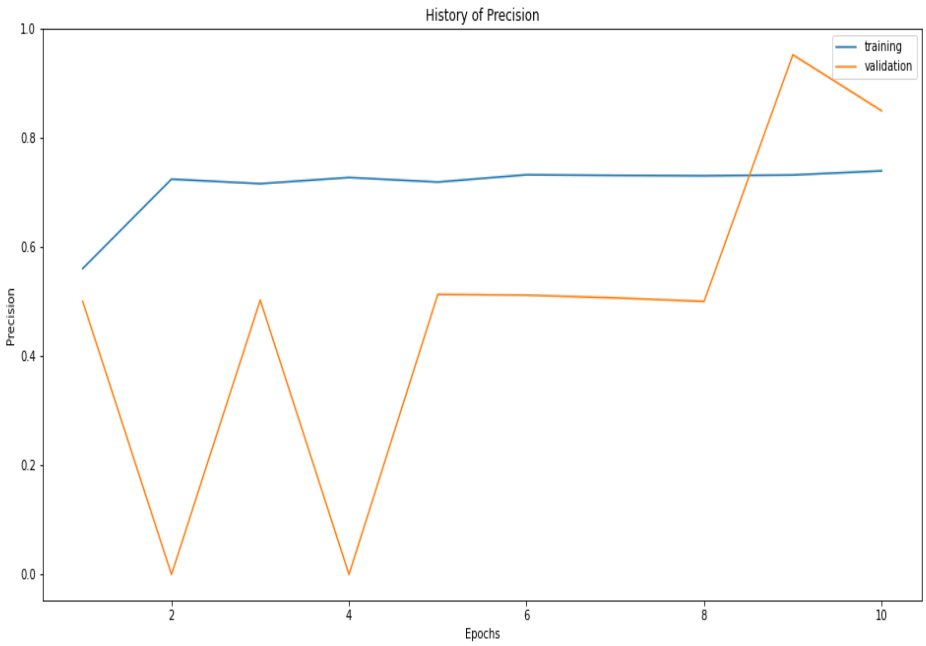

Figure 4.13: Precision of of Recall-Net Based on ResNet50

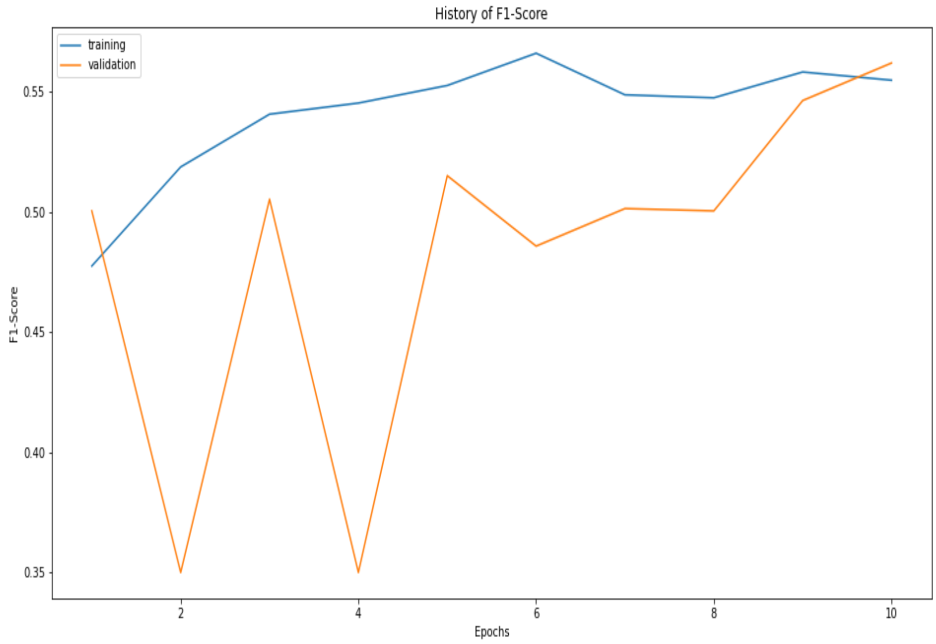

Figure 4.14: F1-Score of of Recall-Net Based on ResNet50

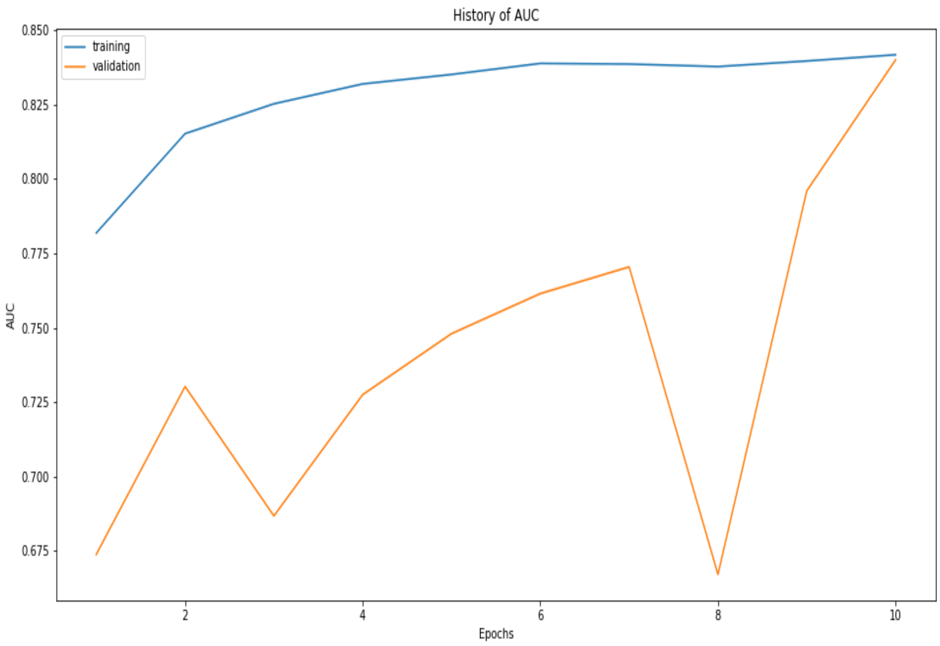

Figure 4.15: AUC of of Recall-Net Based on ResNet50

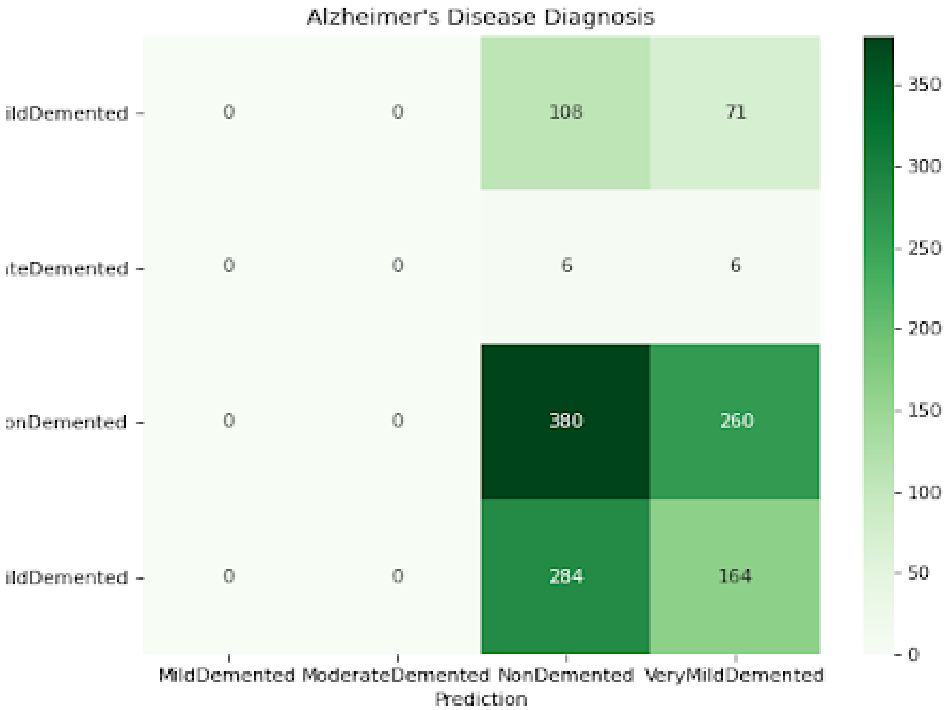

Figure 4.16: Confusion Matrix of ResNet50

4.5 Comparison between Different Variation of Recall-Net

The accuracy, precision, recall, AUC and F1-Score are calculated for the proposed CNN model to analyze and compare the performance with other models. The experimental result of our model in pre-processed ADNI dataset is presented in Table 4.4. The result comparison between raw and pre-processed images of the ADNI dataset has been presented in Table 4.4. There, note that after pre-processing the image accuracy increased from 59.19% to 79.77% along with other metrics like precision, recall, AUC and f1-score. Thus, the pre-processing improved the model's performance a great deal. The comparison between our proposed CNN model the result of our experiment with other models like InceptionV3, VGG16 and VGG19 is displayed briefly in Table 4.5. From the table, we can see that our proposed model based on ResNet50 has the highest accuracy of 79.77% compared to other experimented models based on InceptionV3 and VGG16. The other models based on InceptionV3 and VGG16 have achieved 76.23% accuracy and 75.02% accuracy respectively on the same ADNI pre-processed dataset. data set. Precision, recall, AUC and f1-score metrics for our proposed model based on ResNet50 are also much better than other experimental models.

Table 4.4: Performance Comparison of Recall-Net with & without Pre-processing

	Accuracy	**Precision**	**Recall**	**AUC**	**F1-Score**
No Pre-processing	0.5919	0.4922	0.2365	0.3606	0.3960
Pre-processed Data	0.7977	0.6949	0.3401	0.8412	0.5512

Table 4.5: Performance Comparison of Recall-Net with Different CNN Architectures as Base Model

Model Name	**Accuracy**	**Precision**	**Recall**	**AUC**	**F1-Score**
Recall-Net+ ResNet50	0.7977	0.6949	0.3401	0.8412	0.5512
Recall-Net+ InceptionV3	0.7623	0.5293	0.4449	0.8095	0.5097
Recall-Net+ VGG16	0.7502	0.5004	0.5004	0.7806	0.5004

Table 4.6: Performance Comparison of Recall-Net with Existing Models

Study	**Modality**	**Dataset**	**Model**	**4-class Classification Accuracy**
Liu et al. [14]	MRI, PET	ADNI	MK-SVM, SAE, SMR	51.96%
Amoroso et al. [35]	sMRI	ADNI	DNN, RF, Fuzzy Logic	46.30%
Shi et al. [31]	MRI, PET	ADNI	MM-SDPN, SVM	57%
Liu et al. [16]	MRI, PET	ADNI	MK-SVM, SAE-ZEROMUSK, SAE-CONCAT	43.79%
Liu et al. [6]	MRI, PET	ADNI	SAE, SMR	47.42%
Our Proposed Model	T1-weighted MRI	ADNI	Modified CNN	79.77%

4.6 Compare the Performance of Recall-Net with Other Models

Our goal was to achieve higher accuracy in 4 class classification than other pre-existing CNN models for classifying MRI images of AD patients. We have compared the performance of our proposed model for 4-class classification with the models associated with previous works. In terms of 4-class classification, SVM has been used mostly as the model of previous works. Though CNN is not the most prevalent model so far in terms of 4-class classification, we have got a superior performance from our CNN based model, Recall-Net. From Table IV, we can see that our proposed CNN model, Recall-Net has an accuracy of 79.77% whereas, other existing models have much lower accuracy like 51.96% [14], 46.3% [35], 57% [31], 43.79% [16], 47.42% [6]. From Table 4.3, we can observe that in terms of 4 class classification, our result beat the existing result by quite a margin not just in terms of accuracy, but also in terms of precision, recall, auc and F1-score. In terms of all the evaluation metrics, our work performs better. We have shown a comparison between our model and the models of previous works in the Table 4.6.

Chapter 4 has been about the qualitative and quantitative findings of our research which addresses four objectives of our research. Chapter 5 consists of conclusions, discussion and some recommendations for future research.

Chapter 5

Summary, Conclusion, Discussion and Recommendations

Chapter 5 provides conclusions based on the experiment results from MRI data of Alzheimer's patients collected from ADNI dataset. Moreover it provides discussion and recommendations based on our insights for future research. This chapter will give a overview of the purpose of the study, research question, literature review and findings of the study. It will then present conclusions, discussion of the conclusions and recommendations regarding the area of further research.

5.1 Overview

In this research, we have proposed a deep learning method to classify brain MRI of AD patients into four classes indicating different progressive stages of AD. The necessary brain MRI data along with the patient's possible diagnosis or label was collected from ADNI to train our architecture. The data was pre-processed and then used for training. The details of whole pre-processing and training was discussed in Chapter 3.1. We shall revise the necessary steps as follows,

- Associate the MRI data of the patients with their label from the csv file.
- Slice the 3D MRI data into a 2D axial view of the brain to avoid the curse of dimensionality.
- Extracted the brain portion of the image inside the skull i.e. ROI necessary for the classification.
- Split the whole dataset into train and test dataset.
- Train different variations of our proposed model on the pre-processed train dataset.
- Calculate model's performance metrics associated with each variation from test dataset.
- Analyse and compare the results we obtained from each variation of our model.

To evaluate how good our proposed model is performing, we tested our best variation of the proposed model against our validation dataset. For each image, our proposed model predicted it's possible label and it predicted correctly 79.77% of the time for four different classes of images. This result shows that we can quite confidently predict the different AD stages of a patient.

5.2 Key Insights

In our research, we have applied different variations of our models of CNN to classify four classes of Alzheimer's Disease. From the result we compiled at Chapter 4, we can derive some major insights about the data, model and the research as a whole. The insights are listed below as follows,

- Among the models we have applied, ResNet-50 has shown the best performance as the base model of our proposed CNN model, Recall-Net.
- Before and after pre-processing, model metrics increased by quite a margin.
- The model made most of its mistakes classifying between NC and EMCI.

5.3 Discussion

The key insights mentioned in Chapter 5.2 would be discussed here. The possible reason, analysis and suggestion based on these conclusions are described below,

- Most of the state of the art architecture is trained, tested and validated against the largest structured image dataset "ImageNet"[4]. Regardless of the weights or the biases of these pre-trained architectures, their building blocks and design facilitates a solution for the complex image classification problem. Using these architectures as base can provide a detailed and complex feature map of image which would be nearly impossible to do by handcrafted feature engineering or other by architectures that are not rigorously tested. Using these feature maps generated from these models, we have been able to create our own classifier to accomplish our goals.
- Even though VGG16 and Inception V3 both are excellent models in their own regard, the ResNet-50 architecture as backbone has performed better in our research than the former. The possible reason could be that VGG16 is more basic in design than ResNet-50. Not only ResNet-50 has more trainable parameters, but also it provides a residual approach within layers. For Inception V3 it reduces the computational cost as opposed to ResNet-50 which increases the computational accuracy. However, intuitively Inception V3 should perform similarly to ResNet-50. Another reason why ResNet-50 have performed better than Inception V3 could be

the type of data itself. Further research regarding the efficacy of different architectures and different kinds of image could be done on topic to expand more.

- The reason the model confuses NC and EMCI more than other categories is that a healthy person having some cognitive impairment doesn't mean that he/she is going to become EMCI. A cMCI and ncMCI both look very similar. However, the difference between them is so subtle and still a topic of medical research that it is understandable why the model is making that mistake.

5.4 Recommendation For Practice

Our model could be used for setting up an supplemental online service which will classify the brain MRI of patients uploaded to it as a part of a data pipeline. This would not only help diagnosing patients but also for collecting MRI data of patients for further research. This could have a tremendous commercial value. Additionally, as a part of a system in a hospital it could provide a large enough dataset to do a large sample study in Bangladesh. This research could also be used as a framework regarding other brain disease classification using DL which have not been attempted yet.

5.5 Recommendations For Future Work

Our proposed architecture for classifying 4 different classes of Alzheimer's patients can be further improved in several areas like including more advanced data on the patients which were not available during our time of study, trying out newer experimented CNN architecture and human interpretable causal inference between the data and result. The classification could be further improved by incorporating the whole 3D MRI image of the patient instead of 2D and other information like their routine medical information, their DNA sequence etc. for the detailed effects of this disease on the patients. In section 2.1, we have reviewed some papers which used more advanced imaging techniques such as fMRI, FDG-PET, PET and sometimes combined some of them. Even though most of these advanced imaging techniques aren't widespread that much, our architecture could be modified in the future to integrate these kinds of medical images. In future more advanced brain mapping technology like connectome [12] could be used for definite diagnosis of Alzheimer's diseases. More sophisticated method for pre-processing the brain MRI could be used which would process the image without losing too much information. More recent and better pre-processing method could result in better accuracy along with other metrics. Due to the increase of GPU capability in recent years, the computer vision branch of deep learning also grew tremendously. Nowadays, a lot of new research on general CNN architecture is coming out like Faster R-CNN promising high accuracy on complex image datasets. We haven't been able to test these newer experimented

CNN architectures on our dataset due to time constraints. However, it could potentially cause better model metrics results. This model could be used for setting up an supplemental online service which will classify the brain MRI of patients uploaded to it as a part of a data pipeline. This would not only help diagnosing patients but also for collecting MRI data of patients for further research. This could have a tremendous commercial value. There is an additional scope to study the MRI data not just for classifying but to establish human interpretable causality or causal inference for the classification between different parts of the brain or genetic or other medical records. This could provide key insights in the inner workings of the disease and help both clinicians and pharmaceuticals to better manage the diseases and develop possible cures for the disease.

5.6 Conclusion

The problem of classifying and detecting MRI data of Alzheimer's patients can be effectively solved by using the deep learning technique. We can achieve expert-level accuracy using deep learning for computer vision by leveraging huge pre-trained neural network architecture. This approach proves a great potential for automating medical tasks with human-level accuracy in the biomedical and medical imaging field. Our approach can be applied to a variety of medical imaging problems to achieve a high level of accuracy gain. However, our accuracy and results are limited by only human-level expertise and model we experimented with. Without significant progress in detection by humans, the field will be stuck. Additionally, we lack the knowledge of cause-effect relation in data and classification because of the oblique nature of the neural network.Therefore we can easily detect and classify Alzheimer's patients progression,early detection and also can do classification.

Bibliography

[1] C. Y. Chang and D. H. Silverman, "Accuracy of early diagnosis and its impact on the management and course of alzheimer's disease," *Expert Review of Molecular Diagnostics*, vol. 4, no. 1, pp. 63–69, 2004.

[2] J. Damoiseaux, C. Beckmann, E. S. Arigita, F. Barkhof, P. Scheltens, C. Stam, S. Smith, and S. Rombouts, "Reduced resting-state brain activity in the "default network" in normal aging," in *Cerebral Cortex*, vol. 18, Oxford Academic, 2007, pp. 1856–1864.

[3] E. Burton, R. Barber, E. Mukaetova-Ladinska, J. Robson, R. Kalaria, O'Brien, and J.T., "Medial temporal lobe atrophy on mri differentiates alzheimer's disease from dementia with lewy bodies and vascular cognitive impairment: A prospective study with pathological verification of diagnosis," in *Brain*, vol. 132, Oxford Academic, 2009, pp. 195–203.

[4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, Ieee, 2009, pp. 248–255.

[5] B. Magnin, L. Mesrob, S. Kinkingnehun, M. Pelegrini-Issac, O. Colliot, M. Sarazin, B. Dubois, S. Lehericy, and H. Benali, "Support vector machine-based classification of alzheimer's disease from whole-brain anatomical mri," in *Neuroradiology*, vol. 51, Elsevier, 2009, pp. 73–83.

[6] R. C. Petersen, P. Aisen, L. A. Beckett, M. Donohue, A. Gamst, D. J. Harvey, C. Jack, W. Jagust, L. Shaw, A. Toga, *et al.*, "Alzheimer's disease neuroimaging initiative (adni): Clinical characterization," *Neurology*, vol. 74, no. 3, pp. 201–209, 2010.

[7] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of alzheimer's disease and mild cognitive impairment," in *NeuroImage*, vol. 55, Elsevier, 2011, pp. 586–867.

[8] O. Abdel-Hamid, A.-r. Mohamed, H. Jiang, and G. Penn, "Applying convolutional neural networks concepts to hybrid nn-hmm modell for speech recognition," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2012.

[9] M. Binnewijzend, M. Schoonheim, E. Sanz-Arigita, A. Wink, N. Tolboom, S. Adriaanse, J. Damoiseaux, P. Scheltens, B. Berckel, and F. Barkhof, "Resting-state fmri changes in alzheimer's disease and mild cognitive impairment," in *Neurobiology of Aging*, vol. 33, Elsevier, 2012, pp. 2018–2028.

[10] C.-Y. Wee, P.-T. Yap, D. Zhang, L. Wang, and D. Shen, "Constrained sparse functional connectivity networks for mci classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 7511, Springer, 2012, pp. 212–219.

[11] H.-I. Suk and D. Shen, "Deep learning-based feature representation for ad/mci classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 8150, Springer, 2013, pp. 583–590.

[12] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, K. Ugurbil, W.-M. H. Consortium, *et al.*, "The wu-minn human connectome project: An overview," *Neuroimage*, vol. 80, pp. 62–79, 2013.

[13] L. Deng and D. Yu, "Deep learning: Methods and applications," *Foundations and trends in signal processing*, vol. 7, no. 3–4, pp. 197–387, 2014.

[14] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of alzheimer's disease with deep learning," in *2014 IEEE 11th international symposium on biomedical imaging (ISBI)*, IEEE, 2014, pp. 1015–1018.

[15] H.-I. Suk, S.-W. Lee, and D. Shen, "Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis," in *NeuroImage*, vol. 101, Elsevier, 2014, pp. 569–582.

[16] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. D. Feng, "Multi-phase feature representation learning for neurodegenerative disease diagnosis," in *Australasian Conference on Artificial Life and Computational Intelligence*, Springer, 2015, pp. 350–359.

[17] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," in *Neural and Evolutionary Computing*, vol. 2, 2015, pp. 2–10.

[18] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," in *International Journal of Computer Vision*, vol. 115, Springer, 2015, pp. 211–252.

[19] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, 2015. arXiv: 1409.1556 `[cs.CV]`.

[20] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, *Rethinking the inception architecture for computer vision*, 2015. arXiv: 1512.00567 `[cs.CV]`.

[21] I. Goodfellow, Y. Bengio, and A. Courville, "Convolutional networks," in *Deep Learning*, vol. 2, MIT press, 2016, pp. 326–366.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[23] E. Hosseini-Asl, R. Keynton, and A. El-Baz, "Alzheimer's disease diagnostics by adaptation of 3d convolutional network," in *2016 IEEE international conference on image processing (ICIP)*, IEEE, 2016, pp. 126–130.

[24] A. Ortiz, J. Munilla, J. M. Górriz, and J. Ramírez, "Ensembles of deep learning architectures for the early diagnosis of the alzheimer's disease," in *International Journal of Neural Systems*, vol. 26, SCI Journal, 2016.

[25] S. Sarraf and G. Tofighi, "Deepad: Alzheimer's disease classification via deep convolutional neural networks using mri and fmri," in *Alzheimer's Disease Neuroimaging Initiative*, bioRxiv, 2016, pp. 070–441.

[26] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, IEEE, 2017, pp. 1–6.

[27] A. Farooq, S. Anwar, M. Awais, and S. Rehman, "A deep cnn based multi-class classification of alzheimer's disease using mri," in *2017 IEEE International Conference on Imaging Systems and Techniques (IST)*, 2017, pp. 1–6. DOI: 10.1109/IST.2017.8261460.

[28] M. Hon and N. M. Khan, "Towards alzheimer's disease classification through transfer learning," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2017, pp. 1166–1169.

[29] J. Islam and Y. Zhang, "A novel deep learning based multi-class classification method for alzheimer's disease detection using brain mri data," in *Brain Informatics*, Springer, 2017, pp. 213–222.

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

[31] J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying, "Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of alzheimer's disease," *IEEE journal of biomedical and health informatics*, vol. 22, no. 1, pp. 173–183, 2017.

[32] H.-I. Suk, S.-W. Lee, and D. Shen, "Deep ensemble learning of sparse regression models for brain disease diagnosis," in *Medical Image Analysis*, vol. 37, Elsevier, 2017, pp. 101–113.

[33] J. Wu, "Introduction to convolutional neural networks," *National Key Lab for Novel Software Technology. Nanjing University. China*, vol. 5, no. 23, p. 495, 2017.

[34] K. Aderghal, A. Khvostikov, A. Krylov, J. Benois-Pineau, K. Afdel, and G. Catheline, "Classification of alzheimer disease on imaging modalities with deep cnns using cross-modal transfer learning," in *2018 IEEE 31st international symposium on computer-based medical systems (CBMS)*, IEEE, 2018, pp. 345–350.

[35] N. Amoroso, D. Diacono, A. Fanizzi, M. La Rocca, A. Monaco, A. Lombardi, C. Guaragnella, R. Bellotti, S. Tangaro, A. D. N. Initiative, *et al.*, "Deep learning reveals alzheimer's disease onset in mci subjects: Results from an international challenge," *Journal of neuroscience methods*, vol. 302, pp. 3–9, 2018.

[36] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu, and J. He, "Short-term load forecasting with deep residual networks," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3943–3952, 2018.

[37] Y. Ding, J. H. Sohn, G. K. Micheal, H. Trivedi, R. Harnish, N. W. Jenkins, D. Lituiev, T. P. Copeland, M. S. Aboian, C. M. Aparici, S. C. Behr, and R. R. Flavell, "A deep learning model to predict a diagnosis of alzheimer disease by using f-fdg pet of the brain," in *Radiology*, vol. 290, Radiological Society of North America, 2018, pp. 569–582.

[38] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," in *Machine Learning*, vol. 2, 2018, pp. 1–28.

[39] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and electronics in agriculture*, vol. 147, pp. 70–90, 2018.

[40] T. Koch, L. Liebel, F. Fraundorfer, and M. Korner, "Evaluation of cnn-based single-image depth estimation methods," in *Computer Vision Foundation*, IEEE, 2018.

[41] U. Raghavendra, S. Bhat, A. Gudigar, and R. Acharya, "Automated system for the detection of thoracolumbar fractures using a cnn architecture," in *Future Generation Computer Systems*, vol. 85, Elsevier, 2018, pp. 184–189.

[42] J. Wang and L. Zihao, "Research on face recognition based on cnn," in *IOP Conference Series: Earth and Environmental Science*, vol. 170, 2nd International Symposium on Resource Exploration and Environmental Science, 2018, pp. 032–110.

[43] J. Weller and A. Budson, "Current understanding of alzheimer's disease diagnosis and treatment," *F1000Research*, vol. 7, 2018.

[44] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," in *IEEE Transactions on Industrial Electronics*, IEEE, vol. 7, 2018, pp. 990–998.

[45] Q. Ji, J. Huang, W. He, and Y. Sun, "Optimized deep convolutional neural networks for identification of macular diseases from optical coherence tomography images," *Algorithms*, vol. 12, no. 3, p. 51, 2019.

[46] X. Li, X. Song, and T. Wu, "Aognets: Compositional grammatical architectures for deep learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6220–6230.

[47] C. Nicholson, *Evaluation metrics for machine learning - accuracy, preci- sion, recall, and f1 defined*, 2019. [Online]. Available: https://wiki.pathmind.com/accuracy-precision-recall-f1.

[48] S. Yadav and S. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *Journal of Big Data*, vol. 6, 2019.

[49] A. Dhillon and G. Verma, "Convolutional neural network: A review of models, methodologies and applications to object detection," in *Progress in Artificial Intelligence*, vol. 9, Springer, 2020, pp. 085–112.

[50] F. Ramzan, M. U. G. Khan, A. Rehmat, S. Iqbal, T. Saba, A. Rehman, and Z. Mehmood, "A deep learning approach for automated diagnosis and multi-class classification of alzheimer's disease stages using resting-state fmri and residual neural networks," *Journal of medical systems*, vol. 44, no. 2, pp. 1–16, 2020.

[51] S. Ranjbar, M. Nejad, H. Zakeri, and A. Gandomi, "Computational intelligence for modeling of asphalt pavement surface distress," in *New Materials in Civil Engineering*, Elsevier, 2020, pp. 079–116.

[52] A. W. Salehi, P. Baglat, B. B. Sharma, G. Gupta, and A. Upadhya, "A cnn model: Earlier diagnosis and classification of alzheimer disease using mri," in *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, IEEE, 2020, pp. 156–161.

[53] J. Wen, E. Thibeau-Sutre, M. Diaz-Melo, J. Samper-González, A. Routier, S. Bottani, D. Dormont, S. Durrleman, N. Burgos, O. Colliot, *et al.*, "Convolutional neural networks for classification of alzheimer's disease: Overview and reproducible evaluation," *Medical image analysis*, vol. 63, p. 101 694, 2020.

[54] J. Ren, M. Green, and X. Huang, "From traditional to deep learning: Fault diagnosis for autonomous vehicles," in *Learning Control*, Elsevier, 2021, pp. 205–219.

[55] S.-H. Tsang, *Review: Googlenet (inception v1)— winner of ilsvrc 2014 (image classification).* [Online]. Available: https://medium.com/coinmonks/paper-review-of-googlenet-inception-v1-winner-of-ilsvlc-2014-image-classification-c2b3565a64e7.