Alzheimer's Disease Detection and Classification using Transfer Learning Techniques and Ensembling Operations on Convolutional Neural Networks

by

Alvina Awwal
17101074
Homaira Huda Shomee
17101061
Sayed Us Sadat
17101364
Sadia Nur Amin
17101397

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
January 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:

Alvina Awwal
17101074

Sayed Us Sadat

Sayed Us Sadat
17101364

Homaira Huda

Homaira Huda Shomee
17101061

Sadia

Sadia Nur Amin
17101397

Approval

The thesis/project titled "Alzheimer's Disease Detection and Classification using Transfer Learning Techniques and Ensembling Operations on Convolutional Neural Networks" submitted by

1. Alvina Awwal (17101074)
2. Homaira Huda Shomee (17101061)
3. Sayed Us Sadat (17101364)
4. Sadia Nur Amin (17101397)

Of Fall 2020, has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on January 8th, 2021.

Examining Committee:

Supervisor:
(Member)

Zavid Parvez
Digitally signed by Zavid Parvez
DN: cn=Zavid Parvez, o=Brac University, ou=CSE, email=zavid.parvez@bracu.ac.bd, c=US
Date: 2021.01.14 17:25:56 +06'00'

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

Co-Supervisor:
(Member)

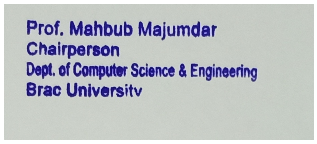
Tanzim 13.1.2021

Md Tanzim Reza
Lecturer
Department of Computer Science and Engineering
BRAC University

Program Coordinator:
(Member)

Md. Golam Rabiul Alam
Assistant Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)



Mahbubul Alam Majumdar
Professor and Dean
Department of Computer Science and Engineering
BRAC University

Abstract

Alzheimer's disease (AD) is a neurological disease that affects the healthy cells of the brain and results in people having long-term memory loss, thinking problems, disorientation, behavioral inconsistencies and finally death. When the disease gets detected, the pathological load is already high and there is no way of coming back from there. This neurodegenerative disease consists of three general stages which we classified in our research that includes very mild (early stage), mild (middle stage) and finally, the moderate stage (late stage). We implemented 5 existing efficient and recent CNN models such as VGG19, Inception-ResNet-v2, ResNet152v2, EfficientNetB5 and EfficientNetB6 including one model of our own. Later, we did ensembling operations thrice with multiple combinations of the models to enhance our outcome and that was achieved since this gave improved accuracy of up to around 96% compared to the individual models where the maximum was 92.2% from EfficientNetB5. The results achieved showed precise detection and classification of AD and its stages even though data was limited and it was a challenge differentiating a healthy brain from that of a subject with AD.

Keywords: Convolutional Neural Network (CNN), Alzheimer's Disease (AD), Transfer Learning, Neural Network (NN), MRI, Deep Learning, Average Ensemble, VGG19, Inception-ResNet-v2, ResNet152v2, EfficientNetB5, EfficientNetB6

Dedication

We dedicate our thesis work to our supportive and loving family and friends.

Acknowledgement

To begin with, we are thankful to the Almighty Allah for letting us continue our thesis even through these dreadful and scary Covid-19 days without any major interruption. We are ever so grateful that we were all in good health, along with our loved ones.

Next, our thesis would have been unorganized and in disarray without the proper guidance and help from our supervisor Dr. Mohammad Zavid Parvez. His availability and kindness were a significant help during these past 3 semesters, of which 2 were online.

Furthermore, we would like to thank our co-supervisor Md Tanzim Reza whose knowledge and expertise was invaluable in conducting our research. His feedback was very much appreciated.

Finally, none of this would have been possible without the support of our beloved parents and family who stood beside us through even the longest of nights and were there to always pick us up.

Table of Contents

Declaration i

Approval ii

Ethics Statement iv

Abstract iv

Dedication v

Acknowledgment vi

Table of Contents vii

List of Figures ix

List of Tables x

Nomenclature xii

1 Introduction 1
- 1.1 Motivation 1
- 1.2 Contribution 2
- 1.3 Organization of paper 2

2 Background Information 3
- 2.1 Literature Review 3
- 2.2 Convolutional Neural Network 7
  - 2.2.1 Convolution Layers 8
  - 2.2.2 Activation Functions 8
  - 2.2.3 Pooling 9
  - 2.2.4 Dropout 10
  - 2.2.5 Batch Normalization 10
- 2.3 Transfer Learning 10
- 2.4 Neural Network Models 10
  - 2.4.1 VGG 10
  - 2.4.2 ResNet 11
  - 2.4.3 EfficientNet 12
  - 2.4.4 Custom 13
  - 2.4.5 Ensemble 15

3 Dataset Description **16**
3.1 Introduction 16
3.2 Dataset Details 16
3.3 Dataset Acquisition Process 16
3.4 Dataset Processing 16
3.4.1 Conversion to NIfTI format 17
3.4.2 Grouping into classes 17
3.4.3 Train-Test split 17

4 Proposed Methodology **18**
4.1 Introduction 18
4.2 MRI Data Acquisition 18
4.3 Data Preprocessing 18
4.4 2D Labeling 18
4.5 Grouping into classes 20
4.6 Classification using models 20
4.6.1 Classification using CNN models 20
4.6.2 Classification using Average Ensembling 20

5 Result, Analysis and Discussion **21**
5.1 Result And Analysis 21
5.2 Discussion 27

6 Conclusion and Future Work **29**
6.1 Conclusion 29
6.2 Challenges/Limitations 29
6.3 Future Work 29

Bibliography **33**

List of Figures

2.1 Architecture of a Convolutional Neural Network [19]. 8
2.2 ReLU Activation Function [21]. 9
2.3 Maxpool Operation in Convolutional Neural Network [24]. 9
2.4 Architecture of VGG19 [34]. 11
2.5 Architecture of Inception-Resnet-v2 [39]. 12
2.6 Architecture of EfficientNetB5 [42]. 13
2.7 Architecture of EfficientNetB6 [42]. 13
2.8 Custom Neural Network. 14
2.9 Average Ensemble of the 6 CNN models. 15

3.1 MRI Image Slices from OASIS Dataset. 17
3.2 Summary of subject demographics and dementia status [53]. 17

4.1 Block diagram of the proposed model. 19
4.2 MRI Images of the grouped classes(from left to right- Non-Demented; Very Mild Demented; Mild Demented; Moderate Demented). 20

5.1 VGG19 Model Accuracy. 21
5.2 VGG19 Loss Function. 22
5.3 Inception-ResNet-v2 Model Accuracy. 22
5.4 Inception-ResNet-v2 Model Loss Function. 22
5.5 EfficientNetB6 Model Accuracy. 23
5.6 EfficientNetB6 Loss Function. 23
5.7 ResNet152v2 Model Accuracy. 23
5.8 ResNet152v2 Loss Function. 24
5.9 Custom Model Accuracy. 24
5.10 Custom Loss Function. 24
5.11 EfficientNetB5 Model Accuracy. 25
5.12 EfficientNetB5 Loss Function. 25
5.13 Training and Validation Accuracies of the 6 individual models. 28
5.14 Test Accuracy of the ensembled models. 28

List of Tables

5.1 Ensemble Result of VGG19 and ResNet152v2 and Inception-ResNet-v2 26
5.2 Ensemble Result of EfficientNetB5 and EfficientNetB6 26
5.3 Ensemble Result of all 6 CNN Models 26

Nomenclature

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

AD Alzheimer's Disease

ADNI Alzheimer's Disease Neuroimaging Initiative

CDR Clinical Dementia Rating

CNN Convolutional Neural Network

DL Deep Learning

DMN Default Mode Network

DNN Deep Neural Network

DTI Diffusion Tensor Magnetic Resonance Imaging

EMCI Early Mild Cognitive Impairement

fMRI Functional Magnetic Resonance Imaging

FMRIB Functional Magnetic Resonance Imaging of the Brain

FSL FMRIB Software Library

GPU Graphics Processing Units

KNN k-Nearest Neighbors Algorithm

LMCI Late Mild Cognitive Impairement

MCI Mild Cognitive Impairement

MRI Magnetic Resonance Imaging

MS Multiple Sclerosis

NIfTI Neuroimaging Informatics Technology Initiative

NN Neural Network

OASIS Open Access Series of Imaging Studies

PD Progressive disease

$pMCI$ Progressive Mild Cognitive Impairement

$ReLU$ Rectified Linear Unit

RF Random Forest

RLO Random Linear Oracle

SGD Stochastic Gradient Descent

$sMCI$ Static Mild Cognitive Impairement

SVM Support Vector Machine

VGG Visual Geometry Group

WHO World Health Organization

Chapter 1

Introduction

Alzheimer's disease is one of the most prevalent reason for dementia in older people which is a long-lasting degenerative problem that slowly reduces and ultimately kills ones thinking abilities. It leads to a huge loss of cognitive and mental strength, behavioral problems as well as language difficulties. Even though it is known to affect people who are 65 or older, it is not necessarily directly proportional to increasing age. However, it is a fact that every 1 in 3 elderly people die of AD or other forms of dementia, which is a rate higher than deaths due to breast cancer and prostate cancer both combined [1]. This disease is one which has no cure at all, therefore the only way to be able to prolong our lives with it is to detect it as early as possible with the accurate stage so that proper precautions can be taken on time. It was estimated that in 2020, this disease along with dementia would have cost USA $305 billion which will eventually increase to as high as $1.1tn by 2050 [2]. In our country Bangladesh, very little research or statistics can be found on these issues but a recent WHO data which was published in 2018 suggested that 1.85% or 14,340 of total deaths in the country was a result of Alzheimer's and dementia related deaths [3]. For our thesis paper on detection and classification of AD and its stage, using proper CNN architectures and transfer learning techniques, we achieved valid results with increased accuracy compared to previous researches. OASIS MRI image dataset have been used in our research which performed very well and gave us very satisfactory results.

1.1 Motivation

Our first and foremost incentive for choosing this topic was to detect this disease with its stage as early and accurately as possible since Bangladesh is a country where very few people are aware about AD or its ramifications. It is naturally assumed that whenever an elderly person has some sort of lapse in memory, it is due to their age and not because of AD or dementia. That is a serious problem since Alzheimer's has no cure and can be potentially fatal. Moreover, the medical sector of our country is not very advanced yet and is usually very wrong in its diagnosis, thus, we wanted to propose a model which would detect this dangerous disease correctly so that life can be prolonged since this ultimately has no cure at all. The scope of improving this sector is potentially endless and a lot of work is yet to be done and a lot of breakthroughs to be reached. Not just our country, but in other big nations such as USA, billions of dollars are spent on AD annually and the papers that we have read so far for our knowledge all show varying methodologies and algorithms that achieve pretty decent outcomes. But, our aim was to build a model which would be a little different from those and give even better accuracy.

1.2 Contribution

Our primary focus was to contribute to the society by first, detecting AD as accurately as possible and then to further classify at which stage of the disease the patient is at. Moreover, we have implemented the most efficient and recent NN architectures in order to enhance our results even more. Finally, ensembling techniques were used which combined various different CNN models together to achieve the optimal output.

1.3 Organization of paper

In this paper, the first chapter is an introduction to the thesis followed by chapter two which contains background information of our research and models. Then comes dataset description and proposed methodology in chapters three and four. Next is results, analysis and discussion in chapter five and finally we have our conclusion which includes future work in our last chapter which is six.

Chapter 2

Background Information

2.1 Literature Review

This paper presents AD classification after detection using data analysis of brain MRI. An ensemble operation of deep CNN is developed and premium performance on the OASIS dataset is demonstrated. DenseNet-121, DenseNet-161, DenseNet-169 has been used as the proposed network architecture for this paper. In a model based on DenseNet, every layer is connected directly to all the subsequent layers. So, from all preceding layers, every layer receives the feature-maps. The interface maps act as the network's global state, where every layer will apply a feature map of their own. The global state is viewed from anywhere in the network, and the network growth rate determines how much each layer will add to. As input, patches from 3 physical planes of imagining such as horizontal or Axial plane, frontal or coronal plane, and median or Sagittal plane are used. The three DenseNets have previously been trained with ImageNet dataset and along with that, transfer learning was applied. Basically how this works is that each model takes an input which is an MRI image and shows us its understood depiction. This leads them to get 4 output classes. The data used are as follows- 70%, 20% and 10% as training data, test data and validation data respectively. For optimization, SGD algorithm along with early-stopping are used as the regularization technique. Therefore, the accuracy of their model is 93.18% which includes 92% recall, 93% precision, 92% f1score and this is mainly because of applying class weights during the training [2].

The paper first tried to identify a population of autoantibodies which is capable of predicting early-stage AD utilizing 50 MCI patients from ADNI, and they all had little CSF AB42 rates associated with continuing AD-related pathology and each had an early MCI or late MCI clinical diagnosis, a differentiation established on educationally changed cut-off marks on the Wechsler rational subscale of memory II. The working diagnostic biomarker panel was chosen from the best 50 uniquely distinguished autoantibodies in the population of the group of MCI. Using the chosen biomarkers, the samples were separated based on age and sex-balanced subjects with an overall predictive precision of 99.6% which were based on 5 replication tests. The testing and training sets were randomly generated with subjects such that both the tests had controls and cases that matched by gender and age. Each had 25 MCI plus 25 control subjects. Selected biomarkers have been checked with the RF algorithm, accuracy of the classification is recorded as OOB error score in an uncertainty matrix and misclassifications. The overall gist is that first, this research showed that a board of autoantibodies may be used to discriminate individuals having MCI from safe gender and age controls with good whole precision consistent with early stage AD pathology. Second, the introduction of separate biomarker detection techniques was perceived producing biomarker panels presenting a large correlation of identified comparable diagnostic and autoantibody biomarkers findings. Then, the chosen board of related biomarkers and their underlying diagnostic rationale were more specific meaning that MCI subjects could be differentiated from persons at a greater level of AD. Fourthly, the biomarkers were often disease precise because they were able to differentiate MCI subjects from others with mild-moderate and early stage PD,MS and premature stage breast cancer. Eventually, they were ineffective in identifying people with complex clinical labels not embedded in substantial variations of the disease. Therefore, this was an unsuccessful attempt [4].

Grading approach based on graph that combined similarities with inter-subject and variability with intra-subject is the new method that was developed in this paper. This new method was tested with 2 separate anatomical scales which are the entire brain structures along with the hippocampal subfields. This was the first contribution. Followed that was, the creation of the fusion of the anatomical scale using an approach which is called cascade of classifiers. They extended the ranking system to the hippocampal subfields and a parceling of the whole brain structures which were focused on maps of multi-scale. Then they compared every component of the graph at every anatomical scale to validate the new, multi-scale grading framework. Then, they contrasted the findings obtained in their research with the outcomes of the state-of - the-art literature methods suggested. Finally, with a set of cognitive results used in medical procedures, they correlated the findings obtained with the imaging-based biomarker. In this research paper, they used ADNI dataset. Two clusters of MCI were considered. The 1st one consists of subjects who have sMCI. The 2nd had patients who had MCI symptoms initially and then in the next thirty six months evolved to AD. The group was termed pMCI. Their multi-scale graph-based rating system demonstrates good results for AD estimation with a region below the curve (AUC) of 81%. In comparison, the alternative approach earned 85% of AUC when paired with cognitive ratings [5].

Pictures are obtained utilizing the image sequence T2. The methodology has a set of quantitative procedures: extraction of features, filtering, feature selection based on Student's t-test, and classification based on KNN. The key objective of the study was to establish a CABD method, focused on T2-weighted brain MRI images, to distinguish regular and AD instances. Another aim of this research was to create a relative study of the feature mining techniques considered, to detect the supreme feature mining techniques promising greater efficiency. The main aim of this research work was to create a comparative analysis of the characteristic mining procedures considered, in order to identify the ideal function mining technique promising superior efficiency. The pre-processing of the images for testing were originally applied with a median filter that aims to improve the images by eliminating the noise. A feature extraction method was then applied to extract the important features of MRI images. A complete relative analysis was conducted in this research of the function selection techniques, which included the Contourlet Transform (CoT), Curvelet Transform (CuT), Complex Wavelet Transform (CWT), Dual Tree Complex Wavelet Transform (DTCWT), Discrete Wavelet Transform (DWT), Empirical Wavelet Transform (EWT), and Shearlet Transform (ST); Then they used Student's t-test to pick a subclass of functions for classification dependent on the p value. Lastly, to classify brain MRI, the classifier KNN was used. The methods ST+KNN gave a precision, accuracy, specificity and sensitivity with the standard MRI data of 100%, 98.48%, 100% and 96.97% correspondingly [6].

This research focused on how to improve the accuracy using small datasets which are done by merging the small local dataset which is data source I including 26 subjects, with a large shared dataset from ADNI which is data source II with 86 subjects. FMRI dataset was used for this paper. An adaptation method was used and it goes as follows: first, via brain network modeling we have extracted weighted relations between various functional regions as original features. Second, a feature selection phase was conducted prior to the domain adaptation, since the actual features were very big. The selected features were spread to 2 separate feature spaces from the two main data sources. Then, we conducted an updated subspace alignment to fit the sample points into the same subspace from the two different feature spaces. Finally, the matched samples for the classifier training and testing were as an integrated dataset in one subspace. The classification function was not done well with the local small data set from the Tongji dataset. Its accuracy was just above 50%, which is the same as random guesses. When we used ADNI's wide shared data set and mixed with Tongji's data set, the classifier's output improved dramatically, reaching 60% accuracy. Thus, the expansion of the training set size actually enhanced the classifier's efficiency. This progress was, however, insufficient relative to the classification findings achieved from the adaptation trial, where the accuracy was above 80%. dissatisfying precision prior to adaption suggests that accurate classification findings cannot be obtained easily by incorporating separate datasets to provide a greater training set. With adaptation, the accuracy was 84.6%, sensitivity was 92% and specificity was 79% [7].

Mild cognitive impairment, MCI, is the initial stage of AD. This paper suggested an approach named voxel-based hierarchical feature extraction (VHFE) for the early identification of AD. Following a template for Automatic Anatomical Labeling (AAL),the whole brain was first segmented into ninety regions of interest, ROIs for short. Using their standard values they picked the instructive voxels from each ROI to separate the uninformative data, and organized them into vectors. Following that, depending on the voxel interrelationship of various groups, the features of the first phase are selected. After that, to know the deep hidden features, each subject's brain feature maps which are constructed from the fetched voxels are passed into the CNN. At the end, to verify the feasibility of the suggested procedure we check it using the subclass of Alzheimer 's Disease Neuroimaging (ADNI) dataset. The test outcome indicates that the suggested technique is robust in contrast with the modern, up to the date methods. They were broken down to 3 categories- MCI VS NC, AD VS NC and finally AD VS MCI. 4 categories of subjects with 6 matched groups were part of ADNI-2. There were EMCI vs AD, NC vs EMCI, EMCI vs LMCI, NC vs LMCI, LMCI vs AD and NC vs AD. The used method got a result of 99.7%(AD vs NC), 97.8%(AD vs NCI) and 97.7%(NC vs MCI)- ADNI-1 dataset.

For ADNI-2, it enhanced the accuracies by 2.66% (AD vs LMCI), 1.99% (AD vs EMCI), 3.16% (EMCI vs LMCI),4.16% (NC vs EMCI), 1.7% (AD vs NC) and 3.97% (NC vs LMCI) [8].

Throughout this article, they suggest an approach based on transfer learning to diagnose Alzheimer's disease from images which are MRI. They believe following a strong and validated framework for natural MRI data and utilizing transfer learning consisting of learned training data collection will not only boost a model's accuracy but also minimize dependence on a big training set. They test their hypothesis with comprehensive studies on the dataset model acquired from ADNI, where 50 subjects' MRI scans from three groups each, which were AD, MCI and NC (total 150 subjects), was taken to achieve results of accuracy. By freezing some layers in their design, they investigate whether layerwise transfer learning has an impact on their program, and they show in-depth findings of connections with their training data size and layerwise transfer learning. In the end, they present comparative outputs with another 6 state-of - the-art approaches, where their proposed model performs better than the others significantly, giving an increase in accuracy of 4 percent and 7 percent over the rest for classification of AD vs. MCI, MCI vs. NC. They announced tests of three way classification as well, attaining remarkable efficiency, demonstrating the strength of the proposed approach. In this paper, they performed the test on 2 categories per subject- first is with 16 images and second with 32 images. The accuracy was greater with the latter category, with AD vs NC being 99.36%, AD vs MCI was 99.2% and MCI vs NC was 99.04%. This resulted in the average being 99.2% [9].

This research used CNN to detect a healthy brain from one having Alzheimer's. The architecture used was the famous LeNet-5 which used fMRI data from ADNI dataset which consisted of 24 females and 19 males, some with the disease and some with normal healthy brains. The topology of CNN uses spatial relationships to minimize the number of learned parameters, thereby enhancing general training for feed-forward back propagation. This network is essentially composed of neurons with learnable weights and biases that form the Convolutional Layer. Other network structures such as Pooling Layer, FullyConnected Layer and Normalization Layer are also included. The data was separated into 3 parts: training (60%), testing (20%) and validation (20%). Because of its convolutional layers, CNN is a good feature extractor that can extract high-level features from a huge number of images. Due to its complicated network architecture, this deep learning approach is also a strong classifier. This current approach led to an accuracy of 96.86% which is far greater than what was achieved using Support Vector Machine (SVM) which was only 84% [10].

Here, they used resting state fMRI results from multicenter analysis to evaluate group division precision along with regional selection of feature reliability and subsequent detection of cortical systems separating AD subjects and controls between logistics regression which is cross-validated regularized with elastic net regression and classic stepwise logistical regression. They theorized that elastic net logistic regression will result in greater generalizable collection of features and bigger robust network detection than classical stepwise logistic regression. In this research they used 53 patients with the clinical probability of having AD and 118 healthy elderly control patients. The dataset used in this research was collected from 4 separate 3.0 Tesla MRI scanners. FMRI dataset processing was done using Data Processing Assistant for rs-fMRI. Two regression models was used to predict group membership such as Bidirectional stepwise un-penalized logistic regression and penalized logistic regression technique with an elastic net penalty. To train this model, 2/3 of the data from random samples were taken. Step-wise regression resulted in only 70% accuracy compared with elastic regression. Their results refer to a broad functional disconnection network, together with the dorsal DMN but often including functional links involved throughout concentration, object detection and dialectal processing. In a study of Alzheimer's and control instances, elastic net regression produced diagnostic precision that was cross-validated which exceeded, however failed to be a standard meaningful AD biomarker (Consensus Group, 1998); analytical methods focused on stepwise regression did not even reach this standard [11].

In the following research paper, various Machine Learning techniques (Pattern recognition, SVM, Naïve Bayes, RF, RLO) have been used to classify the fMRI data for AD diagnosis which is proved to be a very powerful method to identify many regions in the brain. Next, using the mapping toolbox for statistical parameters, the usable image data were preprocessed to generate statistically enabled voxel maps for individual maps. Afterwards, the voxels were selected by applying a fast filter and these voxels were commonly allowed between demented as well as non-demented groupings. Using stochastic methods, the pattern in the fMRI data is recognized. The research seems to have 97.14% accuracy on average. The research is done on both demented patients and non-demented patients. Despite their disparate mathematical origins, numerical results were identical in terms of voxel selectors. This indicates that in the selection process, data quality along with its separability is more important than the actual algorithm. These numerical results prove that in clinical scenarios, classification models are apt. For each combination, the greatest values display a low standard deviation [12].

2.2 Convolutional Neural Network

CNN falls under DNN that is most often intended for examining visual data [13]. It is also known as ConvNet or CNN which is a powerful Neural Network Algorithm for image recognition and classifications. CNN is a feed-forward network which takes an array of pixels as inputs and applies filters to those [14, 15].
A convolutional layer has a number of filters that does some convolutional operation. After an image is passed through a filter, it is called a filtered image and this operation is called convolution [16] . Then, the tensors are maxpooled and the result is sub-sampled to give a smaller image. Due to this above procedure, the number of parameters decrease, reducing the complexity [17].
The convolution was done several times in our model to give subsampled nodes which were flattened. These were fed into a fully connected feed-forward network having shared weights on the edges to compress the fully connected network [18].

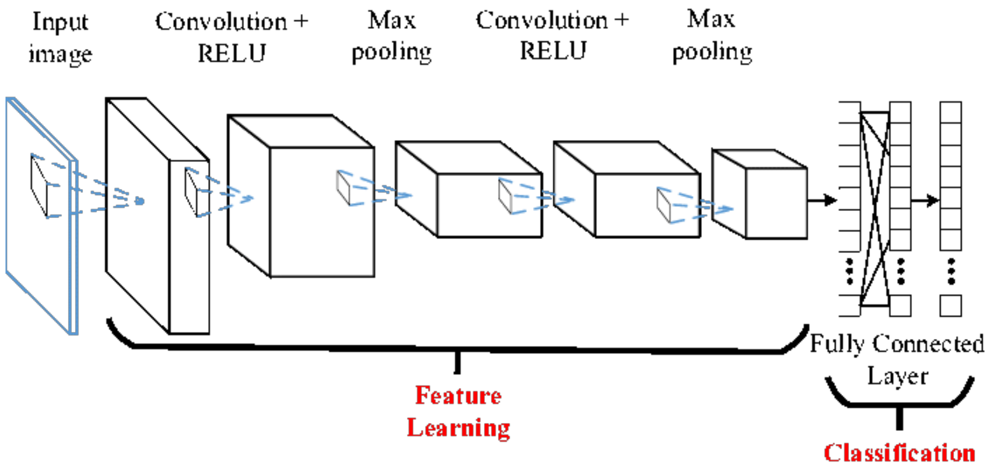


Figure 2.1: Architecture of a Convolutional Neural Network [19].

2.2.1 Convolution Layers

Convolution Layers basically used for feature extraction process which is the first layer of a CNN. For this feature extraction, some filters are applied on the image data which is basically a grid like pattern or matrix. All the filters are convolved with each input image to create a feature map of neurons. Here dot production between filters and inputs take place and the outcome of every convolutional layer gets simpler. Instead of fully connected neural network we get fewer shared parameters which reduces the complexity and increases the efficiency of the model.

2.2.2 Activation Functions

In neural network, activation functions are a very important component. An activation function is needed to decide whether a neuron gets fired or not. It takes output of the previous layer as input and based on the type of activation function we used it gives us the output. It is basically divided into two types which are Linear Activation Function and Non-linear activation function. Rectified Linear Unit or simply known as ReLU is used in CNN as most used activation function since it brings nonlinearity in outputs. It does not take any non-negative values rather converts all negative values to zero which does not activate the neuron. Through this computation gets efficiency since fewer neurons are activated per time. Softmax is another activation function that transforms a number vector into a vector consisting of probabilities which is then used for classification [20].

$$\text{Softmax}(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \tag{2.1}$$

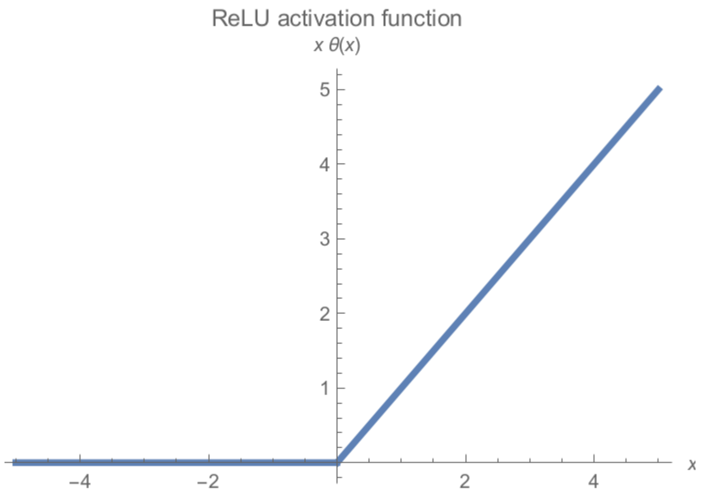


Figure 2.2: ReLU Activation Function [21].

2.2.3 Pooling

The convolutional layers methodically apply trained filters to input to create a feature map that contains all the summarized features of the input. A disadvantage of this feature map output is that the exact location of the features that are in the input are recorded. This means that even the tiniest changes in the location of the features in the input image can cause the entire feature map to shift and therefore change entirely. This could happen with any slight modification such as rotation, cropping, flipping, etc. The most usual way to solve this by using a layer of pooling [22]. All the feature maps are worked on separately by the pooling layer and that results in the same number of new pooled feature maps. The size changes after this filter is added, which is 2 by 2 pixels with a stride of 2, and therefore the feature map is always reduced by a factor of two, thus decreasing the pixel number in each feature map. The pooling function used is maxpooling where the number taken is the maximum value from that particular feature map and that is calculated and the output is sub-sampled to give a smaller image. Due to this above procedure, the number of parameters decrease, reducing the complexity [23].

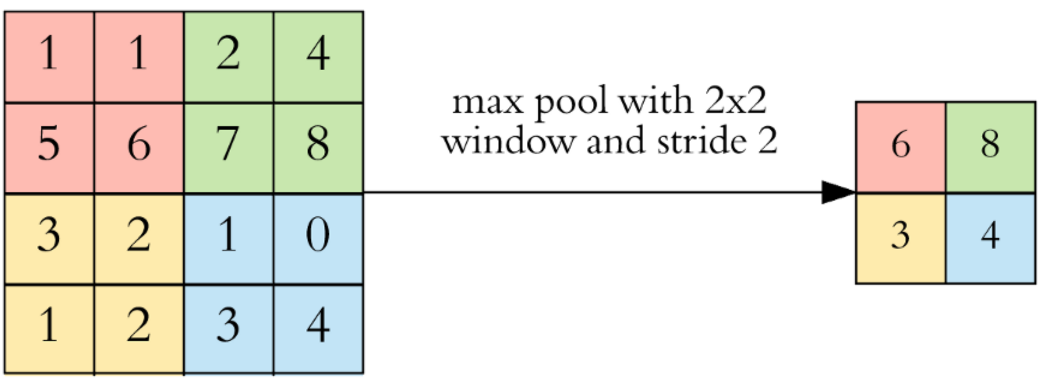


Figure 2.3: Maxpool Operation in Convolutional Neural Network [24].

2.2.4 Dropout

Dropout is dropping out certain units as well as their connections, which include both visible and hidden layers of a NN. In simple terms, it is randomly ignoring certain neurons during the training phase of a batch of neurons. It performs by setting hidden layers' outgoing edges to zero during every modification of training stage. This majorly reduces overfitting and gives better results than other regularization techniques [25].

2.2.5 Batch Normalization

Training DNN is complex because the distribution of inputs from each layers change during the training phase due to the changing previous layers' parameters. That is the reason batch normalization is required. For each mini-batch, the inputs to a later are standardized by batch normalization, which is a method for training DNN [26]. This in turn stabilizes the learning method and significantly decreases the quantity of training cycles which is necessary for deep networks to be trained. It speeds up the training, by reducing the number of epochs by 2 or even better in few cases. It also gives regularization effect, limiting generalization error. Furthermore, every layer independently learns by itself a little bit more than the others [27].

2.3 Transfer Learning

Transfer learning is machine learning research problem that basically reuses a pre-trained model on a new problem which then ultimately helps with training deep neural networks with relatively little data and saves training time as well [28]. It works by initially loading a previously trained model of large dataset. Then the lower convolutional layers are frozen and a custom classifier with several layers of trainable parameters to model are added. The layers are then trained on the available data and finally with the help of fine-tuning the hyper-parameters, layers are unfrozen as required [29, 30].

2.4 Neural Network Models

We have proposed 6 models for our Alzheimer's disease detection and classification and they are described in the following chapters. Those architectures include VGG19, ResNet152v2, Inception-ResNet-V2, EfficientNetB5, EfficientNetB6 and a Custom CNN that we developed on our own [31]. After executing those models individually, we then ensembled thrice with different sets of learning algorithms put together. They were firstly, VGG19, ResNet152v2, Inception-ResNet-v2; secondly the 2 EfficientNet models; and finally a combination of all the previously mentioned 6 NN.

2.4.1 VGG

Among all of the other architectures, VGG Net is a simple and straight forward CNN architecture. The definition of the architecture of VGG is very simple. With increasing filter sizes, we have to stack the convolutional layers. That is, if there are

16 filters in layer-1, then layer-2 must have 16 or more filters. Another interesting aspect is that all filters are sized 3x3 in any VGG architecture.

VGG19

VGG19 is of 19 layers consisting of Convolution, fully connected, MaxPool and SoftMax layers and they are of numbers 16, 3, 5 and 1 respectively [32]. For more than a million images, a previously trained sort of the network can be loaded from the ImageNet database. Consequently, for a large variety of images, the model of resolution 224x224 image inputs, has learned characteristic representations that are rich in nature [33].

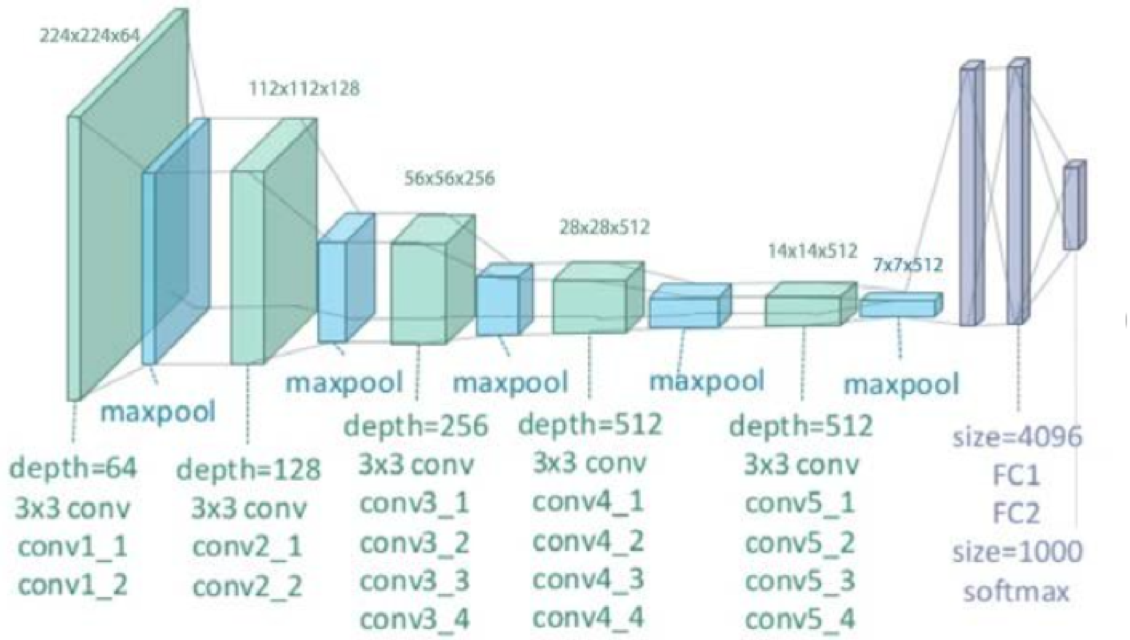


Figure 2.4: Architecture of VGG19 [34].

2.4.2 ResNet

Residual Networks (ResNet), is a convolutional NN used as a foundation for a lot of computer vision applications. ResNet enables training of upto hundreds or even thousands of layers and still achieves compelling results [35]. It solves the problem of vanishing gradient and considerably improves the performance of neural networks [36]. One of ResNet's greatest benefits is that it prevents negative results while increasing network depth. So, we can increase the depth, but we still have faster training and greater accuracy.

Resnet152v2

ResNet-152 introduced the residual learning principle in which the subtraction of the feature is learned by shortcut connections from the input of that layer. The primary merit of this model, which attains the top accuracy among all the versions of the ResNet family, is that it generates improved accuracy of classification without the model becoming increasingly complex [37].

Inception-Resnet-v2

Inception-ResNet-v2 is a Convolutional Neural Network architecture of 164 layers deep and builds on the inception family and integrates residual connections. It combines two architectures together to further boost its performance [38].

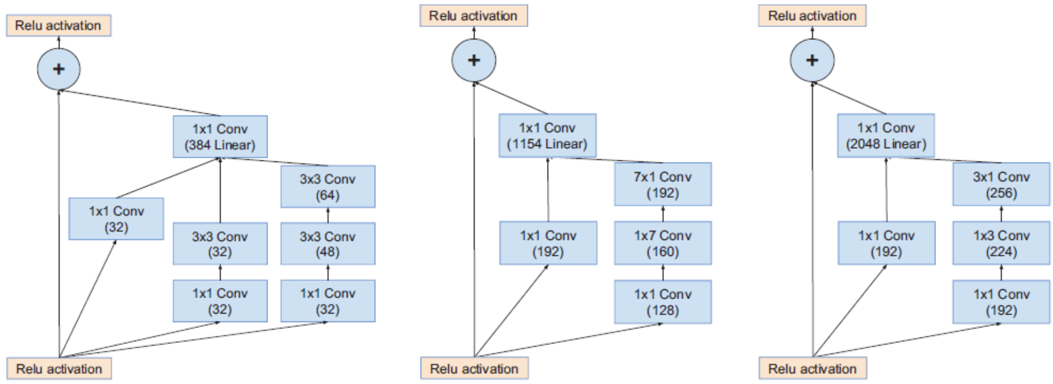


Figure 2.5: Architecture of Inception-Resnet-v2 [39].

2.4.3 EfficientNet

EfficientNet is a CNN architecture and scaling technique that uses a compound coefficient to evenly scales all depth/width/resolution dimensions [40]. The EfficientNet scaling technique evenly weighs network depth, width, resolution with a collection of standard scaling coefficients, unlike conventional practice that randomly measures these variables. This architecture is more accurate and efficient than other ConvNets [41].

EfficientNetB5

This model is one of the EfficientNet models aimed to perform classification on image dataset.

Figure 2.6: Architecture of EfficientNetB5 [42].

EfficientNetB6

This model, with a greater number of parameters, is one of the EfficientNet models also meant to perform classification on image dataset.

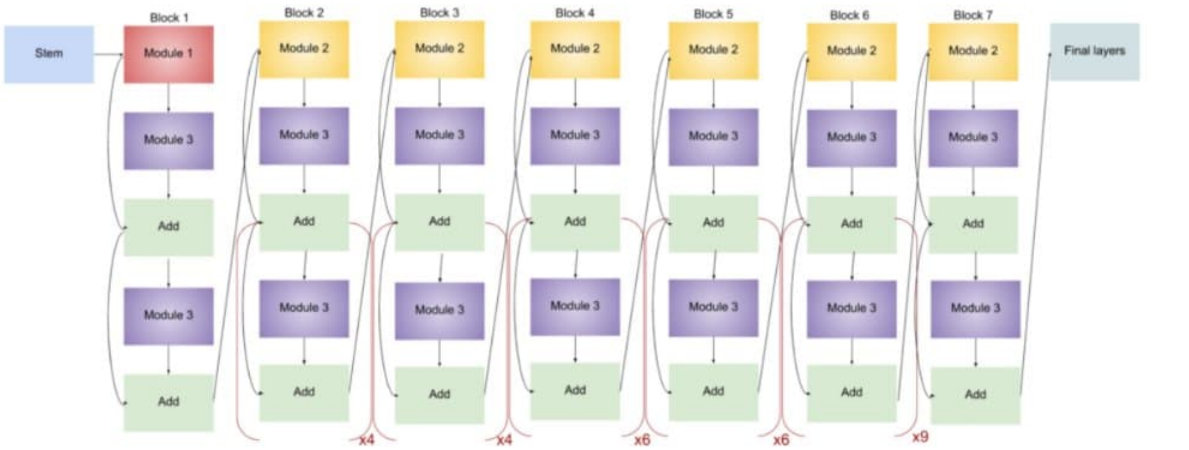


Figure 2.7: Architecture of EfficientNetB6 [42].

2.4.4 Custom

According to our custom architecture, there was an input layer followed by convolutional, maxpooling, sequential and dense layers which were all repeated in various instances.At first we used 2 Conv 2D layers, which will deal with our 2D images, and the activation function that was used was ReLU. This was followed by maxpooling. Furthermore, 3 sequential models, which helps build a model layer by layer, were used which consisted of each of them having 2 separable Conv 2D layers along with batch normalization and maxpooling. This ended with a dropout layer. Next, 2 Conv 2D layers were then again used but this time with LeakyReLU as the activation function. Consecutively maxpooling and dropout was performed.

Then a flatten layer was applied which serves as a link between the dense and the convolutional layer. The 4 sequential models implemented afterwards each called a dense block function which included a dense layer, batch normalization and dropout. Finally, softmax activation function was used for multiclass classification.

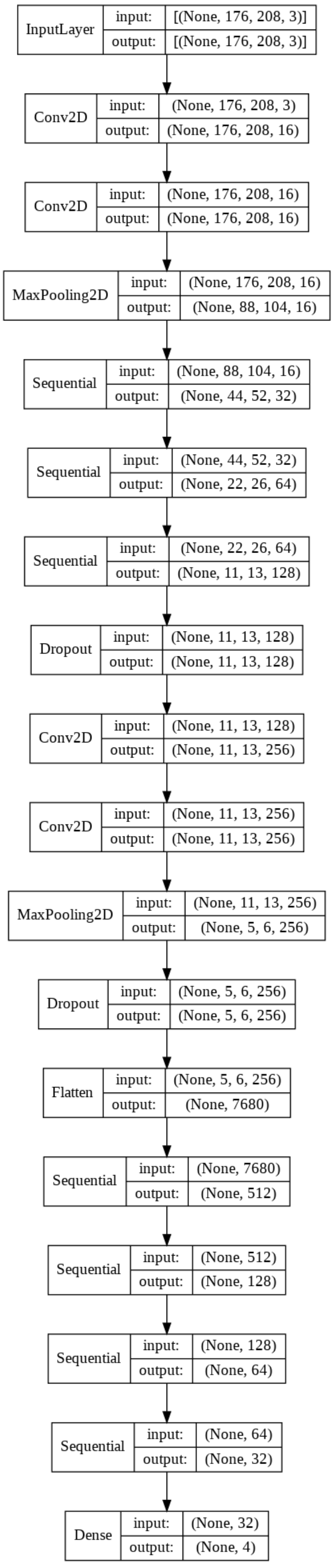


Figure 2.8: Custom Neural Network.

2.4.5 Ensemble

Ensembling is a method used in machine learning and statistics which works with multiple learning algorithms together in order to increase the predictive performance than what would be obtained from a single algorithm [43]. The models combined give better and more accurate results [44]. We used the test data in order to predict the outcome and then used average ensembling on those models in order to get the final output. Average ensembling, also known as model averaging, is an ensemble learning method which groups the predicted result from each model equally with improved performance on the average. We used this technique because it is simple and straightforward.

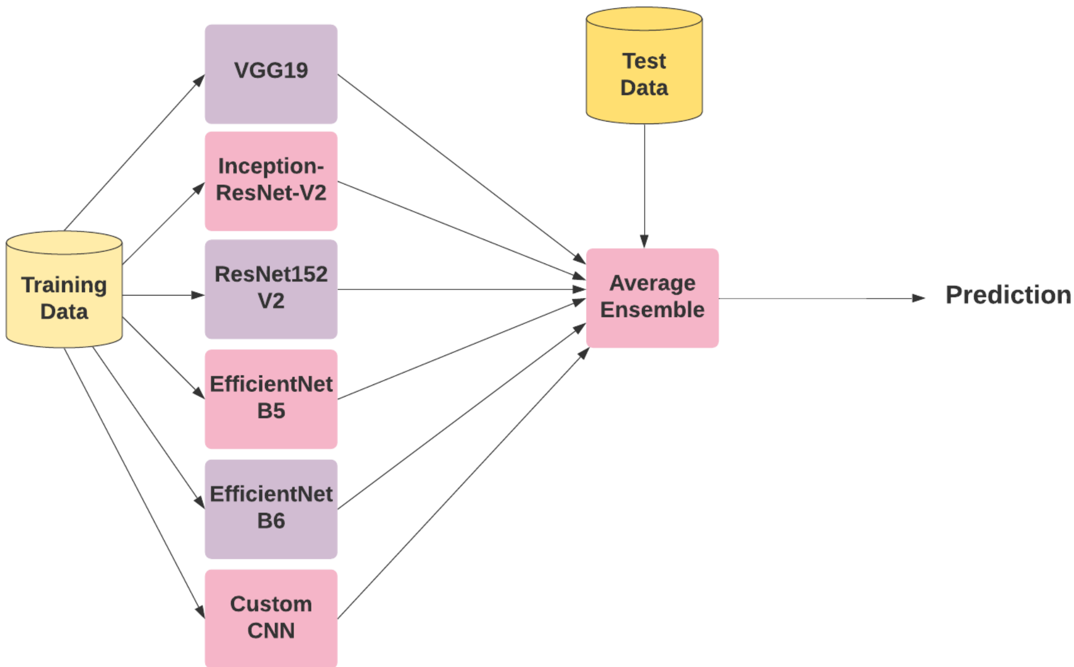


Figure 2.9: Average Ensemble of the 6 CNN models.

Chapter 3

Dataset Description

3.1 Introduction

OASIS stands for The Open Access Series of Imaging Studies, a project specifically designed to make MRI datasets of the brain freely and readily accessible for study and analysis towards the scientific public [45]. This is so that these datasets help in moving the sector of basic and clinical neuroscience forward through new discoveries. The Washington University Alzheimer's Disease Research Centre, the Neuroinformatics Research Group (NRG), the Biomedical Informatics Research Network (BIRN) and Dr. Randy Buckner at the Howard Hughes Medical Institute (HHMI) have made this OASIS available.

3.2 Dataset Details

The OASIS-I dataset that we are using consists of 416 subjects, with each providing 3-4 single T1-weighted MRI scans in one session [46–49]. The patients were aged from 18-96, were of both the genders and right-handed. Out of the 416, 100 of them were aged over 60 and have been identified with very mild to moderate AD. Moreover, 20 non-demented individuals have also been added for reliability [50].

3.3 Dataset Acquisition Process

We first visited the OASIS website because it is an authentic site which contained MRI images for research and analysis. Then we searched for the data we wanted and upon finding it, applied for access to the data so we can use it for our thesis project. After gaining access to the dataset of OASIS-1, we applied it to our research models.

3.4 Dataset Processing

We started processing the data after it had been downloaded. This step helps in improving productivity and profits along with reducing costs, storage convenience, distribution and reporting and finally assessment and evaluation. It is clear that a lot of effort goes into preparing the data. The method of processing is described below.

3.4.1 Conversion to NIfTI format

All the data were in 16-bit big-endian Analyze 7.5 format after the initial download. Although this format can be used, it has the problem of not including orientation information. Therefore, these had to be converted to NIfTI format before work could be done on them [51]. For this particular conversion, a comprehensive library was used called FSL. That library included tools for MRI, FMRI and DTI image data analysis. NIfTI was used because it is one of the most omnipresent file formats for storing neuroimaging data such as a brain MRI. A python package that works with these kinds of data is NiBabel. NiBabel takes NIfTI as input and extracts a 2D slice from the data which was stored as 3D volume. For our research, we have taken 3 slices from each patient which can be seen in Figure 3.1 [52].

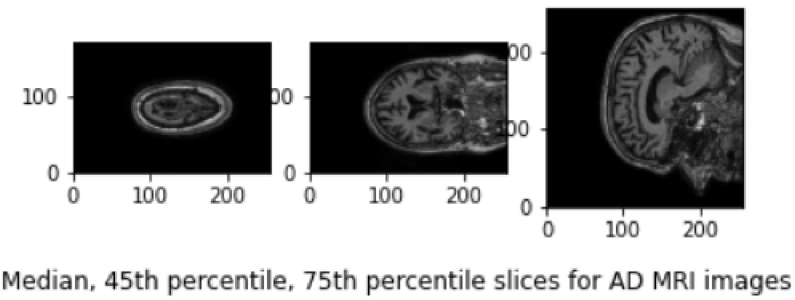


Figure 3.1: MRI Image Slices from OASIS Dataset.

3.4.2 Grouping into classes

The Clinical Dementia Rating (CDR; 0=nondemented; 0.5=very mild dementia; 1=mild dementia; 2=moderate dementia) was provided for all the subjects. From the csv file, we then divided the data into four classes which are Non-demented, Very mild-demented, Mild demented and Moderate demented according to the patient's ID and their CDR.

		Non-Demented				Demented				
Age Group	N	n	mean	male	female	n	mean	male	female	CDR 0.5/1/2
<20	19	19	18.53	10	9	0		0	0	0/0/0
20s	119	119	22.82	51	68	0		0	0	0/0/0
30s	16	16	33.38	11	5	0		0	0	0/0/0
40s	31	31	45.58	10	21	0		0	0	0/0/0
50s	33	33	54.36	11	22	0		0	0	0/0/0
60s	40	25	64.88	7	18	15	66.13	6	9	12/3/0
70	83	35	73.37	10	25	48	74.42	20	28	32/15/1
80s	62	30	84.07	8	22	32	82.88	13	19	22/9/1
≥90	13	8	91.00	1	7	5	92.00	2	3	4/1/0
Total	416	316		119	197	100		41	59	70/28/2

Figure 3.2: Summary of subject demographics and dementia status [53].

3.4.3 Train-Test split

The whole dataset was broken down into 3 parts. The dataset parts were training, validation and testing and there were 60%, 20% and 20% respectively.

Chapter 4

Proposed Methodology

4.1 Introduction

In this chapter, we will be discussing the complete methodology of our proposed models in a summarized version.

4.2 MRI Data Acquisition

Our data was acquired from the OASIS-1 dataset which consisted of MRI images from 416 patients which included subjects with AD as well as without the disease. For every patient, three to four T1 weighted MRI scans were included and a CSV file was included in the dataset which contained the patient ID, gender, age, true labels, etc. which helped us with the classifying process for the train-test split.

4.3 Data Preprocessing

The images that we downloaded from the OASIS website had to be preprocessed according to our requirements and thus a lot of work had to be first done on them to bring them to the correct format. These included multiple conversions before work could be started on the MRI images.

4.4 2D Labeling

The images we acquired from OASIS were in Analyze 7.5 format which we later first converted to NIfTI format using FSL library and then further converted to 2D using the NiBabel python package. These MRI images in 2D were used as our input in the models.

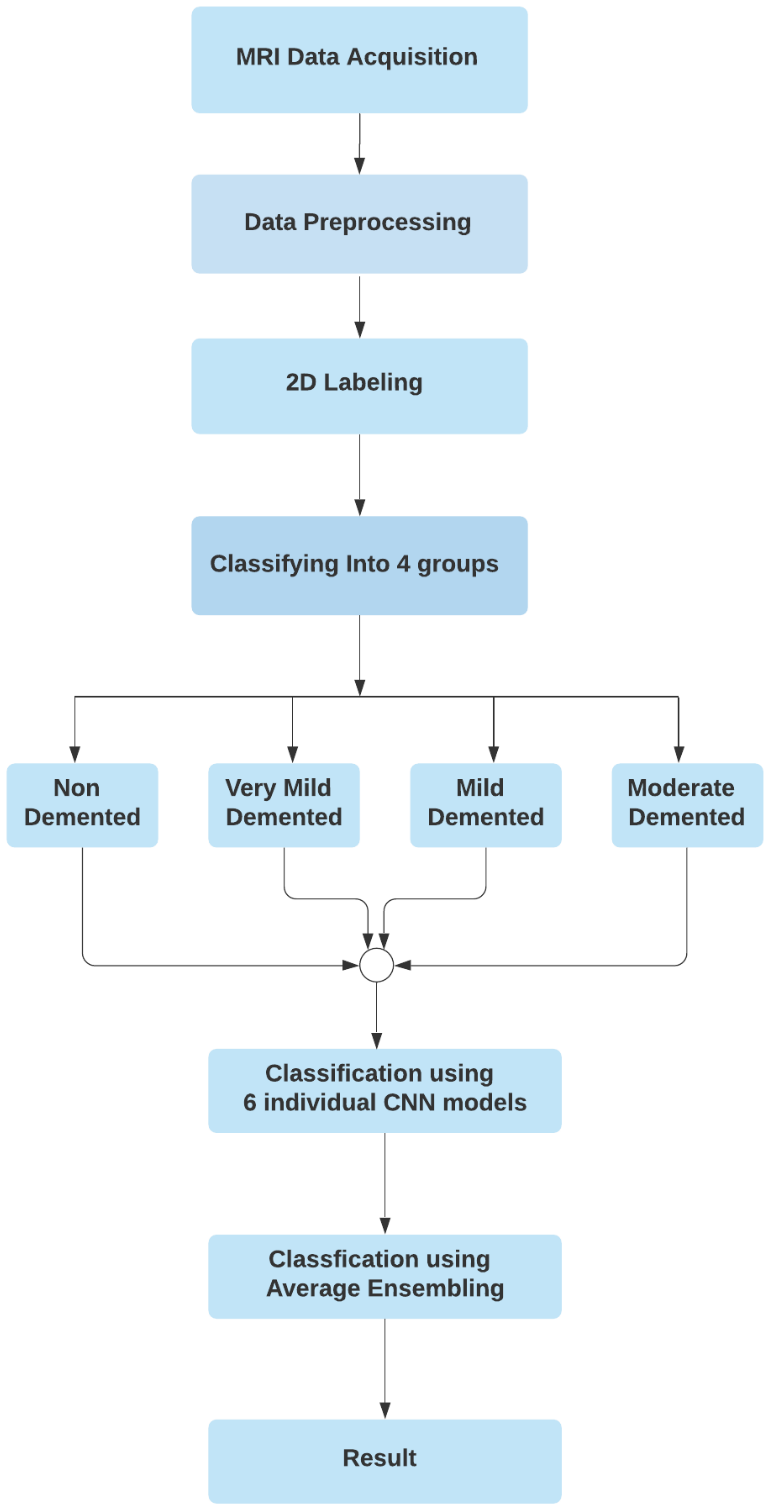


Figure 4.1: Block diagram of the proposed model.

4.5 Grouping into classes

For our thesis, we grouped a patient into one of four classes which included Non-demented, very mild demented, mild demented and moderate demented. This grouping was done primarily with the help of the CDR (Clinical Dementia Rating) from the downloaded csv files. We later split these groups into 60% training, 20% validation and 20% test.

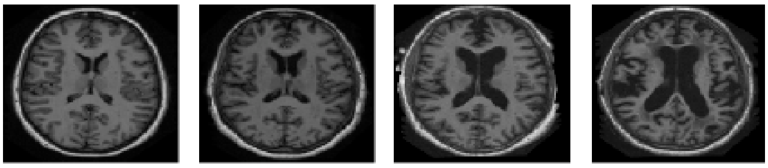


Figure 4.2: MRI Images of the grouped classes(from left to right- Non-Demented; Very Mild Demented; Mild Demented; Moderate Demented).

4.6 Classification using models

4.6.1 Classification using CNN models

We first used a VGG architecture which was VGG19. Compared to other architectures, since VGG-19 is relatively small, we have added some extra convolutional and dense layers towards the end before prediction. Then, ResNet152v2 as well as Inception-ResNet-v2 was used and Since ResNet is already a large architecture, we have fine-tuned both the models by freezing the weights of quite a few convolutional layers and deleting few nodes at the end to better suit our data. Furthermore, EfficientNetB5 and EfficientNetB6 from the EfficientNet family was chosen for our project and we have modified both the models just like what was previously done with ResNet. Finally, a custom CNN model was built which included an input layer followed by convolutional, maxpooling, sequential and dense layers which were all repeated in various instances.

4.6.2 Classification using Average Ensembling

We ensembled thrice, once with VGG19 and ResNet152v2 and Inception-ResNet-v2; then with EfficientNetB5 and EfficientNetB6; and finally with all the 6 architectures together.

Our work was done using Google Colaboratory Notebook with Tensorflow version 2.3.0 which is a free version and comes with 12GB RAM and a limited access to GPU. Hardware limitations played a major factor in us having to use this free version. To train deep learning models, multiple matrix operations can be performed in parallel using GPU. The library Tensorflow was used because it comes with various in-built functions and models which can be easily applied to our project.

Chapter 5

Result, Analysis and Discussion

5.1 Result And Analysis

For all the models except Custom, we have used the ReLU activation function followed by Softmax for classification. The Custom was also run using LeakyReLU along with the traditional ReLU in order to further experiment if that helped improve accuracy in any way. The validation scores (in ascending order) that we achieved by simulating the individual architectures were 82.1% for VGG19, 84.3% for Inception-ResNet-v2, 87.1% for EfficientNetB6, 89.4% for ResNet152v2, 91.1% for Custom CNN, 92.2% for EfficientNetB5.

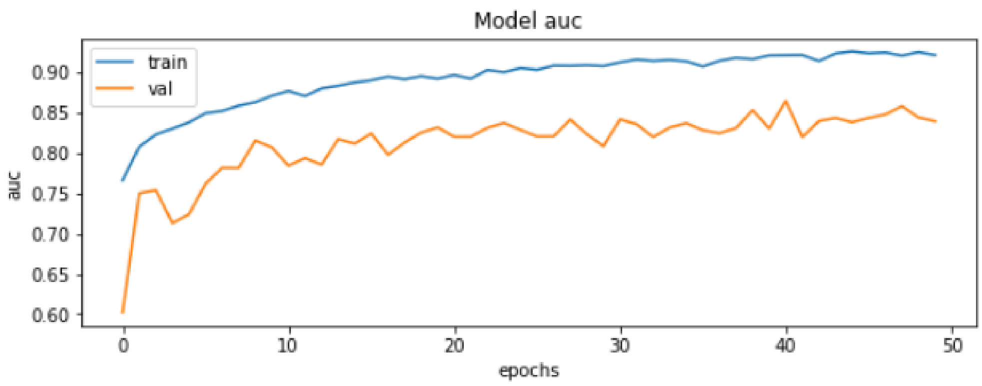


Figure 5.1: VGG19 Model Accuracy.

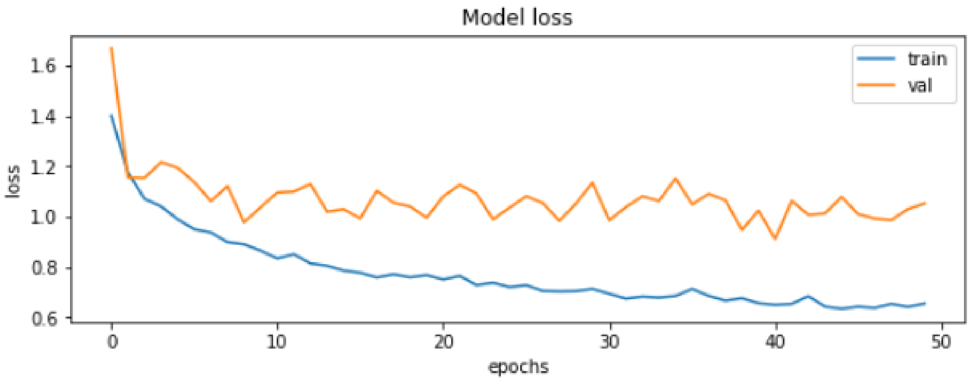


Figure 5.2: VGG19 Loss Function.

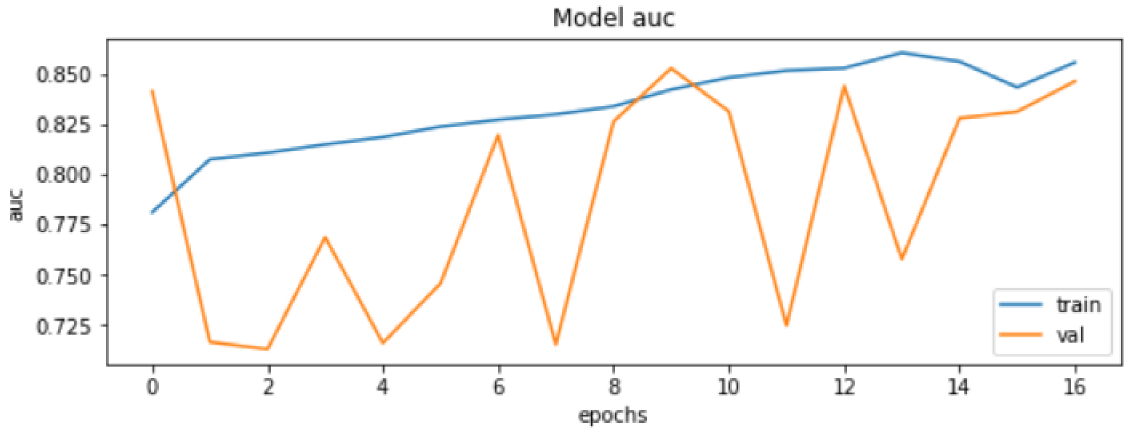


Figure 5.3: Inception-ResNet-v2 Model Accuracy.

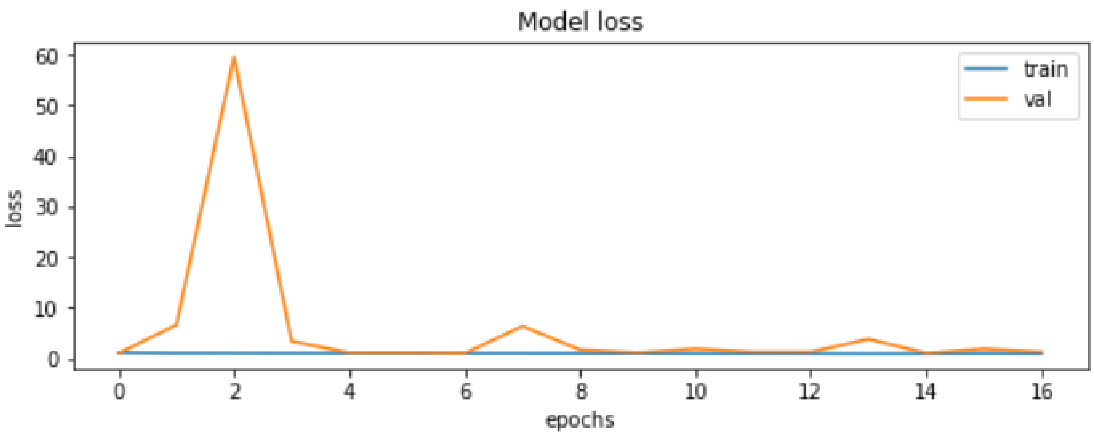


Figure 5.4: Inception-ResNet-v2 Model Loss Function.

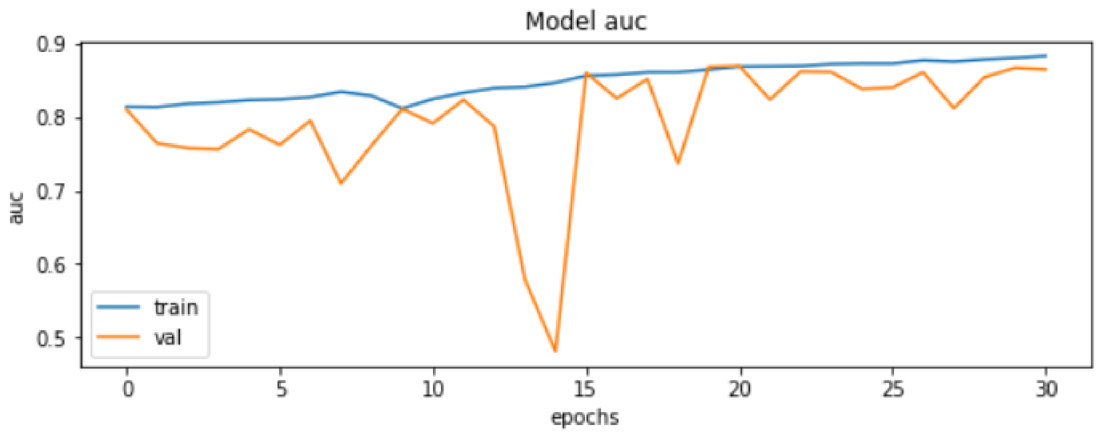


Figure 5.5: EfficientNetB6 Model Accuracy.

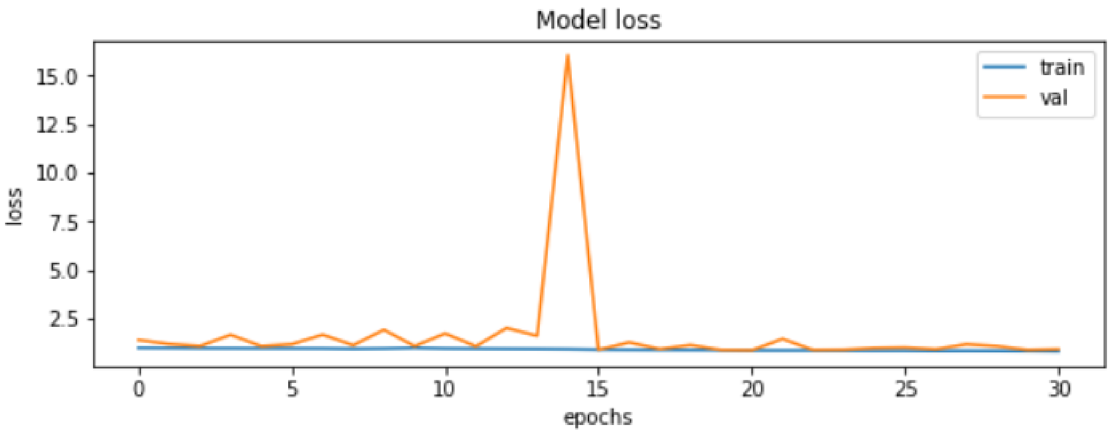


Figure 5.6: EfficientNetB6 Loss Function.

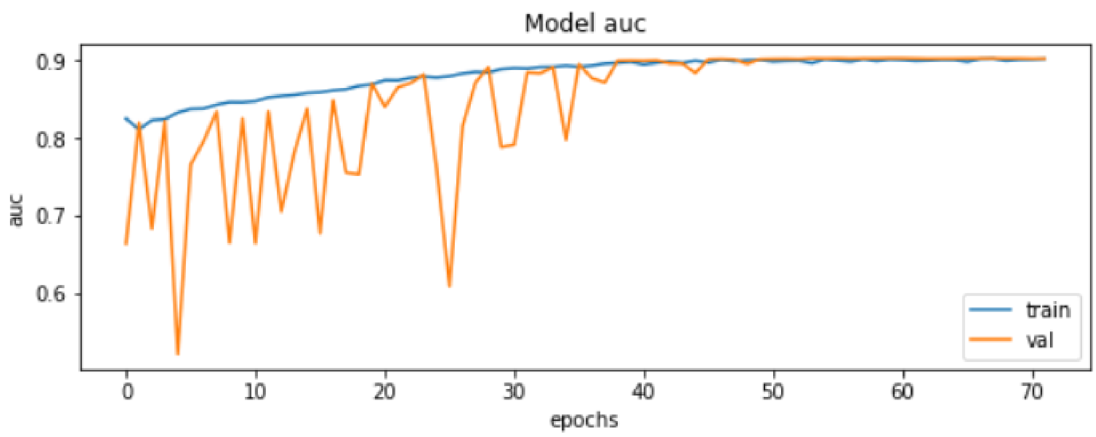


Figure 5.7: ResNet152v2 Model Accuracy.

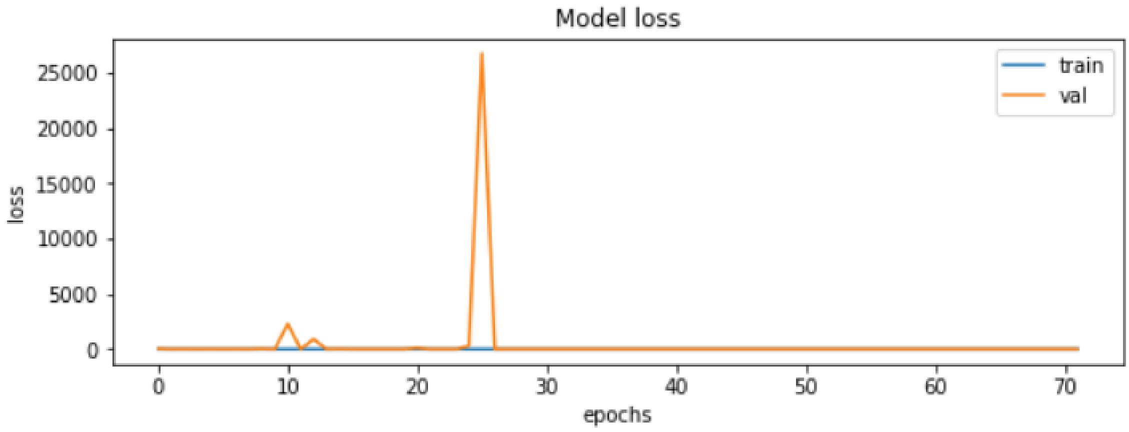


Figure 5.8: ResNet152v2 Loss Function.

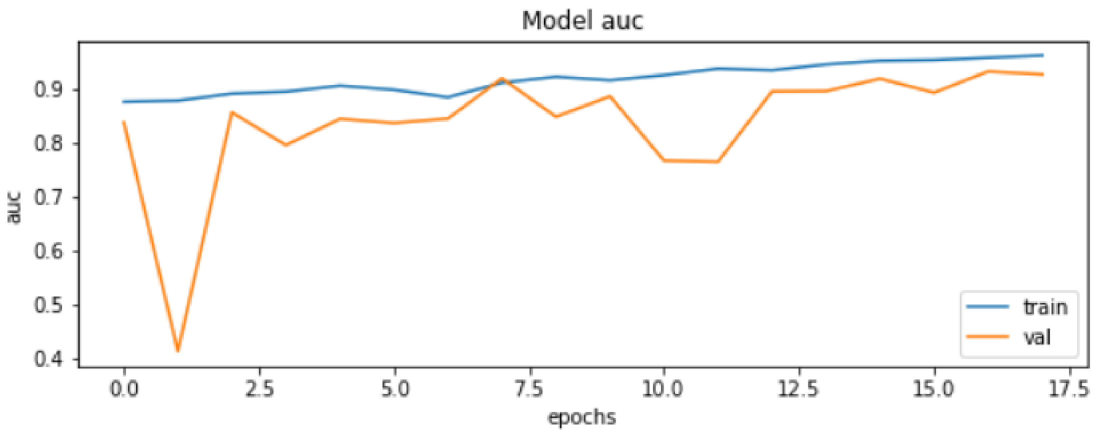


Figure 5.9: Custom Model Accuracy.

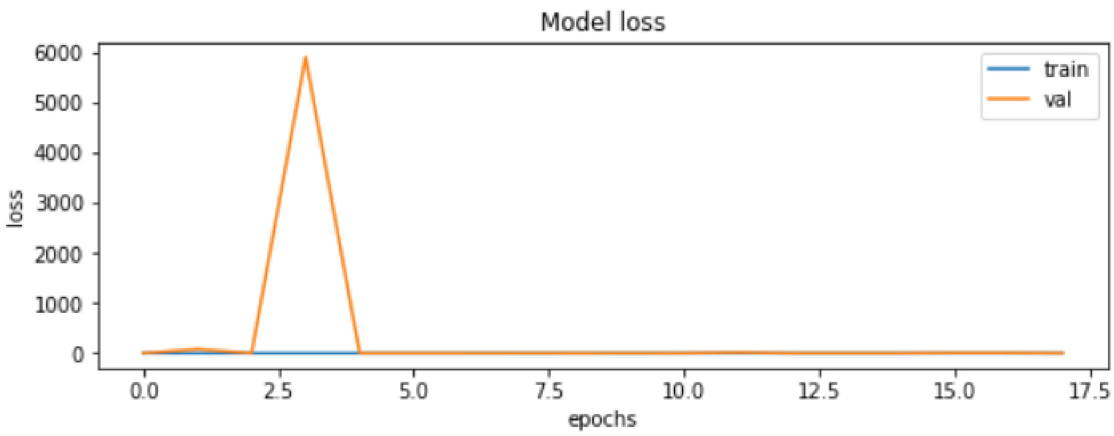


Figure 5.10: Custom Loss Function.

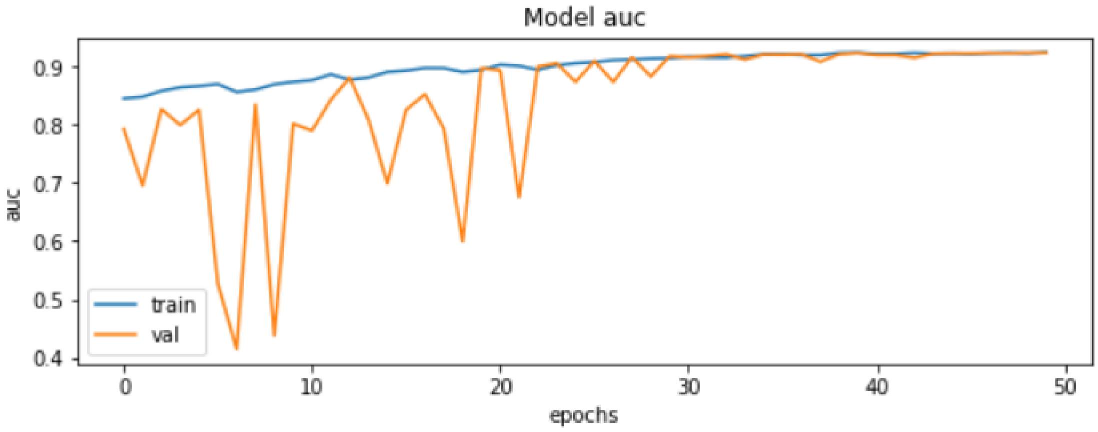


Figure 5.11: EfficientNetB5 Model Accuracy.

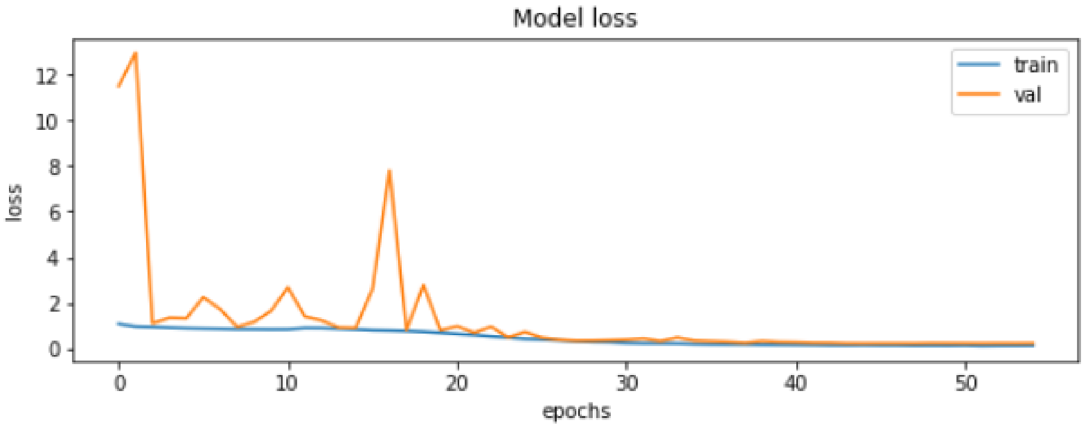


Figure 5.12: EfficientNetB5 Loss Function.

We even tried ensemble techniques on our data in order to further enhance the accuracy. Average ensembling was done and the test accuracies are as follows- 88% for VGG19 and ResNet152v2 and Inception-ResNet-v2; 92% for EfficientNetB5 and EfficientNetB6; 96% for all 6 models altogether.

Table 5.1: Ensemble Result of VGG19 and ResNet152v2 and Inception-ResNet-v2

Class	Precision	Recall	F1-Score
Non-demented	0.57	0.93	0.71
Very Mild	0.67	0.17	0.27
Mild	0.48	0.27	0.35
Moderate	0.60	0.25	0.35
Weighted Average	0.77	0.66	0.71

Table 5.2: Ensemble Result of EfficientNetB5 and EfficientNetB6

Class	Precision	Recall	F1-Score
Non-demented	0.53	0.73	0.61
Very Mild	0.67	0.47	0.55
Mild	0.49	0.28	0.36
Moderate	0.64	0.50	0.56
Weighted Average	0.90	0.86	0.88

Table 5.3: Ensemble Result of all 6 CNN Models

Class	Precision	Recall	F1-Score
Non-demented	0.43	0.63	0.51
Very Mild	0.67	0.57	0.62
Mild	0.79	0.48	0.60
Moderate	0.64	0.80	0.71
Weighted Average	0.92	0.90	0.91

5.2 Discussion

From our findings and results, we can see that we have achieved good results from most of the models. Since we have used so many different kinds of architectures along with ensembling, the outcomes we got were varied. That is also because different epoch sizes were used for different models and the architectures were all modified in dissimilar ways. Overfitting however, was successfully reduced through early stopping which is apparent from the train accuracies. The gap between the training and validation accuracies were very insignificant and that proves that performance was satisfactory even though the training accuracies could have been higher. If we compare between the training and the validation accuracies, some of the architectures have performed really well and their numbers are proof enough since they are very close to each other which is for majority of the cases except VGG19 and Custom CNN. The average ensembling technique also gave decent results on our data, with each model giving an accuracy 4% greater than the former as shown in Figure 5.15, and the combination of all 6 models achieved the best output.

Figure 5.13: Training and Validation Accuracies of the 6 individual models.

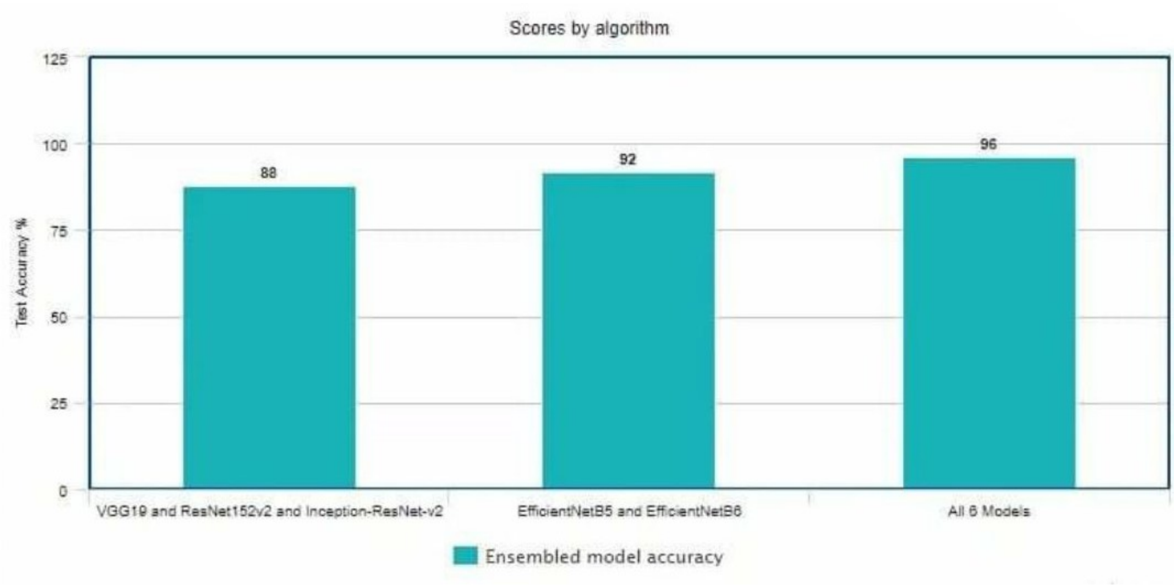


Figure 5.14: Test Accuracy of the ensembled models.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

Alzheimer's disease is among the major causes of deaths in industrialized countries and it is also increasing drastically in the developing ones. Early, effective and precise detection and diagnosis of AD is significant for initiation of proper treatment. In present years, deep learning (DL) has proved to perform significantly well with image data and given results with high accuracy, better than other general machine learning techniques. To enhance our work further, we have used newer and most recent architectures along with transfer learning techniques. Additionally, we have also ensembled different learning algorithms together to make our model even more efficient and accurate.

6.2 Challenges/Limitations

There were quite a few challenges that we faced during this period. Firstly, GPU or hardware was limited since all the training was done on Google Collab and thus this took much longer than anticipated. Secondly, if the dataset that we acquired was bigger than we would not have had to use transfer learning necessarily. After that, a lot of pre-processing, such as slicing of the 3D images, had to be done since the size of each data was very large and hence had to be used in smaller batches which was inefficient which would not have occurred if the formats of the MRI images were standard. Finally, AD is not very visible or apparent in an MRI image of a brain unlike ones with a tumor, for example. Therefore, it posed a serious challenge for image feature extraction.

6.3 Future Work

Currently we have worked with OASIS dataset. But, this data was limited and since our aim is to work with larger datasets in the future, we can move forward and not have to use pre-trained models. We can use more of custom algorithms and not have to necessarily ensemble them since they would give enhanced results on their own. Also, our goal for the future is to use quantum CNN algorithm which will speed up training, improve accuracy and make the performance of our model overall more efficient.

Bibliography

[1] *Facts and Figures.Alzheimer's Association, howpublished = https://www.alz.org/alzheimers-dementia/facts-figures,*

[2] Jyoti Islam and Yanqing Zhang. "An ensemble of deep convolutional neural networks for Alzheimer's disease detection and classification". In: *arXiv preprint arXiv:1712.01675* (2017).

[3] *BANGLADESH: ALZHEIMER'S AND DIMENTIA. WORLD HEALTH RANKINGS, howpublished = https://www.worldlifeexpectancy.com/bangladesh-alzheimers-dementia,*

[4] Cassandra A DeMarshall et al. "Detection of Alzheimer's disease at mild cognitive impairment and disease progression using autoantibodies as blood-based biomarkers". In: *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 3 (2016), pp. 51–62.

[5] Kilian Hett et al. "Multi-scale Graph-based Grading for Alzheimer's Disease Prediction". In: *arXiv preprint arXiv:1907.06625* (2019).

[6] U Rajendra Acharya et al. "Automated detection of Alzheimer's disease using brain MRI images–a study with various feature extraction techniques". In: *Journal of Medical Systems* 43.9 (2019), p. 302.

[7] Wei Li et al. "Detecting Alzheimer's disease on small dataset: A knowledge transfer perspective". In: *IEEE journal of biomedical and health informatics* 23.3 (2018), pp. 1234–1242.

[8] Lulu Yue et al. "Hierarchical feature extraction for early Alzheimer's disease diagnosis". In: *IEEE Access* 7 (2019), pp. 93752–93760.

[9] Naimul Mefraz Khan, Nabila Abraham, and Marcia Hon. "Transfer learning with intelligent training data selection for prediction of Alzheimer's disease". In: *IEEE Access* 7 (2019), pp. 72726–72735.

[10] Saman Sarraf and Ghassem Tofighi. "Classification of alzheimer's disease using fmri data and deep learning convolutional neural networks". In: *arXiv preprint arXiv:1603.08631* (2016).

[11] Stefan J Teipel et al. "Robust detection of impaired resting state functional connectivity networks in Alzheimer's disease using elastic net regularized regression". In: *Frontiers in aging neuroscience* 8 (2017), p. 318.

[12] Rubén Armañanzas et al. "Voxel-based diagnosis of Alzheimer's disease using classifier ensembles". In: *IEEE journal of biomedical and health informatics* 21.3 (2016), pp. 778–784.

[13] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. "Understanding of a convolutional neural network". In: *2017 International Conference on Engineering and Technology (ICET)*. IEEE. 2017, pp. 1–6.

[14] Mr Amir Ebrahimighahnavieh, Suhuai Luo, and Raymond Chiong. "Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review". In: *Computer Methods and Programs in Biomedicine* 187 (2020), p. 105242.

[15] Taeho Jo, Kwangsik Nho, and Andrew J Saykin. "Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data". In: *Frontiers in aging neuroscience* 11 (2019), p. 220.

[16] Keiron O'Shea and Ryan Nash. "An introduction to convolutional neural networks". In: *arXiv preprint arXiv:1511.08458* (2015).

[17] Phil Kim. "Convolutional neural network". In: *MATLAB deep learning*. Springer, 2017, pp. 121–147.

[18] Patrice Y Simard, David Steinkraus, John C Platt, et al. "Best practices for convolutional neural networks applied to visual document analysis." In: *Icdar*. Vol. 3. 2003. 2003.

[19] Patrik Kamencay et al. "A new method for face recognition using convolutional neural network". In: (2017).

[20] Chigozie Nwankpa et al. "Activation functions: Comparison of trends in practice and research for deep learning". In: *arXiv preprint arXiv:1811.03378* (2018).

[21] Michiel Straat. "On-line learning in neural networks with ReLU activations". PhD thesis. 2018.

[22] YA Bachtiar and T Adiono. "Convolutional Neural Network and Maxpooling Architecture on Zynq SoC FPGA". In: *2019 International Symposium on Electronics and Smart Devices (ISESD)*. IEEE. 2019, pp. 1–5.

[23] Jawad Nagi et al. "Max-pooling convolutional neural networks for vision-based hand gesture recognition". In: *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. IEEE. 2011, pp. 342–347.

[24] Arden Dertat. *Applied Deep Learning - Part 4: Convolutional Neural Networks.* URL: https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2. (2018, June 19).

[25] Nitish Srivastava et al. "Dropout: a simple way to prevent neural networks from overfitting". In: *The journal of machine learning research* 15.1 (2014), pp. 1929–1958.

[26] Sergey Ioffe and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift". In: *arXiv preprint arXiv:1502.03167* (2015).

[27] Nils Bjorck et al. "Understanding batch normalization". In: *Advances in Neural Information Processing Systems*. 2018, pp. 7694–7705.

[28] Lisa Torrey and Jude Shavlik. "Transfer learning". In: *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI global, 2010, pp. 242–264.

[29] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. "A survey of transfer learning". In: *Journal of Big data* 3.1 (2016), p. 9.

[30] Rohan Saha. "Transfer Learning–A Comparative Analysis". In: (2018).

[31] Stephen L Thaler. *Neural network-based target seeking system*. US Patent 6,115,701. 2000.

[32] Kun He, Yan Wang, and John Hopcroft. "A powerful generative model using random weights for the deep image representation". In: *Advances in Neural Information Processing Systems* 29 (2016), pp. 631–639.

[33] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556* (2014).

[34] Yufeng Zheng, Clifford Yang, and Alex Merkulov. "Breast cancer screening using convolutional neural network and follow-up digital mammography". In: *Computational Imaging III*. Vol. 10669. International Society for Optics and Photonics. 2018, p. 1066905.

[35] Riaz Ullah Khan et al. "Evaluating the performance of resnet model based on image recognition". In: *Proceedings of the 2018 International Conference on Computing and Artificial Intelligence*. 2018, pp. 86–90.

[36] Sasha Targ, Diogo Almeida, and Kevin Lyman. "Resnet in resnet: Generalizing residual architectures". In: *arXiv preprint arXiv:1603.08029* (2016).

[37] Long D Nguyen et al. "Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation". In: *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE. 2018, pp. 1–5.

[38] Christian Szegedy et al. "Inception-v4, inception-resnet and the impact of residual connections on learning". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 31. 1. 2017.

[39] Christian et al. "Inception-v4, inception-resnet and the impact of residual connections on learning". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 31. 1. 2017.

[40] Mingxing Tan and Quoc V Le. "Efficientnet: Rethinking model scaling for convolutional neural networks". In: *arXiv preprint arXiv:1905.11946* (2019).

[41] Long D Nguyen et al. "Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation". In: *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE. 2018, pp. 1–5.

[42] Vardan Agarwal. *Complete Architectural Details of all EfficientNet Models*. 2020. URL: https://towardsdatascience.com/complete-architectural-details-of-all-efficientnet-models-5fd5b736142.

[43] Thomas G Dietterich. "Ensemble methods in machine learning". In: *International workshop on multiple classifier systems*. Springer. 2000, pp. 1–15.

[44] Jiao Wang and Si-wei Luo. "Exploiting ensemble method in semi-supervised learning". In: *2006 International Conference on Machine Learning and Cybernetics*. IEEE. 2006, pp. 1104–1107.

[45] Shui-Hua Wang et al. "Alzheimer's disease detection by pseudo Zernike moment and linear regression classification". In: *CNS & Neurological Disorders-Drug Targets (Formerly Current Drug Targets-CNS & Neurological Disorders)* 16.1 (2017), pp. 11–15.

[46] Jyoti Islam and Yanqing Zhang. "A novel deep learning based multi-class classification method for Alzheimer's disease detection using brain MRI data". In: *International Conference on Brain Informatics*. Springer. 2017, pp. 213–222.

[47] Debesh Jha et al. "Alzheimer's disease detection using extreme learning machine, complex dual tree wavelet principal coefficients and linear discriminant analysis". In: *Journal of Medical Imaging and Health Informatics* 8.5 (2018), pp. 881–890.

[48] Weifeng Chen et al. "OASIS: A Large-Scale Dataset for Single Image 3D in the Wild". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 679–688.

[49] Bijen Khagi et al. "CNN Models Performance Analysis on MRI images of OASIS dataset for distinction between Healthy and Alzheimer's patient". In: *2019 International Conference on Electronics, Information, and Communication (ICEIC)*. IEEE. 2019, pp. 1–4.

[50] Daniel S Marcus et al. "Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults". In: *Journal of cognitive neuroscience* 19.9 (2007), pp. 1498–1507.

[51] Zalán Rajna et al. "Speeding up the file access of large compressed nifti neuroimaging data". In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2015, pp. 654–657.

[52] Robert Cox et al. "A (Sort of) new image data format standard: NIfTI-1: WE 150". In: *Neuroimage* 22 (2004).

[53] Marcus. *CROSS-SECTIONAL DATA ACROSS THE ADULT LIFESPAN*. https://www.oasis-brains.org/files/oasis_cross-sectional_facts.pdf.