Detecting Different Stages of Alzheimer's Disease from MRI images using Deep Learning and Computer Vision Techniques

by

Fardin Rahman
20101072

Sadman Sharif
20101107

Syed Shams Islam
20301200

Nihad Adnan Shah Tirtho
20101611

Md. Ashir Intheshar
20101041

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 and Engineering

Department of Computer Science and Engineering
School of Data and Sciences
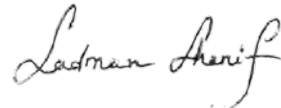Brac University
January 2024

Declaration

It is hereby declared that

1. The thesis submitted is 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 and Signature:

Fardin Rahman	Sadman Sharif
20101072	20101107

Syed Shams Islam	Nihad Adnan Shah Tirtho
20301200	20101611

Md. Ashir Intheshar
20101041

Approval

The thesis titled "Detecting Different Stages of Alzheimer's Disease from MRI images using Deep Learning and Computer Vision Techniques" submitted by

1. Fardin Rahman (20101072)

2. Sadman Sharif (20101107)

3. Syed Shams Islam (20301200)

4. Nihad Adnan Shah Tirtho (20101611)

5. Md. Ashir Intheshar (20101041)

Of Fall, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on January 18, 2024.

Examining Committee:

Supervisor:
(Member)

_____

Dewan Ziaul Karim
Lecturer
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

_____

Dr. Md. Golam Rabiul Alam
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

_____

Dr. Sadia Hamid Kazi
Chairperson
Department of Computer Science and Engineering
Brac University

Abstract

The preliminary and precise diagnosis of Alzheimer's Disease is significant for the speedy management and intervention of the disorder. Numerous valuable tools such as Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET) etc. are used for evaluating the function and structure of brain which could help diagnose Alzheimer's Disease. However, simplifying the MRI images manually is a hectic and long drawn process that is prone to observer variability. The prime prospect of this study is to employ a Computer Vision and Deep learning Based framework that would automatically classify the stage Alzheimer Disease (AD) through the MRI images. A large amount of dataset is used to enhance the effectiveness of the suggested structure. Moreover, this research demonstrates the capability of Computer vision and Deep learning in assisting premature AD detection. It provides a beneficial insight into the enrichment of neurological disease diagnosis using computer-aided technology. The highlight of this study is the introduction of a custom model that outperforms all state-of-the-art Convolutional Neural Network (CNN) models in performance. This novel model has achieved an exceptional accuracy of 96.6%, which showcases a meaningful advancement in the field and also provides a promising direction for future research in neurodegenerative disease diagnosis.

Keywords: Machine Learning; Alzheimer's Disease; Dementia; Magnetic Resonance Imaging; Convolutional Neural Network; Grey Matter; White Matter; Neuroimaging; Computer Vision; Positron Emission Tomography

Acknowledgement

Firstly, we want to address our sincere appreciation to Allah for providing us with the ability to accomplish this research paper effectively. We would also love to convey our heartfelt appreciation to our supervisor Dewan Ziaul Karim for supporting us throughout the whole period of this research. We were able to overcome all the obstacles on our way only for his significant support. Moreover, we would like to appreciate the whole judging panel for giving us the opportunity to show our work and publish it. They provided us with feedback helped us to further improve our work. Lastly, we wish to extend our heartfelt gratitude and profound love to our beloved parents, whose unwavering support and guidance have been the cornerstone of our journey.

Table of Contents

List of Figures

List of Tables

Nomenclature

The following list discusses many symbols and abbreviations that will be used later in the document's body.

AD Alzheimer's Disease

$ADNI$ Alzheimer's Disease Neuroimaging Initiative

ANN Artificial Neural Network

CN Cognitive Normal

CNN Convolutional Neural Network

CSF Cerebrospinal Fluid

DNN Deep Neural Network

DTI Diffusion Tensor Imaging

$EMCI$ Early Mild Cognitive Impair

$FDG - PET$ Fluorodeoxyglucose Positron Emission Tomography

GAN Generative Adversarial Neural Network

GLM General Linear Model

HC Healthy Control

$LMCI$ Late Mild Cognitive Impair

MCI Mild Cognitive Impairment

MRI Magnetic Resonance Imaging

$OASIS$ Open Access Series of Imaging Studies

SVM Support Vector Machine

VAE Variational Auto Encoder

ViT Visual Transformer

WHO World Health Organization

Chapter 1

Introduction

1.1   Background

Alzheimer's Disease, a sort of severe cachexia that slowly mutilates the brain and impacts a large number of individuals globally. It is a prevalent reason for cognitive decline in elderly individuals and is diagnosed through symptoms like loss of memory, cognitive impairment, and conversion in behavior. Prompt identification of AD is essential for enhancing patient results and creating successful therapies[1]. MRI is a convenient tool that detects and monitors the development of Alzheimer [2], [3].

DL and CV methods have demonstrated potential in identifying various phases of AD through MRI images. Deep Learning is a part of ML that trains Artificial Neural Networks (ANN) to recognise the patterns and features from extensive datasets. Computer Vision (CV) is a discipline that concentrates on teaching computers to comprehend and interpret visual information. It is possible to create models for detecting several AD stages through MRI scans by using a combination of these two techniques.

Studies have shown that DL and CV techniques are to categorize the different AD stages from MRI scans [4]. As an instance, some models that are based on deep learning frameworks have been taught to identify Early and Late mild cognitive impairment (EMCI, LMCI) and AD by analyzing MRI scans with great precision. Methods from CV have been applied to study brain structures and identify small alterations in brain areas linked to various phases of AD.

Nonetheless, there are still obstacles linked to the utilization of DL and CV methods for identifying various phases of AD from MRI pictures. A significant obstacle is the requirement for extensive and varied datasets that consist of individuals with and without AD at various stages. Another difficulty is the requirement for models that can be easily understood and clarify the characteristics and trends utilized to identify various phases of AD.

The data which was published by the World Health Organization (WHO) illustrates that about 50,000,000 individuals throughout the whole world are suffering from Dementia. Besides, it is highly expected by the experts that the number would triple by the year 2050 [5]. And among all the cases, 60-70% of instances of dementia

overall have a direct or indirect connection to AD [6]. As a result, this is the most familiar form of Dementia. The United States has around 6.2 million individuals who are the victims of this severe Disease. According to the experts there is a high chance that this number would rise to 13.8 million within the next three decades. Elderly individuals above age 65 are more prone to this disease [6]. Therefore, this results in an increased number of disabilities among the elderly population. Moreover, it may lead to a negative impact on the economic sector as it costs over \$1 trillion dollars worldwide every year because of AD . In addition to that the US government has spent over \$ 355 billion dollars in the fiscal year 2021-22 to provide necessary treatments to the individuals with AD and other sorts of dementia. This amount is more likely to rise up to \$ 1.1 trillion dollars within the upcoming decades [6].

To sum up, there is potential for DL and CV methods to identify various phases of AD through MRI images. More investigation is required to address the difficulties related to these methods and create dependable models for timely identification and monitoring of AD. Models like these could have important implications for enhancing patient outcomes and creating successful treatments for AD.

1.2   Problem Statement

AD is a type of escalating neurodegenerative disability that has given rise to a consequential challenge in the health sector throughout the whole world. Among all the cases related to dementia AD is notably the prevalent kind. The characterization and detection of the stages of AD precisely at an early stage is pivotal for an effective monitorization and progression of this disease. Besides, effective treatment planning and accessing the efficacy of possible intercession is also necessary. For the time being, the diagnosis of the stages of AD largely depend on the medical history of the patients, clinical assessments and rational tests. But these are pretty much subjective and inclined to inter-observer mutability. Therefore, if advanced neuroimaging techniques like Positron Emission Tomography, also known as "PET", MRI etc. are integrated along with the conventional deep learning and computer vision models then there is promising probability of improvement. Some recent work shows that Liu et al. [7] applied CNN to distinguish stages of AD with a success rate of 93.7% on a dataset with over 500 subject matters. In addition to that Tofighi and Sarraf [8] employed a deep learning method that distinguishes the Mild Cognitive Impairment (MCI) and Healthy Control (HC) stages of AD with an accuracy of 86%. Despite all these progress there is much more room for research in this field. Although the work conducted on the existing research mainly focuses on the binary classification, a vigorous algorithm detecting the stages of AD accurately remains a challenging task till now. Therefore the prime goal of this research is to fill in these gaps by creating an ideal Computer Vision and DL methods that is efficient enough to distinguish the AD stages by using the MRI images.This study seeks to obtain a much higher accuracy through leveraging the potential of modern algorithms on a large dataset. Subsequently, This would potentially enhance the capability of early detection of AD stages that can benefit the AD patients through personalized treatment policy.

1.3 Motivation

AD gradually damages memory and cognitive facility, and progressively worsens as time goes on. Since this is a global health disease, it can affect a large number of people yet there is hope that with improved detection and timely treatment, it can lead to significant improvements in the health of the affected. MRI has emerged as a crucial tool for comprehending the composition and operation of the brain. MRI also promises the early identification and precise diagnosis of AD. Deep learning, specifically computer vision techniques, has shown immense potential for analyzing and interpreting MRI scan image data in the medical area. These techniques can be used to provide more precise and effective methods for identifying various AD stages through MRI scanned pictures. Which ultimately leads to patient identification of episodes leading to more personalized treatment plans and improved disease outcomes.Our study intends to advance a robust inefficient model to accurately delineate the different stages of AD from MRI scanned images and make important contributions to medical imaging and disease care.

Chapter 2

Problem and Objectives

2.1 Research Problem

In 1906 a bunch of cases were reported with extreme transformation in brain cell structures. Those brain cells had some major changes compared to a normal cell. It had way too many plaques and $\frac{1}{3}$rd of the cerebral cortex were non functional. In 1910 Kraepelin was first to depict this condition as 'Alzheimer Disease' in his book named "Clinical Psychiatry". Afterwards in 1998 researchers from Max Planck Institute of Neurobiology discovered that specific brain segments are damaged due to amyloid plaques and neurofibrillary cramps [9]. This has been confirmed as the first officially reported case of AD. Till now this research work is being used to diagnose patients with AD. It is currently listed in the top 5 fatalities in the US. Moreover research shows that it may surpass cancer and coronary disease as the principal reason for death in the US. Therefore there is an inevitable need and necessity of predicting AD at its premature stage and preventing it from progressing any further. Diagnosing AD requires a lot of tests, visualization and data analysis. But these are very hectic tasks to do. Advanced image processing techniques like MRI and Positron Emission Tomography (PET) are utilized to diagnose the structural change in Brain Cells.
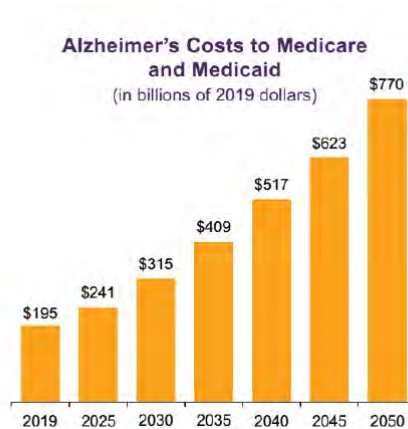
Figure 2.1: Alzheimer's Disease medication cost until year 2050 [10]

4

According to [11], these Neuroimaging techniques are employed to congenitally picture the layout or the pharmacology of brains. These techniques can be further illustrated into 2 sub-branches: functional and structural imaging . Functional imaging mainly provides us with the activities conducted by the brain whereas structural imaging gives us information related to neurons, glial cells, synapses etc. of the brain.

The Magnetic Resonance Image (MRI) employed in article [12], [13] uses radio wave and magnetic fields to develop high-quality 3D images. Structural MRI is most widely used for AD cases because it can measure brain volume and detect the loss of neurons and tissues [13]. Here figure 2a and 2b illustrates the structural and functional MRI respectively.
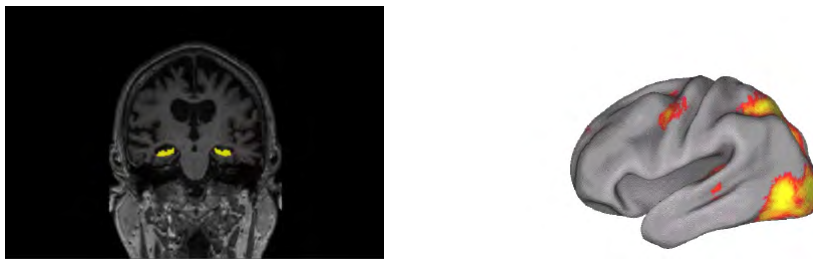
Figure 2.2: (a) Structural MRI [14] (b) Functional MRI [15]

And lastly, Positron Emission Tomography (PET) discussed in paper [16], [17] uses radioactive tracers in order to analyze the radioactive spheres. This technique makes use of tracers like fluorodeoxyglucose and amyloids to diagnose AD. In this process specific things like thinking ability, listening, remembering potential etc are considered to come to a conclusion regarding the diagnosis of AD.
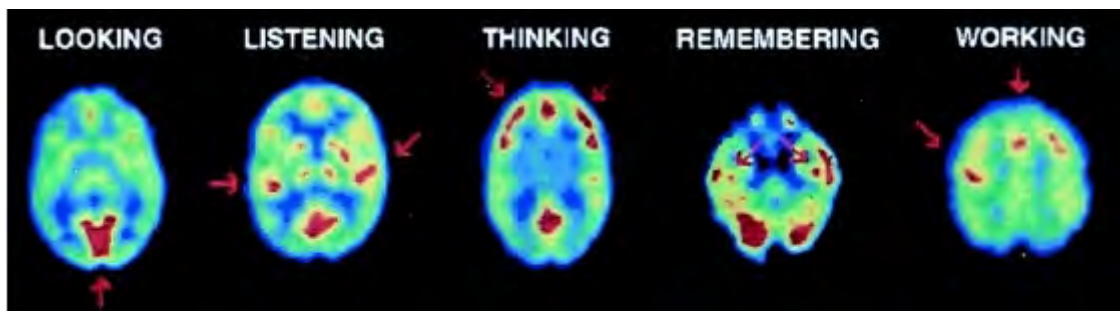
Figure 2.3: PET scan in HC (Healthy Control) [18]

So, till now we have covered some popular and most commonly used diagnosing techniques of AD. But the prime goal of this initiative is to help the neurologists classify or differentiate AD stages using neurological images. According to [19], enormous neurological images have been divided into four binary classes like AD

Cognitive Normal (CN), Healthy Control (HC), Early and Late Mild Cognitive Impair (EMCI, LMCI). Discrete ML methods have been applied including some popular methods like Support Vector Machine (SVM), Random Forest Tree, General Linear Model (GLM) etc. to categorize different AD stages.

Convolutional Neural Network (CNN) is the most popularly utilized DL method that gives favorable outcomes in more or less every medical image analysis. It has accomplished this because of its classification and image analysis capability. As a result, it has sought the attention of researchers all over the world. Although it has shown some excellent diagnosis results, there is still much room for optimization. For Example, the paper [20] employed a CNN framework which extracts features utilizing MRI scans and that has been followed by a SVM and Random Forest Classifier. Additionally, in [21] a deep convolutional feature learning process has been implemented on both supervised and unsupervised data for distinguishing the AD stages. Moreover, in accordance with [22] some extra work has been done before providing it to the CNN. These Extra works include the augmentation of training data through scaling, deforming cropping and rotation of the MRI scans. However, a recent study [23] proposed 3D CNN architecture that is established on ResNet and VGGNet and the main finding from this study was that manual feature extractions are not required at all.

After keeping all this information in mind, the question may arise whether DL and computer vision methods can differentiate AD stages. The answer to this question would be yes because a huge number of researches has been conducted in this field. In [24], [25], several Convolutional Neural Network and ML techniques have been implemented to recognize AD stages. But in most of the cases Mild Cognitive Impairment (MCI) has not been sub-branched. Minimum Importance has been paid to the two sub-branches of MCI which are Early and Late Mild Cognitive Impair (EMCI, LMCI).

As a result, this research is trying to answer the following question:

How productive and constructive are these deep learning methods in detecting and classifying the Alzheimer's Disease phases?

Deep learning has been a proven efficacious tool in aiding informed medical decisions making. According to [26] a computer can make accurate assessments as good as a radiologist. Through many studies where various ML algorithms and frameworks were evaluated, it was concluded that traditional ML approaches are not well equipped to approach complicated problems like AD [27] as successful brain image classification requires discrimination among similar brain patterns. According to [28], [29] 2D natural image classification using DL has benefited studies in medical image analysis domains. Among all DL techniques, CNN has exhibited nominal achievement in organ segmentation and disease detection [30].

2.2 Research Objective

As projected in many studies, the number of AD patients are increasing every year and this trend will continue. Neuroimaging modalities, especially MR images, have already demonstrated its ability in capturing complex brain patterns. As a result, they have been sought after and utilized in ingenious ways in developing automatic systems for diagnosing certain diseases. PET and DTI neuroimaging modalities are relatively new. PET and DTI scans are very promising in disease detection yet they are not as widely available as MR scans. MCI is considered as the most significant phase in AD progression. The literature has provided further division of MCI stage in an attempt to better quantify the likelihood of MCI to dementia conversion. We would like to propose and implement a 2D CNN model which would differentiate in-between the early and late MCI in addition to CN and AD. Moreover, due to patient confidentiality, medical datasets are not widely available which significantly constraints the development of pretrained models for various medical imaging modalities. In our work, we would like to partially address this issue by employing unsupervised pretraining on unlabeled data. The effectiveness of this technique is still not understood completely till today but in practice and empirical results suggests they provide better regularization behavior.Therefore, our study would contribute to the literature in the following way:

- Understand and implement the necessary preprocessing required for MRI images before training a deep model.

- Explore and utilize the options available to address class imbalance in classification tasks.

- Validate the effectiveness of state-of-the-art CNN models in medical image classification.

- Propose a simple yet expressive model capable of making accurate predictions ensuring portability in resource constrained settings.

- Compare performance metrics between state-of-the-art models and proposed custom models.

- Understand and leverage the effectiveness of unsupervised pre-training as a means to address scarcity of labeled data and class imbalance.

- Employing Explainable AI(XI) tools such as GradCam(Gradient-weighted Class Activation Mapping) in understanding model predictions by highlighting important regions in the MR scans that goes into the decision making.

Chapter 3

Literature Review

3.1   Alzheimer's Disease

A set of conditions is referred to as "dementia" which consists of interrelated neurological disability that are primarily distinguished by amnesia and cognitive decline. Contrary to a very common misconception attributing dementia as a definitive consequence of normal aging for its pervasiveness in older adults, healthy aging does not accompany dementia. According to estimates, AD is the reason for 60 to 80 % of all the cases of dementia. It is an irrecoverable growing neurodegenerative disorder identified by localized brain atrophy and deteriorates the neuron's essential biochemical processes, which leads to cognitive impairment. Memory impairment, apraxia, agnosia, damage to visual skills, general dementia, and behavioral and personality abnormalities are the clinical hallmarks of AD [31]. [32], [33] claim that at least 20 years pass before AD-related brain alterations manifest as symptoms. AD is generally divided into three main stages namely CN, MCI and AD. The earliest observable sign of dementia is said to be MCI. As the intermediary phase in-between CN and AD, MCI is called the preambulatory phase of AD. According to the findings in [34], 30–40% of older persons 65 and older have MCI, and of those, 5–6 years later, 30–40% will acquire AD. Typically this conversion time is 18 months but it can range from 6 months to 36 months. Patients with MCI are then further divided into MCI converters and MCI non converters, depending on whether they have recovered. Other MCI stage subtypes, such as early and late MCI, are also present but are infrequently discussed in the literature. According to a standard test, cognitive impairment that that ranges from 1 to 1.5 SD below the standardizing mean is the objective definition of EMCI [35]. According to the report from ADNI [36], 17.5% of LMCI (Late Mild Cognitive Impairment) and 2.3% of EMCI is the annual MCI to dementia conversion rate. Amyloid deposition and changes in brain metabolism [35] are characteristic features of EMCI, along with breakdowns in functional networks [37] and alterations in brain volume [38]. However, there is ongoing discussion regarding the long-term outcomes associated with these changes. EMCI demonstrated increased diversity in its features and an increased probability of exhibiting negative measures associated with the diagnostics of AD [39]. A rapid fall and linear alteration in semantic memory, episodic memory and perceptual speed have been identified as characteristics of LMCI in terms of cognitive decline [40].More crucially, LMCI associated with poor episodic memory is linked to a higher risk of AD and a quicker loss in cognitive function[41].

At the moment, AD diagnosis is carried out by clinical examination which includes a comprehensive interview with the patients as well as their relatives [42], [43] but a conclusive AD diagnosis for obtaining 'ground truth' evidence obtained through autopsy with the aid of tissue pathology of neurofibrillary entangle and beta-amyloid plaques. The beta-amyloid plaques develop in and around the neuron and the neurofibrillary tangles are the abnormal accumulation of the tau protein. Needless to say a benchmark is called for to confirm AD and efforts have been made to this end by the National Institute and the Alzheimer's Association in 2011 by defining relevant biomarkers which includes unusual PET and MRI neuroimaging and abnormal cerebrospinal fluid amyloid and tau biomarkers [44]. It is revised every three to four years to reflect any new information affecting the pathophysiology and course of the disease.[45].

3.2 Neuroimaging Modalities and Tests

MRI neuroimaging can be divided into structural MRI and functional MRI. PET neuroimaging includes fluorodeoxyglucose PET (FDG-PET), amyloid PET. Another non-invasive neuroimaging modality used for AD diagnosis is Diffusion Tensor Imaging (DTI). Although MRI, PET, DTI neuroimaging modalities are effective strategies for diagnosing AD, their availability is limited. The other two most frequently used and widely adopted biomarkers are Mini Mental State Examination(MMSE) [46] and Clinical Dementia Rating(CDR) tests [47]. Clinical accuracies of these tests fare in the range of 70-90% as opposed to the post-mortem diagnosis. Consequently, by no means these tests are considered to be ground truth evidence for AD diagnosis nevertheless they offer significant value considering the limited biomarkers available.

3.3 Deep Models for AD Detection

Artificial Neural Network(ANN) is a subdivision of ML that tries to mimic the functionality of neurons and aims to model the working mechanism of the human brain[48]. Although the human brain is a very complicated structure involving various complex biological processes and interdependent systems, ANN has demonstrated remarkable learning abilities from data. Artificial neurons are the building blocks of ANNs which are connected with one another sparsely or densely based on the ANN architecture which helps the network to adapt to new information in its environment [49]. The artificial neurons share information associated with some weight with each other through the network. A learning algorithm is then applied to the network to learn the weights with the view to minimizing a loss function[50]. Once the network is fitted to a model through training it can generalize to new information and perform various ML tasks. Deep Neural Networks (DNN) are a sort of ANN that has both an input and an output layer with numerous hidden layers in between. Every single layer is made up of a number of artificial neurons. The hidden layer is responsible chiefly for learning a mapping between the input and the output among other objectives. DNNs perform effectively at learning information in complex data such as MR images [51]. Deep learning models are a class of ANNs which tries to simulate the human brain response to learn data representation

through multiple layers of computation. These models are characterized by their depth which is the amount of layers present in the grid. Each layer performs linear and non-linear mathematical operations to uncover the hierarchical structure and representation of the data.

Popular Deep Learning Models With Strengths and Limitations

Models		Strengths	Limitations
AE		• Easier implementation • Works well for dimension reduction • Interpretable data captured by the encoder • Can capture and represent highly non-linear and complex data • Good starting points for CNNs	• Learns as much features as possible but not as much relevant features
RBM		• Known to learn good generative models • Robust to missing data	• Computationally very expensive during training
DNN		• Good at learning complex mapping between inputs and outputs • Can work with large datasets • Works well for vector-based problems	• Poor generalizability • Suboptimal for images • Slow training
DPN		• Efficient at learning feature representation from small samples	• Performance is restricted because of the plain concatenation of learned hierarchical characteristics from distinct levels.
RNN		• Good at sequential data • Suitable for ongoing research like longitudinal studies	• Prevalence of vanishing or exploding gradients meaning weight updates are difficult to learn and unstable
CNN	2D CNN	• Easier and faster to train • Effective local feature extraction in images	• Cannot account for information present in third dimension
	3D CNN	• Effective local characteristics extraction in scans. • Can learn dense 3D information present in 3D brain volume	• Computationally very expensive to train

Table 3.1: Popular Deep Learning Models and their Strengths[52]

3.4   Class Imbalance and Data Augmentation

One of the key obstacle in training a DL model or any machine learning model in that regard is class imbalance. Class imbalance is a scenario where the data distribution of classes or labels in a dataset is significantly skewed. This simply means that one class has a much larger number of samples than the other classes or class. Class imbalance is a recurring theme in real world datasets especially in disease diagnosis and rare event prediction. In order for a model to generalize well to new data and overcome overfitting and underfitting, the training data needs to be representative. The consequence of class imbalances is biased models, poor generalization and misleading evaluation metrics. Even to ensure dataset balance by removing some instances of the oversampled classes resulting in a smaller dataset, the minimized dataset has shown to perform better [53]. Over the years, many methods have been developed to express the problem of class imbalance. Synthetic Minority Oversampling Technique (SMOTE) randomly duplicates the underrepresented class of images to minimize the overfitting problem. This technique has shown success for [54] where their model performance on the validation set increased from 78% to 94%. Image reconstruction and synthesis is another way to address class imbalance which has become quite popular over the years. Generative models like GAN(Generative Adversarial Neural Network) have become very popular in this regard and the literature has extensively made use of this technique to generate augmented samples which closely resembles the data distribution of the original undersampled instances. [55] used GANs to reconstruct neuroimages. The augmented balanced dataset helped them achieve an accuracy of 74%. VAEs are the second most utilized generative model. Although VAEs are usually not used much for image augmentation owing to their proneness to generate fuzzy images, few researches have concluded dependent VAE or hybrid architectures has the potential to increase the standard of the specimen. VAE-GAN has shown to improve on both weaknesses of VAEs and GANs to produce enhanced quality images [56]. VAEs can also work with small datasets because of the availability of an encoder [57]. One of the most promising aspects of VAEs is their capacity to offer disentangled, interpretable and editable latent space[58]. This enables the inspection of the encoder representation of the image that allows to a better understandability and altering with the base data. Improved variants of VAE such as VQ-VAE2, IAF-VAE, Hamiltonian VAE have shown promising results in other domains so they are worth exploring in the medical domain. While GANs are tested in a lot of diverse configurations for medical image augmentation, other new techniques have emerged over the years especially diffusion models. Although DMs are at their infancy, they have already demonstrated their capabilities at producing high quality images[59]. Promising variants of DMs like Denoising Diffusion Implicit Model (DDIM) [60] and Fast Diffusion Probabilistic Model (FastDPM) [61] are worth exploring.

3.5 Related Work

According to [62], it is advantageous to use deep models over typical shallow models for early AD stage detection and this can be done by multi-modality data which consists of visual data, transmitted data, and clinical information. It highlights the prospects of three-dimensional CNN for analyzing MRI image data in Alzheimer's disease analysis and for each single data modality [69] uses the one-vs-one coding SVM, K-Nearest Neighbor algorithm, and random forests algorithm.Considering the outcomes of the suggested model, [62] concludes by signifying the CNN architecture for MRI image data and mentions using combinations, such as decision-level and feature-level as basic models, thus strengthening the reliability of the findings.

The research work [63] suggests a multi-model CNN classification method for AD detection using MR brain images, particularly three-dimensional ones, and multi-scale three-dimensional Computer-Aided Engineering (3D CAEs). [63] also shows the preprocessing of the MR images using the N3 approach. The research first converts MR pictures into compressed advanced features using a deep three-dimensional CNN before creating multiscale 3D CAEs to get features from the MR brain images. [63] has released the results of testing the suggested model, which show that it performs well in terms of classification, achieving an precision of 88.31% for detection of Alzheimer's and an 92.73% of ROC. The study's advantage was that there was no need for segmentation during the transaction of MR images, which reduced the cost of computation.

The research [64], focuses on the use of CV, specifically the techniques of DL during the early stages of AD detection found from MRI scans. The study utilizes the Efficient-Net CNN and U-Net CNN to identify AD in its early stages with accuracy. Early detection is important for effective treatment as it allows patients to become aware of preventive action before irreversible damage of the brain takes place [64]. Previous research has utilized computers for AD diagnosis, but many ways happen to be limited because of inherent data. While early stages of AD can be identified, it cannot be predicted if the disease has not revealed itself [64]. However, DL has become a preferred approach for identifying AD in its prior stages. The paper provides an outline of key research on AD and explores how the techniques of DL can help us in early detection of the disease [64].

The research [65], aims to achieve early detection of AD by developing a novel method which helps to detect the nature of MCIc from MRI scans. MCI is a phase that gradually advances to AD, and its early detection is important for timely action. The proposed method combines a local feature descriptor that utilizes the local binary pattern texture operator and the fast Hessian detector to identify Important aspects and explanations [65]. A basic CNN is employed for categorizing. The outcomes of the research demonstrate promising results for early AD detection. The accuracy for classifying among MCIc and cognitively normal, also known as CN, individuals is achieved at 88.46%, while the accuracy among CN and AD is obtained to be 88.99% [65]. This outcome shows the potential of the suggested approach as a substantial advancement in the early detection of AD, providing a valuable opportunity for individuals to prepare and manage the disease.

The objective of [66] is to create an automated system, which is capable of detecting and classifying AD at an initial stage using brain MRI. The implementation of this approach aims to accurately detect individuals affected by dementia and proficiently classify the advancing phases of the state, encompassing Mild Cognitive Impairment,also known as MCI, it is caused by AD, moderate AD, and severe AD. The proposed system makes use of transfer learning and improving of a convolutional network that has already been trained, specifically Alex-Net [66]. By leveraging the learned features from a large dataset, the system can effectively extract relevant information from MRI images to aid in classification. Both segmented (Grey Matter, White Matter, and CSF) and unsegmented scanned images are used for multi-class and binary categorizing. To evaluate the system's [66] execution, the dataset of OASIS is employed. The method demonstrates optimistics results, achieving total accuracy for multi-class categorization of unsegmented images which was 92.85%. [66]. The accuracy obtained indicates the system's potential for accurate identification and classification of different stages of AD. By providing an efficient and accurate automated detection and classification system, this research contributes to the early diagnosis and intervention of AD, enabling timely medical interventions and improving patient outcomes.

The ultimate goal of [67] is to develop a complete structure for AD detection and classification of medical images. [67] utilizes CNNs, the study concentrates on classifying the four stages of AD and binary image classifications between each stage pair. Two approaches are implemented: the first involves simple CNN architectures applied to 2 dimensional (2D) and 3 dimensional (3D) scans of the brain, while the second utilizes transfer learning with pre-trained models like VGG19. Evaluation based on nine performance metrics reveals promising outcomes. The outcome of this research shows that the CNN frameworks used in the first approach have advantageous properties, such as basic structures that decrease the complexity of computation, overfitting, memory demands and . These architectures achieve high accuracies for 2 dimensional (2D) and 3 dimensional (3D) multi-class stage classifications of 93.61% and 95.17%. of AD. If optimized, the VGG19 pre-trained model achieves a 97% accuracy for detecting multi-class stage classifications of AD [67].

This paper [68] explores the segmentation of AD using DL models and brain MRI. Manual methods for AD classification are time-consuming and not always accurate.Deep learning models, which are modeled after the human brain, are capable of processing data and coming to difficult choices. Using images of MR from the ADNI dataset, the study evaluates various deep learning models. The model of DenseNet-121 achieves a success rate of 88.78%, but it is computationally slow. To improve execution time, the authors propose depthwise CNN layers in place of the CNN layers in DenseNet-121, resulting in an improved performance rate of 90.22%. The paper gives a summary of 20 popular deep learning models, evaluates their performance for AD classification, and discusses the modifications made to DenseNet-121.

A deep learning technique using structural MRI of the brain, AD and MCI could be detected early is presented in the publication [69] titled "Automated Detection of AD and MCI Using Whole Brain MRI". To extract characteristics from brain

MRI data, it combines standard convolution, pointwise convolution and depthwise convolution. The model utilizes a hierarchical transformation of the structural MRI images, generating compact high-level features while reducing computational complexity. Compared to the latest AD classification techniques, the suggested model demonstrates superior performance in terms of accuracy, achieving an impressive 96% accuracy in classifying brain images into CN, AD and MCI groups. The paper emphasizes the contributions of the proposed model, including reduced computational complexity and its potential for practical AD detection.

In addition to that, research has been conducted on mobility data collected by smartphones. According to [70] data of 35 AD patients has been gathered by smartphones from a daycare center. These data listed the accelerometer changes of each patient when they used to perform their daily activities. And depending on these accelerometer changes of the daily activities of the patients they were categorized into mild, medium and late stages of AD. This methodology tracks this time series and executes this using the CNN model to figure out the pattern that classifies AD stages. This model achieved a promising accuracy rate of 90.91 % and F1 score of 89.7 %. This showed much improvement than the output achieved using traditional feature based classifiers. Therefore, mobility data is a great asset that can be used to treat AD patients. This research showed a discrete approach as it did not use the conventional supervised learning models.

Although CNN has shown some strong capability to determine hierarchical structural features, it still has some limitations as it is hard to employ it starightly to functional MRI images. As a result a novel CNN model is proposed in [71] to learn the submerged feature from Brain Functional Networks (BFN). As BFN are constructed with both dynamic and static functional connectivity the fMRI's are decomposed into several static Brain functional Networks. Then to calibrate the BFN dynamics voxel-wise variances in dynamics functional connectivity were used. After that a pair of 3D images that depict both static and dynamic BFN are supplied to a 3D CNN model where the model can hierarchically learn those static and dynamic features. In this way different Brain Neural Networks form a sort of network and help each other out in order to make a better classification. This process played an inevitable role in AD stage diagnosis and showed an improvement of almost 10% than the previous conventional methods.

The study [72] examines regional changes in brain morphology and recommends a deformation-based ML strategy for identifying AD and forecasting the switch from MCI to AD using high-resolution MRI. In the MRI-based categorization, it is difficult to discriminate between progressing MCI and stable MCI but it is really crucial for anticipating conversion in MCI participants. In [72], the researchers used the distortion of the regions of emotion processing, known as the amygdala, and the regions of episodic memory, known as the hippocampus to detect the progression of MCI. Once mild or moderate AD was determined to be caused by extensive morphological shifts throughout the gray matter of the entire brain, these participants were further categorized using a linear SVM. The study also published the results of their suggested technique, demonstrating great performance with 96.5% accuracy in differentiating Alzheimer's patients from aging healthy people, 91.74% accuracy

in differentiating continuous MCI from aging healthy people, and 88.99% accuracy in differentiating continuous MCI from stable MCI.

The study [73] proposes an innovative and effective technique for detecting AD using MRI data of the brain. It includes an extensive CNN that draws inspiration from the Inception-V4 network and uses the OASIS database to demonstrate its potential. The proposed model in this study classifies three main stages of AD and the framework yields outstanding findings. The study [73] displays its efficiency in identifying diseases and categorizing them across various body regions, in addition, it is also efficient in segmenting organs and substructures. The model performs better than conventional techniques, with an accuracy of 73.75% on the OASIS dataset. Although the number is not huge, it has a different significance in comparison to the conventional techniques that were used to classify all the MRI data. The model also relies on hyperparameters from a deep image classifier. And its advantage is, it eliminates the requirement for manual hand-crafting of features.

Through biomarkers and neuroimaging, the thesis paper [74] provides a comprehensive examination of research focusing on the diagnosis of AD early. It emphasizes the significance of early detection and the growing demand for image analysis. The paper delves into various AD datasets, that includes widely used AD's Neuroimaging Initiative (ADNI) dataset with 509 cases, the Harvard Medical School dataset with 613 T2-weighted brain MRI scans, the Max Planck Institute Leipzig Mind-Brain-Body Dataset-LEMON with 228 subjects, the National Health and Aging Trends Study facilitated by the National Alzheimer's Coordinating Center, and the OASIS with 416 cases. These datasets provide valuable resources for understanding AD and developing diagnostic techniques. The study aims to contribute to early-stage neuroimaging research and enhance the detection of AD.

Chapter 4

Methodology and Work Plan

4.1 Methodology

The first and foremost job for this research was to collect diverse MRI images. After collection of Data we followed the below methodologies and work plan:

1. **Data Preprocessing and Preparation:** We made sure that the dataset had all the necessary stages of AD. After that we imputed the data into our Machine Learning Algorithm. But before that we had to process the data. For data preprocessing we used the following methods.

   a) Intensity Normalization.

   b) Realignment.

   c) Skull Stripping.

   d) Denoising.

   e) Resizing.

2. **Data Labeling and Training:** After that we labeled the data into different stages of AD. The stages are mentioned below.

   a) MCI

   b) CN

   c) EMCI

   d) LMCI

   Now, we divide the data into training (80%) and validation (20%) sets in order to assess the models capability and accuracy.

3. **Data Augmentation:** In this step both training and validation set of data are augmented using OpenCV library. Finishing the augmentation process the datasets are again splitted back to augmented training and validation sets.

4. **N-layered 2D CNN Model:** After getting a more generalized dataset from the previous step we implicated the N-layered CNN framework on our data. This framework will incorporate the following layers.

   a) Convolution

   b) Pooling

   c) Normalization

   d) Flatten

   e) Dropout

   f) Dense

   g) Activation

5. **Prediction and Analysis:** Passing through the N-layered CNN framework our model gets ready for predicting the stages of AD. This model would be able to predict if a brain is healthy or it falls under any stages of AD.

6. **Performance Assessment and Comparative Analysis:** Standard evaluation criteria including Precision, Accuracy, Recall, and F1-score are taken into consideration while assessing performance. Following that, these findings are contrasted with those obtained using currently used techniques.

4.2   Work Plan

Figure 4.1: Work Plan

4.3 Input Data Management

When working with different neuroimaging modalities, various input data management is taken into consideration. They have their own strengths and weaknesses which are studied intensively throughout the literature. A comparative analysis for Input data management for different neuroimaging modalities have been summarized [52]. Our first challenge of this research was to augment the dataset to avoid class imbalance problem.
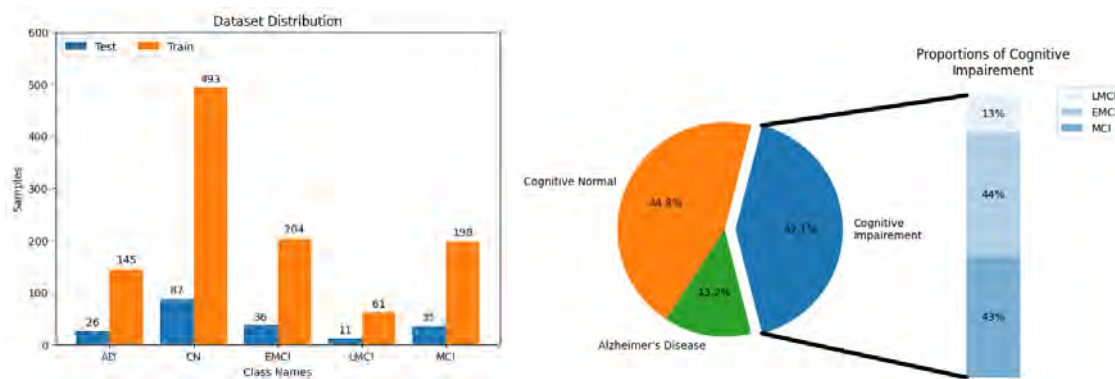

Figure 4.2: Class Imbalance Before Oversampling

After oversampling and augmentation we get total 1437, 2009, 1680, 1008, 1610 images for Alzheimer's Disaease, Cognitive normal, Early mild cognitive impairment, Late mild cognitive impairment and Mild cognitive impairment respectively. This step helped us to overcome the challenges of overfitting due to the limited amount of data.
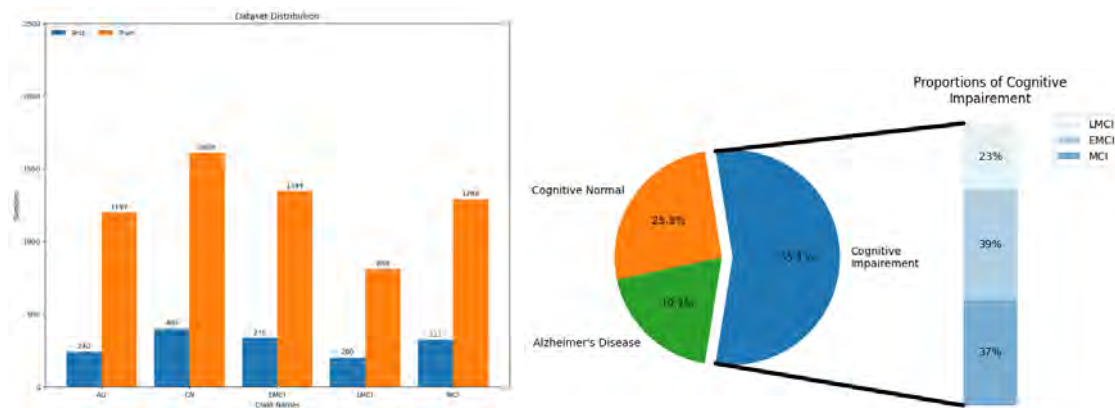
Figure 4.3: Balanced Dataset After Oversampling

To this day, a complete cure for AD still awaits discovery. Given that an AD patient typically lives 3 to 9 years[75], it is imperative that this disease is detected at its early stage so as to effective clinical intervention can be made possible, disease progression can be alleviated[76] and the severe effects of dementia can be delayed. As the onset of AD is marked by MCI, classification of CN or MCI from AD is generally not as important as predicting MCI conversion.

4.4 Input Data Preprocessing

Although preprocessing is an important component of machine learning, the architecture of deep learning models has rendered many preprocessing steps less critical [77] ,[78]. Nonetheless, preprocessing methods such as intensity normalization, motion correction, skull stripping and realignment are still used in the majority of studies. Normalization ensures similar structures have similar intensity.[79] For this purpose, the N3 non-parametric non-uniform intensity normalization algorithm is most usually utilized. [80] The spatial alignment of image scans to reference anatomical space is known as realignment. Skull stripping is the removal of bone structures of the skull. It may include neck as well. And motion correction is the suppression of motion artifacts from neuroimages.
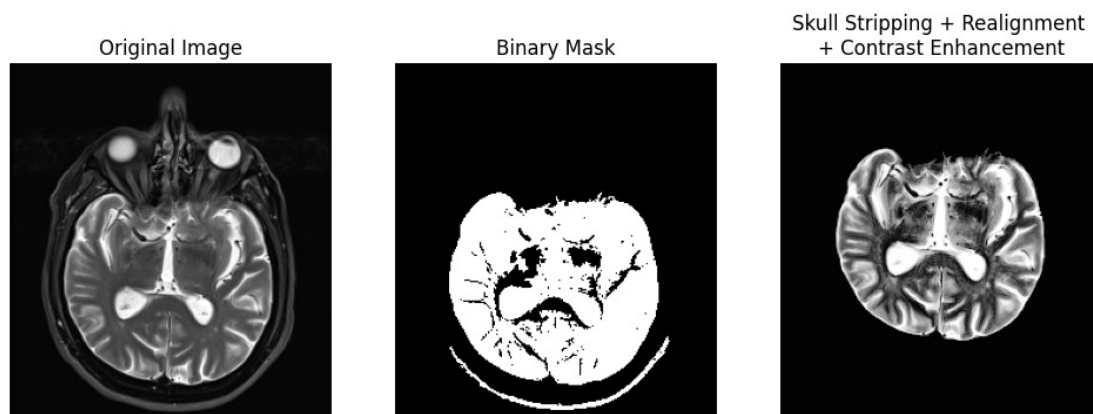
Figure 4.4: Skull Stripping and Realignment

During most of our model training we underwent a series of preprocessing steps to ensure compatibility with the chosen deep learning models. Initially we resize the image using bilinear interpolation. Image dimensions varied from model to model. Subsequently, we normalized and centered the pixel values using function from the respective architectures. It ensured the pixel value to range from 0 to 1, promoting stable convergence while training. Additionally, shuffling of the data was done to eliminate any sort of inherent order, and split was used to divide it into testing and training sets. And finally, we employed one-hot encoding to convert class labels into categorical format which is suitable for training deep Neural Networks.

Chapter 5

CNN Model Implementation and Results

5.1 Efficient Net B1

5.1.1 Model Implementation

EfficientNet originated from a baseline network which was created through neural architecture search using AutoML MNAS. It employs a mobile inverted bottleneck convolution akin to MobileNet V2 but on a larger scale, primarily by increasing FLOPS.

Moreover, it uses a technique called 'compound coefficient' for scaling. Instead of randomly adjusting depth, width and resolution, this model uniformly scales all dimensions using fixed coefficients. Through this approach and AutoML,it creates seven models of varying size that outperforms other state of art Models [81].
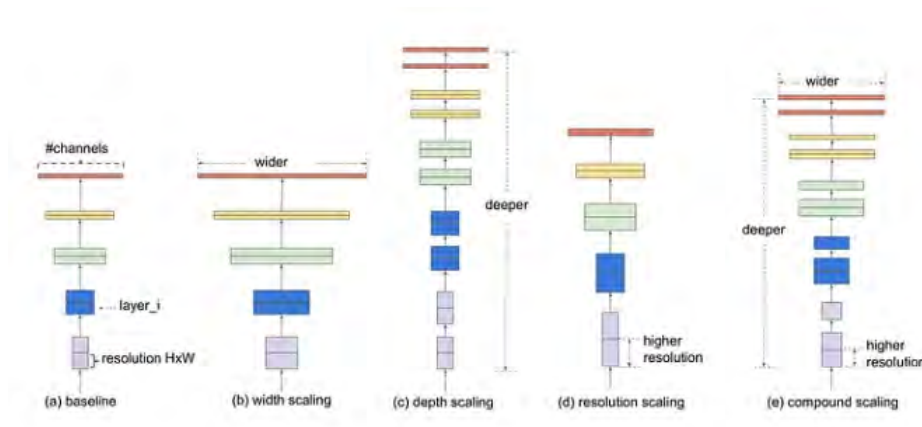

Figure 5.1: Figure: Compound Scaling [81]

This above mentioned model is based on balancing the dimensions by scaling with a constant ratio. The mathematical equations are mentioned below.

$$\text{Depth } d = \alpha^\phi, \text{Width } w = \beta^\phi, \text{Resolution } r = \gamma^\phi, (1)$$

$$\text{such that } \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$$

$$\alpha \geq 1, \beta \geq 1, \gamma \geq 1$$

The values of α, β, and γ are determined through a grid search algorithm, while Φ represents the increase in computational resources. This parameter is customizable but constrained by the condition that $\alpha \cdot \beta^2 \cdot \gamma^2$ is close to 2. The computational cost in CNNs is primarily influenced by convolutional operations. Scaling the network according to the provided equation results in a FLOPS increase of $(\alpha \cdot \beta^2 \cdot \gamma^2)^\Phi$. With $\alpha \cdot \beta^2 \cdot \gamma^2$ approximately equal to 2, the total FLOPS rises by 2^Φ, for any new Φ. The EfficientNet-B0 model is obtained with $\alpha = 1.2$, $\beta = 1.1$, and $\gamma = 1.15$, maintaining $\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$. Subsequently, models EfficientNet B1 to B7 are derived by scaling the baseline network with fixed values of α, β, and γ. While better performance could potentially be achieved with a more extensive search, it would come at a higher cost. To mitigate this, the search is conducted from scratch using a smaller network [81].

5.1.2 Results
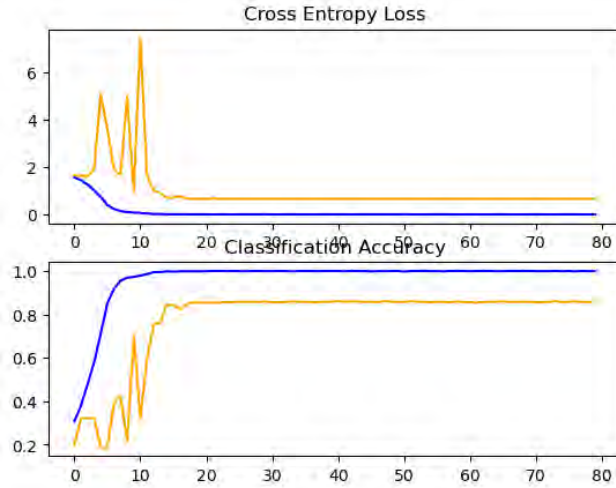
Figure 5.2: Accuracy and loss

The Efficientnet-B1 architecture attained a notable classification accuracy of 85%. The accuracy chart above showcases the model's learning journey, highlighting its ability to make accurate predictions. Simultaneously, the cross-entropy loss chart illustrates the model's optimization process, marked by a declining trend that signifies effective training. This result indicates that Efficientnet-B1 effectively extracted essential features from the MRI scans, resulting in precise classification.

5.2   Inception V3

5.2.1   Model Implementation

The Inception V3 stands as a DL model rooted in CNN, primarily employed for classifying images. It represents an enhanced iteration of the foundational Inception V1 model, originally introduced as GoogLeNet by a team at Google in 2014. In the realm of Inception V1, the issue of overfitting arose when multiple deep convolutional layers were incorporated into the model. To mitigate this concern, Inception V1 adopted a novel approach: instead of deepening the model, it broadened it by employing multiple different filters of varying sizes at the same level. Consequently, Inception models moved away from layers which are deep in favor of parallel layers [82].

Inception V3, serves as an advanced variant of Inception V1, implementing multiple techniques for enhancing network performance:

- Enhanced efficiency.

- Deeper network architecture, yet maintaining speed.

- Reduced computational complexity.

- Inclusion of auxiliary classifiers for regularization.

Inception V3 made its debut in 2015, boasting 42 layers. Let us visualize the key optimizations that elevate Inception V3's performance:

1. **Factorization into Smaller Convolutions:** Inception V1 already reduced dimensions generously, but Inception V3 further improved efficiency by breaking down larger convolutions into smaller ones.For instance, 3×3 layer replaced the costly 5×5 convolutional layer, reducing computational demands.This factorization resulted in a 28% reduction in parameters and computational costs.

2. **Spatial Factorization into Asymmetric Convolutions:** Instead of further reducing to a 2×2 convolution, Inception V3 opted for asymmetric convolutions, represented as n×1 or 1×n. By replacing 3×3 convolutions with 1×3 followed by 3×1 convolution. It achieved the same receptive field as a 3×3 convolution while being 33% more cost-effective for equivalent input and output filter numbers.

3. **Utility of Auxiliary Classifiers:** Auxiliary classifiers were introduced for enhancing the convergence of DNN and combat the vanishing the problems that are gradient in very deep networks. While these classifiers did not yield early-stage improvements, they significantly boosted accuracy toward the training's conclusion, effectively acting as regularizers within the Inception V3 model.

4. **Grid Size Reduction:** Departing from traditional max pooling and average pooling techniques, Inception V3 efficiently reduced feature map grid sizes by enlarging the activation dimensions of filters of networks. This involved blocks of parallel convolution and pooling, followed by concatenation. For example, a d×d grid having k filters would, after reducing, result in a $\frac{d}{2} \times \frac{d}{2}$ grid with 2k filters.

After completing all optimizations, the final Inception V3 model looks like this:
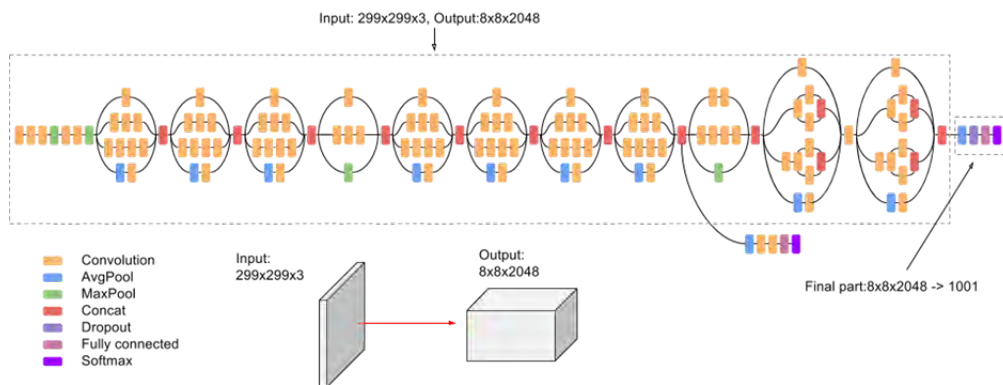

Figure 5.3: Inception V3 Model [82]

In its entirety, the Inception V3 model comprises 42 layers, exceeding the layer count of its previous generations, Inception V1 and V2. Nevertheless, the power of this model stands as truly remarkable. We will visualize its efficiency shortly. But first, let us meticulously examine the constituent elements that constitute the Inception V3 model.
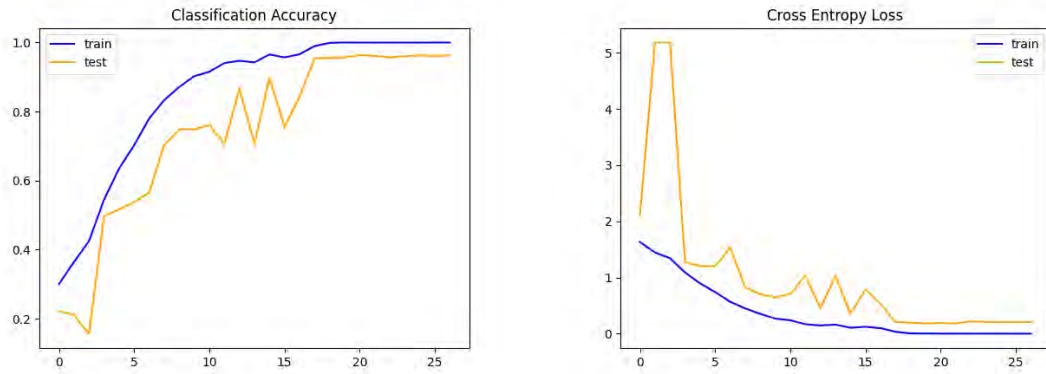
5.2.2   Results

Figure 5.4: (a) Classification Accuracy   (b) Cross Entropy Loss

The Inception-V3 architecture achieved an impressive classification accuracy of 95%. The accuracy graph above illustrates the model's learning progress, emphasizing its capability to make precise predictions. Concurrently, the cross-entropy loss graph depicts the model's optimization process, characterized by a decreasing trend that signifies effective training. This outcome indicates that Inception V3 successfully extracted crucial features from the MRI scans, leading to accurate classification.

	precision	recall	f1-score	support
AD	0.963265	0.967213	0.965235	244.000000
CN	0.936061	0.908189	0.921914	403.000000
EMCI	0.944118	0.958209	0.951111	335.000000
LMCI	0.970874	1.000000	0.985222	200.000000
MCI	0.946708	0.946708	0.946708	319.000000
accuracy	0.949367	0.949367	0.949367	0.949367
macro avg	0.952205	0.956064	0.954038	1501.000000
weighted avg	0.949183	0.949367	0.949177	1501.000000
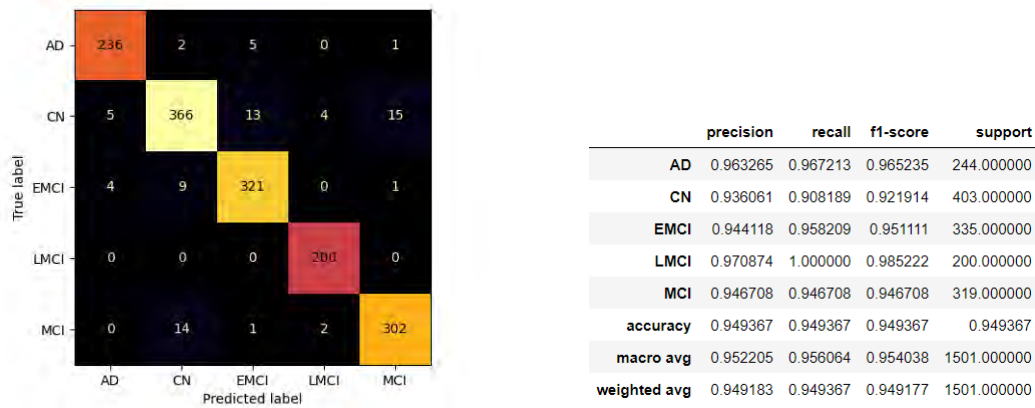
Figure 5.5: (a) Confusion Matrix   (b) Metrics

The figures provided above display the confusion matrix along with important metrics, including Precision, F1-score and Recall. The F1-score creates a balance between precision and recall, by giving a comprehensive measure of accuracy that takes into account false predictions. Precision reflects the positive predictions accuracy, whereas recall evaluates the model's capacity to obtain every positive instances. These metrics together assess the model's performance, emphasizing its accuracy and its ability to make positive predictions.

5.3 Resnet

5.3.1 Model Implementation

ResNet-50 is a groundbreaking convolutional neural network distinguished by its remarkable depth of 50 layers. It falls under the ResNet category, an acronym for Residual Networks, which has become a staple in various computer vision applications. The pivotal achievement of ResNet was its capacity to facilitate the training of exceptionally DNN, exceeding 150 layers [83].
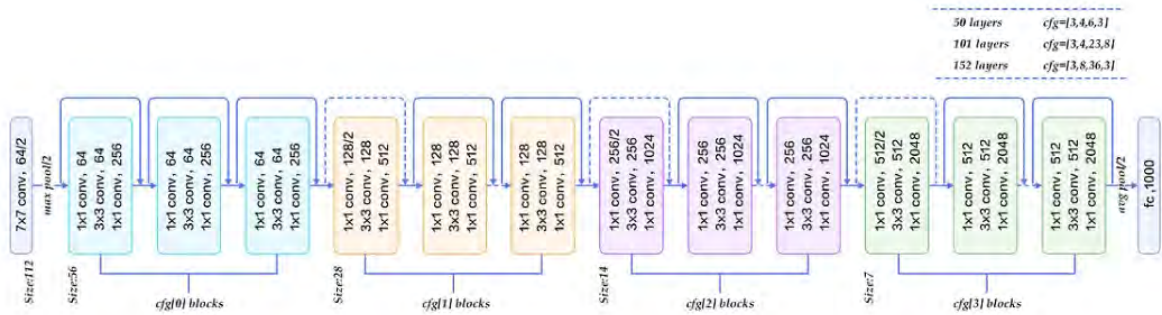
Figure 5.6: Resnet-50 Architecture [83]

One of the central predicaments in CNNs is the 'Vanishing Gradient Problem.' During the phase of backpropagation, gradients diminish significantly, resulting in minimal weight adjustments. To surmount this issue, ResNet employs a technique known as 'SKIP CONNECTION'.
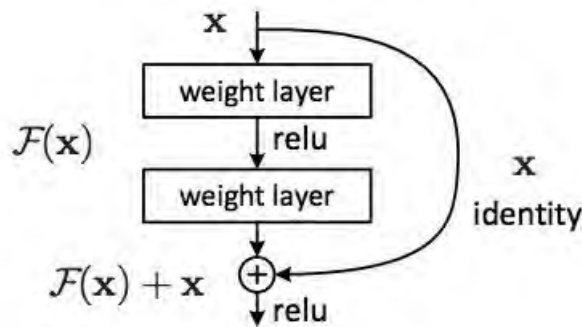
A 'SKIP CONNECTION' is a direct link that circumvents certain layers within the model. This skip connection introduces a deviation in the output. Without skip connection, input 'X' encounters multiplication of weight followed by the addition of bias term, passing through the activation function 'F', ultimately yielding 'F(X).' However, with the skip connection method, the output becomes 'F(X) + X.

It has two primary types of building blocks:

1. **Identity Block:** Here the output layer is added to value of 'X' only if input size matches the size of output. If they do not match, a 'convolutional block' is incorporated into the shortest path to equalize the input and output sizes.

2. **Convolutional Block:** To achieve size compatibility between the input and output, a convolutional block is employed in cases where the input size differs from the output size. This alignment is accomplished through two methods: padding the input volume or applying 1x1 convolutions.
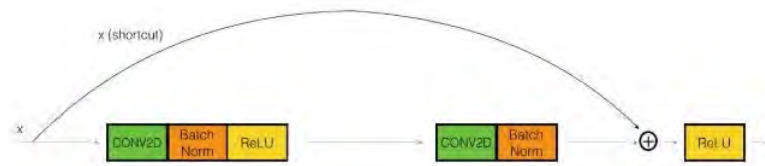
Figure 5.7: Identity Block [83]
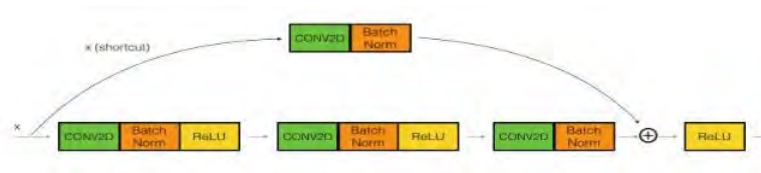
Figure 5.8: Convolutional block [83]

In CNNs, pooling is employed to reduce the image size, typically with a stride value of 2.

In summary, ResNet-50 is a pioneering neural network architecture that addresses vanishing gradient problem through skip connections by enabling training of deep networks. It employs identity and convolutional blocks to handle input and output size discrepancies, making it a powerful tool in the realm of computer vision.

5.3.2 Results
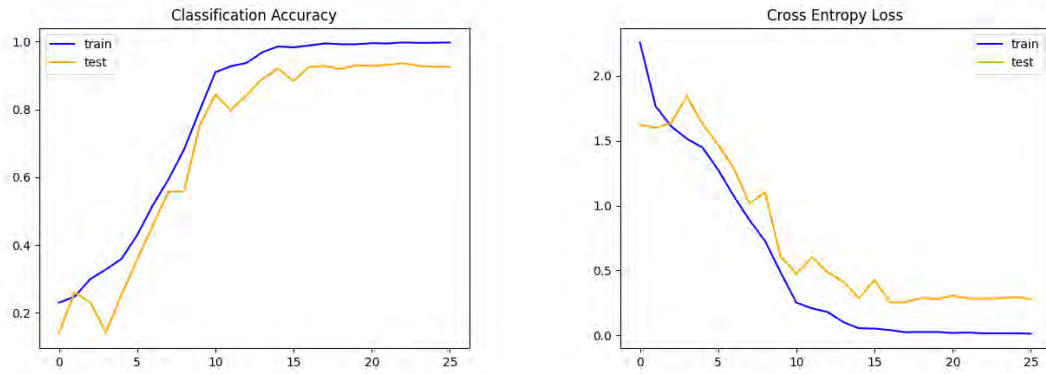
Figure 5.9: (a) Classification Accuracy   (b) Cross Entropy Loss

The ResNet-50 architecture achieved an impressive classification accuracy of 92%. The accuracy graph provided above illustrates the model's learning progress, emphasizing its capacity for precise predictions. At the same time, the cross-entropy loss graph reflects the model's optimization process, showing a consistent decrease, indicating successful training. These results indicate that ResNet-50 effectively captured essential features from the MRI scans, leading to accurate classification.
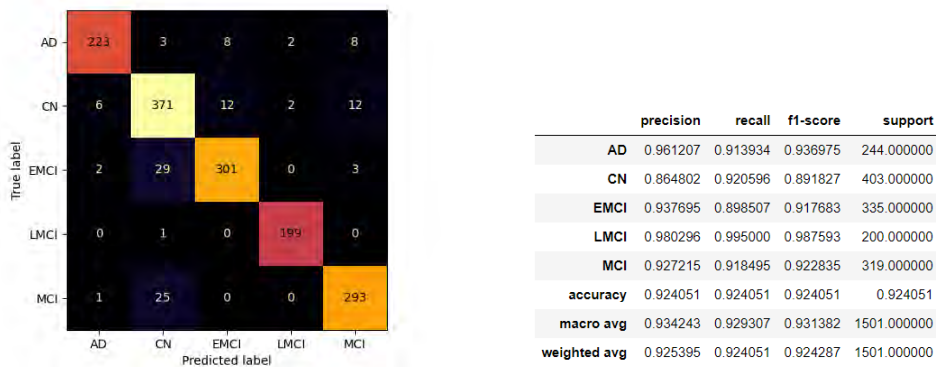
Figure 5.10: (a) Confusion Matrix   (b) Metrics

The figures mentioned above display the confusion matrix along with essential metrics: Precision, F1-score and recall. F1-score provides balanced assessment of precision and recall, offering an overall accuracy that considers false predictions. Precision reflects the positive prediction accuracy, whereas recall evaluates the model's capacity to identify every positive instances. These metrics together assess the model's efficiency, emphasizing its accuracy and its ability to make positive predictions.

5.4 VGG 16 and VGG 19

5.4.1 Model Implementation

In order to handle large-scale pictures, the VGG model investigates the depth of layers using a relatively relatively small (3 x 3) convolutional filter size. A collection of VGG models with layer lengths ranging from 11 to 19 includes the VGG16 model [84].

VGG16 is characterized by its deep architecture, which consists of:

1. **Convolutional Layers:** Each of the 13 convolutional layers in VGG16's architecture uses a tiny 3x3 convolutional filter. After each layer, padding size one is applied for preserving feature maps' spatial dimensions.

2. **Max-Pooling Layers:** Spatial resolution of feature mapping is decreased by using max-pooling layer with 2x2 kernel and a stride value of 2 after each block of convolutional neural layers.

3. **Fully Connected Layers:** First two hidden layers of 3 fully interconnected layers of VGG16 each comprise 4096 neurons. The output layer, which corresponds to the categories in the ImageNet dataset, is composed of 1000 neurons.
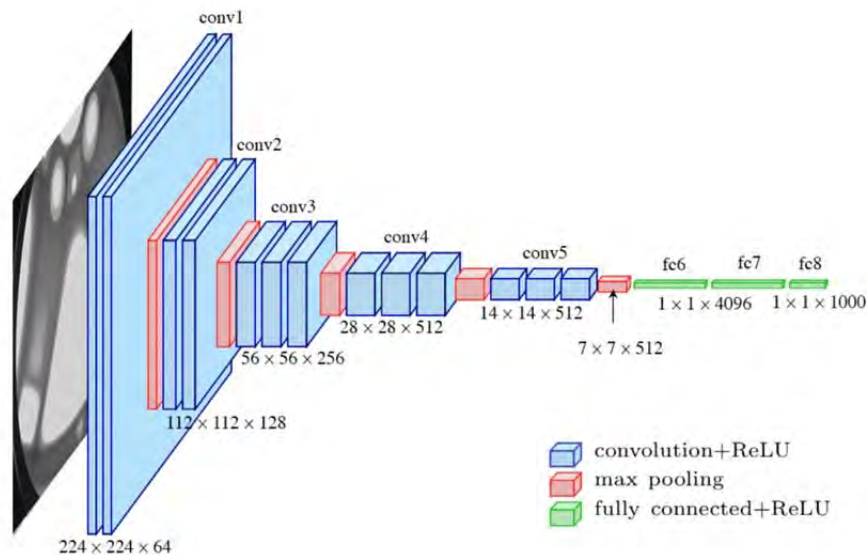
Figure 5.11: VGG16 Architecture [84]

More than one layer makes up the VGG architecture. Then it processes input images with sizes (224, 224, and 3). The process begins with two convolutional layers, each of which uses the same padding and has 64 filters having size of (3,3). And max-pooling with a stride of (2,2) after each convolutional layer. A max-pooling layer with a stride of (2, 2) is followed by two further layers with 128 filters having size of (3, 3). And it is also followed by two sets of three convolutional layers, each set of which uses 512 filters having size of (3, 3), with same number of padding. The network produces a (7, 7, 512) feature map through these convolutional and max-pooling , which has been flattened into a (1, 25088) features of vector. The following 3 connected layers

29

receive this vector. The initial 2 fully connected layers each produce an output vector of (1, 4096), whereas the third fully interconnected layer creates an output vector of (1, 1000) that can be classified using the softmax function into 1000 classes. The ReLU activation function is used in all hidden layers to speed up learning and prevent vanishing gradient issues. The architecture
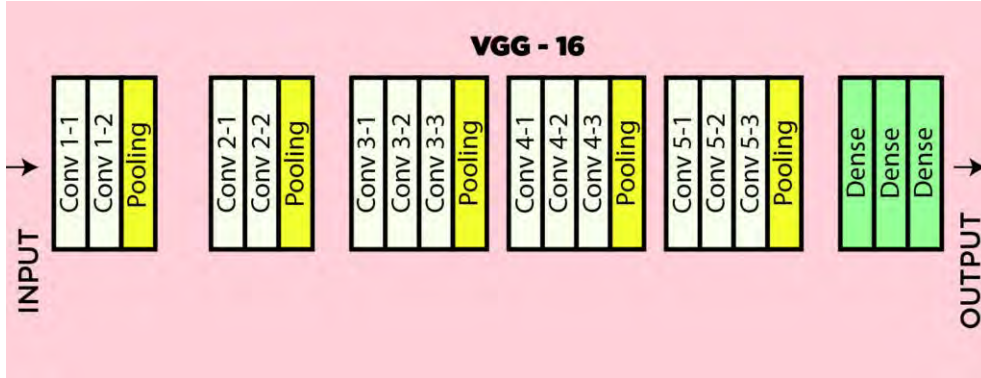
Figure 5.12: Architecture Map [84]

of the model, the operations in each layer, and the quantity of parameters in each layer must all be specified in order to implement the VGG16 model mathematically. The VGG16 model produced outstanding outcomes in image classification tests. Convolutional, max-pooling, and fully connected layers make up its architecture, which is renowned for its richness and capacity to capture complex information.

5.4.2 Results
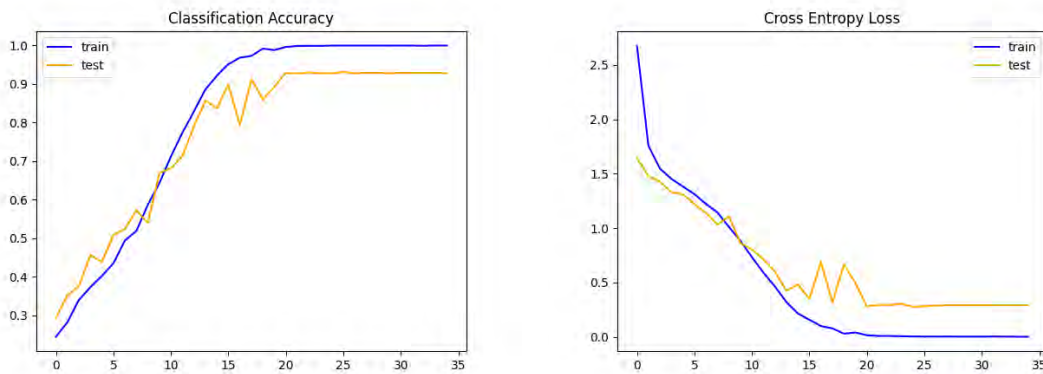
Figure 5.13: (a) Classification Accuracy   (b) Cross Entropy Loss

VGG-16 architecture gave us a commendable classification accuracy of 92%. The above accuracy graph shows the model's learning progression, highlighting its ability to make accurate predictions. Simultaneously, the cross entropy loss graph demonstrates the models optimization process with a decreasing trend indicating effective training. This output signifies that Alexnet extracted vital features from the MRI scans, resulting in accurate classification.

	precision	recall	f1-score	support
AD	0.958506	0.946721	0.952577	244.000000
CN	0.877990	0.910670	0.894032	403.000000
EMCI	0.923780	0.904478	0.914027	335.000000
LMCI	0.970149	0.975000	0.972569	200.000000
MCI	0.932907	0.915361	0.924051	319.000000
accuracy	0.924717	0.924717	0.924717	0.924717
macro avg	0.932667	0.930446	0.931451	1501.000000
weighted avg	0.925249	0.924717	0.924856	1501.000000

Figure 5.14: (a) Confusion Matrix   (b) Metrics

Above mentioned figures show us confusion matrix and crucial metrics: Precision, F1-score and recall. F1-score balances precision and recall, giving an overall accuracy considering false predictions. Precision indicates accurate positive predictions, while recall measures the ability to capture all positives. These metrics collectively evaluate the model performance, through highlighting its accuracy and ability to make positive predictions.
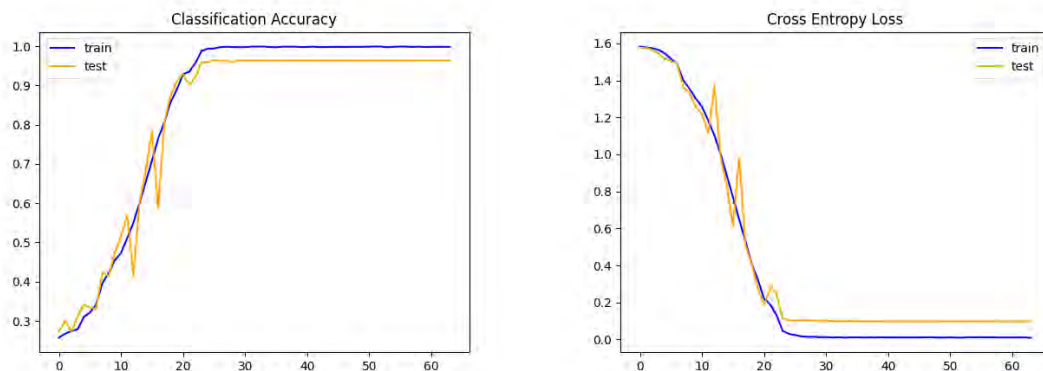
Figure 5.15: (a) Classification Accuracy   (b) Cross Entropy Loss

Alongside, we also implemented VGG-19. This architecture achieved an impressive classification accuracy of 95%. The accuracy graph above illustrates the model's learning process, emphasizing its capability to make precise predictions. At the same time, the cross-entropy loss graph depicts the model's optimization journey, characterized by a decreasing trend that indicates successful training. This outcome indicates that VGG-19 effectively extracted crucial features from the MRI scans, leading to accurate classification.

31

	precision	recall	f1-score	support
AD	0.975207	0.967213	0.971193	244.000000
CN	0.921569	0.933002	0.927250	403.000000
EMCI	0.956923	0.928358	0.942424	335.000000
LMCI	0.947867	1.000000	0.973236	200.000000
MCI	0.952381	0.940439	0.946372	319.000000
accuracy	0.948035	0.948035	0.948035	0.948035
macro avg	0.950789	0.953803	0.952095	1501.000000
weighted avg	0.948231	0.948035	0.947971	1501.000000
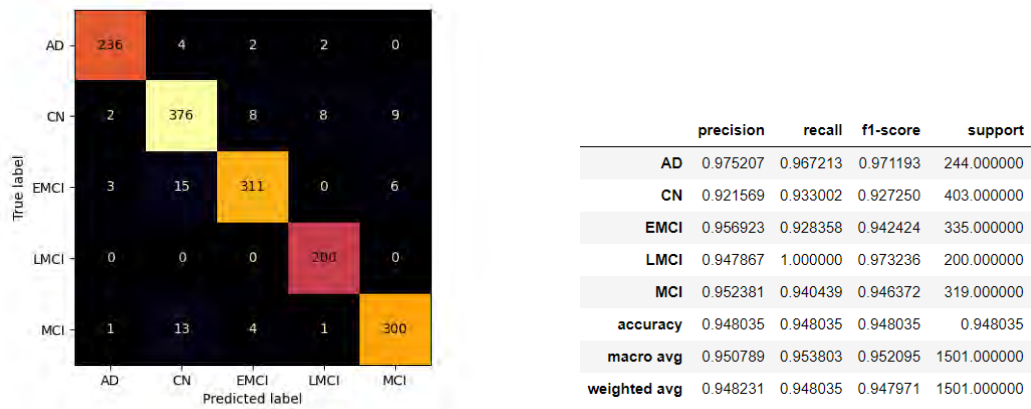
Figure 5.16: (a) Confusion Matrix   (b) Metrics

The figures shown above presents a confusion matrix alongside essential metrics: F1-score, Precision and recall. F1-score finds an equilibrium between precision and recall, offering a comprehensive gauge of accuracy that considers false predictions. Precision signifies the validity of positive predictions, whereas recall evaluate model's capability of identifying every positive cases. All these evaluating factors collectively predicts the model's performance, underscoring its accuracy and its aptitude for making positive predictions.

5.5 Alex Net

5.5.1 Model Implementation

Alexnet is a CNN architecture that comprises a total of eight weighted layers, among which there are 5 convolutional and 3 fully interconnected layers. After each layer except the final one, ReLU activation is applied. To produce a probability distribution across 1000 class labels softmax activation is applied at the final layer. Dropout regularization is utilized in the initial 2 fully interconnected layers. After $1st$, $2nd$ and $5th$ convolutional layers, the Max-pooling is performed, as illustrated in the figure. The kernels in the $2nd$, $4th$ and $5th$ convolutional layers are exclusively connected to kernel maps. Conversely, the kernels in the $3rd$ convolutional layer are interconnected to all kernel maps in the $2nd$ layer. And lastly, in the fully connected layers, establish connections with all the neurons in the preceding layer [85].
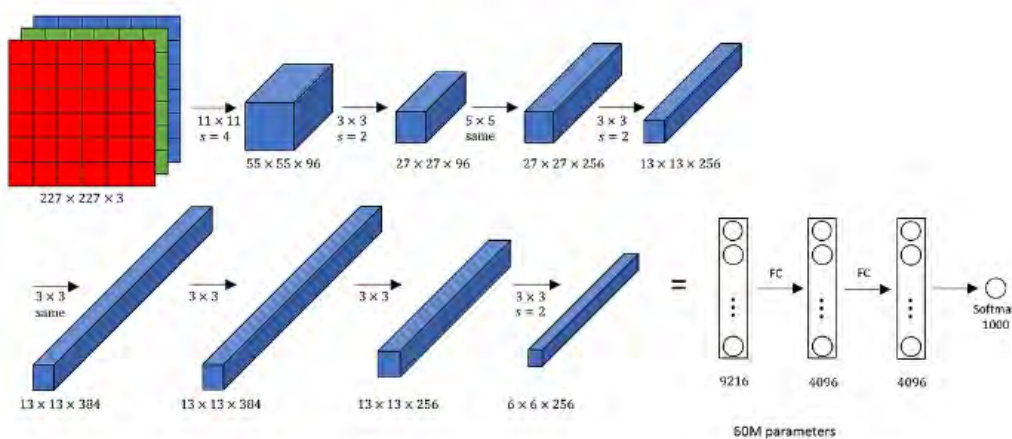
Figure 5.17: AlexNet Architecture [85]

Some key features of this model are:

- Use of 'ReLU' as an activation function instead of 'tanh'.

- Using 128 batch size.

- Dropout and Data augmentation (cropping, flipping).

A significant milestone of AlexNet is its adoption of the ReLU (Rectified Linear Unit) nonlinearity. Traditionally, neural networks were trained using activation functions like sigmoid or Tanh. But AlexNet illustrated that by employing ReLU non-linearity, deep CNNs can be trained at a significantly higher rate compared to slower training incorporated with saturating activation functions such as Tanh or sigmoid.

5.5.2 Results
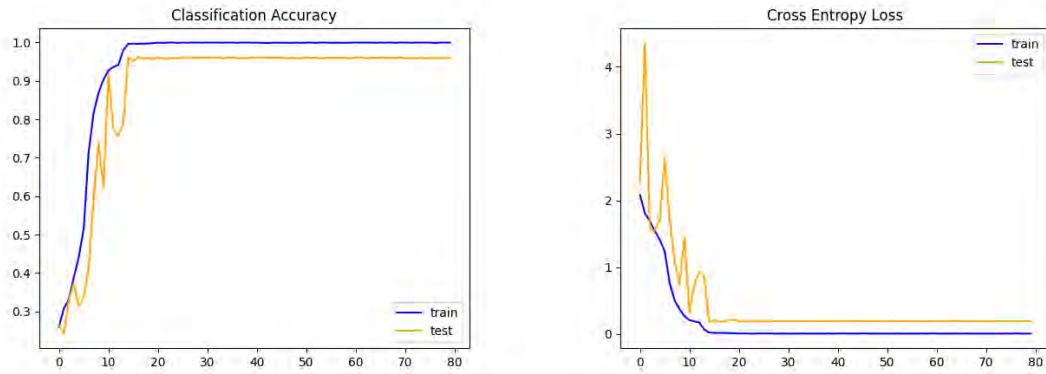
Figure 5.18: (a) Classification Accuracy   (b) Cross Entropy Loss

The training of the image data using the AlexNet architecture yielded a commendable classification accuracy of 95%. The accompanying classification accuracy graph showcased the model's learning progression, highlighting its ability to make accurate predictions. Simultaneously, the cross-entropy loss graph demonstrated the model's optimization process, with a decreasing trend indicating effective training. This outcome signifies that AlexNet effectively learned and extracted features from the image data, resulting in a robust and accurate classification model.
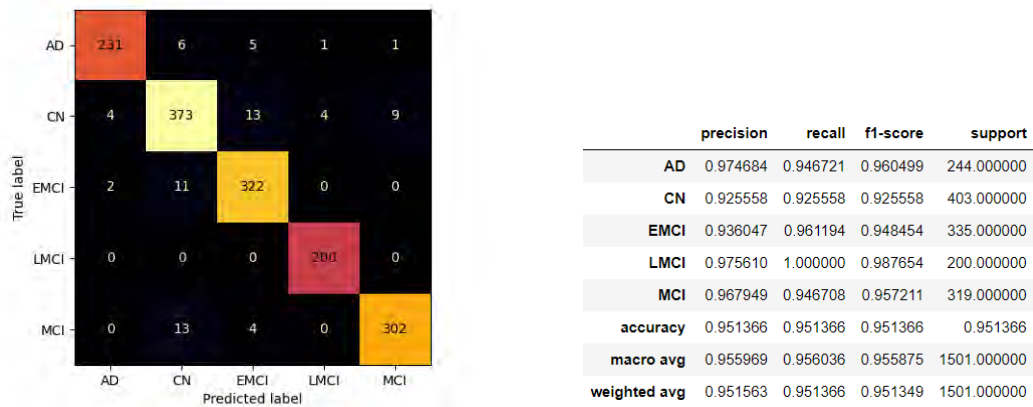
Figure 5.19: (a) Confusion Matrix   (b) Metrics

The above figures illustrate the confusion matrix and critical metrics such as Precision, F1-score, and recall. F1-score serves as a harmonious blend of precision and recall, delivering a holistic measure of accuracy that considers false predictions. Precision measures the positive predictions, whereas recall evaluate model's capability to detect every positive instances. Together, all these measuring factors gives a comprehensive evaluation of the architecture, underlining its accuracy and its proficiency in making positive predictions.

5.6   DenseNet-121

5.6.1   Model Implementation

Densely Connected Convolutional Networks, or DenseNets, represent advanced deep learning architecture by simplifying the connectivity patterns between layers. In traditional CNNs, information flows sequentially between layers, each receiving input from its previous layer and passing it to next layer. But as it becomes more profound, they face the "vanishing gradient" problem, where information can get lost as it travels through numerous layers, hampering training. DenseNets tackles this issue by introducing a modified architecture called Densely Connected Convolutional Networks. In DenseNets, every layer is directly interconnected to each and every other layer, creating dense network of connections. By creating $\frac{L(L+1)}{2}$ direct connections of L layers, this connectivity pattern ensures effective information flow and solves the vanishing gradient issue [86].

DenseNet-121 is a specific configuration of DenseNets designed for the ImageNet dataset. Let us break down the key components of DenseNet:
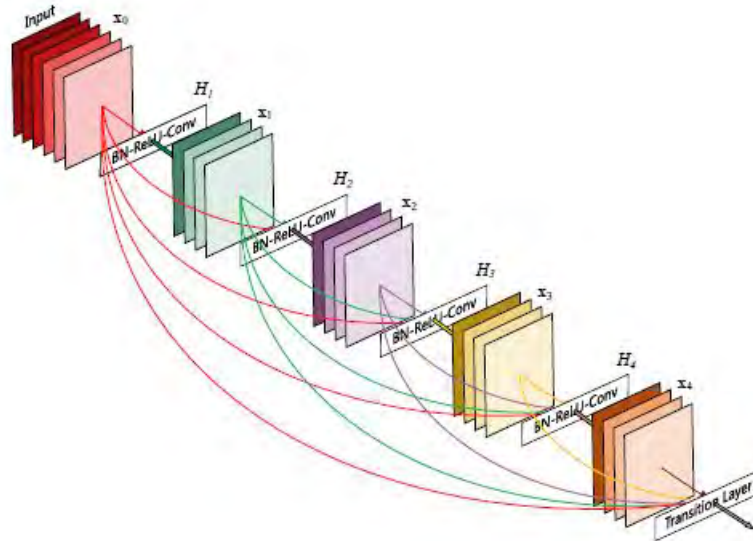
Figure 5.20: DenseNet-121 Architecture [86]

1. **Connectivity:** In traditional CNNs, layers are connected sequentially. Unlike it, DenseNets introduces dense connectivity. Every layer is directly interconnected to each and every other layer, developing feature reuse and reducing vanishing gradient problems.

$$\mathbf{x}_\ell = H_\ell([\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{\ell-1}]).$$

Here, [x0,x1,...,xl-1] is a concatenation of the feature maps. The multiple inputs of H are merged into a single tensor to make the implementation easy [86].

35

2. **DenseBlocks:** DenseNets consist of multiple dense blocks and in a dense block, layers maintain exact same feature map size while varying filter numbers. This enables the concatenation of feature maps from all preceding layers, promoting information flow.

3. **Growth Rate:** The rise is denoted as 'K,' which determines how many new feature maps every layer contributes. It regulates the amount of information added to the network at each layer. Each function H should result in k feature maps, thus

$$k_l = k_0 + k * (l - 1)$$

4. **Bottleneck Layers:** To improve computational efficiency, bottleneck layers (1x1 convolutions) can be introduced before 3x3 convolutions. These layers reduce the number of input channels.

   DenseNet-121, is a dynamic neural network that can be used for image classification and various other applications in the field of computer vision.

5.6.2 Results
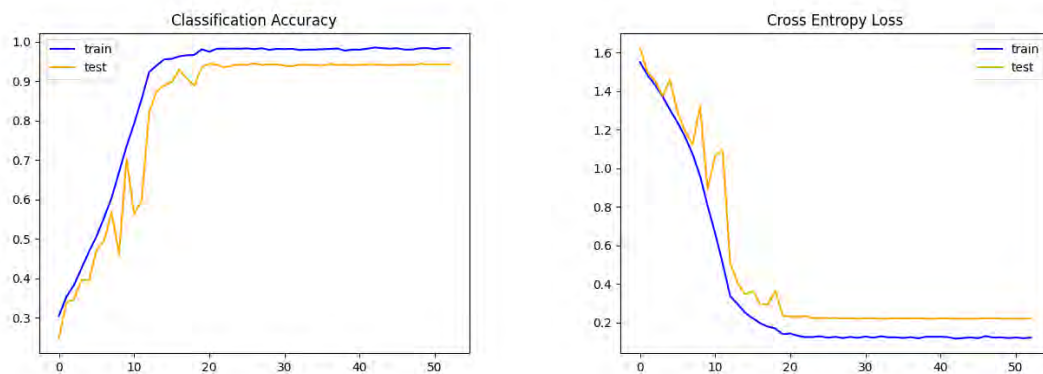
Figure 5.21: (a) Classification Accuracy   (b) Cross Entropy Loss

The DenseNet-121 architecture, when trained on image data, achieved an impressive classification accuracy of 93%. The accuracy graph provided above illustrates the model's learning journey, emphasizing its capability to make precise predictions. Additionally, the cross-entropy loss graph showcased the model's optimization process, marked by a decreasing trend, which indicates effective training. This result indicates that DenseNet-121 proficiently acquired and extracted features from the image data, resulting in a strong and accurate classification model.

	precision	recall	f1-score	support
AD	0.970711	0.950820	0.960663	244.000000
CN	0.877404	0.905707	0.891331	403.000000
EMCI	0.948012	0.925373	0.936556	335.000000
LMCI	0.956938	1.000000	0.977995	200.000000
MCI	0.951613	0.924765	0.937997	319.000000
accuracy	0.934044	0.934044	0.934044	0.934044
macro avg	0.940936	0.941333	0.940908	1501.000000
weighted avg	0.934699	0.934044	0.934160	1501.000000
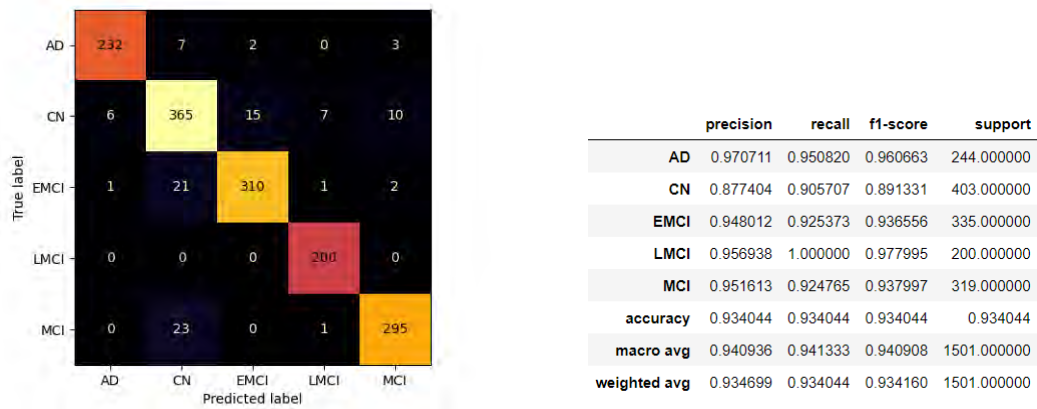
Figure 5.22: (a) Confusion Matrix (b) Metrics

The figures provided above depict the confusion matrix and essential metrics, including F1-score, Precision and recall. F1-score creates a balance between precision and recall, offering a well-rounded measure of accuracy that accounts for false predictions. Precision assesses the positive predictions accuracy,whereas recall assesses the model's ability to detect all positive instances. These combined metrics together provide a thorough assessment of the architecture's performance, emphasizing its accuracy and its effectiveness in making positive predictions.

5.7 Xception

5.7.1 Model Implementation

Xception is a CNN architecture built by Google researchers. It utilizes Depth wise Separable Convolutions, which is an extension of the Inception modules found in CNN. In this context, Depth-wise Separable Convolution is akin to an Inception module with a large tower numbers. This insight led to the creation of a new CNN architecture inspired by Inception. However, instead of Inception modules, Depth-wise Separable Convolutions are implemented. This model consists of an entry flow, followed by a middle flow which is repeated 8 times, and finishes with the flow of exit. Finally, batch normalisation is applied to all Separable Convolution and Convolution layers [87].
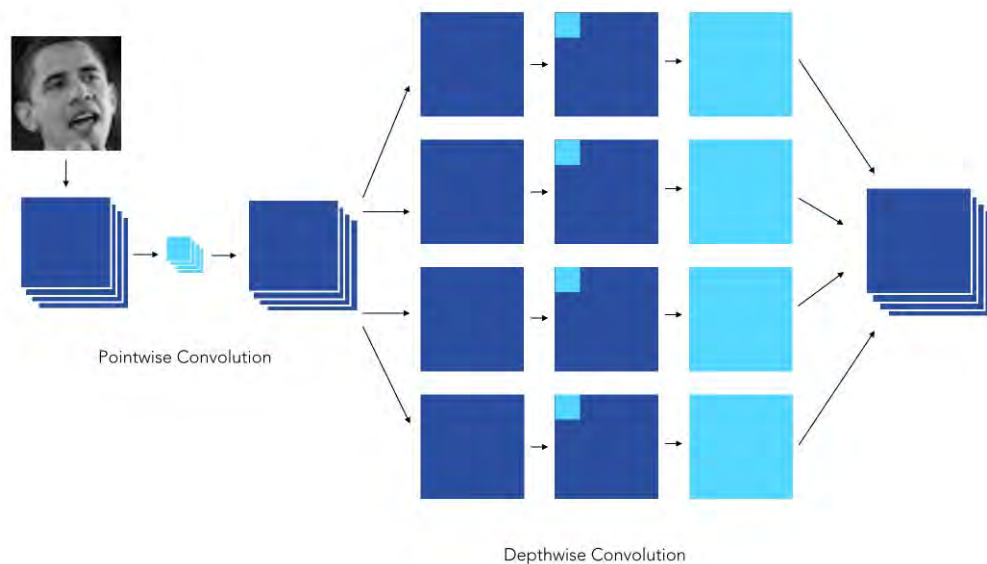
Figure 5.23: Implementation of Xception [87]

The Xception architecture is built on two key principles:

1. **Depthwise Separable Convolution:** This type of convolution operates channel by channel and uses an n×n spatial convolution.

2. **Pointwise convolution:** This involves a 1×1 convolution aimed at altering dimensions.

Unlike traditional convolution, there's no need for convolution across all channels. This leads to fewer connections and an overall lighter model. XCeption presents an architecture composed of Depth wise Separable Convolution blocks along with Maxpooling, interconnected by shortcuts similar to ResNet implementations. What makes XCeption different is that it reverses the order of operations compared to the typical Depth-wise Convolution followed by Pointwise Convolution, as shown in the figure below:

5.7.2 Results
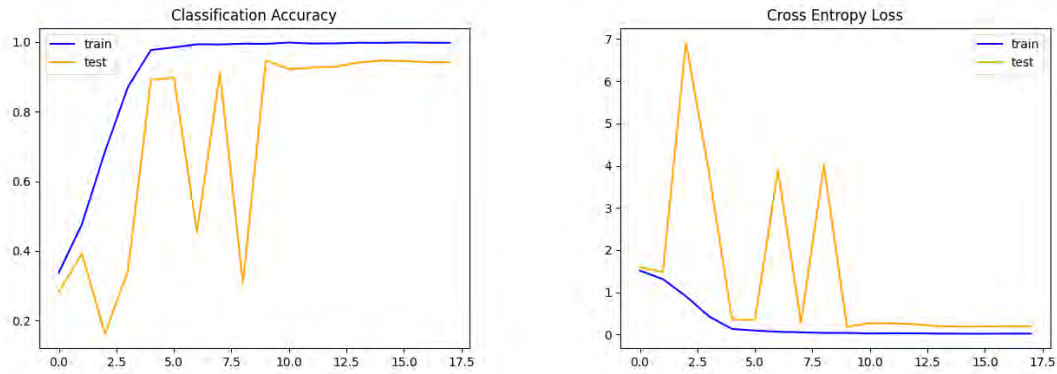
Figure 5.24: (a) Classification Accuracy   (b) Cross Entropy Loss

When trained on image data, the Xception architecture attained an outstanding classification accuracy of 93%. The accuracy chart above shows the model's learning process, underscoring its ability to make accurate predictions. Furthermore, the cross-entropy loss chart demonstrates the model's optimization journey, characterized by a consistent decline. This outcome highlights Xception's adeptness in acquiring and extracting features from the image data which ultimately leads to a robust classification model.


	precision	recall	f1-score	support
AD	0.958678	0.950820	0.954733	244.000000
CN	0.905852	0.883375	0.894472	403.000000
EMCI	0.921687	0.913433	0.917541	335.000000
LMCI	0.966184	1.000000	0.982801	200.000000
MCI	0.905199	0.927900	0.916409	319.000000
accuracy	0.926049	0.926049	0.926049	0.926049
macro avg	0.931520	0.935105	0.933191	1501.000000
weighted avg	0.925873	0.926049	0.925848	1501.000000

Figure 5.25: (a) Confusion Matrix   (b) Metrics

The above diagrams illustrate the confusion matrix and key metrics, including F1-score, Precision and recall. F1-score serves as a balanced measure of accuracy, taking into account both precision and recall. Precision evaluates the positive predictions accuracy, whereas recall evaluates the model's capability to detect all positive instances. Every metrics collectively offer a comprehensive evaluation of model's efficiency, highlighting its accuracy and its capability in positive predictions.

Chapter 6

Vision Transformer

6.1 Model Implementation

Vision Transformer is a neural network model that is used for tasks like, image classification in computer vision. It is built on the Transformer architecture which was originally created for natural language processing (NLP). VIT's modify this architecture to suit its image processing needs. A major adaptation can be seen in image representation. Unlike NLP Transformers, which treat text as word sequences, ViT's break down images into several sequences of patches. These patches are small 16x16 pixel rectangular sections of the image. This approach is crucial for adapting the model to handle visual data effectively.
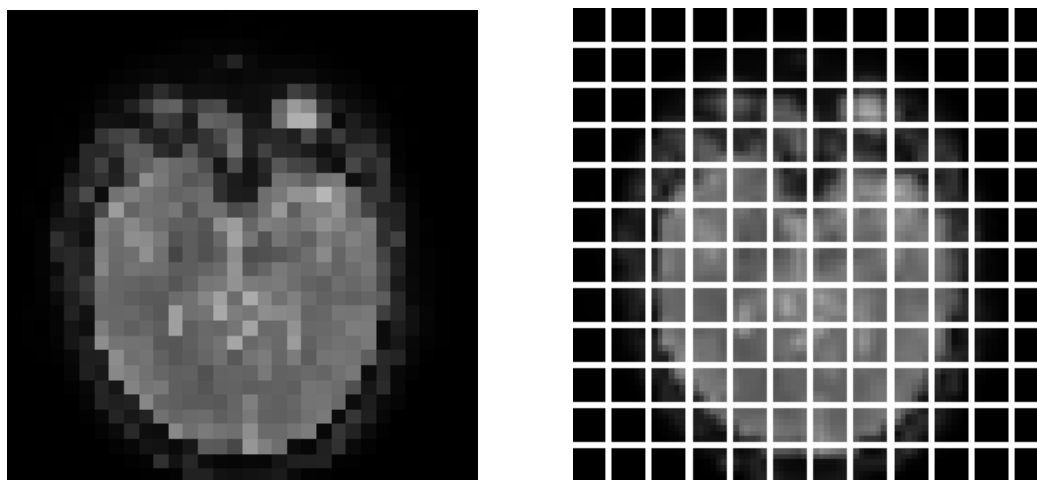
Figure 6.1: 144 patches per MRI scan

In Vision Transformers (ViT's), an image is initially segmented into several patches. Each of these patches is then converted into a vector that encapsulate the patch's features. Normally, these features are derived using CNN, which is trained to identify characteristics vital for classifying images. Following this, each vector representing a patch is passed into a Transformer encoder. This encoder is comprised of several layers of self-attention mechanisms.
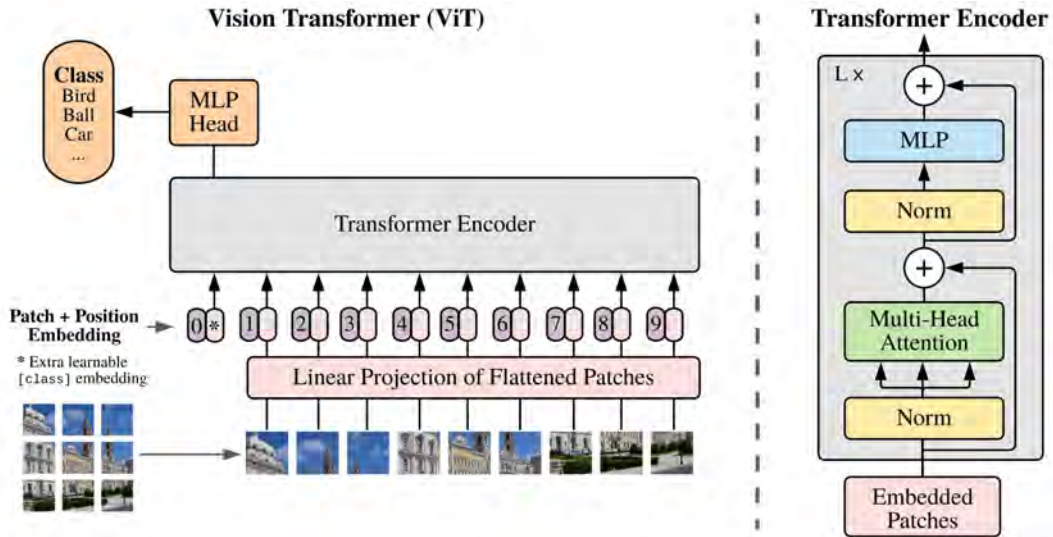
Figure 6.2: ViT Architecture [88]

These self-attention layers enable the model to understand and learn the relationships and dependencies between distant patches within the image. This learning is essential for image classification, as it helps the model comprehend how different segments of an image collectively determine its overall classification. And lastly, the Transformer encoder outputs a series of vectors. These vectors are a representation of the image's features, which are further utilized in the classification of the image.

6.2 Results

Vision Transformers (ViTs) for image classification offers several advantages. It is capable of learning the global characteristics of images, as it can focus on any section of the image irrespective of its position. These features are beneficial for activities like object detection and comprehending scenes. Moreover, it has reduced sensitivity to data augmentation compared to Convolutional Neural Networks (CNNs). Additionally, ViTs are versatile in their application to various types of image classification tasks, such as object detection, scene interpretation and detailed classification.
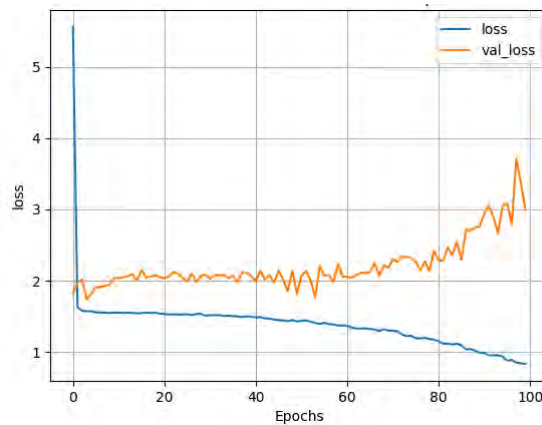
Figure 6.3: Train and Validation Loss Over Epochs

Nevertheless, when applied to our dataset, there were multiple challenges associated with using Vision Transformers (ViTs). The training of ViTs is resource-intensive because of their substantial number of parameters. This high computational cost is a common challenge in case of image-related tasks, which often involve processing large-sized pixel data. Additionally, ViTs is a bit inefficient compared to CNN. This inefficiency arises because ViTs need to process every part of the image, regardless of its relevance to the specific task. Furthermore, transformer-based models, including Vision Transformers (ViT), are typically trained on a substantial volume of image data, ranging from 14 million to 300 million images, to achieve effective generalization on new unseen data [88]. In comparison to CNNs, transformers generally have fewer inbuilt biases and are highly dependent on extensive datasets for large-scale training. In our case, the dataset comprised merely 8000 images. This limited dataset size resulted in a low accuracy rate of 32%. The volume of data is crucial, as it influences the ability of transformers to generalize effectively and maintain robustness across various computer vision applications.

Chapter 7

Proposed Custom Model

7.1 Unsupervised Pre-training

We trained a denoising convolutional autoencoder on a different Alzheimer's dataset comprising four classes based on CDR(Clinical Dementia Rating) scores. The goal was not image reconstruction, rather we wanted to capture the characteristics of MR images related to Alzherimer's disease via the latent representation in the bottleneck layer of the encoder network. After training the autoencoder network, we utilized the encoder network with the learned weights and used it as feature extractor in the top layers of our proposed custom architecture. This way the kernels associated with the top convolutional layers of our custom model were not initialized randomly or via some defined initializer, rather the parameters in the kernel were in a good proximity for optimization from where the network benefitted from faster convergence. That way the encoder network served as an effective scalable regularizer. This technique is called unsupervised pretraining. The choice of using denoising autoencoder was to force the network to learn a meaningful representation of the data by constraining the network with the reconstruction of images from its noisy corruption. This is similar to masked autoencoders which are typically used in VITs. The convolutional autoencoder is pretty simple, the encoder layer resembles the first two convolutional layers of our custom architecture and the decoder is a mirror of the encoder featuring transposed convolution operations. We trained the autoencoder using MSE(L2 Norm).

7.2   Model Implementation

Our custom model comprises of 5 Convolutional blocks. The first, second and fifth convolution incorporates downsampling layers using strided (s=2) average pooling. Each convolution operation is followed by ReLU activation then batch normalization. The first two convolutional blocks come from the pretrained encoder network. He normal initialization was used for the kernels of the last three convolutional blocks. The network gradually decreases the input dimension while increasing the depth. For the classification head, five fully connected dense layers were used with four hidden layers and one output layer. 4096 neurons were used in the first hidden layer and 2048 in the second and third layer and 1024 neurons in the last hidden dense layer. Needless to say, there were five neurons in the output layer for the classification in five classes. We used 80 epochs, a batch size of 32, train test split of 80:20, Adam optimizer with an initial learning rate of 1e-4 with categorical cross entropy for loss. We also used reduce learning rate on plateau as the callback to adjust learning rate with factor 0.3, min delta 1e-3 and patience of 2.
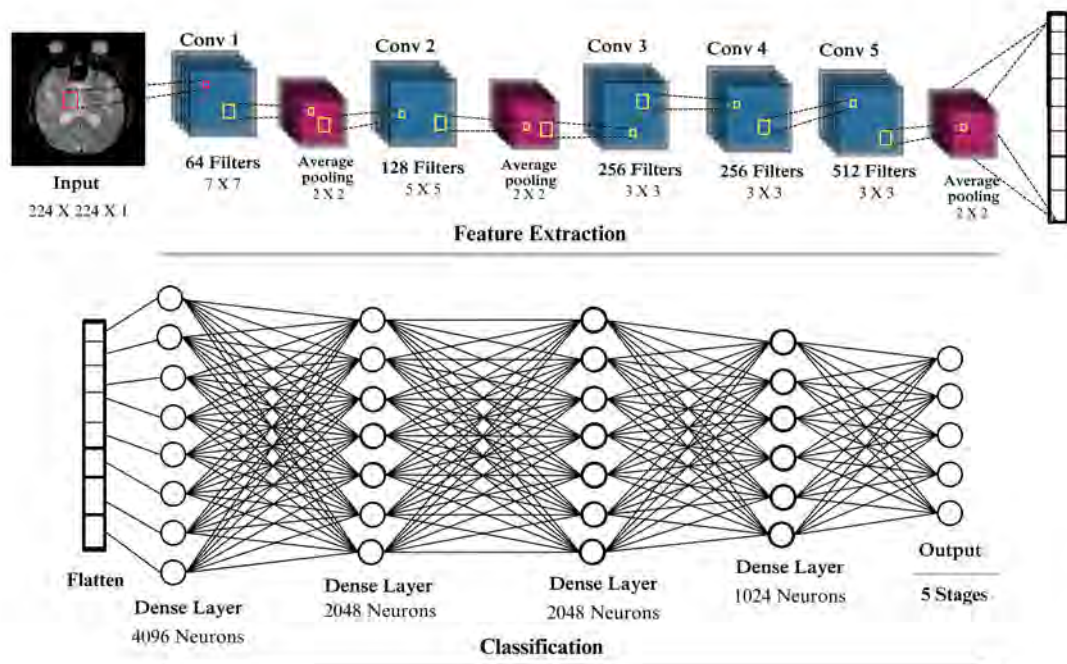

Figure 7.1: Custom Model Architecture

7.2.1 Feature Extraction

In this architecture, convolutional layers form the foundation for feature extraction from input images, which is essential for identifying complex patterns. The first convolutional layer employs 64 filters, each having a size of 7x7, and a stride of 2 pixels. This configuration helps the model to capture the basic visual elements such as edges and color gradients from the raw MRI scans. Through using a large stride, the layer successfully reduces the spatial resolution of the feature maps, which eventually diminishes the sensitivity to the precise locations of features within the MRI scan.

Following the initial convolution, we applied a BatchNormalization. This layer normalizes the feature maps produced by the Conv2D layer. It ensures that the mean activation remain around 0 and the standard deviation around 1. This process is important because it combat the issue of internal covariate shift, where the distribution of activations changes during training. Therefore, it stabilizes and accelerates the learning process. After that, we incorporated an AveragePooling2D layer with a 2x2 pooling window and a stride of 2. This pooling layer down samples the feature maps and summarizes the extracted features into a more compact form, which significantly reduced the computational load on the network. Thus, it contributed to make the representation more robust to small translations in the input image.

Layer Type	Size/ Unit	Kernel	Strides	Padding	Activation	Other Attributes
Conv2D	64	7 x 7	2	Same	ReLU	Input shape, Encoder learned kernel
Batch Normalization	-	-	-	-	-	-
Average Pooling2D	-	2 x 2	2	Valid	-	-
Conv2D	128	5 x 5	1	Same	ReLU	Encoder learned kernel
Batch Normalization	-	-	-	-	-	-
Average Pooling2D	-	2 x 2	2	Valid	-	-
Conv2D	256	3 x 3	1	Same	ReLU	he_normal initializer
Batch Normalization	-	-	-	-	-	-
Conv2D	256	3 x 3	1	Same	ReLU	he_normal initializer
Batch Normalization	-	-	-	-	-	-
Conv2D	512	3 x 3	1	Same	ReLU	he_normal initializer
Batch Normalization	-	-	-	-	-	-
Average Pooling2D	-	2 x 2	2	Valid	-	-

Table 7.1: Feature Extraction Layers

In the second block we included a Conv2D layer with a substantial increase in the number of filters to 128, with a smaller kernel size of 5x5, that operates with a stride of 1 and padding 'same'. This configuration allows the model to detect higher-level features and also maintain the spatial dimensions of the feature maps. Again, this is followed by a BatchNormalization and AveragePooling2D layer. It further condenses the feature information and decreases dimensionality to prepare the data for even more complex pattern recognition.

The third block is consisted of a Conv2D layer with 256 filters and a 3x3 kernel size, maintaining a stride of 1. This layer captures more intricate and abstract features from the already processed data, providing a richer representation of the MRI scan. Like all the previous Conv2D layers, it is also followed by a BatchNormalization layer to ensure the consistency of data distribution that are moving forward.

The fourth and fifth blocks is similar to the structure of the third, with Conv2D layers continuing to refine the features, followed by BatchNormalization. For the fourth block the number of filters remains at 256 but returns to 512 in the fifth block, still using 3x3 kernels and strides of 1. These layers are crucial for developing a deep and detailed understanding of the data. Finally, a AveragePooling2D layer in the fifth block reduces the spatial dimensions for the last time and produces a set of feature maps which encapsulates the most salient characteristics of the MRI scan.

7.2.2 Classification

The transition of convolutional layers is marked by the introduction of a Flatten layer. The Flatten layer transforms the multi-dimensional output of the convolutional layers into a one-dimensional vector. This step is crucial because fully connected layer does not preserve the spatial structure of the input data and therefore it requires a flattened input vector. The flattened vector is then passed through a series of fully connected (Dense) layers. Each of these layers is comprised of 4096 neurons which allows the network to learn complex relationships in the data. The ReLU (Rectified Linear Unit) activation function is applied at these layers to introduce non-linearity to the learning process because it is essential for modeling the non-linear relationships inherent in real-world data. Following each Dense layer, a BatchNormalization step is applied to maintain numerical stability because it normalizes the output of the neurons and ensures that no single neuron's output can disproportionately influence the subsequent layers.

Moreover, to combat overfitting we used a regularization technique known as Dropout. It is employed after each BatchNormalization step to mitigate overfitting by randomly deactivating a subset of neurons during each training iteration (30% in this case). This prevents the network from being dependent on any specific neuron and encourage robust learning of features. This pattern of a Dense layer followed by BatchNormalization and Dropout is again repeated 2 times but this time with neuron number 2048. And lastly, we used another dense layer comprised of 1024 neurons to solidify the network's ability to establish more sophisticated associations between the learned features and the output predictions.

Layer Type	Size/ Unit	Kernel	Strides	Padding	Activation	Other Attributes
Flatten	-	-	-	-	-	-
Dense	4096	-	-	-	ReLU	-
Batch Normalization	-	-	-	-	-	-
Dropout	-	-	-	-	-	30% Dropout rate
Dense	2048	-	-	-	ReLU	-
Batch Normalization	-	-	-	-	-	-
Dropout	-	-	-	-	-	30% Dropout rate
Dense	2048	-	-	-	ReLU	-
Batch Normalization	-	-	-	-	-	-
Dropout	-	-	-	-	-	30% Dropout rate
Dense	1024	-	-	-	ReLU	-
Batch Normalization	-	-	-	-	-	-
Dropout	-	-	-	-	-	30% Dropout rate
Dense (Output Layer)	Num of Classes (5)	-	-	-	Softmax	-

Table 7.2: Fully Connected Layers

The network's learning process is maximum in the final layer, which is a Dense layer with neurons similar to the number of classes to be identified. This layer uses the softmax activation function, which converts the output of the network into a probability distribution over the predicted classes, allowing it to predict properly. Finally, the model is compiled with the Adam optimizer to handle sparse gradients and adjust the learning rates adaptively. This optimization algorithm is particularly adept at navigating the complex landscape of high-dimensional weight space. Furthermore, the loss function employed is categorical_crossentropy, which is a standard choice for multi-class classification problems because it can measure the discrepancy between the predicted probability distribution and the true distribution. The metric of accuracy is also tracked to provide a clear measure of the model's performance during training.
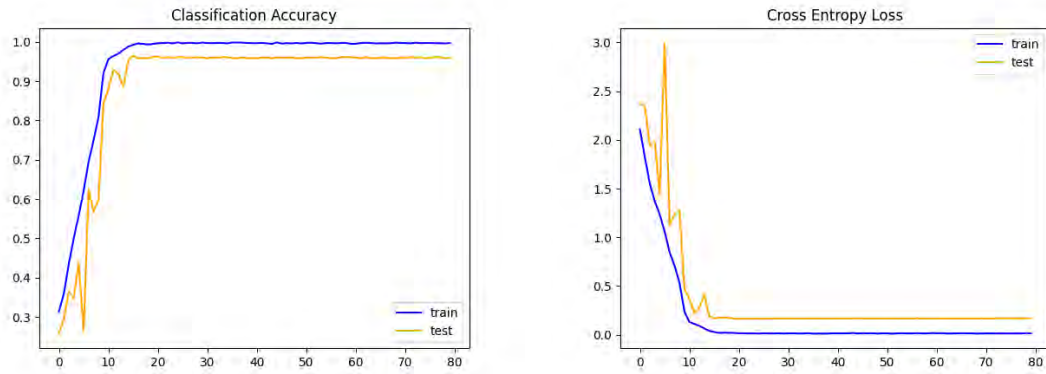
7.2.3 Results

Figure 7.2: (a) Classification Accuracy   (b) Cross Entropy

Our custom architecture has achieved an impressive classification accuracy of 96.6%, outperforming all other state-of-the-art models. The accuracy graph provided above illustrates the model's learning progress, emphasizing its capacity for precise predictions. At the same time, the cross-entropy loss graph reflects the model's optimization process, showing a consistent decrease, indicating successful training. These results indicate that our custom model has effectively captured essential features from the MRI scans, leading to accurate classification.

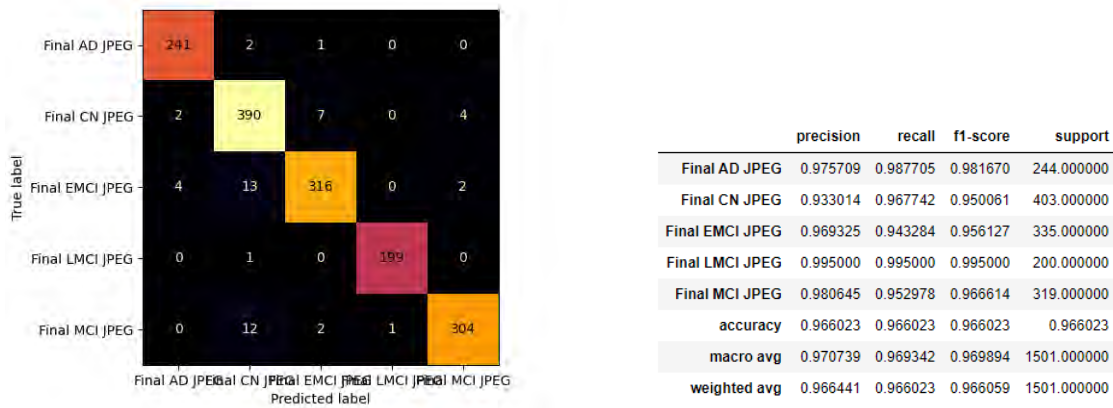|	precision	recall	f1-score	support
Final AD JPEG	0.975709	0.987705	0.981670	244.000000
Final CN JPEG	0.933014	0.967742	0.950061	403.000000
Final EMCI JPEG	0.969325	0.943284	0.956127	335.000000
Final LMCI JPEG	0.995000	0.995000	0.995000	200.000000
Final MCI JPEG	0.980645	0.952978	0.966614	319.000000
accuracy	0.966023	0.966023	0.966023	0.966023
macro avg	0.970739	0.969342	0.969894	1501.000000
weighted avg	0.966441	0.966023	0.966059	1501.000000

Figure 7.3: (a) Confusion Matrix   (b) Metrics

The figures mentioned above display the confusion matrix along with essential metrics: Precision, F1-score and recall. F1-score provides balanced assessment of precision and recall, offering an overall accuracy that considers false predictions. Precision reflects the positive prediction accuracy, whereas recall evaluates the model's capacity to identify every positive instances. These metrics together assess the model's efficiency, emphasizing its accuracy and its ability to make positive predictions.
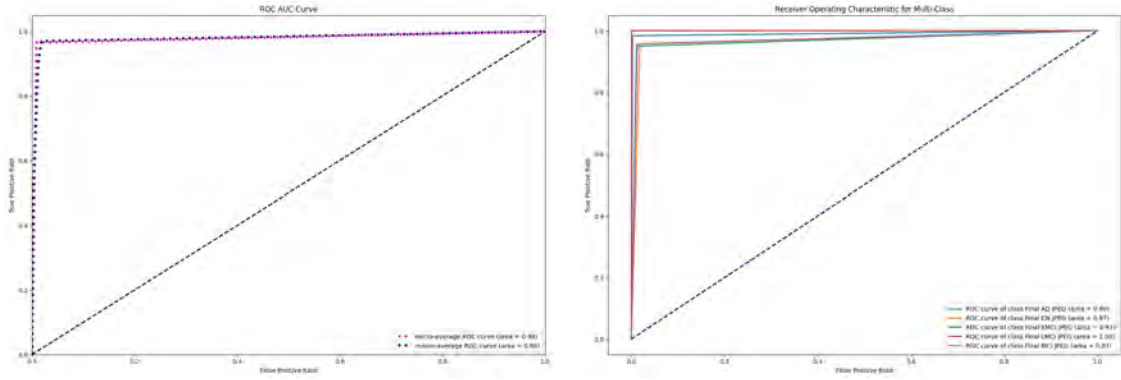
Figure 7.4: (a) ROC AUC Curve (b) Multi-class ROC AUC Curve

The first image below shows an ROC curve with an AUC of 0.98, indicating that the model almost perfectly distinguishes between two classes. The second image illustrates ROC curves for a multi-class problem with one class having a perfect AUC of 1.0, showing exceptional model accuracy in class differentiation.

7.3 Grad-CAM Application

Grad-CAM, short term used for Gradient-weighted Class Activation Mapping, is a well-known technique used in deep learning for visualizing and understanding the decision-making process of convolutional neural networks (CNNs). It is basically a class activation mapping method, which assists in identifying the specific areas of an input image that greatly affects the output of a CNN.
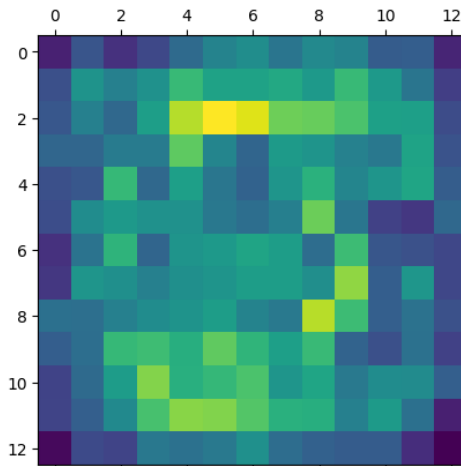

Figure 7.5: Heat-Map

In deep learning, we have seen that Convolutional Neural Networks (CNNs) are commonly used for image classification tasks. We have also worked on a custom model, which is based on CNN architecture. Our model learned to distinguish pertinent details from the input MRI scans and then used this information to predict the content of the image. Grad-CAM has allowed us to visualize the specific areas of the input MRI scans that our custom CNN model relied on to make its decisions.
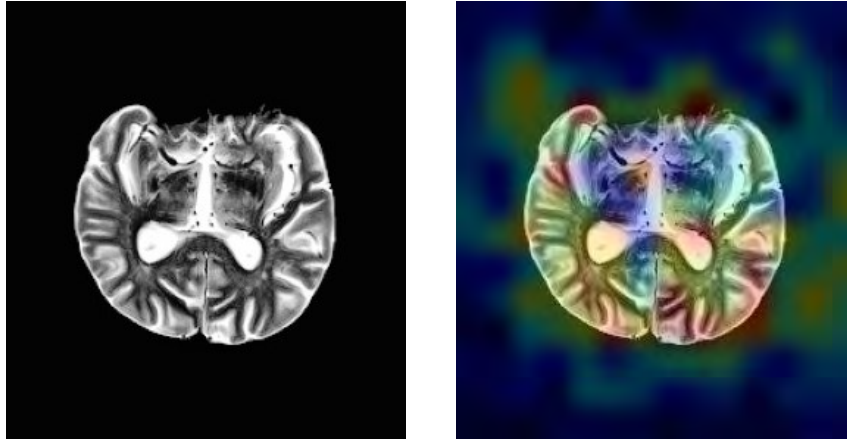
Figure 7.6: Overlaying Heat-map onto Original MRI Scan

Grad-CAM identified the output class score through computing the gradient of the class score with respect to the feature maps of the final convolutional layer ('conv2d_4') of our custom CNN model. These feature maps are then provided with weights based on these gradients, resulting in a heatmap which highlighted the crucial regions of the input MRI scan. And lastly, by overlaying this heatmap onto the original input MRI scan has allowed us to visualize the most significant areas which are responsible for the decision-making process of our custom CNN model.
Using Grad-CAM in this predictive process of a CNN has enabled us to gain a deeper understanding of our model's strengths and weaknesses. Grad-CAM proves especially beneficial in tasks related to object localization, where identifying specific regions of an object or image is crucial.

7.3.1   Base Line Model Summary
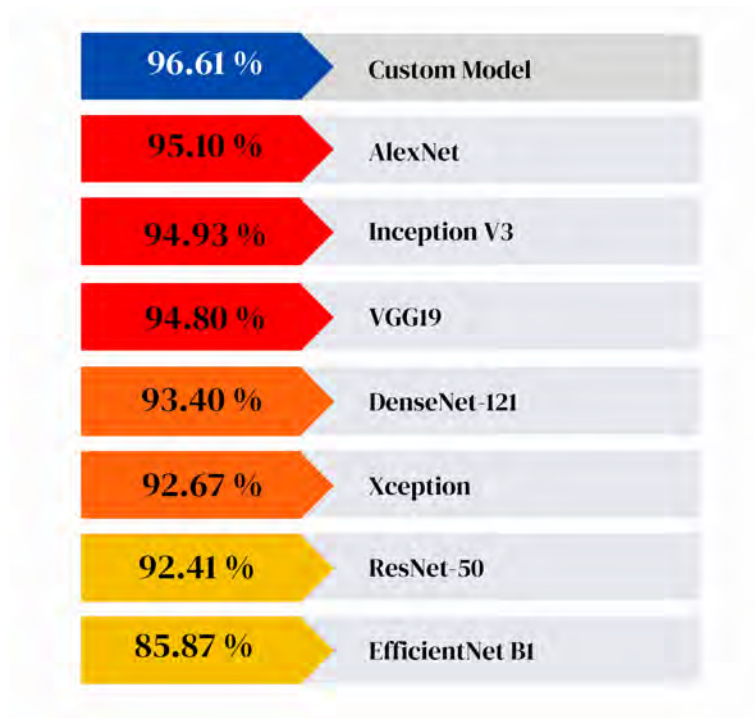
Chapter 8

Conclusion and Future Plan

Deep Learning (DL) and Computer Vision (CV) techniques show promising results in detecting and categorizing different stages of AD using MRI scans. AD, a prevalent well-known disorder, requires early identification for better patient outcomes and effective treatments. DL has demonstrated accurate identification of MCI and AD by recognizing patterns from extensive datasets. Moreover, understanding the different stages of MCI namely EMCI and LMCI should provide further insight for better understanding the MCI conversion. CV techniques help study brain structures and identify subtle changes linked to AD stages. The global prevalence of dementia, with AD being the most common form, demands urgent attention. The projected rise in AD cases will lead to increased disability and economic burden. As a result, implementing the newly improved custom model can immensely help to determine the patient outcomes and therefore advancing AD treatments. We are aiming to optimize this custom model further, using the following methods:

- **Model Optimization:** Application of techniques like Ensemble methods, Masked Autoencoder for VIT, Segmentation Masks, Deep Clustering methods for automatic instance labeling.

- **Large-Scale Clinical Validation:** Conduct extensive clinical validation studies involving diverse patient populations to assess the model's real-world performance and generalizability.

- **Educational Initiatives:** Develop educational programs to train healthcare professionals in the use of AI-based tools for AD diagnosis, promoting wider adoption of ML in the medical community.

Bibliography

[1] *What is Alzheimer's? — alz.org*, https://www.alz.org/alzheimers-dementia/what-is-alzheimers, [Accessed 23-May-2023].

[2] *Alzheimerapos;s disease - Diagnosis and treatment - Mayo Clinic — mayoclinic.org*, https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/diagnosis-treatment/drc-20350453, [Accessed 23-May-2023].

[3] H. R. Roth, C. Shen, H. Oda, *et al.*, "Deep learning and its application to medical image segmentation," *Medical Imaging Technology*, vol. 36, no. 2, pp. 63–71, 2018.

[4] *Alzheimer's disease*, https://www.radiologyinfo.org/en/info/alzheimers.

[5] *Dementia — https*, https://www.who.int/news-room/fact-sheets/detail/dementia, [Accessed 23-May-2023].

[6] Author(s), "Title of the article," *Journal Name*, vol. Volume, no. Issue, Page Range, Year. DOI: 10.1002/alz.13016. [Online]. Available: https://doi.org/10.1002/alz.13016.

[7] M. Liu, D. Zhang, D. Shen, and A. D. N. Initiative, "View-centralized multi-atlas classification for alzheimer's disease diagnosis," *Human brain mapping*, vol. 36, no. 5, pp. 1847–1865, 2015.

[8] P. C. Petersen and R. W. Berg, "Lognormal firing rate distribution reveals prominent fluctuation–driven regime in spinal motor networks," *Elife*, vol. 5, e18805, 2016.

[9] L. Keuck, "History as a biomedical matter: Recent reassessments of the first cases of alzheimer's disease," *History and Philosophy of the Life Sciences*, vol. 40, no. 1, p. 10, 2018.

[10] S. Al-Shoukry, T. H. Rassem, and N. M. Makbol, "Alzheimer's diseases detection by using deep learning algorithms: A mini-review," *IEEE Access*, vol. 8, pp. 77 131–77 141, 2020.

[11] N. L. Hill and J. Mogle, "Alzheimer's disease risk factors as mediators of subjective memory impairment and objective memory decline: Protocol for a construct-level replication analysis," *BMC geriatrics*, vol. 18, pp. 1–8, 2018.

[12] M. Waser, T. Benke, P. Dal-Bianco, *et al.*, "Neuroimaging markers of global cognition in early alzheimer's disease: A magnetic resonance imaging- electroencephalography study," *Brain and behavior*, vol. 9, no. 1, e01197, 2019.

[13] K. Trojachanec, I. Kitanovski, I. Dimitrovski, and S. Loshkovska, "Longitudinal brain mri retrieval for alzheimer's disease using different temporal information," *IEEE Access*, vol. 6, pp. 9703–9712, 2017.

[14] R. Cuingnet, E. Gerardin, J. Tessieras, *et al.*, "Automatic classification of patients with alzheimer's disease from structural mri: A comparison of ten methods using the adni database," *neuroimage*, vol. 56, no. 2, pp. 766–781, 2011.

[15] I. Lajoie, S. Nugent, C. Debacker, *et al.*, "Application of calibrated fmri in alzheimer's disease," *NeuroImage: Clinical*, vol. 15, pp. 348–358, 2017.

[16] A. Drzezga, D. Altomare, C. Festari, *et al.*, "Diagnostic utility of 18f- fluorodeoxyglucose positron emission tomography (fdg-pet) in asymptomatic subjects at increased risk for alzheimer's disease," *European journal of nuclear medicine and molecular imaging*, vol. 45, pp. 1487–1496, 2018.

[17] V. L. Villemagne, V. Doré, S. C. Burnham, C. L. Masters, and C. C. Rowe, "Imaging tau and amyloid-β proteinopathies in alzheimer disease and other conditions," *Nature Reviews Neurology*, vol. 14, no. 4, pp. 225–236, 2018.

[18] A. D. Cohen and W. E. Klunk, "Early detection of alzheimer's disease using pib and fdg pet," *Neurobiology of disease*, vol. 72, pp. 117–122, 2014.

[19] M. M. Monti, "Statistical analysis of fmri time-series: A critical review of the glm approach," *Frontiers in human neuroscience*, vol. 5, p. 28, 2011.

[20] I. R. Silva, G. S. Silva, R. G. de Souza, W. P. dos Santos, and R. A. d. A. Fagundes, "Model based on deep feature extraction for diagnosis of alzheimer's disease," in *2019 international joint conference on neural networks (IJCNN)*, IEEE, 2019, pp. 1–7.

[21] 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.

[22] S. Basaia, F. Agosta, L. Wagner, *et al.*, "Automated classification of alzheimer's disease and mild cognitive impairment using a single mri and deep neural networks," *NeuroImage: Clinical*, vol. 21, p. 101 645, 2019.

[23] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3d brain mri classification," in *2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)*, IEEE, 2017, pp. 835–838.

[24] F. U. R. Faisal and G.-R. Kwon, "Automated detection of alzheimer's disease and mild cognitive impairment using whole brain mri," *IEEE Access*, vol. 10, pp. 65 055–65 066, 2022.

[25] Y. Kazemi and S. Houghten, "A deep learning pipeline to classify different stages of alzheimer's disease from fmri data," in *2018 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, IEEE, 2018, pp. 1–8.

[26] S. Klöppel, C. M. Stonnington, J. Barnes, *et al.*, "Accuracy of dementia diagnosis—a direct comparison between radiologists and a computerized method," *Brain*, vol. 131, no. 11, pp. 2969–2974, 2008.

[27] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future," *Classification in BioApps: Automation of Decision Making*, pp. 323–350, 2018.

[28] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *Ieee Access*, vol. 6, pp. 9375–9389, 2017.

[29] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, vol. 19, pp. 221–248, 2017.

[30] G. Litjens, T. Kooi, B. E. Bejnordi, *et al.*, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[31] Z. Zhao, J. H. Chuah, K. W. Lai, *et al.*, "Conventional machine learning and deep learning in alzheimer's disease diagnosis using neuroimaging: A review," *Frontiers in Computational Neuroscience*, vol. 17, p. 10, 2023.

[32] P. Benefits, "Alzheimer's disease facts and figures includes a special report on the financial and personal benefits of early diagnosis," 2018.

[33] F. Falahati, E. Westman, and A. Simmons, "Multivariate data analysis and machine learning in alzheimer's disease with a focus on structural magnetic resonance imaging," *Journal of Alzheimer's disease*, vol. 41, no. 3, pp. 685–708, 2014.

[34] S. Klöppel, C. M. Stonnington, C. Chu, *et al.*, "Automatic classification of mr scans in alzheimer's disease," *Brain*, vol. 131, no. 3, pp. 681–689, 2008.

[35] P. Aisen, R. Petersen, M. Donohue, A. Gamst, R. Raman, R. Thomas, *et al.*, "Clinical core of the alzheimer's disease neuroimaging initiative: Progress and plans. alzheimers dement. 2010; 6 (3): 239–46," Epub 2010/05/11. doi: 10.1016/j. jalz. 2010.03. 006 PMID: 20451872, Tech. Rep.

[36] M. W. Weiner, D. P. Veitch, P. S. Aisen, *et al.*, "2014 update of the alzheimer's disease neuroimaging initiative: A review of papers published since its inception," *Alzheimer's & dementia*, vol. 11, no. 6, e1–e120, 2015.

[37] E.-S. Lee, K. Yoo, Y.-B. Lee, *et al.*, "Default mode network functional connectivity in early and late mild cognitive impairment," *Alzheimer Disease & Associated Disorders*, vol. 30, no. 4, pp. 289–296, 2016.

[38] X. Hua, C. R. Ching, A. Mezher, *et al.*, "Mri-based brain atrophy rates in adni phase 2: Acceleration and enrichment considerations for clinical trials," *Neurobiology of aging*, vol. 37, pp. 26–37, 2016.

[39] E. C. Edmonds, C. R. McDonald, A. Marshall, *et al.*, "Early versus late mci: Improved mci staging using a neuropsychological approach," *Alzheimer's & Dementia*, vol. 15, no. 5, pp. 699–708, 2019.

[40] J. Bennett, "Bach," *The Musical Times and Singing Class Circular*, vol. 32, no. 576, pp. 85–85, 1891.

[41] N. Aggarwal, R. Wilson, T. Beck, J. Bienias, and D. Bennett, "Mild cognitive impairment in different functional domains and incident alzheimer's disease," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 76, no. 11, pp. 1479–1484, 2005.

[42] J. P. Lerch, J. Pruessner, A. P. Zijdenbos, *et al.*, "Automated cortical thickness measurements from mri can accurately separate alzheimer's patients from normal elderly controls," *Neurobiology of aging*, vol. 29, no. 1, pp. 23–30, 2008.

[43] E. Gerardin, G. Chételat, M. Chupin, *et al.*, "Multidimensional classification of hippocampal shape features discriminates alzheimer's disease and mild cognitive impairment from normal aging," *Neuroimage*, vol. 47, no. 4, pp. 1476–1486, 2009.

[44] G. M. McKhann, D. S. Knopman, H. Chertkow, *et al.*, "The diagnosis of dementia due to alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease," *Alzheimer's & dementia*, vol. 7, no. 3, pp. 263–269, 2011.

[45] M. C. Carrillo, R. A. Dean, F. Nicolas, *et al.*, "Revisiting the framework of the national institute on aging-alzheimer's association diagnostic criteria," *Alzheimer's & Dementia*, vol. 9, no. 5, pp. 594–601, 2013.

[46] M. F. Folstein, S. E. Folstein, and P. R. McHugh, ""mini-mental state": A practical method for grading the cognitive state of patients for the clinician," *Journal of psychiatric research*, vol. 12, no. 3, pp. 189–198, 1975.

[47] J. C. Morris, "The clinical dementia rating (cdr): Current version and scoring rules," *Neurology*, vol. 43, no. 11, pp. 2412–2414, 1993.

[48] S.-C. Wang, *Interdisciplinary computing in Java programming.* Springer Science & Business Media, 2003, vol. 743.

[49] J. Pagel and P. Kirshtein, *Machine dreaming and consciousness.* Academic Press, 2017.

[50] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: A methodology review," *J. Biomed. Inf.*, vol. 35, no. 5–6, pp. 352–359, 2002. DOI: 10.1016/s1532-0464(03)00034-0. [Online]. Available: https://doi.org/10.1016/s1532-0464(03)00034-0.

[51] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on mri," *Z. Medizinische Phys.*, vol. 29, no. 2, pp. 102–127, 2019. DOI: 10.1016/j.zemedi.2018.11.002. [Online]. Available: https://doi.org/10.1016/j.zemedi.2018.11.002.

[52] M. A. Ebrahimighahnavieh, S. Luo, and R. Chiong, "Deep learning to detect alzheimer's disease from neuroimaging: A systematic literature review," *Computer methods and programs in biomedicine*, vol. 187, p. 105242, 2020.

[53] S. Afzal, M. Maqsood, F. Nazir, *et al.*, "A data augmentation-based framework to handle class imbalance problem for alzheimer's stage detection," *IEEE access*, vol. 7, pp. 115528–115539, 2019.

[54] S. Murugan, C. Venkatesan, M. Sumithra, *et al.*, "Demnet: A deep learning model for early diagnosis of alzheimer diseases and dementia from mr images," *IEEE Access*, vol. 9, pp. 90319–90329, 2021.

[55] S. Hu, W. Yu, Z. Chen, and S. Wang, "Medical image reconstruction using generative adversarial network for alzheimer disease assessment with class-imbalance problem," in *2020 IEEE 6th international conference on computer and communications (ICCC)*, IEEE, 2020, pp. 1323–1327.

[56] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther, "Autoencoding beyond pixels using a learned similarity metric," in *International conference on machine learning*, PMLR, 2016, pp. 1558–1566.

[57] J. M. D. Delgado and L. Oyedele, "Deep learning with small datasets: Using autoencoders to address limited datasets in construction management," *Applied Soft Computing*, vol. 112, p. 107 836, 2021.

[58] A. Kebaili, J. Lapuyade-Lahorgue, and S. Ruan, "Deep learning approaches for data augmentation in medical imaging: A review," *Journal of Imaging*, vol. 9, no. 4, p. 81, 2023.

[59] Z. Dorjsembe, S. Odonchimed, and F. Xiao, "Three-dimensional medical image synthesis with denoising diffusion probabilistic models," in *Medical Imaging with Deep Learning*, 2022.

[60] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," *arXiv preprint arXiv:2010.02502*, 2020.

[61] Z. Kong and W. Ping, "On fast sampling of diffusion probabilistic models," *arXiv preprint arXiv:2106.00132*, 2021.

[62] J. Venugopalan, L. Tong, H. R. Hassanzadeh, and et al., "Multimodal deep learning models for early detection of alzheimer's disease stage," *Sci Rep*, vol. 11, p. 3254, 2021. DOI: 10.1038/s41598-020-74399-w. [Online]. Available: https://doi.org/10.1038/s41598-020-74399-w.

[63] F. Li, D. Cheng, and M. Liu, "Alzheimer's disease classification based on combination of multi-model convolutional networks," in *2017 IEEE International Conference on Imaging Systems and Techniques (IST)*, Beijing, China, 2017, pp. 1–5. DOI: 10.1109/IST.2017.8261566. [Online]. Available: https://doi.org/10.1109/IST.2017.8261566.

[64] *Efficient Alzheimer's disease detection using deep learning technique - Soft Computing — doi.org*, https://doi.org/10.1007/s00500-023-08434-z, [Accessed 23-May-2023].

[65] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile networks and applications*, vol. 19, pp. 171–209, 2014.

[66] M. Maqsood, F. Nazir, U. Khan, *et al.*, "Transfer learning assisted classification and detection of alzheimer's disease stages using 3d mri scans," *Sensors*, vol. 19, no. 11, p. 2645, 2019.

[67] H. A. Helaly, M. Badawy, and A. Y. Haikal, "Deep learning approach for early detection of alzheimer's disease," *Cognitive computation*, pp. 1–17, 2021.

[68] R. A. Hazarika, D. Kandar, and A. K. Maji, "An experimental analysis of different deep learning based models for alzheimer's disease classification using brain magnetic resonance images," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 10, pp. 8576–8598, 2022.

[69] F. U. R. Faisal and G.-R. Kwon, "Automated detection of alzheimer's disease and mild cognitive impairment using whole brain mri," *IEEE Access*, vol. 10, pp. 65 055–65 066, 2022.

[70] S. Bringas, S. Salomón, R. Duque, C. Lage, and J. L. Montaña, "Alzheimer's disease stage identification using deep learning models," *Journal of Biomedical Informatics*, vol. 109, p. 103 514, 2020.

[71] T.-E. Kam, H. Zhang, Z. Jiao, and D. Shen, "Deep learning of static and dynamic brain functional networks for early mci detection," *IEEE transactions on medical imaging*, vol. 39, no. 2, pp. 478–487, 2019.

[72] X. Long, L. Chen, C. Jiang, L. Zhang, and K. Chen, "Disease neuroimaging initiative. prediction and classification of alzheimer disease based on quantification of mri deformation," *Alzheimer's*, vol. 12, e0173372, 2017.

[73] 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: International Conference,BI2017, Beijing, China, November 16-18, 2017, Proceedings*, Springer, 2017, pp. 213–222.

[74] S. Al-Shoukry, T. H. Rassem, and N. M. Makbol, "Alzheimer's diseases detection by using deep learning algorithms: A mini-review," *IEEE Access*, vol. 8, pp. 77 131–77 141, 2020.

[75] A. Khan and S. Zubair, "An improved multi-modal based machine learning approach for the prognosis of alzheimer's disease," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 6, pp. 2688–2706, 2022.

[76] G. Livingston, J. Huntley, A. Sommerlad, *et al.*, "Dementia prevention, intervention, and care: 2020 report of the lancet commission," *The Lancet*, vol. 396, no. 10248, pp. 413–446, 2020.

[77] D. Cheng and M. Liu, "Classification of alzheimer's disease by cascaded convolutional neural networks using pet images," in *Proceedings of the International Workshop on Machine Learning in Medical Imaging*, 2017, pp. 106–113.

[78] M. Liu, D. Cheng, K. Wang, Y. Wang, and A. D. N. Initiative, "Multimodality cascaded convolutional neural networks for alzheimer's disease diagnosis," *Neuroinformatics*, vol. 16, pp. 295–308, 2018.

[79] Y. Chen, H. Jia, Z. Huang, and Y. Xia, "Early identification of alzheimer's disease using an ensemble of 3d convolutional neural networks and magnetic resonance imaging," in *Advances in Brain Inspired Cognitive Systems: 9th International Conference, BICS 2018, Xi'an, China, July 7-8, 2018, Proceedings 9*, Springer, 2018, pp. 303–311.

[80] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in mri data," *IEEE transactions on medical imaging*, vol. 17, no. 1, pp. 87–97, 1998.

[81] A. Sarkar, "Understanding EfficientNet — The most powerful CNN architecture," May 2023. [Online]. Available: https://medium.com/mlearning-ai/understanding-efficientnet-the-most-powerful-cnn-architecture-eaeb40386fad%7D,.

[82] A. N. T, "Inception V3 Model Architecture," *OpenGenus IQ: Computing Expertise Legacy*, Oct. 2021. [Online]. Available: https://iq.opengenus.org/inception-v3-model-architecture/%7D,.

[83] A. Rastogi, "ResNet50 - Dev Genius," Mar. 2022. [Online]. Available: https://blog.devgenius.io/resnet50-6b42934db431%7D,.

[84] GeeksforGeeks, "VGG 16 CNN model," *GeeksforGeeks*, Jan. 2023. [Online]. Available: https://www.geeksforgeeks.org/vgg-16-cnn-model/%7D,.

[85] S. Bangar, "AlexNet Architecture Explained - Siddhesh Bangar - Medium," Jun. 2022. [Online]. Available: https://medium.com/@siddheshb008/alexnet-architecture-explained-b6240c528bd5%7D,.

[86] A. Ahmed, "Architecture of DenseNet-121," *OpenGenus IQ: Computing Expertise Legacy*, Aug. 2021. [Online]. Available: https://iq.opengenus.org/architecture-of-densenet121/#:~:text=Using%20the%20DenseNet-121%20architecture,to%20perform%20the%20convolution%20operation.

[87] *XCeption model and depthwise separable convolutions*, Mar. 2019. [Online]. Available: https://maelfabien.github.io/deeplearning/xception/#the-depthwise-convolution.

[88] C. Bhalerao, "Vision Transformers [ViT]: A very basic introduction," Dec. 2023. [Online]. Available: https://medium.com/@BH_Chinmay/vision-transformers-vit-a-very-basic-introduction-6cd29a7e56f3.