Enhancing Alzheimer's Disease Detection Using MRI Scans Through Transfer Learning Approach

Author

MUHAMMAD ALI AFTAB

MS-20 (EE)

Regn Number

00000327399

Thesis Supervisor

DR. Shahzad Amin Sheikh

DEPARTMENT OF ELECTRICAL ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD, PAKISTAN

July 2024

THESIS ACCEPTANCE CERTIFICATE

It is certified that final copy of MS/MPhil thesis written by Mr. Muhammad Ali Aftab (Registration No. 00000327399) Entry-2020, of (College of E&ME) has been vetted by the undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors, and mistake and is accepted as partial fulfillment for award of MS degree. It is further certified that necessary amendments as pointed out by GEC member of the scholar have also been incorporated in the said thesis.

Signature: Shahzad

Name of Supervisor Dr. Shahzad Amin Sheikh

Date: 25-Jul. 2024

Signature (HoD):

Dr. Qasim Umar Khan

Date: 25-Jul-2024

Signature (Dean) Rashid

Brig Dr. Nasir Rashid

Date: 25 JUL 2024

DEDICATION

This work is dedicated first and foremost to Almighty ALLAH. And then to my beloved parents, for all their love, patience, kindness, and support. Without their encouragement, this project would not have been made possible.

ACKNOWLEDGEMENTS

In the name of Allah, the Most Gracious, the Most Merciful. With the grace of ALLAH Almighty, who empower me to recognize my fatal and making me able to ample this thesis something valuable. I am thankful to my beloved parents and the respected supervisor Dr. Shahzad Amin Shiekh for his most cooperative attitude, valuable comments, constructive suggestions and step-by-step guidance. And to my GEC members Dr. Qasim Umar Khan, Dr. Fahad Mumtaz Malik, and Kamran Aziz. Lastly, I would like to express my sincere appreciation to my sisters and brother Baila Aftab and Kanwal Aftab, and Usman Aftab for helping, encouraging, and supporting me throughout the study.

ABSTRACT

Alzheimer's Disease (AD) is a pervasive neurodegenerative ailment, affecting a vast population globally and posing challenges for early and precise diagnosis within medical image analysis. Although machine learning and deep learning have emerged as competent methodologies for AD detection, several obstacles persist, especially with imbalanced datasets and convolutional effectiveness. This research thesis is using deep learning models empowered with transfer learning to efficiently detect the Alzheimer’s disease classes. More precisely, by employing fine tuning a pre-trained VGG16 and Inception V3 model is investigated for multi-class classification. The study's paramount objective is to enhance AD detection via a custom fine-tuning framework. In this study CNN deep learning models VGG 16 (sequence 1) and Inception V3 (sequence 1 & 2) are proposed for classifying AD in to four stages i.e. Non-demented, Very Mild-demented, Mild-demented and Moderate-demented using brain MRI scans. Then, their performance is evaluated using certain metrics such as accuracy, loss, precision, recall, f1-score, Matthew’s correlation coefficient and balanced accuracy. The results showed that the proposed models, VGG16 (sequence 1) and Inception V3 (sequence 1 & 2) outperformed many of the state-of-the-art models by achieving testing accuracies of 93.9% and (93.87%, 93.09%) for Kaggle MRI dataset.

Table of Contents

LIST OF FIGURES

List of Tables

List of Abbreviations

AD	Alzheimer’s Disease
CNN	Convolutional Neural Network
DL	Deep Learning
ML	Machine Learning
MRI	Magnetic Resonance Imaging
SMOTE	Synthetic Minority Oversampling technique
VGG	Visual Geometry Group

Chapter 1

Introduction

1.1 Background of Study

Alzheimer's disease is a neurodegenerative ailment which is terminal and progresses over time. But this progression of neural damage can be prevented by detecting this syndrome at an early stage [1, 2]. The rapid and accurate determination of (AD) based on structural (MRI) has triggered significant interest among researchers, driven by deep learning techniques due to its satisfactory implementation in medical image analysis [3]. Deep learning techniques includes many techniques such as CNN, RNN, DRN and many more [4, 5]. In this project, a convolutional neural network is implemented which is pre-dominantly popular and gives more precise results.

As per the investigation the major source behind dementia is Alzheimer (AD), almost 60% to 80% of the cases that were reported under dementia accounted as AD. It affects not only the brain tissues but causes gradual memory loss, impairment in thought process and decision-making skills, and even impedes the daily routine activities [6]. As per the records of an international association (ADI) for Alzheimer, over 50 million people around the globe are suffering from this disease and a gradual increase is estimated by the end of 2050. Whereas, by that time the number of patients is expected to triple, marking it to reach 152 million patients, meaning that every 3 seconds, someone will be diagnosed with dementia. The anticipated cost of dementia is $1 trillion that is to be expected to upsurge twice the current amount by 2030 [7].

The unavailability of a significant effective treatment for Alzheimer is one of the potential hurdles faced by the professionals and medics till this date. For early AD diagnosis

Machine learning techniques are employed which is typically a traditional approach of feature extraction, but that too is complex, time consuming, and demands greater technical expertise [9]. Wherein these feature extractions are majorly categorized in two types, i.e. ROIs (region of interest) and voxel-based features [8]. Collectively, these conventional methods are not found to be effective in diagnosing and treating Alzheimer.

Wherein deep learning CNNs are a potential opportunity that offers an effective solution to traditional AD diagnosis challenges. Convolutional neural networks boost efficiency by automatically extracting features from data, eliminating the need for manual feature engineering [10]. This has led to promising results in AD diagnosis, offering potential for improved accuracy and efficiency.

A framework focusing on (CNN) is established for an end-to-end AD detection and classification. Specifically transfer learning and multitask learning approach are being used. Different stages of AD would be multi classified. Two methods, “Transfer learning and multi-task learning” through a pretrained VGG model would be used. After detecting AD stages, precautionary measures could be advised according to its AD stage.

1.2 Problem Statement

Alzheimer's Disease (AD) is a pervasive neurodegenerative ailment, affecting a vast population globally and posing challenges for early and precise diagnosis within medical image analysis. Several obstacles persist, especially with imbalanced datasets and inter-model variability, early-stage detection, multi-classification, convolutional effectiveness, and accuracy [11, 12]. So, the current study is extending the deep learning and transfer learning model as it has emerged as one of the competent methodologies for AD detection [13].

Moreover, based on existing gaps in literature this research focuses on utilizing transfer learning through a pre-trained VGG16 model [8, 14], Inception-V3 model, fine-tuning, and multi-task learning to classify MRI scans for AD diagnosis [15, 16, 8, 17] for early, precise and effective Alzheimer detection through [18, 19, 20, 21]. Wherein the dataset that would be used is an Open source Kaggle dataset consisting of 6400 mri images for four classes [22]. Comprehensively, the study's paramount objective is to detect and enhance AD progression via a transfer learning framework.

1.3 Significance of the Study

The aim of this study is to add valuable insights to the existing body of knowledge in the academy and have real world, practical applications. So, theoretically, it will add new valuable knowledge to the existing body of literature. By exploring the implications of CNN models to detect AD stages. Wherein, it could be a fundamental study in Pakistan with respect to policy implications regarding AD diagnosis. Where it can provide some contextual guidelines for local and national development. Henceforth practically, it has potential applications in numerous fields ranging from medical field to AI, Neuroimaging, machine learning, and public health.

1.4 Research objectives

The goal of this research is to highlight how the VGG16 model's inherent knowledge enables extracting pertinent features from medical imagery, while fine-tuning refines its AD detection capabilities. The holistic approach promises not just to identify AD but to also shed light on its progression, fostering advanced detection strategies and offering invaluable insights for patient-specific trajectories. Wherein, the list of core objectives of this research is given below.

1. To detect and analyze the stages of AD through simple transfer learning approach with respect to VGG16 and Inception V3 models
2. To enhance the detection of AD by introducing a custom fine-tuning framework concerning VGG16 and Inception V3
3. To compare the efficiency of VGG16 and inception V3 models through performance metrics (categorial accuracy class, AUC, balanced accuracy, and Mathew's correlation coefficient, loss, confusion matrices, precision, F1 score, and recall)

1.5 Research Questions

RQ 1: How transfer learning approach through CNN enhances the detection of AD as compared to traditional machine learning approaches?

RQ 2: Which CNN model (VGG16 or Inception V3) is effective in detecting multi stages of AD?

RQ 3: To what extent does a custom fine-tuning framework enhances the accuracy of AD detection?

1.6 Thesis Outline

The contents of this research thesis consist of six chapters in total. Starting with the Introduction as Chapter 1 that sheds light on the background of Alzhiermer disease, prevailing challenges, and deep learning as an opportunity to diagnose AD efficiently. Chapter 2 is about the literature review which presents a detailed account of an already existing body of knowledge and previous studies in this regard. This chapter reviews the existing analysis situs about deep learning approaches and models used to diagnose AD along with the drawbacks and limitations. Which lays the foundation for Chapter 3 named as Proposed design of the study. This chapter is all about the dataset description like strategy of performing tests, selected dataset and details of model's classification that how VGG16 and Inception-V3 models are being handled throughout to achieve the desired results. Wherein, Chapter 4 presents the detailed discussion of results and Analysis of proposed methodology, accuracy rates achieved, and visualization of the findings. This Chapter is all about discussion of the presented results and findings to back them up from the literature chapter to either validate them or to provide new perspectives regarding the field of study. Lastly, Chapter 5 covers the holistic conclusion of the whole thesis from proposed objectives till depicted inferences. Moreover, it entails the limitations of the study along with the future recommendations as well.

Chapter 2

Literature review

This section covers the concept of Alzheimer disease, its symptoms, types of AD, approaches to diagnose AD, that how machine learning and deep leaning methodologies assist the early and effective AD diagnosis and its contribution towards the said field of analysis. It also includes comprehensive details regarding the models being studied, that shows how different approaches contributes the efficient and progressive AD diagnosis. The researcher considering the current literature develops an understanding regarding the list of metrics in detecting AD through CNN.

2.1 Alzheimer's disease (AD)

The systematic functioning of a human body relies upon one of the fundamental organs i.e. the brain. Which is responsible for numerous operational activities, for instance decision-making, critical and analytical thinking, information retention, and keeping record or memory. The whole concept of attaining, perceiving, and retaining information, life experiences, and knowledge restores into memory blocks which enables a human to recognize the environment and circumstances around him/her [23]. Any hinderance in this process causes several disorders like memory loss, inability to recall and identify things or even people. The disorder of dementia resonates with the mentioned inabilities, a human brain goes through, more precisely the Alzheimer disease, a common form of degenerative memory disorder where the patient at early stages suffers from Mild cognitive impairment (MCI) and forgets to recall the recorded information and at advanced stages lose the connection with the closed family members and even forgets the basic instincts like swallowing, breathing, sneezing etc. [23, 24, 25, 26]. Henceforth, the patients who have the symptoms of MCI are more prone to Alzheimer, making this ailment a multi-stage progressive neural deterioration [27].

Consequently, this neurological disorder is becoming that common that half a million people around the globe is suffering from this disease which is further expected to surplus with the addition of 100 million cases by the end of 2050 and the expenditure on health care facilities accounts equivalent to the 18^{th} largest economy of the world [28, 29, 30]. However, despite numerous efforts made towards its effective and accurate diagnosis, still there are complexities in early-stage detection due to its corresponding similarities with aging and other forms of dementia [31]. Like, one of the closest disorders resembling AD is the Vascular dementia (a syndrome where the tissue of brain gets weaken due to vascular disease leading to strokes etc.) making it difficult to categories the AD specifically [30, 32]. Wherein for this purpose most of the studies has put emphasis on Magnetic Resonance Image (MRI) technique to account the size and number of cells, because the early detection is crucial for effective prevention and treatment. Nonetheless, the prompt detection remains ambiguous and challenging requiring further empirical investigation [30, 31].

2.2 Types of Alzheimer Disease

AD is mainly categorized in 4 progressive types, starting with the premature stage of non-dementia to early stages of very-mild and mild dementia advancing towards the moderate dementia [33, 34]. These categories are being highlighted in a recent study in a more comprehensive manner. Starting with the “Preclinical stage”, which accounts for trivial memory problems like difficulty in remembering small details or sometimes the patients at this stage do not show any specific symptoms of dementia. The second stage is “MCI (mild cognitive impairment)”, dictating more visible indications of memory dysfunction but to the extent where patients can execute daily routine activities. While the intensity increases in the third stage of "Mild dementia", when the affected individuals start facing complexity in fulfilling routine tasks and even communicating efficiently due to uncertainty and bafflement. Last two stages "Moderate and Severe dementia" accounts for the advance progressive form of dementia where the severity intensifies to the level where the patient not only suffers from memory deterioration but forget about loved ones, family, communication skills and becomes completely dependent on caretakers [35, 36].

The cruciality and condition of AD symptoms would be indicating the specific stage, but because of lack of accuracy in primary discovery of AD in patients, the identification of early stages becomes uncertain [33]. This in surn makes the process of diagnosis challenging resulting in inefficient treatment that compromises the chances of early preventions from progression as the AD is more receptive to preliminary handling [37] and advances with the passage of time [38]. There are three streams of diagnosis in this regard, mainly one is conducted by neural psychologists for minor initial stages, clinical examination along with non-automated assessment for advanced analysis. Nevertheless, the obstructions in accounting AD at premature stages are still prevailing, one of the reasons of this uncertainty is the familiarity of AD at initial stages with the conventional neural diseases like temporary memory loss or difficulty with the usage of language [36, 39, 40]. Holistically, as discussed by multiple studies, despite having technological advancement in the current time of artificial intelligence there is absence of a reliable and competent framework to curb the spread of this disease.

2.3 Approaches to diagnose AD

One of the prevailing approaches to diagnose Alzheimer is a cognitive or neural imaging procedure, through which it can be diagnosed way before any evident signs/symptoms [41, 42]. For instance, the earliest stage of "Preclinical dementia" can also be identified, whose duration varies to years, for example it can last up to more or less than ten years. So, how this premature form of AD can be diagnosed through neural imaging method? By utilizing advanced imaging equipment to detect and analyze a specific protein “amyloid beta deposit”. The drop in its levels is responsible for Alzheimer regardless of the invisibility of any indication [43]. Henceforth, the development of such technologies would be useful in curtailing severe repercussions from cognitive dysfunction because of its non-intrusive examination. However, the efficacy of this diagnostic approach progressed over time with constant technological advancement. Previously, the threatening illness of AD can only be detected after the demise of patient but the recent development in artificial intelligence equipped the medical experts with multiple methods not only for the diagnosis but for the medication as well in form of neural imaging procedures [44]. These procedures include “magnetic resonance imaging (MRI), Positron emission tomography (PET), Functional magnetic resonance imaging (fMRI), and Computed tomography (CT) [38, 45].

A recent study identified three methods for an effective and efficient diagnosis of AD. Those approaches are mainly (i a traditional systematic inquiry (ii clinical biomarker (iii neuroimaging sensory systems (just like MRI). Wherein the first approach is inefficient and the main source of gathering information is through mainstream technique of manual assessment. Which is its biggest drawback, reason being that the results get influenced by the subjective nature of the procedure, maximizing the probability of error [46, 47]. The second method emphasizes the diagnosis through examining a protein named “amyloid-beta” which is directly linked to the dysfunction of brain. However, the methodology is objective in nature but due to its extensive and demanding techniques it is not admired as a regular strategic framework for premature AD diagnosis [48, 46]. So, the third technique of visualizing brain imaging is an effective technique as compared to others due to its capability of providing visualized in-depth brain analysis. This technique thoroughly examines the structure of the brain, its shape, tissue, substances/chemicals etc. Any slightest change can be visualized through this procedure to indicate the presence of Alzheimer [23]. The advantageous edge of MRI improves when used with other neuro-cognitive equipment as it helps to classify AD from other brain related syndrome/dementia [49, 50, 36]. However, along with opportunities MRI technique has certain limitations/shortcomings like data gathering is easy but the interpretation and comprehension of that visual data/image is complex [48, 47].

2.4 Deep and Machine Learning as an Approach

In addition to these available techniques there is an evident shift towards the implication of machine and deep learning methods. The advancement in brain imaging technology has emphasized the efficacy of deep and machine learning approaches in terms of identifying and treatment of AD. The precision in diagnosis and predictions with these approaches increases as they are believed to acquire precise information accurately [23, 51]. Both the approaches have pros and cons accordingly for instance machine learning methodology is more feasible because of its open and easily accessible database from which dataset for research and study purposes can be retrieved [52]. Whereas DL approach is more favourable because of its precision in results which led its applicability in medical sciences as well [46] specifically in context of Alzheimer diagnosis. The utilization of DL in AD is getting popular since 2013 when a study examined the progression of Alzheimer disease with the help of "stacked auto-encoder and support vector machine classifiers" [53]. Though, there is still ambiguity regarding which approach is viable for AD. Henceforth, there is need to analyse the body of knowledge to assess the effectiveness of deep learning as compared to machine learning approach.

2.4.1 Machine learning Approach

Currently AI is progressing immensely and affecting every field, may it be smart gadgets, programming tools, entertainment sector or medical science. This progression contributed to the medical field by developing innovative machine learning equipment for forecasting diseases. Where machine learning techniques have been employed in form of multi-classifiers, SVM etc. [38, 54] by following three stages i) identifying the brain's ROIs ii) characteristics selected from ROIs iii) formulating and assessing classifiers (models), to detect AD and acquire decent results. The main aim of such modules was to formulate a structural framework to examine brain for any abnormalities, dysfunction, or defect and furthermore to identify the AD symptoms from other brain related issues. In accordance with it, different versions of diagnostic models have been formulated and tested. Such as an approach of multi-classification was introduced namely "Inherent Structure-based Multiview Learning" in which the function of feature selection was enhanced by stratifying voxels and segmenting the brain tissue (white and grey) and applied on MRI baseline data with 93.83% accuracy [55]. The experiment showed a successful precision rate with less margin of error but still needed improvement to cover this margin.

For this purpose, another study projected the same concept of multi-layer model based on the notion of "fuzzy logic" programmed on MRI and PET. The functioning of the model was categorized in three steps, the first step accounts for the preliminary management work where the white and grey matter were being fragment and passive voxels were assorted. The second stage narrates the collection and positioning of features to lessen the amount of ROI and last stage was regarding the fuzzy classification. The overall functioning was observed under AUC (receiver operating characteristics). Consequently, the results showed only 89.59 percent accuracy -rate of this combined module [56]. To further check the precision, a "computer-aided diagnosis" and image processing model were introduced specifically for

Alzheimer. The fundamental purpose of these systematic frameworks was to recognize and classify the AD patients and normal ones. The procedure follows the same methodology of fragmentation/segmentation and multi-classification [57, 58, 23]. However, instead of decreasing the margin of error increased as the accuracy rate was between 73 to 75 percent for MRI and PET which was less than previously mentioned approaches [57].

Holistically, the recorded results highlighted the biggest hurdle experts faced regarding the utilization of traditional learning approaches for AD was manual data gathering and comprehension that consequently devalues the model's functioning. To address this shortcoming So, to enhance the competency of the outcome experts shifted their interest to Deep learning approach as an alternative strategy in diagnosing AD [23].

2.4.2 Deep Learning and Alzheimer Diagnosis

The loophole in ML modules' application can be taken over by employing deep learning models because of their efficacy in reading, detecting, and analyzing smallest change in brain functioning. Moreover, the simulation possesses the capability to diagnose Alzheimer disease with maximum precision in results as compared to the traditional mainstream techniques [23]. As advocated by certain studies this approach is viable and has limited variations, for instance CNN (convolutional neural networks) is contemplated to be the most popular and used version. These systematic frameworks are trained to work on 2-dimentional image processing however, it can also work on 1-dimentional or even extent to 3-dimentional data. The fundamental aspect required to run such neural networks efficiently, is to train them on large set of data [38, 59]. As evident in studies conducted to explore the effective technique to treat AD, the module developed on the basis of deep learning approach under which multi-sensory procedures were involved like SoftMax logistic regressor and the end result showed approximately 91 percent success rate [60].

Wherein on the other hand functionality of DL was estimated by including three dimensional convolutional networks along with SoftMax to scan brain through MRI [61]. For a more advanced approach, instead of implementing simplified Support vector machine technique, a progressive strategy of trained CNN proves to be more efficient, reliable, and accurate. As advocated in 2017 study, results from experimenting with SVM technique show only 84.4 percent precision whereas with CNN deep learning approach the exactitude reaches 96 percent. Which proves the proposed assumption that deep learning neural networks work the best in detecting and treating early and progressive stage Alzheimer [62]. As mentioned earlier Machine learning techniques are not consistent when it comes to feature selection and distinguishing AD from other brain related issues/dementia. To fulfil this gap a CNN model was programmed to classify AD mechanically that was utilized along with different systems like AlexNet for selecting feature, PCA (principal component analysis) for selecting feature in a sequence and SVM for classification and the obtained percentage was 90% [63]. In addition to it, similar method was investigated but with fine-tuning framework which achieved 91.7% accuracy level [64].

Further exploration of existing body of literature underlined another research [65] where the three-dimensional CNN network was programmed by organizing multiple layers for different purposes, among these layers 5 were trained for the selection of features and 3 interlinked layers were used for classifying AD. To investigate the core elements behind the higher functioning of the model, four components were assessed mainly hyper parameters for selection, preliminary processing, data segmentation, and the proportion of dataset. wherein the module was pre-trained with the 60 percent MRI images in training segment and 20 percent data was used in validation stage and finally for testing stage remaining 20 percent data was consumed. After evaluation the model acquired nearly 99% accuracy (98.74%) which ultimately indicates the higher level of efficacy in detecting and curing AD. Similar results of “98.59% for AD vs. NC, 97.65% for AD vs. MCI” were encouraged in [66] as well but with the different conceptual technique where the experts divided the dataset in to three sub-sets and then changed the size of the images to 224 x 224 while assembling them in groups. Afterwards 20 slices out of these assembled groups were assigned to train the module, three classifiers (ResNet, NASNet, MobileNet) were used along with collective learning technique to boost and improve the process. In contrast to such accuracy a for ResNet framework, a more recent study [67] concluded that an improved Resnet-50 module along with Soft NMS achieved lesser accuracy rate of 84.3% overall.

Moreover, this convolutional approach was again experimented in terms of resizing the images of dataset, inclusion of 3 layers where after each layer a max pooling coating was applied for two-fold/binary categorization of Alzheimer’s Disease. To contemplate the efficiency of this model, the performance was assessed based on 3 testing trials with different sizes of images mainly standardized to 128 x 128 and 64 x 64 with option of with or without dropout. Henceforth, the result showed the maximum accuracy with the 128 x 128 size but without dropout. To further check the reliability and validity of the results a cross-validation was performed with the set range of 0.1 to 0.5. Which indicated that the precision rate is directly proportional to the set batch range/size but with one condition that batch size should not exceed more than 64 it is when the accuracy would be decreasing. The end results regarding the model’s functioning in terms of binary classification reached to 95.6 percent accuracy [38, 68].

A similar model with few variations was investigated and the level of accuracy exceeded nearly 100% by attaining 99.5% precision after evaluation. The variations introduced were regarding the image size and activation operation. Where the dataset was chosen from OASIS instead of ADNI with the 200 x 200 image and 0.2 dropouts to encounter the overcompensating problem. For higher performance rate densely compacted units (which were 121) were placed within layers while training the systematic module Al-Khuzaie et al. So, to compare the two versions of CNN deep neural networks, one study contrasted ResNet with AlexNet to contemplate the difference between their performances and to explore which version would work perfectly. The former model consisted of 177 layers in total and five pooling layers with 5 x 3 whereas the later model was assigned 34 along with same 5 pooling layers with the 4 x 4 size image. Out of the dataset eighty percent was utilized in giving instructions and training the model while the evaluation was done with the rest of twenty percent. Accumulatively the outcome suggested that the AlexNet showed 94.5 % precision and works the best in diagnosing AD [69].

For a more comprehensive understanding a recent study explored and analysed 29 models under convolutional networks of deep learning where the final inferences put emphasis on the efficacy of EfficientNet modules whether in terms of pretraining strategy or comparative analysis regarding Alzheimer diagnosis, the precision reached 94 to 97 percent [70]. In line with the above-mentioned comparison, other two modules i.e. VGG16 and VGG19 were being compared regarding the treatment and detection of Alzheimer. Prior to the training stage the brain was augmented, and the image was of 224 x 224 size. Then for first mentioned model sigmoid with 64 and 128 filter kernel sizes was used while for the second softmax with an addition 256 size was used for the purpose of activation. Both versions were not proved to be optimal in diagnosing AD accurately, for instance the accuracy rate for VGG16 was 81 percent and for VGG19 was 84 percent only [71]. In contrast to these levels a latest study highlighted higher precision rates of VGG16 and 19 models which were pre-trained and fine-tuned. They acquired 97 percent accuracy in diagnosing Alzheimer and its progressive stages. However, standard VGG module is not quite efficient with respect to different datasets for instance according to [72] the accuracy rates for standard VGG 16 are fluctuating and not promising as the study investigate VGG16 with 2 datasets with minor changes, where it accomplished 90.4% for dataset 1 and 71.4% for dataset 2. So, VGG is considered to be appropriate for routine usage because of its simplified features with less complicated computational tasks, overfitting issue and adherence to minimal memory usage along with progressive/temporal adaptation [38, 24]. Conclusively, based on these presented facts there is still a need to further explore and investigate the effectiveness of VGG models specifically while dealing with multi-classification to diagnose AD.

Because standard conventional frameworks are not proved to be effective in their predictions. For instance, as per studies in literature like [4, 73] standard version of Inception-V4 and Landmark-based extraction did not achieve decent accuracy rates (73.75% for Inception-V4 and 79.02% for landmark-based extraction. The stance is reinsured by [74] where the investigators performed AD detection on Kaggle Dataset which contained four classes of AD severity. The models they used were DenseNet-169 and VGG-19 with the accuracy fluctuating between 80% to 82%. In line with this [33] utilized open-source dataset for DenseNet-169 and ResNet-50 while achieving 88% accuracy for DenseNet-169 and 82% for ResNet-50. In this regard, few studies have investigated certain approaches and proposed custom alterations according to the achieved accuracy. Such as a study [17] investigated ResNet-50 with different classifiers (like Softmax, SVM, and RF) highlighting accuracy rate of (99%-96%) for softmax, (92%-90%) for SVM, and (85.7%-84.8%) for RF. Wherein a more recent study working with ADNI dataset anticipated to project an improved ResNet-50 with an overall 84.37% [67].

2.5 Summary of the Study's Base Papers

Following is the summary of the base papers utilized as a foundational basis for this thesis.

Table 1

Summary of base papers for the current study

Papers	Year	Dataset Used	Techniques	Achieved Accuracy	Gaps
Islam et al. (Springer)	**2017**	OASIS	• Inception- V4	• 73.75%	• Inefficient accuracy rate
Zhang et al. (IEEE)	**2017**	Longitudinal MRI Scans	• Landmark-based extraction	• 79.02%	• Landmark-based feature extraction is inefficient
Pradhan, A. et al. *(IJERT)*	**2021**	Kaggle Dataset	• DenseNet-169 • VGG19	• Fluctuating between 80% t0 82.6%	• Lesser Accuracy rates
Ghazal et al. (*Computers, Materials & Continua*)	**2021**	Kaggle	• AlexNet Fine Tuned	• 91.7%	• All convolutional layers are not fine-tuned
Sharma et al. *(Front Comput Neurosci)*	**2021**	Kaggle	• Standard VGG16 • (Dataset 1 & 2)	• 90.4% Dataset 1 • 71.1% Dataset 2	• Computational complexity • (14.7M trainable parameters)
Al Shehri W. *(PeerJ Computer Science)*	**2022**	Open-source dataset	• ResNet-50 • DenseNet-169	• 82% for ResNet-50 88% for DenseNet-169	• Lesser Testing accuracy • Different measures needed to detect the system's accuracy
AlSaeed, D. et al. (Sensors)	**2022**	ADNI, MIRIAD (MRI datasets)	• ResNet50-Softmax • ResNet50-SVM • ResNet50-RF • Multiple classifiers used	• Softmax (99%-96%) • SVM (92%-90%) • RF (85.7%-84.8%)	• Small datasets with only 750 subjects approx.
Helaly, H.A. et al (Cogn Comput)	**2022**	ADNI Dataset	• 2D and 3D CNN architectures used	• 2D CNN (93.6%) • 3D CNN (95.1%)	• Small datasets with only 300 subjects

			• Fine tuning using pre-trained VGG19	• VGG19 (97%)	
Yusi C. et al. ***(JRRAS)***	**2024**	ADNI1 Dataset	• Soft NMS • Improved RESNET-50	• 84.37% overall	• Unlabeled dataset affecting efficacy
Alsubaie, M.G. et al. ***ML Knowl Extr***	**2024**	(CNNs, RNNs, and GANs) from 2018 to 2024	• Comparative analysis. • Review Survey (CNNs, RNNs, and GANs) from 2018 to 2024		• Advancements and refining in model architectures and training methodologies are required to enhance the generalizability

2.6 Research Gaps

Based on the above-mentioned review of previous studies, following research gaps in literature are being highlighted.

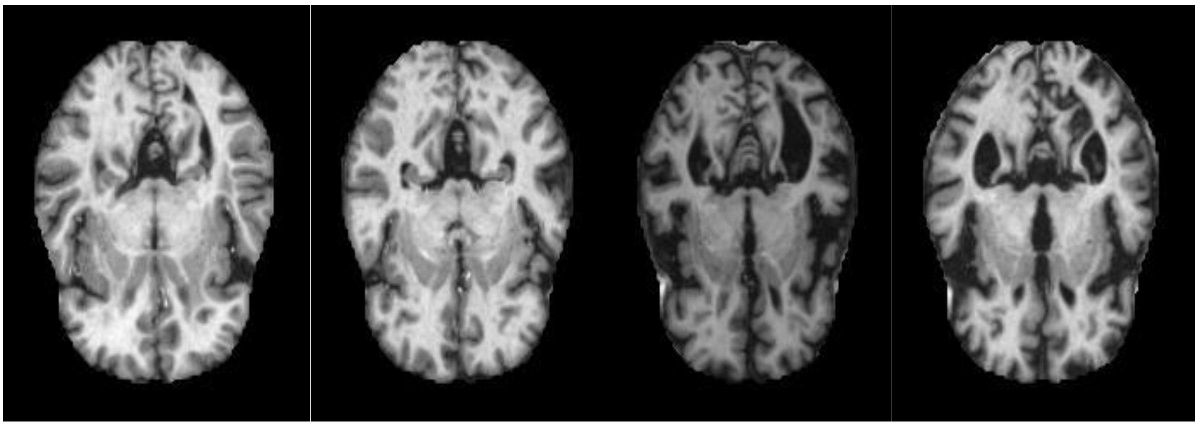
1. Imprecise convolutional effectiveness
2. Classification of limited AD classes
3. Unlabeled datasets result in inefficient accuracy rates.
4. Computational complexity with high number of trainable parameters.

Chapter 3

Proposed Methodology

3.1 Data Set

The focus of this study is to enhance the classification of AD through transfer learning approach with respect to VGG-16 and Inception-V3 models. For this purpose, the methodology of the current research relies on different stages like pre-processing, classification, and application of multiple strategies. To start with these strategies the first step is to select the dataset, so the dataset that was being used is an open source Kaggle dataset which contains a total of around 6400 MRI images, each segregated into the severity of Alzheimer's Classes i.e (Non-Demented – 3200 images , Very-Mild-Demented – 2240 images , Mild-Demented – 896 images and Moderate-Demeneted – 64 images) as shown in Figure 1 below.



Non-Demented Very-Mild Demented Mild Demented Moderated Demented

Figure 1: Segregation of Alzheimer's Classes based on severity

3.2 Synthetic Minority Oversampling Technique (SMOTE)

After the sample's selection and importation, pre-processing was done and dataset that includes Refining and pre-processing of dataset Image Normalization, resizing, reducing class imbalancement. Later, 2D images of 176 X 176 size were used. There are different techniques available for over sampling, which use random sampling to stack the copies of minority class to make it equivalent to the majority class by increasing the copies. However, this technique is not effective, so another approach (SMOTE) is established for this purpose. Therefore, the technique used for reducing class imbalancement in this research, is synthetic minority oversampling technique (SMOTE) as it is one of the efficient techniques being used for balancing dataset [75]. This technique is an algorithm that generates samples from minority class to deals with the imbalacement. Wherein at first the minority class with limited occurrences was being identified by this synthetic algorithm, later it detects the k-nearest neighbors within the identified minority class. In general, to assess and evaluate the similarity Euclidean distance is used. Which in turn assist in interpolating between these k-nearest neighbors and chosen samples by creating synthetic samples through random selection of one of the k-nearest neighbors. To achieve the desired balance repeat and redo the steps 2-4, as shown in Figure 2.

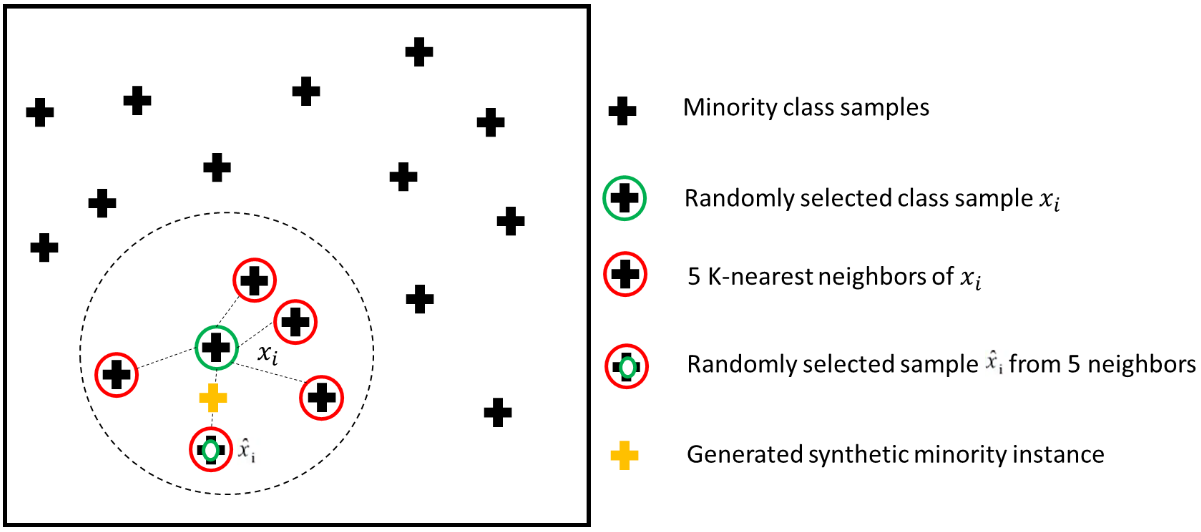


Figure 2: Steps to generate Synthetic samples by suing SMOTE

Ultimately, to attain a balanced dataset, SMOTE technique combines the original minority class samples with the generated synthetic samples. As evident in Table 2 and Figure 3 & 4 below the class imbalancement is greatly reduced and the distribution is in a balanced form.

Table 2

Comparison of Class distribution before and after SMOTE

AD Classes	Before SMOTE		After SMOTE	
	Training	Testing	Training	Testing
Non-Demented	2560	640	2036	654
Very Mild Demented	1792	448	2061	628
Mild Demented	717	179	2056	647
Moderate Demented	52	12	2039	631

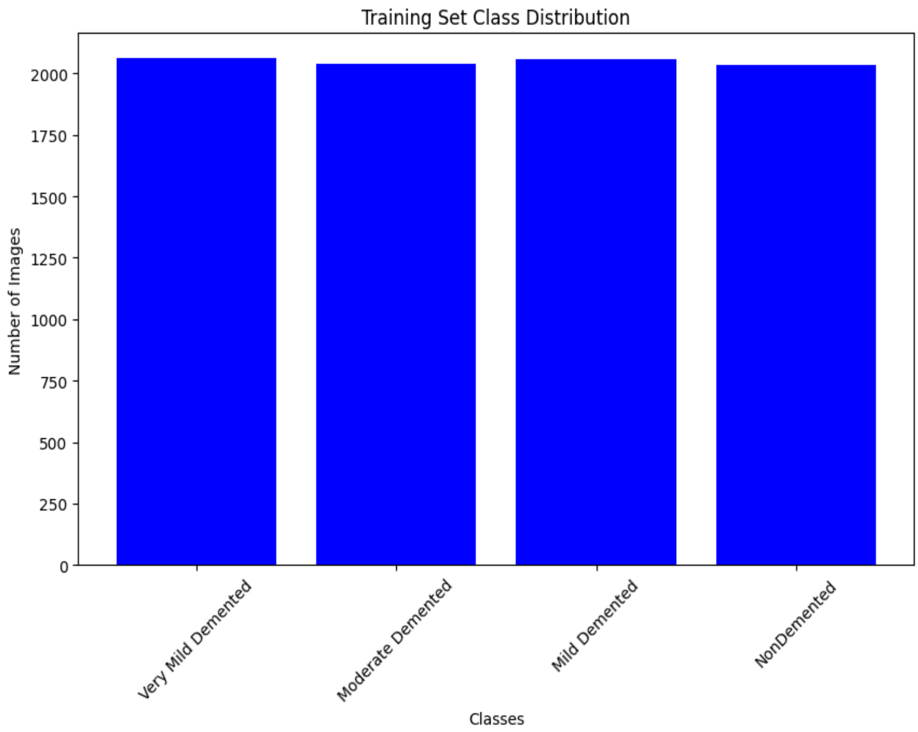


Figure 3: Cass distribution of training images after SMOTE

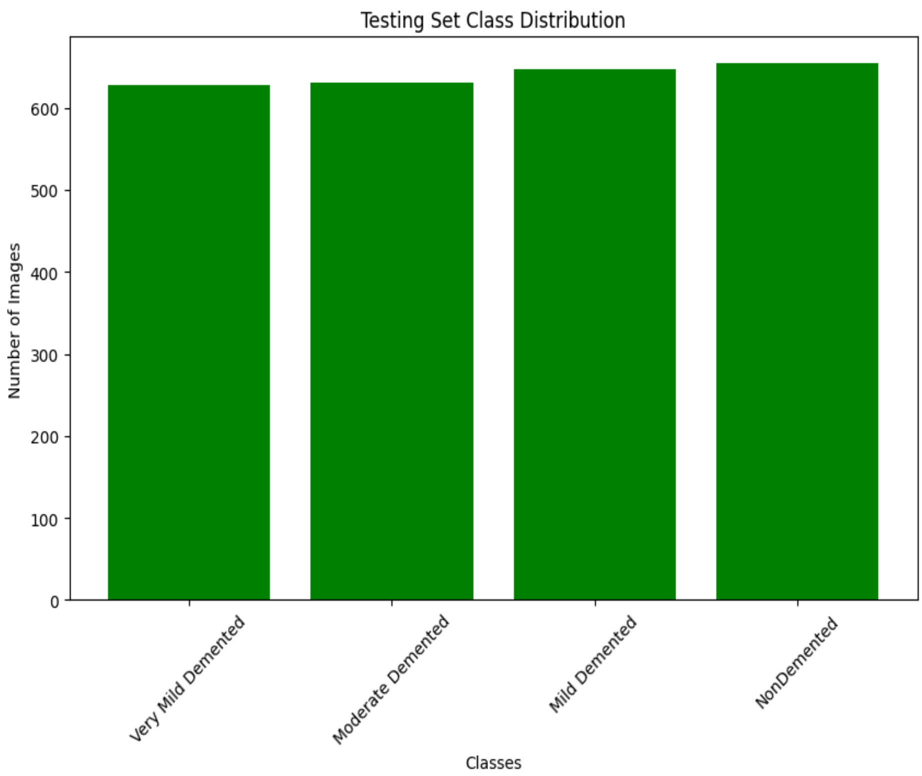


Figure 4: Class distribution of testing images after SMOTE

3.3 Image Classification

Henceforth, after applying smote our data samples were balanced and increased from 6400 to 12800 for 4 classes. And the dataset of this study was ready to be given to the models (VGG-16, Inception-V3) for classification. The data is segmented into train, test, and validation sets. In addition to this process, the layers of VGG-16 and Inception-V3 models were freeze and used two approaches, mainly:

1. Firstly, medical image classification using standard models is performed.
2. Secondly, the researcher performed this similar classification (of medical image) by using simple fine tuning.
3. Thirdly, the same process was done by introducing custom fine-tuning frameworks.

3.3.1 Custom fine-tuning Framework

The addition of following layers in custom fine-tuning has improved the efficiency of the selected models:

- Dropout (reduces overfitting)
- Global average pooling 2D (reduced number of parameters)
- Flatten (prepares data for dense layer)
- Batch Normalization (improves training speed)
- Dense (classification)

In custom fine-tuning framework, firstly 3D activation feature maps are generated by using pre-trained models. Then the Dropout, Global Average pooling 2D, and Batch Normalization layers are used to reduce the computational complexity, for conversion of activation maps into 1D vector, and finally to convert the feature vector into a normal distribution as shown in Figure 5.

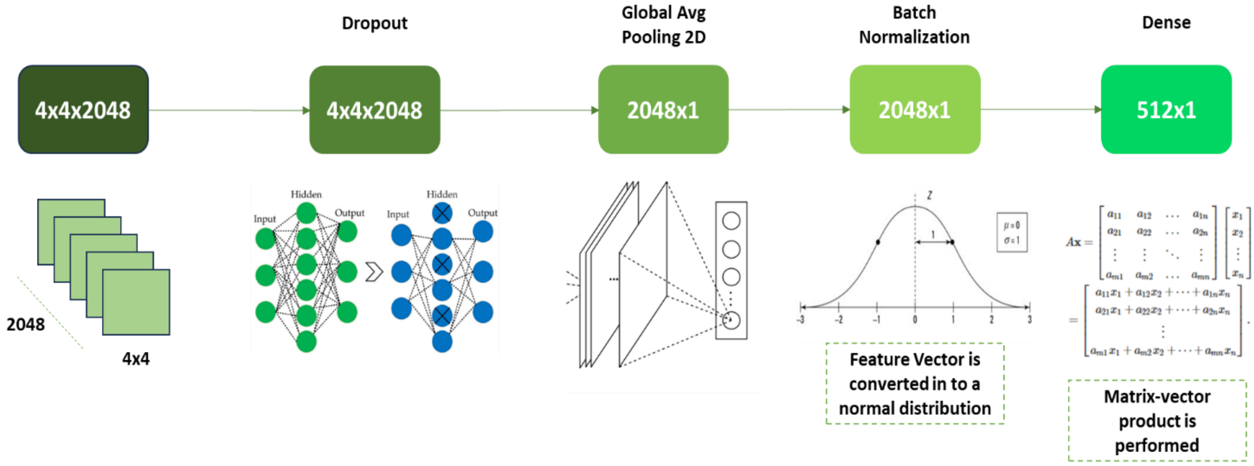


Figure 5: Functioning of layers in custom fine-tuning framework

In continuation of the above-mentioned figure, the remaining steps of feature reduction are performed. For instance, the feature reduction is applied to the 1D vector using a set of five dense layers as evident in the following Figure 6.

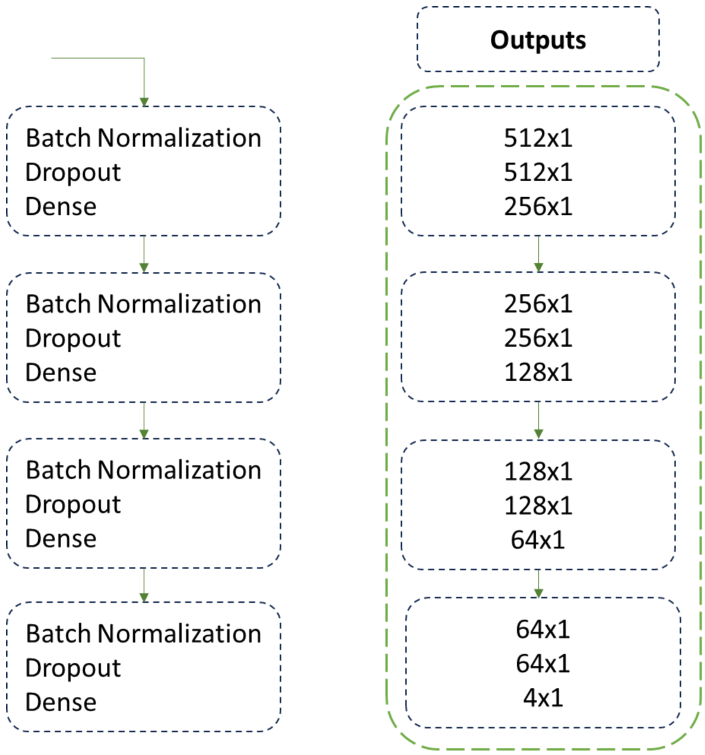


Figure 6: Feature reduction over 1D vector using a set of five dense layers

By incorporating these layers, the functioning and efficiency of the Deep Learning models (VGG16, Inception V3) for image classification tasks have improved massively. After the layer's addition, layers are then trained on the target data. AD is classified into 4 classes and test image is given to the output of the model to calculate testing accuracy of predictions.

For feature reduction different approaches for combination of dense layers are used. For example, 1/8th of previous dense layer for the next layer, similarly, 1/4th and 1/2nd of the previous layer for the next dense layer is used. Out of these layers, two sequences are established for custom fine-tuning frameworks. Wherein sequence 1 is the combination of 1/4th and 1/2nd sequences, likewise, sequence 2 is established by combining 1/8th and 1/2nd sequences. Both sequences are presented in the following Figure 7 and Figure 8.

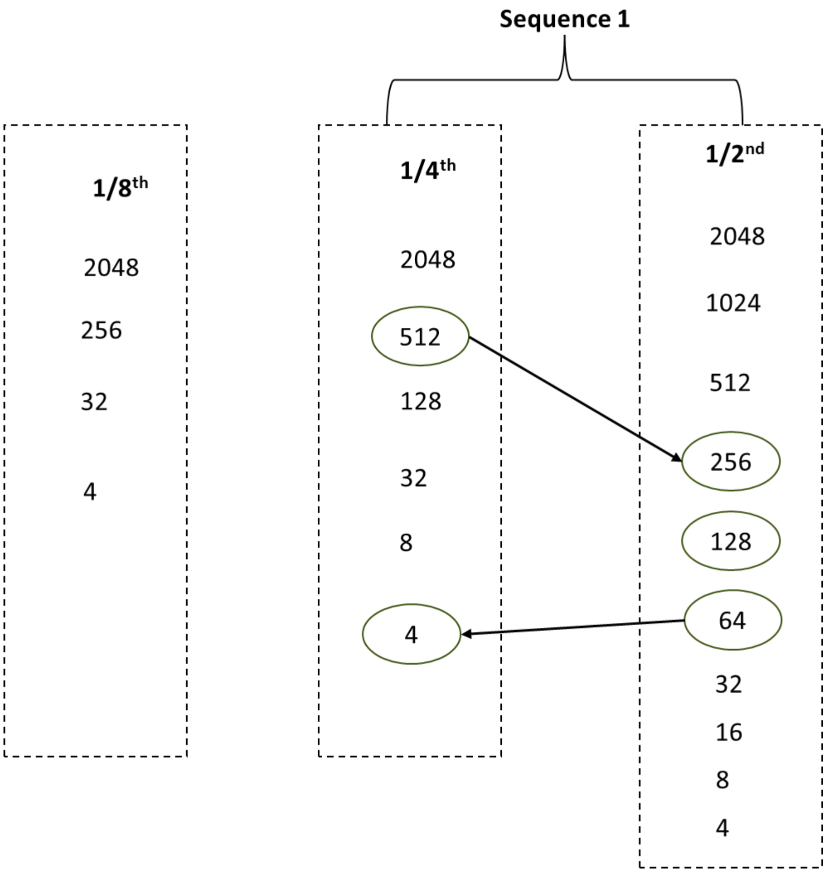


Figure 7: The combination of 1/4th and 1/2nd sequences for Sequence 1

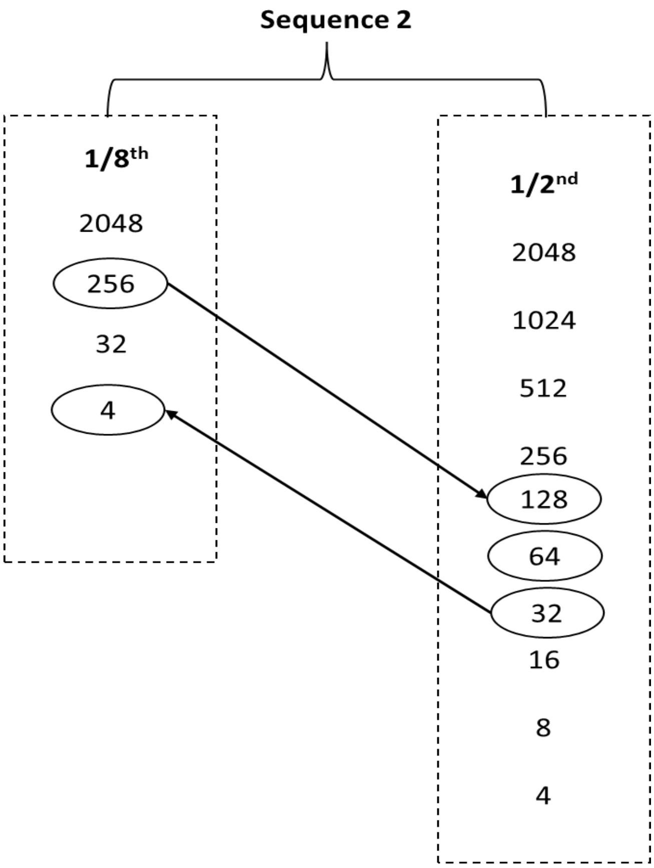


Figure 8: The combination of 1/8th and 1/2nd sequences for Sequence 2

3.4 Performance Metrics

Following performance metrics would be used to compare the efficiency of the said models (VGG16 and Inception V3):

- Validation accuracy and loss
- Confusion matrix
- F1-score
- Categorical accuracy class
- Area under curve (AUC)
- Balanced accuracy
- Mathews's correlation coefficient (MCC)

Since the transfer learning approach is being performed, the pretrained weights from IMAGENET are being incorporated, so that the transfer of information from similar classification tasks can take place. A custom call back function is integrated to stop training our model when accuracy is 99% or more. So, a technique is applied to intervene and slow-down the learning process when a particular metric would not be further improving for longer than the allowed number of patients. Henceforth, the learning rate is retained on the same levels as long as it increases and enhances the metric quantity but is reduced when the results run into stagnation.

3.5 Proposed Architectural Framework

A detailed map is presented below in this regard in Figure 9, explaining the architectural flow for the proposed methodologies and frameworks.

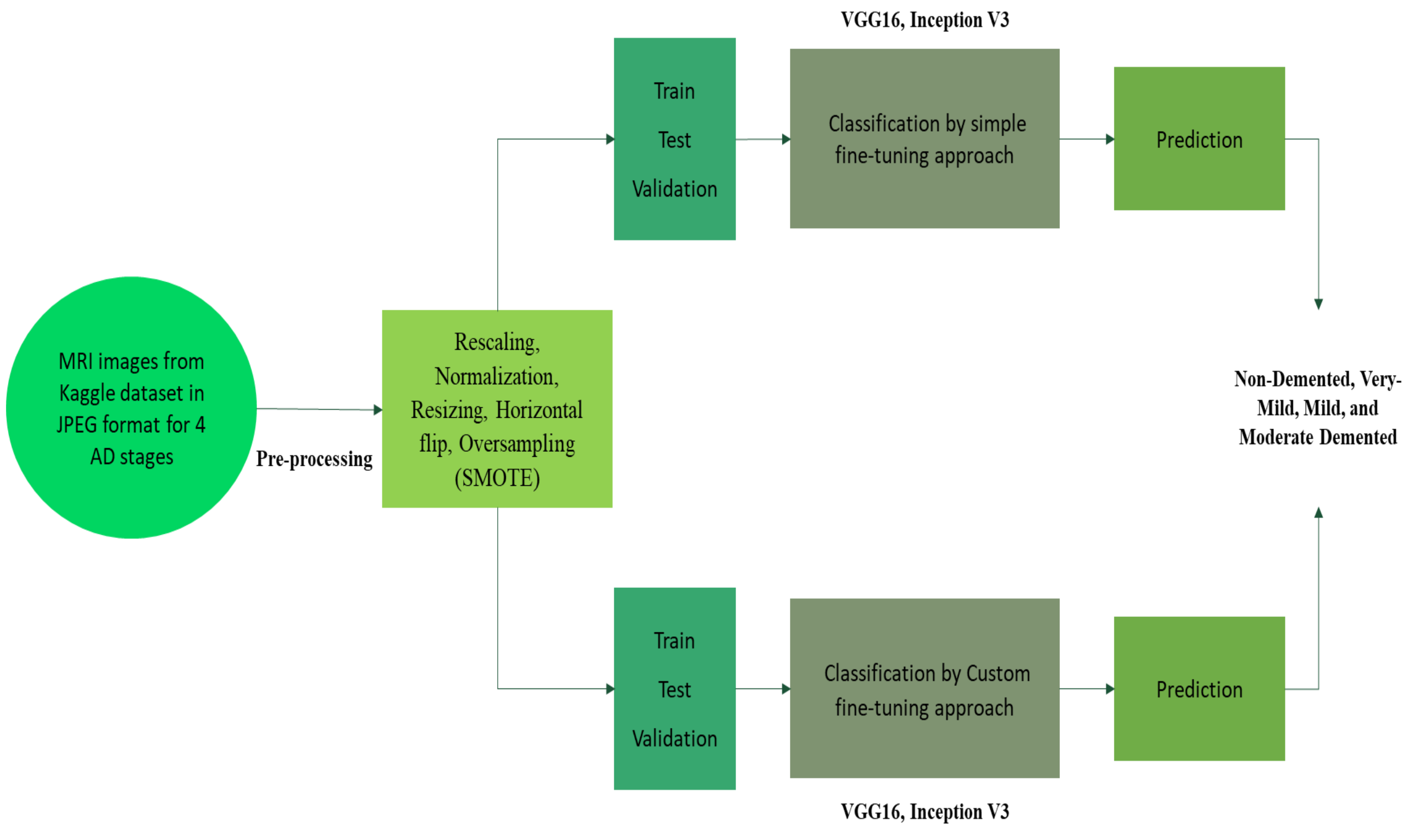


Figure 9: Architectural Map for the Transfer Learning Frameworks

After proposing the mentioned architectural map, the study used pre-trained VGG16 and Inception-V3 models in which their standard layers are frozen for accomplishing transfer learning. Custom fine-tuning frameworks are proposed which are represented below in Figures 10 and 11. Whereas the transfer learning means knowledge is transferred from similar classification tasks to our model's classification tasks in the form of pre-trained weights. By using this knowledge our models classify their specific tasks without being trained from scratch like in Figure 12.

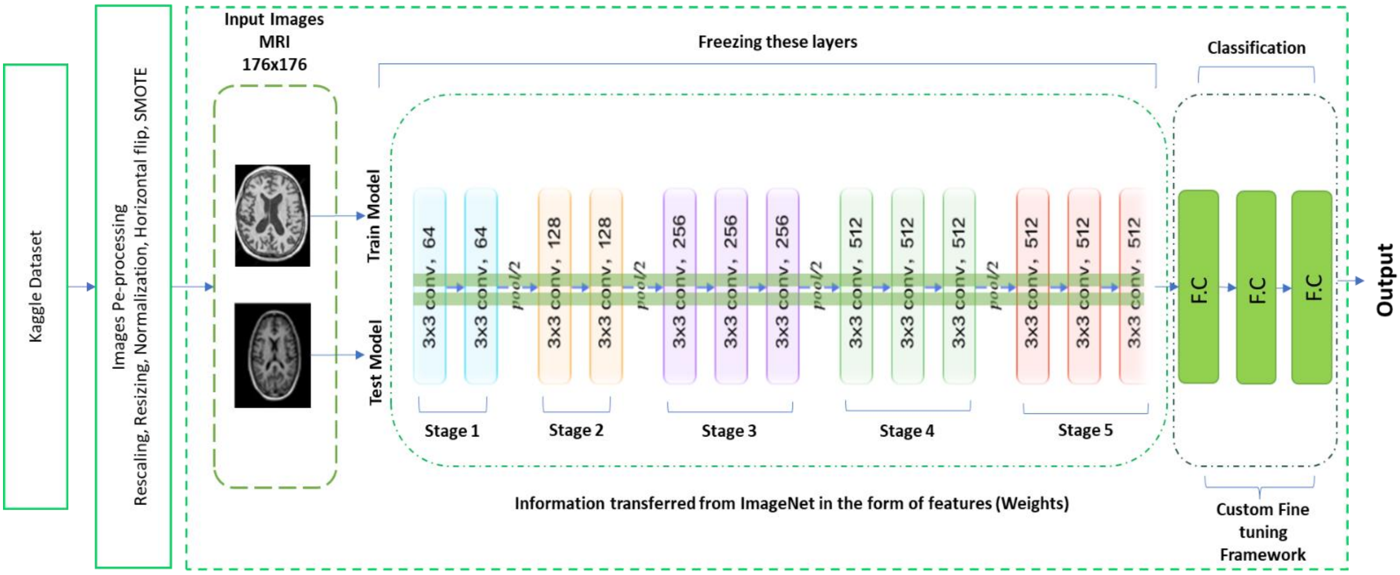


Figure 10: Classification using custom fine-tuned VGG16

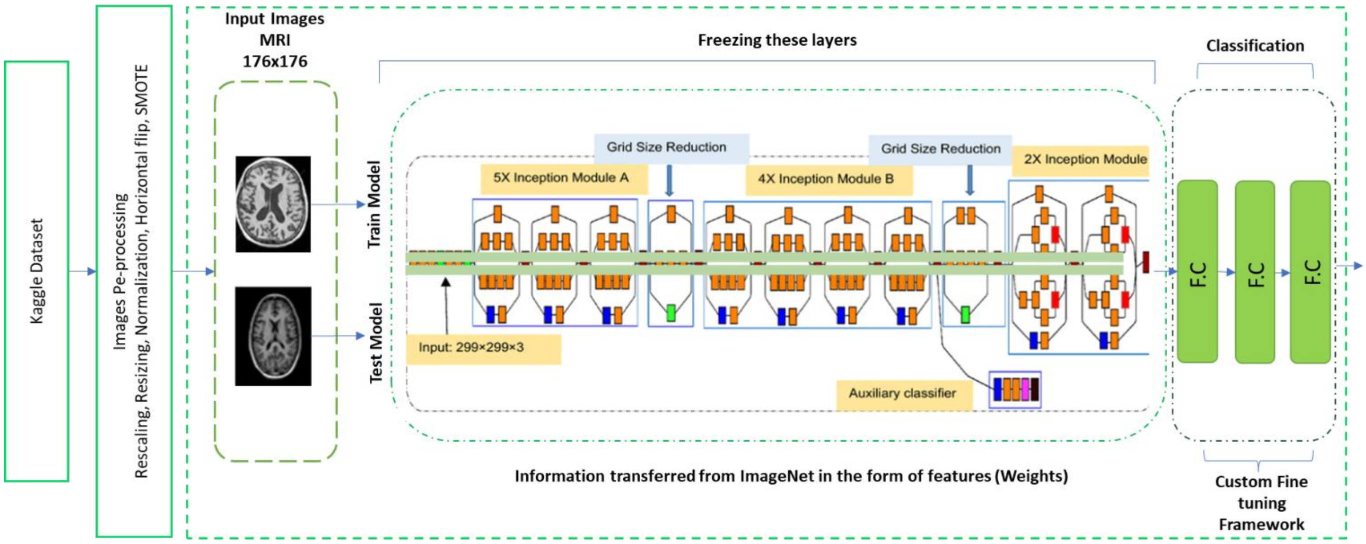


Figure 11: Classification using custom fine-tuned Inception-V3

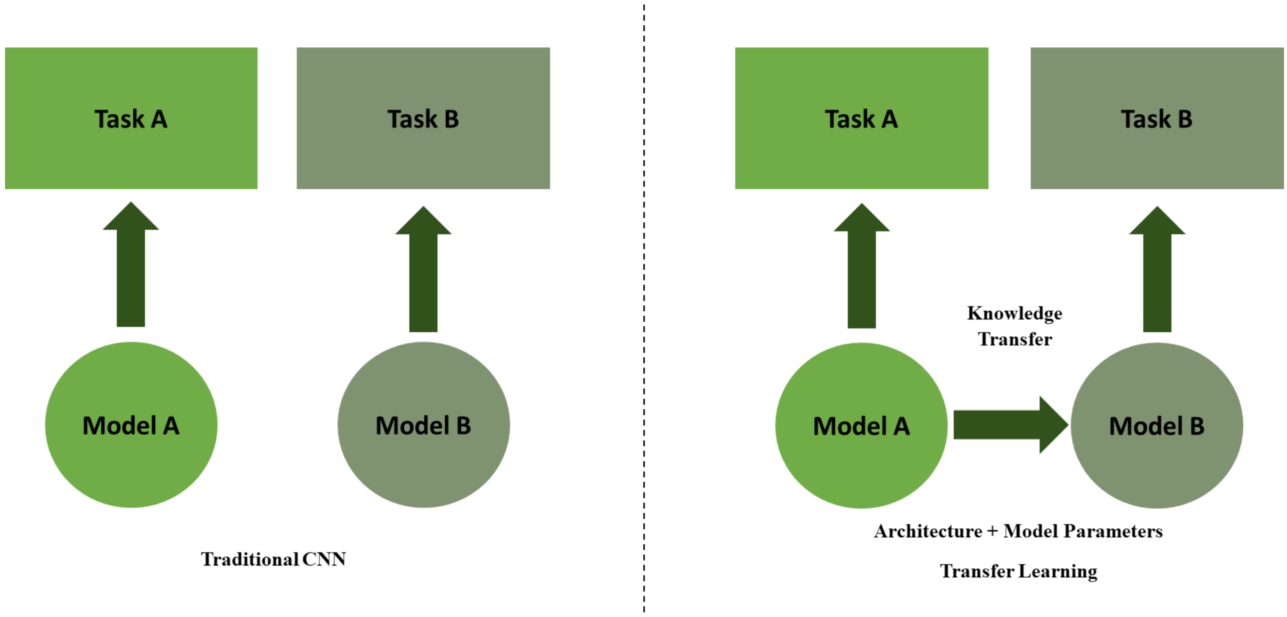


Figure 12: Block diagram for Transfer learning approach

Chapter 4

Discussion & Analysis of Results

4.1 Performance Metrics

This section provides the results, findings, analysis of multi-class classification and discussion with respect to the AD detection through transfer learning approach focused on VGG16 and inception V3 model. These results are analyzed based on the proposed performance metrics that are accuracy, loss, confusion matrix, F1-Score, Area under Curve (AUC), Balanced Accuracy, Testing Accuracy and Matthews Correlation Coefficient. Following are the equational expressions of these mentioned performance metrics, starting with the "Accuracy" which highlights the correct number of predictions from the total number of predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where this expression talks about TP, TN, FP and FN that represents true positive, true negative, false positive and false negative values. The next performance metric "Loss" used for multi-classification is categorical cross-entropy loss also known as SoftMax loss which is a combination of SoftMax activation and cross entropy loss. For the better performance of model "F1 score" metric is integrated which is basically the harmonic mean of precision and recall, having the range of [0 1] which employs that the model performance and F1 score are directly proportional means that the higher the F1 Score is, the better the model performance is. The expression for F1 score is as follows:

$$F1 = \frac{2TP}{2TP + FP + FN}$$

For the purpose of accessing the correct positive result amount to all relevant sample amount "Recall" is used. Additionally, to highlight this correct positive result amount to the positive predicted amount by classifier "Precession" is used. The expressions of these metrics are:

$$Recall = \frac{TP}{TP + FN}$$

$$Precission = \frac{TP}{TP + FP}$$

The last two metrics calculated in this study are "Matthews Correlation Coefficient (MCC)" and "Confusion matrix" wherein the MCC emphasizes the relation between true and predicted values to underline the model's predictions. On the other hand, the confusion matrix provides the holistic depiction of the model's performance. It provides the visual description of how an algorithm performs by displaying true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions, in comparison of actual outcomes. The expression for MCC and visual depiction (Figure 13) of confusion matrix is as follows:

$$MCC = \frac{TPXTN - FPXFN}{\sqrt{((TP + FP) * (TP + FN) * (TN + FP) * (TN + FN))}}$$

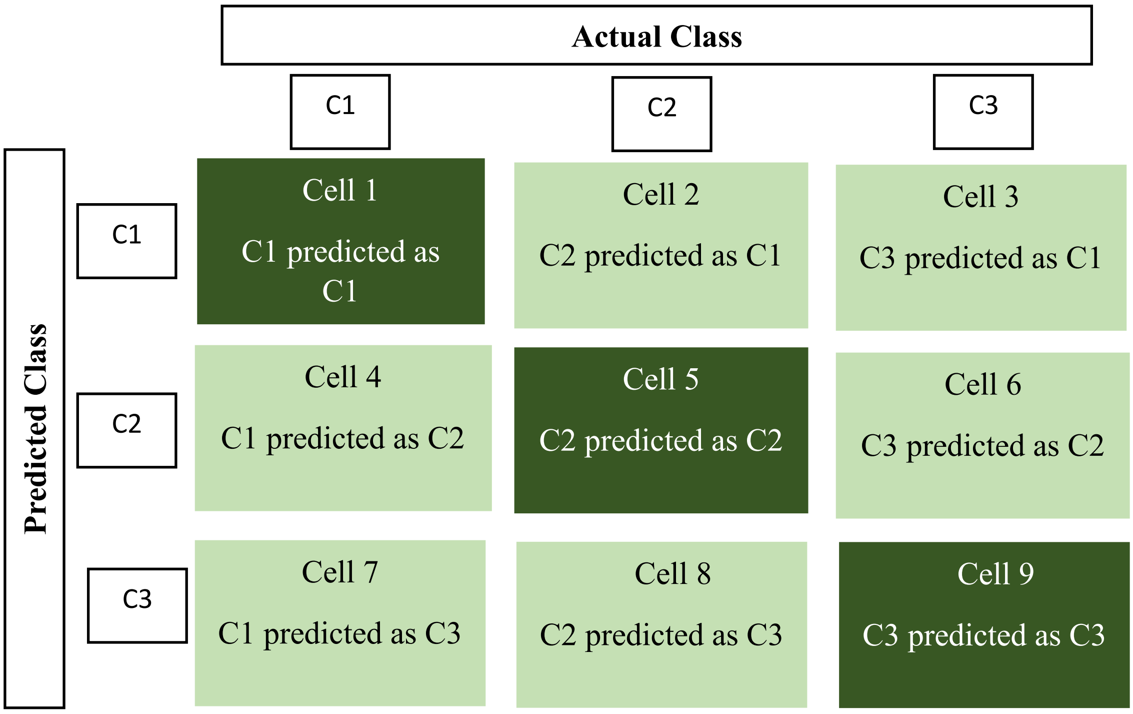


Figure 13: Confusion Matrix for Multiclass Classification

4.2 Evaluation of VGG16 and Inception V3

The present study used transfer learning approach under which a custom fine-tuning approach for multi-class medical image classification is evaluated. . Firstly, simple fine-tuning is incorporated for VGG 16 and Inception V3 to calculate all the performance metrics. Later on the second stage proposed custom fine-tuning framework (sequence 1) for VGG16 and custom fine-tuning framework (sequence 1& 2) for Inception V3 are analyzed. Lastly, the study compares the proposed custom frameworks with the simple frameworks. The upcoming discussion provides the detailed insight of this comparison. Wherein the manner of the discussion is catering the simple fine tuning of VGG16 followed by custom fine-tuning. Likewise, the Inception V3 model is discussed in a similar manner.

4.2.1 Evaluation of VGG16

4.2.1.1 Standard Approach for VGG 16

Firstly, classification is performed using standard VGG16 model in which traditionally it has been trained from scratch for feature extraction and AD is classified in to four classes. By using this standard approaach, VGG16 has achieved training and validation accuracies of 97% and 78%, training and validation auc of 99% and 89%, training and validation loss of 0.13 and 2.41, Mathew's correlation coefficient and balanced accuracie of 66.04% and 76.48% Whereas the following Figure 14 is showing Training and validation accuracies, Auc's and losses.

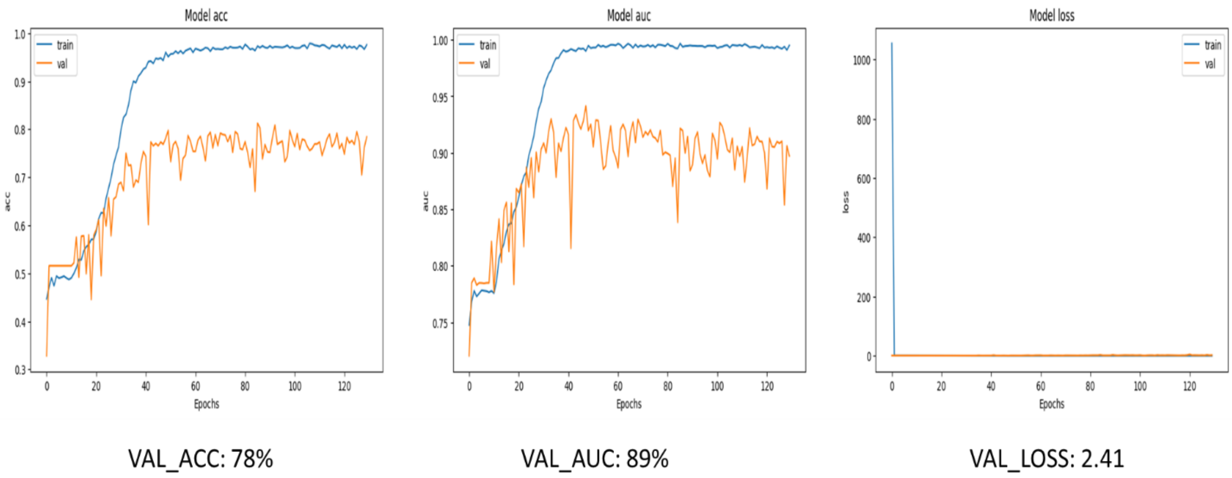


Figure 14: Training and validation accuracies, Auc's and losses for standard VGG16

4.2.1.2 Simple Fine-Tuning Approach for VGG 16

By using simple fine tuning, VGG16 16 has achieved training and validation accuracies of 89.15% and 88.495, training and validation auc of 98.5%and 98%, training and validation loss of 0.273 and 0.32, Mathew's correlation coefficient and balanced accuracie of 85.48% and 89.13%. Whereas the following Figure 15 is showing Training and validation accuracies, Auc's and losses.

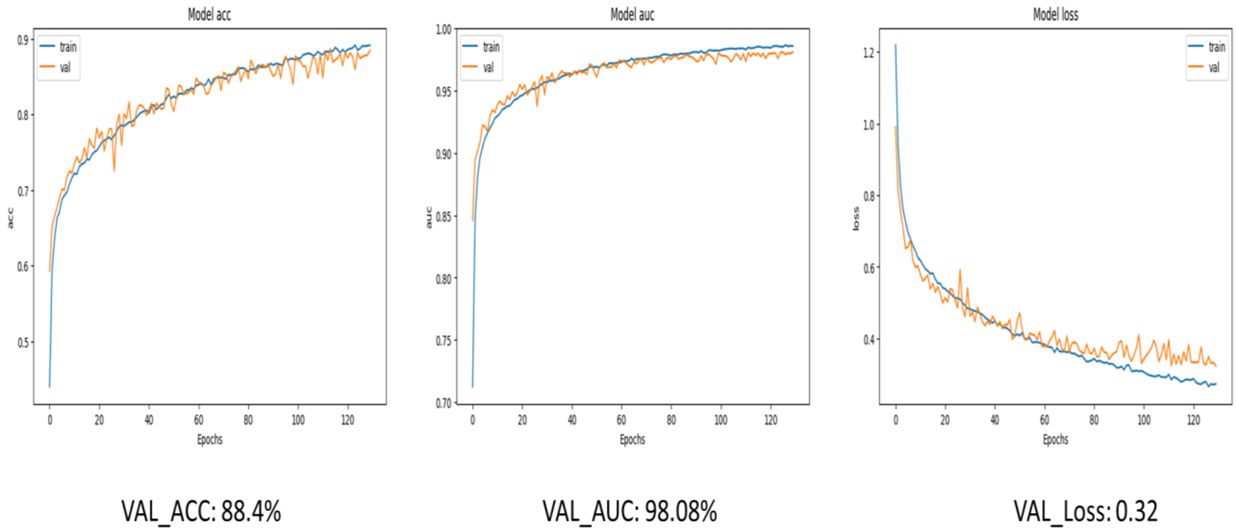


Figure 15: Training and validation accuracies, Auc's and losses for simple fine-tuned VGG16

Wherein the blue line in the above-mentioned figure shows the performance metrics when training dataset is given during training whereas the red line indicates the metrics when validation dataset is given during training.

4.2.1.3 Custom Fine-Tuning Approach for VGG16

When custom fine-tuning framework (sequence 1) is incorporated VGG16 has achieved prominent increase in training and validation accuracies of 90.64% and 94%, training and validation auc of 98.64% and 99.47%, training and validation loss of 0.266 and 0.166, and Mathew's correlation coefficient and balanced accuracie of 91.88% and 93.96%. As shown in the graphs preseted below in Figure 16.

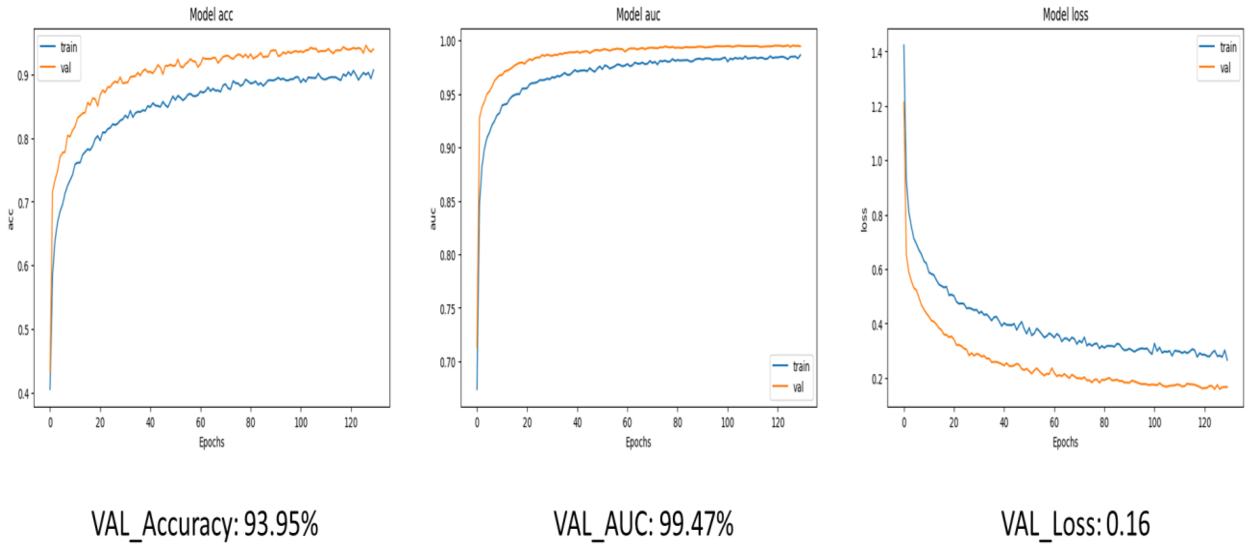


Figure 16: Training and validation accuracies, Auc's and losses for custom fine-tuned VGG16 (sequence 1)

4.2.1.4 Comparative Analysis of Standard, Simple and Custom Fine-Tuning for VGG16

The increase in the efficacy of the performance metrics is clearly visible as depicted in Table 3 below. The tabular data clearly shows that after incorporating custom fine tuning framework validation accuracy has prominently increased by 4.5%, validation loss has decreased by approximately 50% and testing accuracy has increased by 5%.

Table 3

Comparative Analysis of Standard, Simple and Custom Fine-tuning (sequence 1) for VGG16

Parameters	**Standard VGG16**	**Simple Fine-tuned VGG-16**	**Custom fine-tuned VGG16 (sequence 1)**
Loss	0.13	0.273	0.266
Accuracy	97%	89.15%	90.64%

AUC	99%	98.56%	98.64%
F1_Score	0.96	0.89	0.90
Val_loss	2.41	0.32	0.166
Val_Acc	78%	88.48%	94%
Val_AUC	89%	98%	99.47%
Val_F1_score	0.65	0.882	0.938
Testing Accuracy	79.45%	89.10%	93.91

The above-mentioned comparison is followed by the stage level comparison of AD classes i.e. “non-demented, very mild demented, mild demented, moderate demented”, presented below in Table 4 for simple and custom fine tuning, with respect to some performance metrics mainly precision, recall, f1-score and support.

Table 4

Stage level comparison of AD classes w.r.t simple VGG16 and custom fine-tuning VGG16 (sequence 1)

	Simple Fine-Tuning (VGG16)				Custom Fine-Tuning (VGG16) (Sequence 1)			
	Precision	**Recall**	**F1 Score**	**Support**	**Precision**	**Recall**	**F1 Score**	**Support**
Non-Demented	0.89	0.93	0.91	648	0.94	0.96	0.95	621
Very Mild Demented	0.99	1.00	1.00	634	1.00	1.00	1.00	637
Mild Demented	0.87	0.84	0.85	622	0.93	0.90	0.92	662
Moderate Demented	0.81	0.80	0.81	656	0.89	0.89	0.89	640
Micro Avg	0.89	0.89	0.89	2560	0.94	0.94	0.94	2560
Macro Avg	0.89	0.89	0.89	2560	0.94	0.94	0.94	2560
Weighted Avg	0.89	0.89	0.89	2560	0.94	0.94	0.94	2560
Samples Avg	0.89	0.89	0.89	2560	0.94	0.94	0.94	2560

4.2.1.5 Confusion Matrices (VGG16)

The graphical representation of confusion matrix for standard VGG16 model is depicted in Figure 17 below, which shows a prominent confusion between classes as the number of predictions for non-demented and very mild-demented are too low. Followed by simple fine tuning in Figure 18 given below, which is screening the correct number of predictions diagonally while upper and lower entries are depicting confused predictions. For instance, the first entry of first row shows that 600 predictions for non-demented are predicted as non-demented while third and fourth entry shows that 13 and 35 predictions for non-demented are predicted as mild demented and moderate demented which is basically the confusion for non-demented predictions with mild demented and moderate demented.

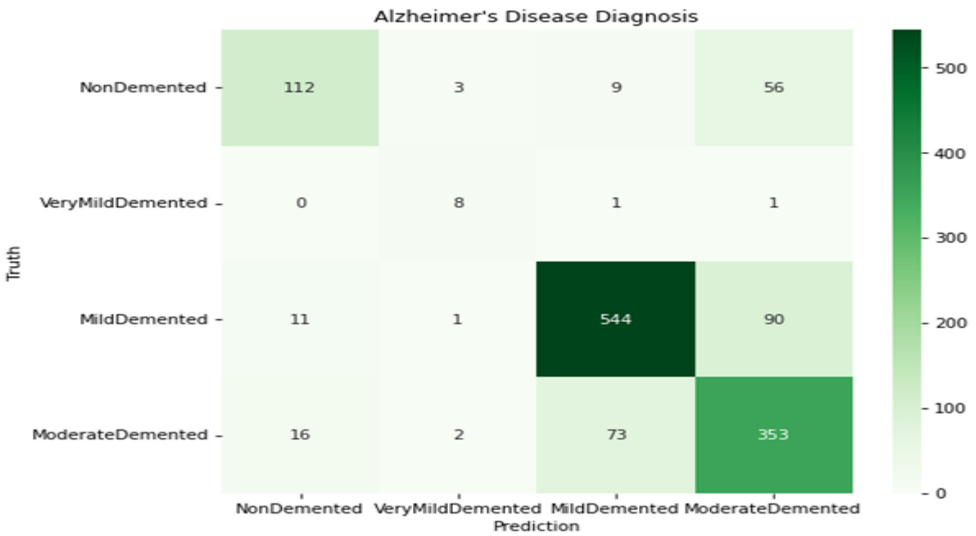


Figure 17: The confusion matrix for standard VGG16 in the model's predictability

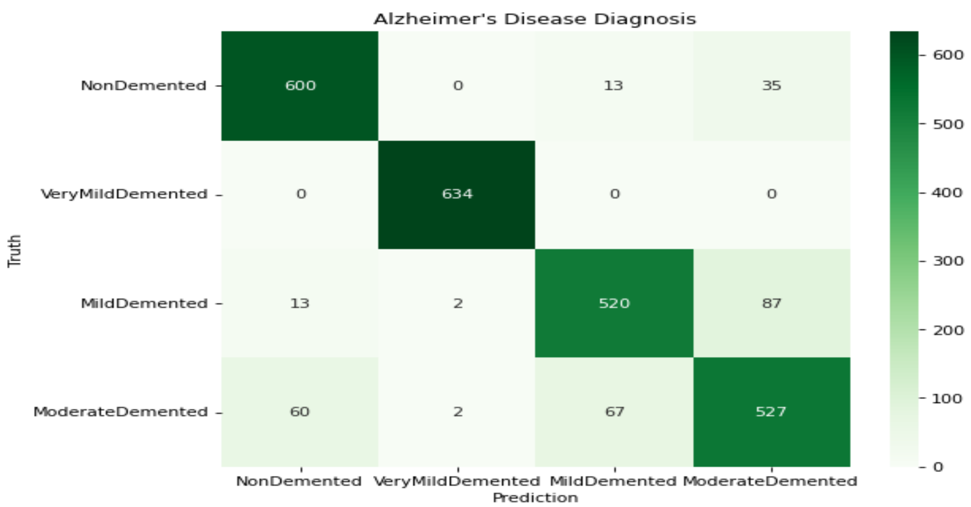


Figure 18: The confusion matrix for simple fine-tuning in the model's predictability (VGG16)

However, after the integration of custom tuned framework (sequence 1), it improved the overall confusion between the classes, as evident in Figure 19 that the confusion of non-demented is 18. Confusion in mild and moderate demented has decreased from 13 to 3 and from 35 to 18. Confusion of mild demented with moderate demented has decreased from 87 to 56 and the confusion of “moderate demented” with non-demented and mild demented has decreased from 60 to 28 and 67 to 41.

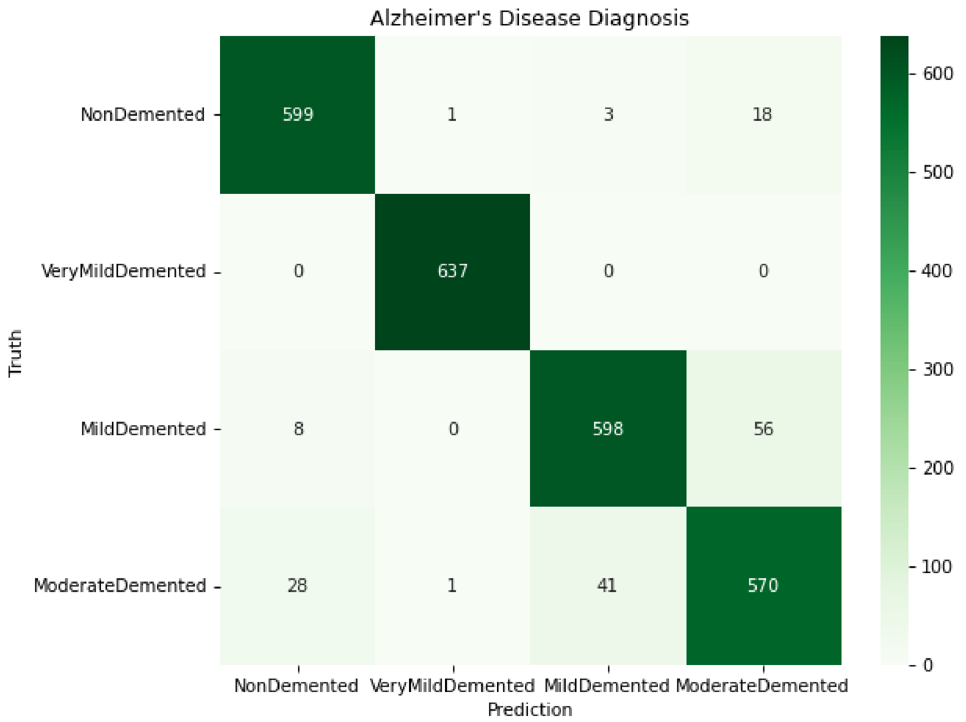


Figure 19: The confusion matrix for custom fine tuning in the model’s predictability (VGG16) sequence 1

4.2.2 Evaluation of Inception V3

4.2.2.1 Standard Approach for Inception V3

The classification is done by using inception V3 model by means of similar method discussed under standard VGG16 in which it is typically trained from scratch for feature extraction and categorizing AD in to four classes. Consequently, in this process the standard

Inception-V3 approach has achieved the following accuracy levels. For instance, it attained the training and validation accuracies of 99% and 71%, training and validation auc of 99%and 88%, training and validation loss of 0.02 and 1.77, Mathew's correlation coefficient and balanced accuracy of 58.83% and 68.15%. The visualization of these levels are shown in Figure 20.

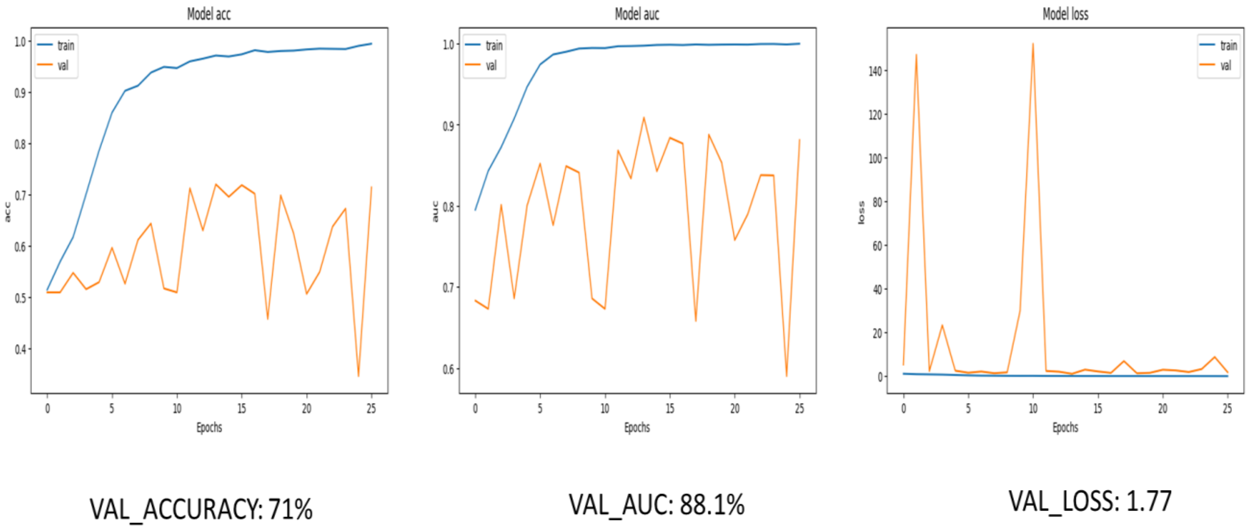


Figure 20: Training and validation accuracies, Auc's and losses for standard Inception V3

4.2.2.2 Simple Fine-Tuning Approach for Inception V3

By using simple fine tuning Inception V3 has achieved training and validation accuracies of 90.7% and 84.96%, training and validation auc of 99%and 97.28%, training and validation loss of 0.254 and 0.38, Mathew's correlation coefficient and balanced accuracy of 83.17% and 77.43%. The following Figure 21 is illustrating Training and validation accuracies, auc's and losses for Inception V3.

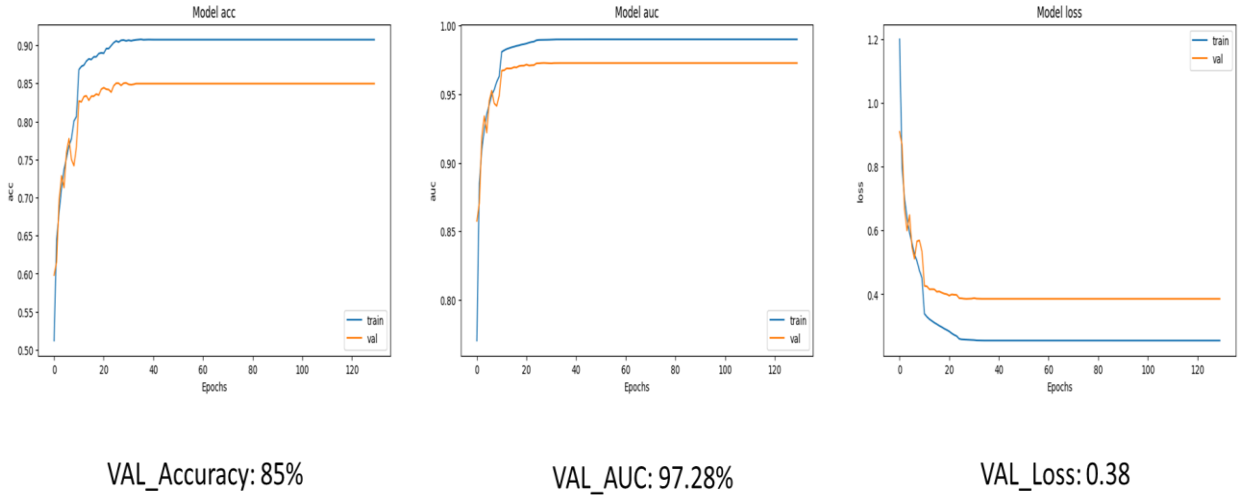


Figure 21: Training and validation accuracies, Auc's and losses for simple fine-tuned Inception V3

4.2.2.3 Custom Fine-Tuning Approach for Inception V3:

When custom fine-tuning framework is integrated Inception V3 has achieved an outstanding increase in training and validation accuracies of 95% and 93.9%, training and validation auc of 99.5% and 99.45%, training and validation loss of 0.15 and 0.166, Mathew's correlation coefficient and balanced accuracie of 91.83% and 93.89% with 1.22 million trainable parameters. Which is refered as sequence 1 for custom fine-tuning framework. So, below mentioned Figure 22 for Training and validation accuracies, auc's and losses regarding custom tuned Inception V3, is showing these results w.r.t sequence 1.

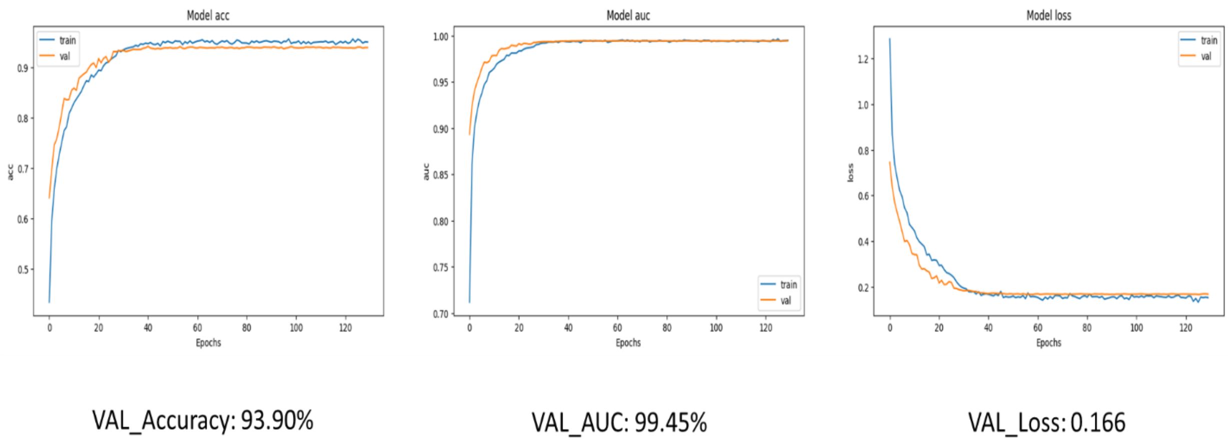


Figure 22: Training and validation accuracies, Auc's and losses for custom fine-tuned Inception V3 (Sequence 1)

The parameters (1.22M) in case of sequence 1 are still computationaly complex as compared to 438K ofVGG16 (sequence 1). Henceforth, the study established another sequence for feature reduction in case of inception V3, named as sequence 2. As a result of which the parameters of Inception V3 (sequence 1) model are reduced to over 50% from 1.22M to 573K for sequence 2. Howvere, by doing this the accuracy rates are not being majorly compromised. The accuracy rate for sequence 2 is 93.09%, which means there is only 0.78% fluctuation in accuracy rates, as presented below in Figure 23.

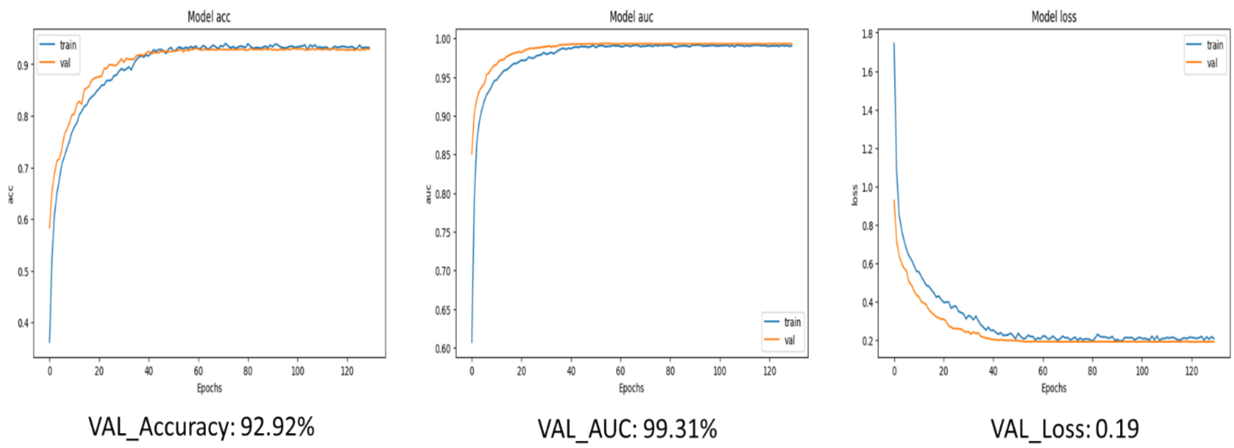


Figure 23: Training and validation accuracies, Auc's and losses for custom fine-tuned Inception V3 (Sequence 2)

4.2.2.4 Comparative Analysis of Simple and Custom Fine-Tuning for Inception V3

The upsurge in the efficacy of the performance metrics is clearly visible as described in table 5. The tabular data clearly shows that after incorporating custom fine tuning framework validation accuracy has prominently increased by 9%, validation loss has decreased by more than 50% and testing accuracy has increased by 10.8%

Table 5

Comparative Analysis of Standard, Simple and Custom Fine-Tuning for Inception V3

Parameters	Standard Inception-V3	Simple Fine-tuned Inception-V3	Custom Fine-tuned Inception V3 (Sequence 1)	Custom Fine-tuned Inception V3 (Sequence 2)
Loss	0.02	0.254	0.15	0.20
Accuracy	99%	90.7%	95%	93.1%
AUC	99%	99%	99.5%	99.4%
F1_Score	0.99	0.90	0.95	0.93
Val_loss	1.77	0.38	0.166	0.19
Val_Acc	71%	84.96%	93.9%	92.92%
Val_AUC	88%	97.28%	99.45%	99.31%
Val_F1_score	0.56	0.848	0.938	0.928
Testing Accuracy	74.1%	83.05%	93.87%	93.9%

In addition to the above-mentioned contrast, for further insightful analysis, the stage level comparison of AD classes "(non-demented, very mild demented, mild demented, moderate demented)" is presented below in Table 6 for simple and custom fine tuning, with respect to some performance metrics mainly precision, recall, f1-score and support.

Table 6

Stage level comparison of AD classes w.r.t simple fine-tuning and custom fine-tuning (Inception V3) for sequence 1 and 2

	Simple Fine-Tuning (Inception V3)				Custom Fine-Tuning (Inception V3) Sequence 1				Custom Fine-Tuning (Inception V3) Sequence 2			
	Precision	**Recall**	**F1 Score**	**Support**	**Precision**	**Recall**	**F1 Score**	**Support**	**Precision**	**Recall**	**F1 Score**	**Support**
Non-Demented	0.83	0.89	0.86	623	0.96	0.98	0.97	633	0.97	0.98	0.87	635
Very Mild Demented	0.99	1.00	0.99	637	1.00	1.00	1.00	641	1.00	1.00	1.00	640
Mild Demented	0.78	0.74	0.76	663	0.91	0.88	0.89	648	0.88	0.87	0.88	631
Moderate Demented	0.72	0.70	0.71	637	0.89	0.89	0.89	638	0.87	0.88	0.88	654
Micro Avg	0.83	0.83	0.83	2560	0.94	0.94	0.94	2560	0.93	0.93	0.93	2560
Macro Avg	0.83	0.83	0.83	2560	0.94	0.94	0.94	2560	0.93	0.93	0.93	2560
Weighted Avg	0.83	0.83	0.83	2560	0.94	0.94	0.94	2560	0.93	0.93	0.93	2560
Samples Avg	0.83	0.83	0.83	2560	0.94	0.94	0.94	2560	0.93	0.93	0.93	2560

4.2.2.5 Confusion Matrices (Inception V3)

The standard Inception-V3 approach forecasting the prominent confusion among the classes during the prediction process. It is evident in below-mentioned Figure 24 that non-demented and very-mild demented classes are not predicted accurately and there is an acute and noticeable confusion among all classes. Whereas the confusion matrix for simple fine tuning in Figure 25 shows the correct number of predictions slantwise while upper and lower entries are representing confused predictions. Thus, the first entry of first row shows that 533 predictions for non-demented are predicted as non-demented while third and fourth entry shows that 21 and 49 predictions for non-demented are predicted as mild and moderate demented which is basically highlighting the confusion for non-demented predictions with mild and moderate demented.

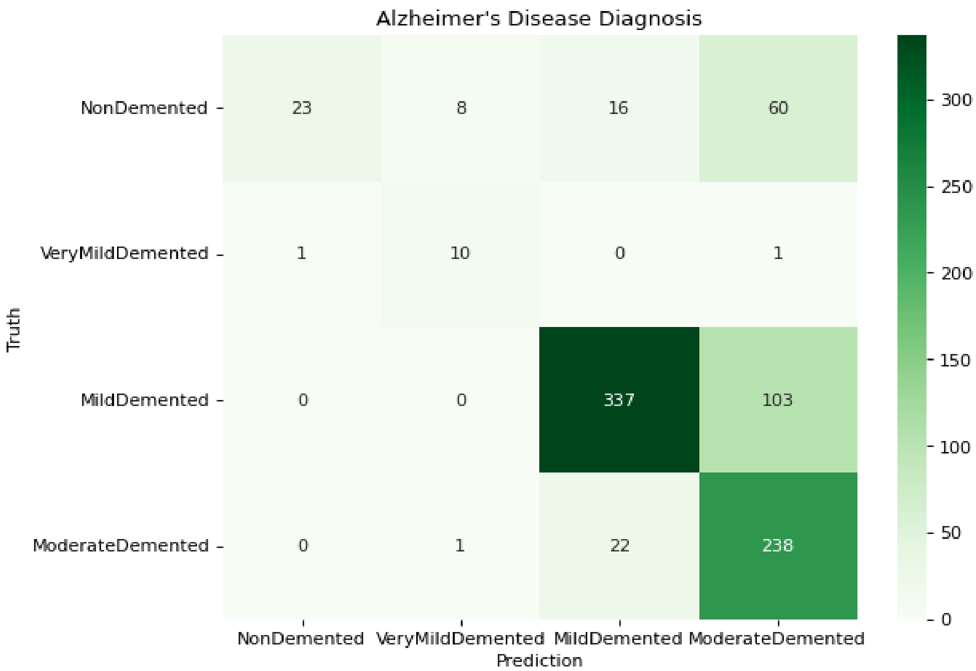


Figure 24: The confusion matrix for standard Inception V3 in the model's predictability

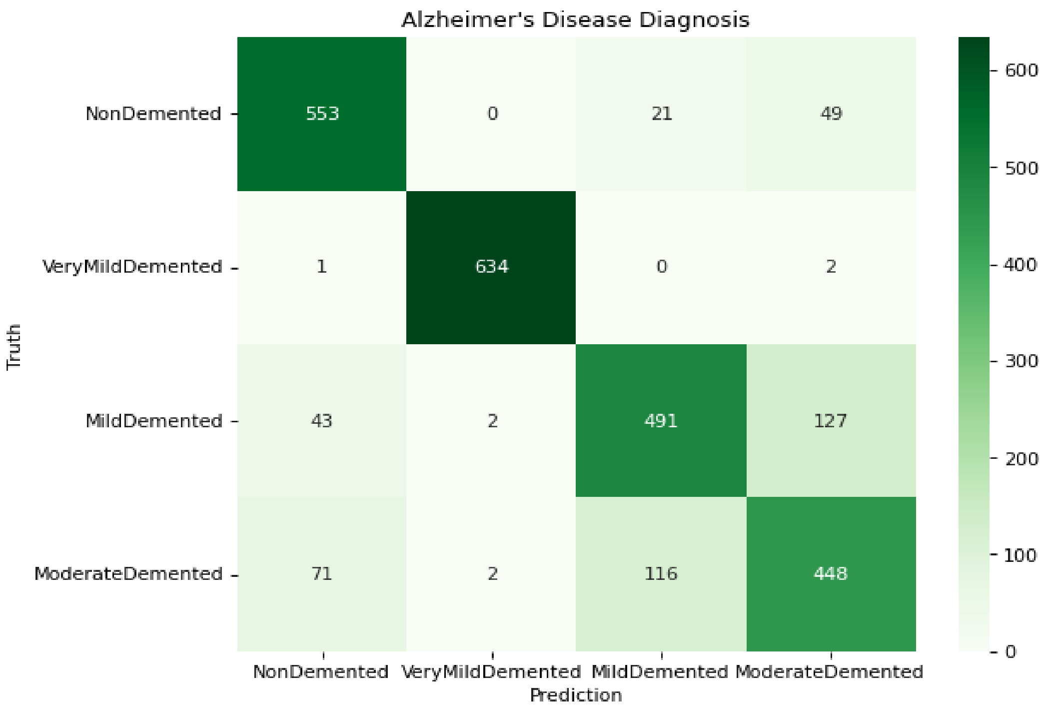


Figure 25: The confusion matrix for simple fine-tuning in the model's predictability (Inception V3)

Whereas the custom fine-tuned framework of sequence 1 improved the overall confusion between the classes, as evident in Figure 26 that the confusion of non-demented class with mild and moderate demented has decreased from 21 to 6 and from 49 to 6. Confusion of mild demented with moderate demented has decreased from 127 to 67 and the confusion of moderate demented with non demented and mild demented has decreased from 71 to 16 and 116 to 52.

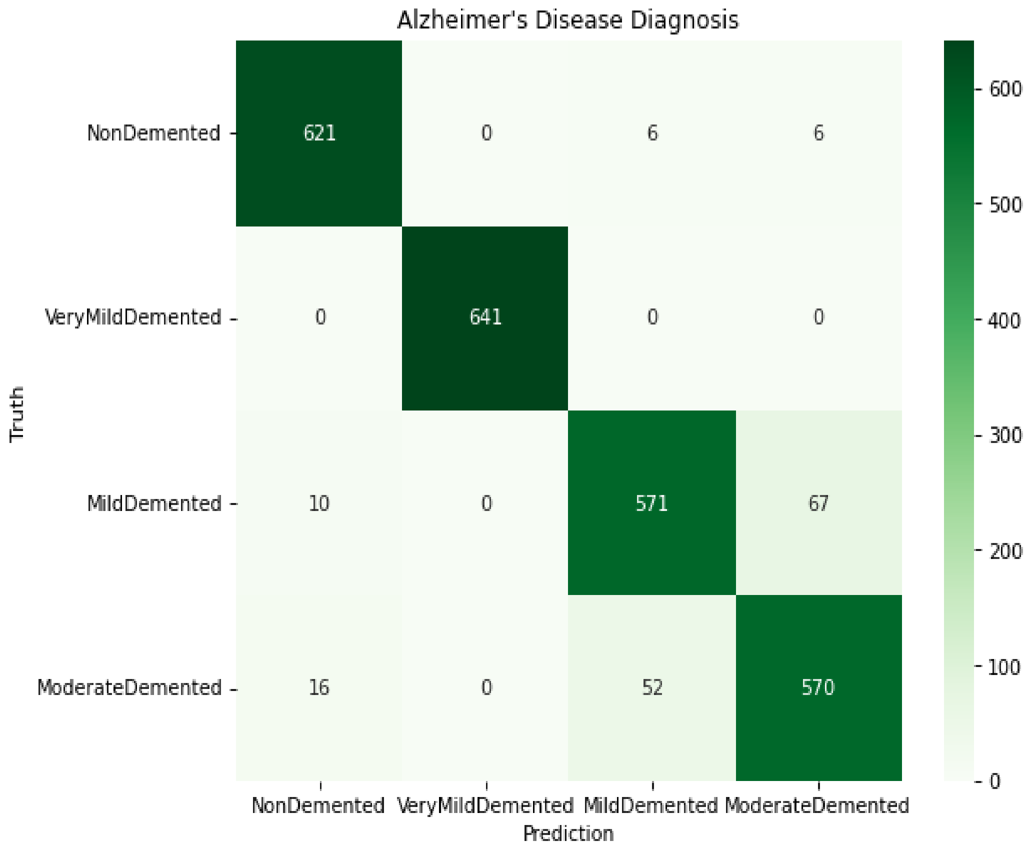


Figure 26: The confusion matrix for custom fine tuning in the model's predictability (Inception V3) sequence 1

In line with this, the confusion matrix of sequence 2 for custom fine-tuned (Inception V3) shows few fluctuations as compared to sequence 1. For further insight refer to the below Figure 27.

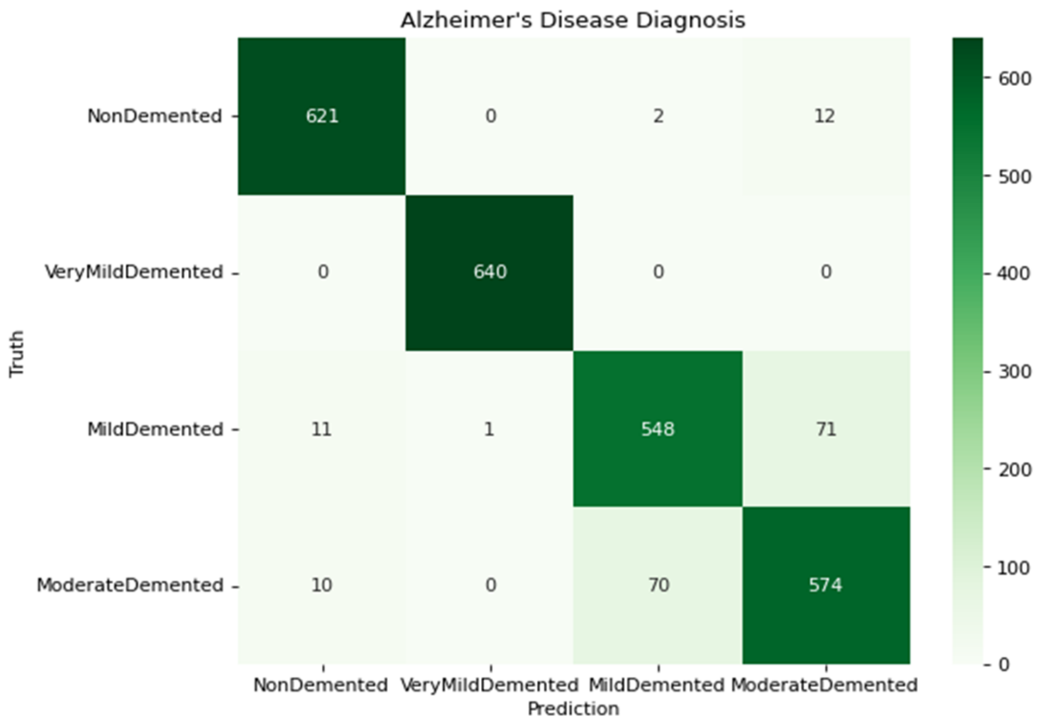


Figure 27: The confusion matrix for custom fine tuning in the model's predictability (Inception V3) sequence 2

4.3 Discussion on Results

In this study various techniques such as SMOTE, transfer learning and custom fine-tuning framework are incorporated in-order to enhance AD detection using MRI scans as dataset for classification. First synthetic minority oversampling technique is used for reducing class imbalancement because it's proven to be one of the best techniques that are used for balancing datasets. As discussed in literary study [76] SMOTE outperforms other techniques by achieving 99.08% accuracy. Therefore, the present study employed SMOTE to reduce class imbalancement. Where after applying it, the dataset size was increased from 6400 to 12800 MRI scans.

Then the balanced dataset is classified using different approaches as discussed below. Firstly, classification is accomplished with respect to standard VGG16 and inception V3 models in which conventionally they are trained from scratch for feature extraction and AD is classified in to four classes. The testing accuracies achieved are 79.5% and 74% which are not so accurate. The finding can be backed up by the literature [74, 33] where they have also trained ResNet-50, DenseNet-169 and Vgg19 models from scratch and achieved accuracies only between 80% to 82.5%. Secondly, transfer learning is introduced in which pre-trained VGG16 and Inception-V3 models are used. Features are transferred from IMAGENET and a simple fine-tuning framework is utilized. The testing accuracies in this approach are enhanced to 89% for VGG16 and 83% for Inception V3, also computational complexity is reduced. Lastly, when custom fine-tuning framework of sequence 1 is employed, the accuracies are further enhanced to almost 94% for both the models. The point to ponder is that the parameters for VGG16 are already 438K w.r.t to sequence 1 but for Inception V3 are 1.22M, which still shows computational complexity in contrast to VGG16. Therefore, to decrease this complexity sequence 2 for custom fine-tuned framework (Inception V3) is introduced by reducing parameters from 1.22M to 573K with testing accuracy rate of 93.09%. Conclusively, fine-tuning approach along with transfer learning is proven to be the best for AD detection which is evident in various previous studies of literature [17, 64, 24] where these studies have achieved outstanding accuracies, for instance first mentioned study attained 91.7%, second accomplished 97% and third study accounted 96% accuracy levels for AD detection through the incorporation of fine-tuned models.

4.4 Contrast of Study's Test Performance with Literature

Several methodologies have been investigated in previous studies regarding the detection of Alzheimer, however, the transfer learning approach has proved to be one of the outstanding approaches. The proposed methodology of the current study has achieved high accuracy levels in detection of AD classes through VGG16 and Inception V3. The below-mentioned table 7 represents the comparison of proposed fine-tuned method's performance with the previous studies (i.e. body of Literature). From the table it is evident that the proposed models outperformed majority of the published approaches while achieving closest accuracy levels with respect to approaches.

Table 7

In comparison with the existing body of knowledge (Literature) for Fine-tuned VGG16 and Inception V3

Literature	Models	Validation Accuracy	Miss rate
Islam et al. (2017)	Inception V4 Network	73.75%	26.25%
Zhang et al. (2017)	Landmark-based extraction	79.02%	20.98%
Pradhan, A. et al. (2021)	DenseNet-169 VGG16	80% 82.6%	20% 17.4%
Ghazal et al. (2021)	AlexNet (Fine tuned)	91.7%	8.3%
Al Sehri W. (2022)	ResNet-50 DenseNet-169	82% 88%	18% 12%
Al Saeed, D. et al. (2022)	Resnet 50-SoftMax ResNet 50-SVM ResNet 50- RF	99% 92% 85.7%	1% 8% 14.3%
Helaly, H.A. et al. (2022)	2D CNN 3D CNN VGG19 (Fine tuned)	93.6% 95.1% 97%	6.4% 4.9% 3%
Yusi C. et al. (2024)	Soft NMS Improved ResNet 50	84.3% Overall	15.7%
Proposed Method	**Proposed Fine tuned VGG16**	**94%**	**6%**
Proposed Method	**Proposed Fine tuned Inception V3**	**93.9%**	**6.1%**

Chapter 5

Conclusions and Future Work

The contents of this chapter consist of proposed objectives listed in chapter 1 which are reassessed along with the deep learning frameworks that are used to attain these objectives. These set aims directed the study by suggesting novel methods to answer the problem under discussion. Moreover, the current chapter also lists the limitations and short comings faced in this thesis to provide further future recommendations.

5.1 Discussion on Objectives

The objectives of this study are concluded one by one.

1. To detect and analyze the stages of AD through simple transfer learning approach with respect to VGG16 and inception V3 models

In this study a simple fine-tuning framework that will detect and analyze AD stages is presented. It was found that at first through incorporating using VGG 16 the simple fine-tuned framework was able to classify AD stages (non-demented, very mild demented, mild demented and moderate demented) with training and validation accuracies of 89.15% and 88.495, training and validation auc of 98.5%and 98%, training and validation loss of 0.273 and 0.32, Mathew's correlation coefficient and balanced accuracie of 85.48% and 89.13%. Similarly by using Inception V3 this framework classified AD stages with training and validation accuracies of 90.7% and 84.96%, training and validation auc of 99%and 97.28%, training and validation loss of 0.254 and 0.38, Mathew's correlation coefficient and balanced accuracy of 83.17% and 77.43%.

2. To enhance the detection of AD by introducing a custom fine-tuning framework concerning VGG 16 and Inception V3

In this thesis, in line with the first objective, custom fine tuning-frameworks ae proposed which contributes towards the enhancement of the detection of AD classes. The proposed custom fine-tuning framework (sequence 1) when used with VGG16 achieved training and validation accuracies of 90.64% and 94%, training and validation auc of 98.64% and 99.47%, training and validation loss of 0.266 and 0.166, and Mathew's correlation coefficient and balanced accuracie of 91.88% and 93.96%. Clearly, this custom fine-tuned VGG16 outperformed the simple fine-tuned VGG16. Similarly the custom Inception-V3 (sequence 1 & 2) outperformed the simple Inception-V3 by achieving an outstanding increase for sequence 1, in training and validation accuracies of 95% and 93.9%, training and validation auc of 99.5% and 99.45%, training and validation loss of 0.15 and 0.166, Mathew's correlation coefficient and balanced accuracy of 91.83% and 93.89%. And for sequence 2 it achieved 93.15% training and 92.92% validation accuracy rates, 99.4% training auc and validation auc 99.31%, 0.20 training loss and validation loss 0.19 along with MCC 90.78% and balance accuracy 93.1%. Henceforth, these objectives are in line with the recent predictions about the transfer learning approach [17, 64, 24].

3. To compare the efficiency of VGG16 and Inception-V3 models through performance metrics (categorial accuracy class, AUC, balanced accuracy, and Mathew's correlation coefficient, loss, confusion matrices, precision, f1-score and recall).

If we compare the custom fine-tuned VGG 16 (sequence 1) with custom fine-tuned Inception V3 (sequence 1 & 2) model in terms of performance metrics, Inception-V3's both sequences outperform VGG 16 in terms of training accuracy and training loss. The training accuracy of Inception V3 (sequence 1) is 95% (sequence 2) is 93.15% which is higher than that of VGG 16's 90.6% and training loss of Inception-V3 (sequence 1) is 0.15, (sequence 2) is 0.20, better than that of VGG16's 0.266. In terms of trainable parameters VGG16 (438k) outperforms Inception V3 sequence 1 (1.22M) and sequence 2 (573K). Consequently, except for minute differences in computational complexities almost both of the models are found to be efficient in terms of predicting AD classes.

5.2 Limitations of the study

Possible limitations of this study are:

1. limited dataset and less balanced classes.
2. The moderate demented class has only 52 brain scans for training and 12 scans for testing.
3. Although SMOTE technique has been applied to reduce class imbalancement in our dataset, this needs to be considered in future in order to enhance AD detection.

5.3 Future Recommendations

There are several ways in which the work done under this research could be extended. Since this is a supervised learning approach in future different unsupervised learning techniques such as VAEs, GANs can also be explored. These techniques can be very crucial when labeled datasets are limited. Different datasets like ADNI and OASIS can also be used along with architectural improvements and advancements to access the performance of deep learning approaches. Different transfer learning approaches and optimization processes could be integrated in order to further enhance the effectiveness of this proposed model.

References

[1] J. Islam and Y. Zhang, "Early Diagnosis of Alzheimer's Disease: A Neuroimaging Study with Deep Learning Architectures," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, UT, USA, 2018.

[2] S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. R. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga and L. Beckettj, "Ways toward an early diagnosis in Alzheimer’s disease: The Alzheimer’s Disease Neuroimaging Initiative (ADNI)," *Alzheimers Dement,* pp. 55-66, 2005.

[3] M. Park and W.-J. Moon, "Structural MR Imaging in the Diagnosis of Alzheimer’s Disease and Other Neurodegenerative Dementia: Current Imaging Approach and Future Perspectives," *Neuroimaging and Head & Neck,* pp. 827-845, 2016.

[4] J. Islam and Y. Zhang, "“A novel deep learning based multi-class classification method for Alzheimer’sdisease detection using brain mri data,”," *Lecture Notes in Computer Science,* vol. 10654, pp. 213-222, 2017.

[5] S. Shamshirband, M. Fathi, A. Dehzangi, A. T. Chronopoulos and H. Alinejad-Rokny, "A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues," *ScienceDirect,* 2021.

[6] M. S. Uddin and G. M. Ashraf, "Introductory Chapter: Alzheimer’s Disease—The Most Common Cause of Dementia," in *Advances in Dementia Research*, IntechOpen, 2018.

[7] A. Wimo, G. C. Ali, M. Guerchet, M. Prince, M. Prina and Y. T. Wu, "The World Alzheimer Report," Alzheimer's Disease International , 2015.

[8] H. A. Helaly, M. Badawy and A. Y. Haikal, "Deep Learning Approach for Early Detection of Alzheimer's Disease.," *Cognitive computation,* vol. 14, no. 5, pp. 1711-1727, 2022.

[9] R. Hegde, K. Prasad, H. Hebbar and B. M. k. Singh, "Feature extraction using traditional image processing and convolutional neural network methods to classify white blood cells: a study," *Researchgate,* 2019.

[10] M. Jogin, Mohana, G. D. Divya, M. S. Madhulika, R. K. Meghana and S. Apoorva, "Feature Extraction using Convolution Neural Networks (CNN) and Deep Learning," in *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, 2018.

[11] P. Kishore, C. U. Kumari, M. Kumar and T. Pavani, "Detection and analysis of Alzheimer's disease using various machine learning algorithms," *Science Direct,* vol. 45, pp. 1502-1508, 2021.

[12] A. Shukla, R. Tiwari and S. Tiwari, "Review on Alzheimer Disease Detection Methods: Automatic Pipelines and Machine Learning Techniques," *Sci,* vol. 5, no. 1, 2023.

[13] A. B. Tufail, Y. K. Ma and Q. N. Zhang, "Binary Classification of Alzheimer's Disease Using sMRI Imaging Modality and Deep Learning.," *Journal of digital imaging,* vol. 33, no. 5, p. 1073–1090, 2020.

[14] D. A. Arafa, H. E.-D. Moustafa, H. A. Ali, A. M. T. Ali-Eldin and S. F. Saraya, "A deep learning framework for early diagnosis of Alzheimer’s disease on MRI images.," *Multimed Tools Appl,* 2023.

[15] w. A. Shehri, "Alzheimer's disease diagnosis and classification using deep learning techniques.," *PeerJ.Computer science, 8, e1177,* 2022.

[16] S. Sharma, S. Gupta, D. Gupta, S. Juneja, A. Mahmoud, S. El–Sappagh and K.-S. Kwak, "Transfer learning-based modified inception model for the diagnosis of Alzheimer's disease.," *Front Comput Neurosci. 2022,* 2022.

[17] D. AlSaeed and S. Omar, "Brain MRI Analysis for Alzheimer’s Disease Diagnosis Using CNN-Based Feature Extraction and Machine Learning.," *Sensors ,* 2022.

[18] A. Ebrahimi and S. Luo, " Convolutional neural networks for Alzheimer's disease detection on MRI images.," *J Med Imaging (Bellingham),* 2021.

[19] V. Patil, M. Madgi and A. Kiran, "Early prediction of Alzheimer's disease using convolutional neural network: a review.," *Egypt J Neurol Psychiatry Neurosurg 58,* 2022.

[20] S. Fathi, A. Ahmadi, A. Dehnad, A.-D. Mostafa and M. Sadegh, "A deep learning-based ensemble method for early diagnosis of Alzheimer's disease using MRI images.," *Research Square,* 2023.

[21] A. El-Latif, S. Chelloug, M. Alabdulhafith and M. Hammad, "Accurate Detection of Alzheimer’s Disease Using Lightweight Deep Learning Model on MRI Data.," *Diagnostics ,* 2023.

[22] Kaggle, "Alzheimer's Dataset (4 class of Images)," 2019. [Online]. Available: https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images.

[23] D. AlSaeed and S. F. Omar, "Brain MRI Analysis for Alzheimer's Disease Diagnosis Using CNN-Based Feature Extraction and Machine Learning.," p. 2911, 2022.

[24] H. Helaly, M. Badawy and A. Haikal, "Deep Learning Approach for Early Detection of Alzheimer's Disease," *Cogn Comput,* p. 1711–1727, 2022.

[25] J. Wen, E. Thibeau-Sutre, M. Diaz-Melo, J. Samper-González, A. Routier, S. Bottani, D. Dormont, S. Durrleman, N. Burgos and O. Colliot, "Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation," *Medical Image Analysis,* vol. 63, 2020.

[26] A. E, P. K and B. B, "Detection of Alzheimer ' s disease and dementia states based on deep learning from MRI images: a comprehensive review," *Journal of the Institute of Electronics and Computer,* pp. 39-53, 2020.

[27] Y. Yang, X. Li, P. Wang, Y. Xia and Q. Ye, "Multi-Source transfer learning via ensemble approach for initial diagnosis of Alzheimer's disease," *IEEE J Transl Eng Heal Med,* pp. 1-10, 2020.

[28] C. Patterson, "World Alzheimer Report 2018—The State of the Art of Dementia Research: New Frontiers," New Frontiers; Alzheimer's Disease International, London UK, 2018.

[29] J. Huang, P. C. M. V. Zijl, X. han, C. M. Dong, G. W. Y. Cheng, K. -H. Tse, L. Knutsson, L. Chen, J. H. C. Lai, E. X. WU, J. Xu and K. W. Y. Chan, "Altered d-

glucose in brain parenchyma and cerebrospinal fluid of early Alzheimer's disease detected by dynamic glucose-enhanced MRI," *Science Advances,* vol. 6, no. 20, 2020.

[30] G. Castellazzi, M. G. Cuzzoni, M. C. Ramusino, D. Martinelli, F. Denaro, A. Ricciardi, P. Vitali, N. Anzalone, S. Bernini, F. Palesi, E. Sinforiani, A. Costa, G. Micieli and E. D. et.al, "A Machine Learning Approach for the Differential Diagnosis of Alzheimer and Vascular Dementia Fed by MRI Selected Features," *Front. Neuroinform,* vol. 14, 2020.

[31] E. Heerema, "How MRI Is Used to Detect Alzheimer's Disease," 12 September 2019. [Online]. Available: https://www.verywellhealth.com/can-an-mri-detect-alzheimers-disease-98632.

[32] W. C. Groves, J. Brandt., M. Steinberg, A. Warren, A. Rosenblatt, A. Baker and Constantine G. Lyketsos, "Vascular dementia and Alzheimer's disease: is there a difference? A comparison of symptoms by disease duration.," *J. Neuropsychiatry Clin. Neurosci.,* pp. 305-315, 2000.

[33] W. A. Shehri, "Alzheimer's disease diagnosis and classification using deep learning techniques," *PeerJ Comput Sci.,* 2022.

[34] A. Kumar, J. Sidhu, A. Goyal and J. W. Tsao., Alzheimer disease, Treasure Island: StatPearls Publishing, 2018, pp. 1-27.

[35] R. Kalaria, "Similarities between Alzheimer's Disease and Vascular Dementia," *J. Neurol. Sci.,* pp. 29-34, 2002.

[36] A. A. A. El-Latif, S. A. Chelloug, M. Alabdulhafith and M. Hammad, "Accurate Detection of Alzheimer's Disease Using Lightweight Deep Learning Model on MRI Data.," *Diagnostics (Basel),* 2023.

[37] E. Sarasso, A. Gardoni, N. Piramide, M. A. Volontè, E. Canu, A. Tettamanti, M. Filippi and F. Agosta, "A Multiparametric MRI Study of Structural Brain Damage in Dementia with Lewy Bodies: A Comparison with Alzheimer's Disease.," *Park. Relat,* pp. 154-161, 2021.

[38] D. A. Arafa, H. E.-D. Moustafa, H. A. Ali and A. Ali-Eldin, "A deep learning framework for early diagnosis of Alzheimer's disease on MRI images," *Multimedia Tools and Applications,* pp. 1-33, 2023.

[39] A. U. R. Butt, W. Ahmad, R. Ashraf, M. Asif and S. A. Cheema, "Computer Aided Diagnosis (CAD) for Segmentation and Classification of Burnt Human skin," *International Conference on Electrical, Communication, and Computer Engineering (ICECCE),* pp. 1-5, 2019.

[40] A. W. Salehi, P. Baglat and G. Gupta, "Alzheimer's disease diagnosis using deep learning techniques.," *International Journal of Engineering and Advanced Technology.,* pp. 874-880, 2020.

[41] D. A. Arafa, H. E.-D. Moustafa, H. A. Ali and A. Ali-Eldin, "A deep learning framework for early diagnosis of Alzheimer's disease on MRI images," *Multimedia Tools and Applications,* pp. 1-33, 2023.

[42] B. Dubois, H. Hampel, H. H. Feldman, P. Scheltens, P. Aisen, S. Andrieu, H. Bakardjian, H. Benali, L. Bertram, K. Blennow, K. Broich, E. Cavedo, S. Crutch, J.-F. Dartigues and C. Duyckaerts, "Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria.," in *Alzheimers Dement*, Washington DC, USA, 2016.

[43] K. Punsky, "Mayo Clinic Researchers Find Way to Prevent Accumulation of Amyloid Plaque, a Hallmark of Alzheimer's Disease," Mayo Clinic , 31 March 2016. [Online]. Available: https://newsnetwork.mayoclinic.org/discussion/mayo-clinic-researchers-find-way-to-prevent-accumulation-of-amyloid-plaque-a-hallmark-of-alzheimers-disease/. [Accessed 10 February 2024].

[44] J. V. Shanmugam, B. Duraisamy, B. C. Simon and P. Bhaskaran, "Alzheimer's disease classification using pre-trained deep networks," *Biomedical Signal Processing and Control,* 2022.

[45] G. D. Gregorio, D. Desiato, A. Marcelli and G. Polese, "A Multi Classifier Approach for Supporting Alzheimer's Diagnosis Based on Handwriting Analysis," in *Pattern Recognition. ICPR International Workshops and Challenges*, 2021, pp. 559-574.

[46] S. Fathi, A. Ahmadi, A. Dehnad, M. Almasi-Dooghaee, M. Sadegh and A. D. N. Initiative, "A Deep Learning-Based Ensemble Method for Early Diagnosis of Alzheimer's Disease using MRI Images.," *Neuroinformatics,* 2023.

[47] H. Sun, A. Wang, W. W. and C. Liu, "An Improved Deep Residual Network Prediction Model for the Early Diagnosis of Alzheimer's Disease," *Sensors,* vol. 21, no. 12, 2021.

[48] S. Basheera and M. S. S. Ram, "Deep learning based Alzheimer's disease early diagnosis using T2w segmented gray matter MRI," *International Journal of Imaging Systems and Technology,* p. 1692–1710, 2021.

[49] F. Caso, F. Agosta, P. G. Scamarcia, S. Basaia, E. Canu, G. Magnani, M. A. Volontè and M. Filippi, "Multiparametric MRI Study of Structural Brain Damage in Dementia with

Lewy Bodies: A Comparison with Alzheimer's Disease.," *Park. Relat. Disord.,* p. 154–161, 2021.

[50] G. Simic, M. G. Stanic, N. Mladinov, I. Jovanov-Milosevic and P. R. H. Kostovic, "Does Alzheimer's Disease Begin in the Brainstem?," *Neuropathol. Appl. Neurobiol.,* pp. 532-554, 2009.

[51] H. Ji, Z. Liu, W. Q. Yan and R. Klette, "Early Diagnosis of Alzheimer's Disease Using Deep Learning," in *In Proceedings of the 2nd International Conference on Control and Computer Vision*, New York, USA, 2019.

[52] E. Pellegrini, L. Ballerini, M. D. C. V. Hernandez, F. M. Chappell, V. González-Castro, D. Anblagan, S. Danso, S. Muñoz-Maniega, D. Job, C. Pernet, G. Mair, T. J. MacGillivray, E. Trucco and W. J. M., "Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: A systematic review," *Alzheimers Dement (Amst),* pp. 519-535, 2018.

[53] H. -I. Suk and D. Shen, "Deep learning-based feature representation for AD/MCI classification.," in *International Conference on Medical Image Computing and Computer-Assisted Intervention,*, Nagoya, Japan, 2013.

[54] G. D. Gregorio, D. Desiato, A. Marcelli and G. Polese, "A Multi Classifier Approach for Supporting Alzheimer's Diagnosis Based on Handwriting Analysis," in *Pattern Recognition. ICPR International Workshops and Challenges*, Researchgate, 2021, pp. 559-574.

[55] M. Liu, D. Zhang, E. Adeli and D. Shen, "Inherent Structure-Based Multiview Learning With Multitemplate Feature Representation for Alzheimer's Disease Diagnosis," *IEEE Transactions on Biomedical Engineering,* pp. 1473-1482, 2016.

[56] I. Krashenyi, A. Popov, J. Ramirez and J. M. Gorriz, "Fuzzy computer-aided diagnosis of Alzheimer's disease using MRI and PET statistical features," in *IEEE 36th International Conference on Electronics and Nanotechnology (ELNANO)*, Kyiv, UKraine, 2016.

[57] L. Lazli, M. Boukadoum and O. A. Mohamed, "Computer-Aided Diagnosis System for Alzheimer's Disease Using Fuzzy-Possibilistic Tissue Segmentation and SVM Classification.," in *IEEE Life Sciences Conference (LSC)*, Montreal, QC, Canada, 2018.

[58] K. T. N.P. and D. Varghese, "A Novel Approach for Diagnosing Alzheimer's Disease Using SVM.," in *In Proceedings of the 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)* , Tirunelveli, India, 2018.

[59] N. Zeng, H. Li and Y. Peng, "A new deep belief network-based multi-task learning for diagnosis of Alzheimer's disease," *Neural Comput. Appl,* p. 11599–11610, 2021.

[60] S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, D. Feng and M. J. Fulham, "ADNI Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease.," *IEEE Trans. Biomed. Eng.,* p. 132–1140, 2015.

[61] S. Korolev, A. Safiullin, M. Belyaev and Y. Dodonova, "Residual and plain convolutional neural networks for 3D brain MRI classification.," in *In Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging*, Melbourne, VIC, Australia,, 2017.

[62] K. A. N. N. P. Gunawardena, R. N. Rajapakse and N. D. Kodikara, "Applying convolutional neural networks for pre-detection of alzheimer’s disease from structural MRI data.," in *In Proceedings of the 2017 24th International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*, Auckland, New Zealand, 2017.

[63] L. Lin, B. Zhang and S. Wu, "Hybrid CNN-SVM for Alzheimer’s Disease Classification from Structural MRI and the Alzheimer’s Disease Neuroimaging Initiative (ADNI).," in *International Conference on Biomedical Engineering, Machinery and Earth Science*, UK, 2018.

[64] T. M. Ghazal, S. Abbas, S. Munir, M. A. Khan, M. Ahmad, G. F. Issa, S. B. Zahra, M. A. Khan and M. K. Hasan, "Alzheimer Disease Detection Empowered with Transfer Learning," *Computers, Materials & Continua,* 2021.

[65] K. Bäckström, M. Nazari, I. Y.-H. Gu and A. S. Jakola, "An efficient 3D deep convolutional network for Alzheimer’s disease diagnosis using MR images.," in *In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging*, Washington, DC, USA, 2018.

[66] H. Ji, Z. Liu, W. Q. Yan and R. Klette, "Early Diagnosis of Alzheimer’s Disease Using Deep Learning.," in *In Proceedings of the 2nd International Conference on Control and Computer Vision, Jeju, Korea*, New York, NY, USA, 2019.

[67] Y. Chen, L. Wang, B. Ding, J. Shi, T. Wen, J. Huang and Y. Ye, "Automated Alzheimer's disease classification using deep learning models with Soft-NMS and improved ResNet50 integration," *Journal of Radiation Research and Applied Sciences,* 2024.

[68] Y. AbdulAzeem, W. M. Bahgat and M. Badawy, "A CNN based framework for classification of Alzheimer’s disease," *Neural Computing and Applications,* p. 10415–10428, 2016.

[69] Al-Adhaileh, "Diagnosis and classification of Alzheimer’s disease by using a convolution neural network algorithm.," *Soft Comput,* p. 7751–7762, 2022.

[70] S. Savaş, "Detecting the Stages of Alzheimer’s Disease with Pre-trained Deep Learning Architectures," *Arabian Journal for Science and Engineering,* vol. 47, no. 2, p. 2201–2218, 2022.

[71] F. Antony, H. B. Anita and J. A. George, "Classification on Alzheimer’s Disease MRI Images with VGG-16 and VGG-19," in *IOT with Smart Systems. Smart Innovation, Systems and Technologies*, Singapore, 2023.

[72] S. Sharma, K. Guleria, S. Tiwari and S. Kumar, "A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer Disease using MRI scans," *Measurement: Sensors,* 2022.

[73] J. Zhang, L. A. M. Liu, Y. Gao and D. Shen, " “Alzheimer’s disease diagnosis using landmark-basedfeatures from longitudinal structural MR images,”," *IEEE Journal of Biomedical and Health Informatics,* vol. 21, no. 6, p. 1607–1616, 2017.

[74] A. Pradhan, J. Gige and M. Eliazer, "Detection of Alzheimer's Disease (AD) in MRI Images using Deep Learning," *International Journal of Engineering Research & Technology (IJERT),* 2021.

[75] S. Hussain, S. S. Oyelere, J. H. Joloudari, A. Marefat and M. A. Nematollahi, "Effective Class-Imbalance learning based on SMOTE and Convolutional Neural Networks," *Artificial Neural Network,* pp. 1-38, 2022.

[76] S. Hussain, S. Oyelere, J. Hassannataj Joloudari, A. Marefat and M. A. Nematollahi, "Effective Class-Imbalance learning based on SMOTE and Convolutional Neural Networks.," *Researchgate,* pp. 1-38, 2022.