Multi-Classification based Alzheimer's Disease Detection with Comparative Analysis from Brain MRI Scans using Deep Learning

by

Azmain Kabir
18101576
Farishta Kabir
18101697
Md. Abu Hasib Mahmud
18101189
Sanzida Alam Sinthia
18101219
S. M. Rakibul Azam
18101123

A thesis submitted to the Department of Computer Science and Engineering in partial fulfillment of the requirements for the degrees of Bachelor of Science in Computer Science and Engineering or Bachelor of Science in Computer Science

Department of Computer Science and Engineering
School of Data and Sciences
Brac University
October 2021

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing the undergraduate 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 that has been accepted or submitted for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Azmain Kabir

Azmain Kabir
18101576

Farishta Kabir

Farishta Kabir
18101697

Md. Abu Hasib Mahmud
18101189

Sanzida

Sanzida Alam Sinthia
18101219

S. M. Rakibul Azam
18101123

Approval

The thesis/project titled "Multi-Classification based Alzheimer's Disease Detection with Comparative Analysis from Brain MRI Scans using Deep Learning" submitted by

1. Azmain Kabir (18101576)
2. Farishta Kabir (18101697)
3. Md. Abu Hasib Mahmud (18101189)
4. Sanzida Alam Sinthia (18101219)
5. S. M. Rakibul Azam (18101123)

Of Summer 2021 has been accepted by the Department of Computer Science and Engineering as satisfactory partial fulfillment of the requirements for the degrees of Bachelor of Science in Computer Science and Engineering or Bachelor of Science in Computer Science on October 2021.

Examining Committee:

Supervisor:
(Member)

Zavid Parvez
Digitally signed by Zavid Parvez
DN: cn=Zavid Parvez, o=Brac University, ou=CSE, email=zavid.parvez@bracu.ac.bd, c=US
Date: 2021.10.05 11:15:50 +11'00'

Dr. Mohammad Zavid Parvez
Former Assistant Professor
Department of Computer Science and Engineering
School of Data and Sciences
Brac University

Co-Supervisor:
(Member)

Emtiaz Hussain
Software Development Engineer, ZF Group
CEO, HashTech360

Program Coordinator:
(Member)

Dr. Golam Rabiul Alam
Associate Professor
Department of Computer Science and Engineering
School of Data and Sciences
Brac University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi
Associate Professor
Department of Computer Science and Engineering
School of Data and Sciences
Brac University

Abstract

The neurodegenerative Alzheimer's Disease is the most widely recognized cause of 'Dementia' and was allegedly the 7th highest cause of death globally. Nevertheless, there is still no conclusive test for distinguishing Alzheimer's disease. Our proposed model eliminates these challenges in an effective manner. The technique fits and analyzes different classes in a single setting and requires significantly less previous apprehension. Several handcrafted or predefined machine learning and deep learning models have been implemented in this field of study. Our proposed multi-classification model is primarily implemented based on the Open Access Series of Imaging Studies (OASIS) data and suggests an 18-layer architecture. We have implemented a unique preprocessing approach of all three anatomical planes of the MRI scans in a single sequential model, which was also evaluated afterward. The research also explores a comparative study among multiple and binary classes in terms of performance and efficiency. Predefined models such as InceptionV3 and VGG19 have also been brought to comparison to measure the model's reliability. Our multiclass setting shows an accuracy of over 80%, which is higher than most of the existing multi-classification models in this dataset. Moreover, the in-depth comparative study using binary classification shows a significant accuracy of over 92%, which ensures the overall efficacy of the model.

Keywords: Alzheimer's Disease, CNN, Multi-class, Binary Class, MRI, Deep Learning, Early Detection, Comparative Analysis, 18-layer, 3D Scans, OASIS-1.

Dedication

We dedicate our work for the improvement and the betterment of Alzheimer's Disease research. We hope small contributions like this might bring significant changes in the future.

Acknowledgement

First and foremost all glory and praise to the Almighty Allah for whom our thesis have been completed without any major interruption.

Secondly, to our Supervisor Dr. Mohammad Zavid Parvez Sir, for his kind advice and support in our research. Also, our Co-Supervisor Emtiaz Hussain. Their guidance paved our path from the very beginning till the conclusion.

Thirdly, the utmost gratitude to our parents without their selfless support we couldn't have reached this far in the long run.

Last but not least, the admirable team harmony has helped us to make it to the end.

P.S: The research has been conducted on the OASIS-1 dataset. The grants are P50 AG05681, P01 AG03991, P01 AG026276, R01 AG021910, P20 MH071616, U24 RR021382.

Table of Contents

List of Figures

List of Tables

Nomenclature

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

AD Alzheimer's Disease

$ADNI$ Alzheimer's Disease Neuroimaging Initiative

CDR Clinical Dementia Ratio

CNN Convolutional Neural Network

MCI Mild Cognitive Impairment

MRI Magnetic Resonance Imaging

$OASIS$ Open Access Series of Imaging Studies

VGG Visual Geometry Group

Chapter 1

Introduction

1.1 Introduction

Alzheimer's disease is a neurodegenerative disease that progresses over time, mostly in older adults. Alzheimer's is regarded as the most common cause of dementia. An umbrella term for cognitive degradation that causes memory loss slowly results in the loss of ability to do daily tasks, thinking, and analytical ability or even as easy tasks as carrying out a conversation. The researchers have identified different patterns in terms of gender, age, and genetic factors. The disease is named after German psychiatrist and pathologist Alois Alzheimer. He was the one to first identify the condition in 1906 [1].

Studies tell us that approximately 29.8 million people were diagnosed with the disease within 2015 [2]. Moreover, the number is about 50 million as of 2020 [3]. Most people over the age of 65 are diagnosed with the disease [4]. This disease is not only the 7th highest reason for death globally, with a shocking number of 29.8 million people [5]. The disease is also said to be one of the most expensive diseases regarding therapy and treatment, along with the long process of diagnosis [6] .

The most unfortunate fact is, the cause of Alzheimer's disease is not deciphered yet. Many factors are thought to play a significant role in terms of the disease. About 70% of which is regarded as genetic. Other factors that increase the disease's chances are severe head injuries, chronic depression, hypertension, and other cognitive disorders. According to the authors Ballard C, Gauthier S, Corbett A, Brayne C, Aarsland D, Jones E (2011) of the journal "Alzheimer's disease" the process of the disease is deeply associated with the elements amyloid-beta (A) plaques, along with some neurofibrillary tangles of the brain [7]. The disease shows the symptoms more severely as it progresses over time. It creates barriers in essential communication, language usage, disorientation, unsynchronized physical activities, and deep cognitive thinking. Mood swings, loss of motivation, inferiority complex, self-harm, insomnia, anxiety attacks, behavioral issues are pretty common also. Many people are even forced to get abandoned by their families and society as the situation declines.[8] Gradually, the person loses the ability to conduct physical and cognitive activities totally and falls into the pit of death. Even though the progression speed varies from person to person, the typical life expectancy is 3 to 9 years of an average patient [9][10].

Usually, cognitive and neurological tests are done along with in-depth medical history research and brain imaging scans like MRI/fMRI, C.T., or PET. Nevertheless, there is no proper conclusive test to detect or analyze the disease pattern in a single setting with higher efficiency and accuracy. So, there is a thriving need for a modernized, efficient automatic solution for addressing these issues only to make diagnosing Alzheimer's disease more fruitful.

1.2 Research Problem

The biggest concern for detecting and analyzing the pattern of Alzheimer's Disease is efficiently conducting the whole process through a single setting without any manual help. A significant drawback of this neurodegenerative disease is that the claims of Alzheimer's can only be certainly understood, and the patterns can be thoroughly analyzed after the person's death through autopsy. The diagnosis generally includes manual assumption-based cognitive tests, neurological tests, and scans. However, most of the time, the scans do not provide conclusive data, and it is hard to detect the problem just by observing the scans individually automatically early. Different existing approaches have their shortcomings and weaknesses that we have tried to identify and work upon. Also, keeping in mind what problems we need to go through while conclusively addressing Alzheimer's Disease.

It is already acknowledged that CNN (Convolutional Neural Networks) performs better than any other approach to address this particular problem, but it requires an enormous amount of data. The authors of the 'Transfer Learning with Intelligent Training Data Selection for Prediction of Alzheimer's Disease' Khan, N. M., Abraham, N., Hon, M. (2019) stated that most feature extraction techniques most transfer learning is the core of their methodology. It focuses on storing knowledge gained while using a machine learning model with a dataset and applying it to a different. However, related problem [11]. Therefore, they used transfer learning here by using a pre-trained model to learn new image representations. The problem was that the vast dataset hindered efficiency, which was a scope to work on to improve the accuracy with the minimal dataset. Authors Sivakani, R., and Ansari, G. A. (2020) proposed that they employed the OASIS longitudinal MRI dataset for classification, which comprises a dataset of brain reports from all ages of AD afected people [12]. The data must be pre-processed by addressing issues such as missing data, inconsistent data, noisy data, incorrect data, outlier data. The research implies that improving categorization could be done more effectively by focusing on feature extraction.

In terms of the existing models, many models do not target different age groups while proposing a model. This situation creates a barrier while making the model suitable for all. S. Sarraf, G. Tofighi (2016) have proposed a model where they have used a DL based CNN model to detect the disease [13] early. However, they kept a scope to extend their research and make enhancements in their model by generalizing this method to predict the disease for different age groups.

Billones, C. D., Demetria, O. J. L. D., Hostallero, D. E. D., and Naval, P. C. (2016)

addressed the model that could achieve accuracies without dividing the gray matter, white matter, and cerebrospinal fluid. This suggests that this characterization problem may not be affected by previous area data and that educated neighborhood highlights can be isolated to illustrate differences between arrangements [14]. Nevertheless, this also leaves room for skepticism, and the work can be extended by focusing on separately classifying the brain based on their sections and identifying how the disease affects the brain. This can show us the pattern of the damage caused in the brain that can help with the early detection for the following cases.

Another big issue with the existing models is data loss or degradation of the brain fMRI/MRI scans that causes less accuracy. For example, authors Mahmood, R., Ghimire, B. (2013) had used the principal component analysis, and artificial neural networks for their proposed approach [15]. The accuracy could have been increased by improving the dimensions classifications of the vector space and increasing the dataset sample. The multi-level automated system offers a lot of room for improvement in making an early detection system that is more easily accessible and accurate. Individual characteristics and brain parts could also be investigated to see how Alzheimer's disease affects a specific part of the brain. However, they did propose that they work on the ratio of ventricle enlargement to hippocampal loss as a better way to identify Alzheimer's disease progression, which could be a great way to improve the current technique.

Following the deep learning-based neural network-related initiatives, feature extraction techniques are becoming increasingly popular. For example, Acharya, U. R., Fernandes, S. L., WeiKoh, J. E., Ciaccio, E. J., Fabell, M. K. M., Tanik, U. J., and Yeong, C. H. (2019) claimed a remarkable accuracy in their approach published in the journal 'Automated Detection of Alzheimer's Disease Using Brain MRI Images– A Study with Various Feature Ex When compared to previous approaches, this one uses four techniques [16]. This is more accurate than current automated Alzheimer's disease diagnosis and analysis procedures. This provides more accuracy than existing automated detection and analysis processes of Alzheimer's Disease. Nevertheless, instead of employing KNN techniques, other classification methods such as SVM, neural networks, random forest, and AdaBoost can provide better results. Additionally, using the right deep learning technology can improve accuracy and performance.

Observing most of the related works shows that Faster R-CNN is better among all object detection predefined models. As R-CNN takes more training time than all other models, they claimed faster R-CNN as best in their model. Authors Ahmad, I., Pothuganti, K. (2020) analyzed why Faster R-CNN is better than R-CNN where they compared training time and testing time as performance parameters [17]. Fast RCNN is also good, but to calculate regional proposals, it uses selective search . However, this approach also falls short in one sector. That is, it consumes more time, and as a result, the process becomes inefficient for regular use. So, the further points to be worked on can be how to minimize the algorithm's execution time and keep the accuracy mark higher, which can be improved by training more relevant data.

Therefore, the core problems we have been able to address are: improving the existing algorithm of the approaches to increase the efficiency, significantly decreasing the data loss, focusing more on the individual section for identifying a pattern on the brain caused by the disease.

1.3 Research Objectives

The objective of our research is to provide a suitable model that will accurately detect the disease and provide a clear pattern through which the disease can be better understood at a primary stage. The work can be extended through the slow progression of the patient's condition by following the changes in the pattern. This can help the manual diagnosis of Alzheimer's disease in a significant way, and along with other neurological tests and observations, this can help the health sector to detect such a complex neurological disease in the early stages for better treatment. So, the primary goals we are trying to reach are:

- Detecting the Alzheimer's disease in from brain MRI scans without any direct manual assistance
- Analyzing a clearer classification on the severity level of the brain caused by the disease
- Improving the efficiency of the existing approaches in terms of time and accuracy
- Providing a generalized model fit for all cases of Alzheimer's disease
- Conducting a comparative study to obtain a deep understanding of the relevant factors to decipher the disease better

Chapter 2

Literature Review

Alzheimer's disease being one of the most threatening causes of death that is slowly progressing over time among the elderly demographic, a better understanding and research study are needed to be conducted in this field. This will help us understand the complexity of the disease better and help us with the appropriate diagnosis that can save valuable lives in this era of modern science.

2.1 Alzheimer's Disease

Alzheimer's is a disease that causes changes in the brain at a microscopic level that starts happening way before the first symptom occurs. Technically this is allegedly most usual cause of dementia which is about 60-70%. The human brain consists of approximately 100B neurons connected for our entire body execution, memory keeping, and regulating physical and cognitive activities. These cells conduct different work to operate different missions, including managing our senses and producing energy. To keep this synchronization, the neurons need ample amounts of oxygen and fuel. The researchers and scientists found out that Alzheimer's attacks the neurons for conducting such operations well by creating a hindrance in their oxygen supply, and this causes the nerves to break down. This breakdown shows as a symptom primarily as a memory loss. Later on, it moves into having trouble overdoing the simplest regular tasks. Even though this often happens among older adults, the damage it creates in the brain can be fatal over time, and it is also one of the biggest reasons for death in the world.

2.1.1 Alzheimer's Disease Classification

Alzheimer's disease is arguably one of the most complex understand the neurological disease that completely shatters the ability of the brain to work smoothly. Unfortunately, there is no conclusive or advanced efficient way to detect the disease in an early stage that can be used for mass use without manual help or prior knowledge. As long as the symptoms occur, the doctors or neuro-specialists usually check the history and conduct a manual interview session with the patient's close ones. Then cognitive testing and petrophysical testing are done. Laboratory tests are also conducted to rule out any other disease that shows similar symptoms. Ultimately the brain image scans are taken to provide an in-depth view of the problem. Usually, the most popular one is MRI or (Magnetic resonance imaging) and fMRI (Functional Magnetic resonance imaging) that uses powerful radio waves and magnetic fields to provide a detailed view of the brain. Then comes Computerized tomography (CT), which uses X-rays to achieve cross-sectional images of the brain. Also, PET (Positron emission tomography) is used. A PET scan is conducted through the radioactive properties as a tracer to detect any anomaly in the body part. After all of these physical tests, it is decided whether the patient has Alzheimer's or not. Nevertheless, all of this cannot diagnose or detect the disease at an early age with much accuracy. It is also said that the pattern of the disease can only be deciphered after the proper autopsy of the brain once the patient is dead.

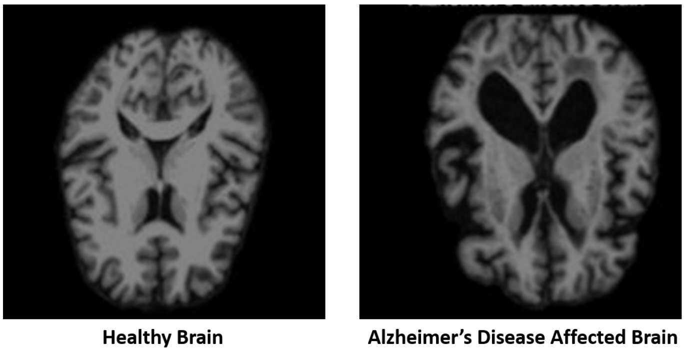

Figure 2.1: A OASIS Data Sample (MRI Scan of a Healthy and AD Affected Brain)

2.2 Related Works

A great deal of research has been done on the subject and many technological ways to improve accuracy and efficiency. All of these have their own set of capabilities and potential research areas for expanding the study. Some highlight the flaws of various models and methodologies, as well as the limitations of existing research. We concentrated on automatic detection and analysis of relevant components. Diagnosis of the disease from neuroimaging data like MRI, using ML has recently been an intense topic of interest. However, the reliance on many training images and the cautious application of machine learning models are major roadblocks for such methods. The authors attempted to solve these issues in the paper 'Transfer Learning with Intelligent Training Data Selection for Prediction of Alzheimer's Disease,' in which a different architecture is initialized with a pre-trained model built up from a large benchmark dataset consisting of natural images [11]. Experiments on the ADNI dataset revealed that the training size is decreased by 10 to 20 times when compared to current approaches. They achieved state-of-the-art performance in the classification issues. Finally, they provided class activation maps (CAM), which show how the suggested model focuses on nerve pathologically important images and can aid the healthcare practitioner in deciphering the model's decision-making process.

The proposed method by Acharya, U. R., Fernandes, S. L., Weikoh, J. E., Ciaccio, E. J., Fabell, M. K., Tanik, U. J., Yeong, C. H. in their paper was dependent on a Computer Aided Brain Diagnosis system where early detection of Alzheimer's can be possible through MRI classification with different feature extraction techniques [16]. The objective is to detect the severity of the brain damage caused by

Alzheimer's Disease in an automated and non-manual way. Here they have investigated T2 weighted brain MRI the axial images by comparing data with and without AD. This provides an in-depth comparison with the normal and affected brain. The dataset consisted of 66 two-dimensional (2D) test images of dimension 256×256 pixels. Then the pre-processing was done to remove noise and other defects from the test dataset, and it was done by implementing a median filtering algorithm. After that, the features were collected through data mining, and the primary features were investigated with a p-value from the student's test. Lastly, the KNN classifier process was done to classify the dataset images. Here, the proposed system had been validated with the help of a few factors such as the percentage values of the metrics. The ST + KNN technique provided 94.54% accuracy, 88.33% precision, 96.30% sensitivity and 93.64% specificity. With the benchmark MRI database, the tool provided the specificity of 98.48%, 100%, 96.97%, and 100%, respectively.

Authors Sivakani, R., & Ansari, G. A. wrote in their literature about a Machine Learning framework that is essential for feature extraction, feature selection process, and classification of Oasis longitudinal dataset [12]. Dr. Alois Alzheimer first discovered Alzheimer's disease signs in a woman in 1906. That Alzheimer's patient had memory loss, as well as language issues and erratic behavior. Her brain was found to have anomalous clusters and a collection of fibers after the tests. Later, researchers looked into it, and the disease was given the term Alzheimer's. Amyloid plaques are the aberrant aggregates, and neurofibrillary fibers are the fibers. They also stated that the sickness begins in the hippocampus and progresses throughout the brain. As a result, neurons die, and brain tissues decrease, resulting in complete memory loss. MCI is the first stage of AD. The subsequent two phases will progress to AD, which is extremely difficult to treat. They used the EM Algorithm to cluster the data in this research. After replacing the missing values, the EM method was applied to the entire training set data. Then there is the matter of choosing features. This is a method of selecting a necessary feature from a list of extracted features. The best-first search strategy was used, and the cfssubseteval evaluator was used to evaluate the subset. Finally, they used the Gaussian process approach to classify the Oasis dataset and created the Linear Regression Model. The Gaussian process algorithm is used to check the Mean Absolute Error (MAE), Correlation Coefficient (CC), and Root Relative Squared Error (RRSE). Using the Decision stump technique, they classified the Oasis data set as a tree structure.

Authors Sarraf, S., & Tofighi, G., used Convolutional Neural Network (CNN) to classify Alzheimer's brain from the standard healthy brain [13]. Besides, they used a famous network called LeNet-5. They also stated two critical concepts about resting-state fMRI. Firstly, for data acquisition and pre-processing, they selected AD patients and 15 healthy adults consisting of 24 female and 19 male from the ADNI dataset. They pre-processed the fMRI data using classic modules of 'FMRIB Software Library v5.0'. The Brain Extraction Tool removed non-brain tissue from T1 anatomical pictures as part of their pre-processing stages for the anatomical data. Motion correction (MCFLRIT), spatial smoothing and the skull stripping were used as pre-processing stages for functional data (Gaussian kernel of 5-mm FWHM). They used high-pass temporal filtering to remove the low-level noise. CNNs are similar to classic neural networks and are inspired by human visual systems. This architecture, they believe, was primarily created with the explicit premise that raw data is 2D images, allowing them to encode certain features and limit the number of hyperparameters. A small image area (double a local receptive field) is treated as input to the hierarchical structure's lowest layer in CNNs. In their model, the class scores are computed by a Fully-Connected Layer (FC) layer, which results in numerous classes. According to their article, the Convolutional Layer is a critical component of CNN design and its main building block. Finally, using the Neuroimaging package Nibabel and Python OpenCV, pre-processed fMRI 4D data in Nifi format were concatenated across the z and t axes and converted to a stack of 2D images in JPEG format. The photos were then labeled for binary categorization of Alzheimer's vs. Dementia. Data that is typical. When it came to binary picture categorization, LeNet was the tool of choice. They separated the data into three categories: training (60%), validation (20%), and testing (10%). They used 270900 samples to train LeNet and validated and tested 90300 pictures. They conducted the cross-validation method five times to ensure uniformity. Their Deep Learning LeNet model had a success rate of 96.8588 percent accuracy. According to the approach, the method can also be generalized to anticipate different phases of Alzheimer's disease for different age groups.

Authors Lodha, P., Talele, A., and Degaonkar, K. employed a machine learning model to accurately predict a person's Alzheimer's illness based on specified characteristics that included many cognitive and medical aspects [14]. RAVLT tests, MOCA and FDG scores, and other predictors were used. They chose ADNI2 from the Alzheimer's Disease Neuroimaging Initiative for the training data (ADNI). They looked at the data from the questionnaire using different metrics. Before looking at the transcripts, they focused on the coding structure. Different studies have accounted for gender differences in subcortical sizes, including enormous quantities in females. Men were shown to have more severe age-related volume loss in both cortical and subcortical areas than women. Separate investigations were conducted on men and women. They took MRI scans in this study and processed them to obtain numeric data, then analyzed using machine learning methods. The five machine learning methods were Vector Machine, Gradient Boosting Algorithm, Neural Network, K-Nearest Neighbor (KNN), and Random Forest. Other algorithms like Support Vector Machine showed an accuracy of 97.56 percent, Gradient Boosting Algorithm's accuracy was 97.25 percent, and K-Nearest Neighbor (KNN) showed an accuracy of 95 percent. In contrast, Neural Network and Random Forest showed much better acts inaccuracy than other methods, with respectively 98.36% and 97.86% in all five algorithms.

The application of CNN for the computer-aided diagnosis of AD and MCI is the topic of this study by authors Billones, C. D., Demetria, O. J., Hostallero, D. E., and Naval, P. C [18]. Mild cognitive impairment (MCI) is a condition in which a person's reasoning capacity shows some little changes that others can easily notice when close to the influential person. They worked with the ADNI dataset. Cortical reproduction and volumetric segmentation were performed utilizing the FreeSurfer image analysis package, which is recorded and publically accessible for download online, to minimize the needless details of the MRI brain scans that may cause the network to learn superfluous highlights. The scans were separated into three categories: healthy controls, mild cognitive impairment, and Alzheimer's disease. They adapted a VGG16 architecture for this experiment, using a pre-processing data pipeline to tackle the AD vs. MCI vs. HC classification challenge. DemNet was used to train the suggested CNN architecture in this work. It tested both 3 manner and binary classifiers, ultimately ranking HC's AD and MCI on the ADNI dataset with 91.85% accuracy, beating other CNN architectures on the ADNI dataset. In addition, the binary classifier achieved an accuracy of 98.33%, 93.89% and 91.67% for AD vs HC, AD vs MCI and MCI vs HC respectively. In addition, the study found that 17 slices in the central part of the brain were sufficient for classification. They achieved these results without zoning cerebrospinal fluid or gray and white matter, implying that this characterization problem probably won't be affected by previous regional data. To show the difference between the arrangements, it is possible to separate the highlights of educated neighbourhood.

Alzheimer's cannot be detected with much accuracy until the condition reaches a certain moderate level. This paper suggests an approach that improves the current Alzheimer's Detection methods that rely on cognitive impairment and helps to detect the disease early. The approach is based on mathematical and image processing methods [15]. Generally, the vector spaces are reduced to three-dimensional spaces, and the images get clustered together. But the MRIs being high dimensional vector spaces, when it gets reduced to the three-dimensional vector spaces, the accuracy deteriorates, and data get lost. So, this literature suggests reducing the MRI vector spaces specifically to 150 dimensions with the help of Principal Component Analysis. To detect the progression of Alzheimer's Disease, they have used reduced dimensions from PCA for the categorization process and employed a multiclass neural network. The neural network is based on initially diagnosed MRI scans from the OASIS MRI database, and about 230 data have been used. The system has been trained on 457 MRIs in total, and the AD diagnosis and classification accuracy is about 90%.

Accurate early-stage detection of Alzheimer's disease is obligatory for fruitful treatment and healing. Therefore, the precise identification of Alzheimer's disease is a substantial research problem. It was pointed out that different researchers use various techniques to detect Alzheimer's disease effectively; however, the accuracy of these models is not up to the mark. On that note, research by authors Memon, M. H., Li, J., Haq, A. U., titled 'Early Stage Alzheimer's Disease Diagnosis Method,' few commonly used models were demonstrated to compare the results, but it still showed a lack of expected accuracy [19]. To resolve the issue, a logistic regression model was used here. Logistic regression was used in their approach, a supervised machine learning procedure utilized to anticipate continuous values. A definitive objective of the regression algorithm is to plot a best-fit line or a bend between the data. The model was used on the Alzheimer's Disease Neuroimaging Initiative (ADNI) data set for experimental work. Ultimately, the performance gained an optimal accuracy of 98.12%.

Then comes a significant point by authors Ahmad, I., & Pothuganti, K. where they compared different machine learning languages such as Fast RCNN, Faster RCNN, Support Vector Machine (SVM) to diagnose Alzheimer's disease in their research [17]. They also analyzed the training time and testing time of different object detection algorithms. Machine learning approaches were followed to diagnose the disease. In their work, they used the ADNI Dataset. Along with AD, they had NC and MCI subjects in the dataset. Also, they categorized the data into four parts which are: 1. Mild Demented, 2. Very Mild Demented, 3. Non-Demented, 4. Moderate Demented. After training the model, they used images of the dataset for prediction. For the classification and regression problems, they used SVM, which gives good generalization performance. For classifying non-linearly separable data, they used kernel functions. They also used a selective search approach to merge similar regions into larger regions using a greedy algorithm. They got output from the CNN output layer for feature extraction, which extracts features from a given image. After that, those characteristics are given to SVM, which identifies the abnormal areas in the MRI image and predicts the name of the disease. From their paper, we can see that the Faster RCNN algorithm works significantly better among all the object detection algorithms. As it takes minimum time to diagnose, this kind of ML algorithm can be very effective for the diagnosis of AD.

Authors Islam, J., & Zhang, Y. (2018) proposed a deep convolutional neural network for Alzheimer's disease diagnosis using brain MRI data analysis [20]. The special feature of this model is that it can identify different stages of Alzheimer's disease and achieve outstanding performance for early diagnosis. Compared to most of the approaches, it is advanced because most of the approaches follow binary classification. They conducted many tests to prove that their proposed model passed the benchmarks on the Open Access Series of Imaging Studies dataset. They used the OASIS dataset, which was split into a training and test dataset with a ratio of 4:1. To manage imbalances in the dataset, they used a training method that was sensitive to cost. The proposed model is a 2D architecture. Since the input MRI is made up of 3D data, for each MRI data they generate patches from three physical image planes. They optimized the individual models using the Stochastic Gradient Descent (SGD) algorithm. For regularization, an early breakpoint was used in their model, and the training dataset was split into a training set and a cross-validation set at a ratio of 9:1. The accuracy of their proposed model is 93.18% with an accuracy of 94%, a recall of 93% and a score of 92%. As their proposed model gives highly accurate results, it can be said that their proposed approach is an effective approach to diagnose AD by analyzing brain MRI data. Furthermore, their model is advanced compared to the majority of works because their model has brought a significant improvement to multi-glass classification. This model can also be used for other medical classification problems.

Chapter 3

Proposed Methodology

3.1 Workflow

We will be using the following materials and models for this research work. As the input dataset, we have considered the Open Access Series of Imaging Studies (OASIS) as the primary source of our MRI image scans. The data set is divided into two parts, where one part is for the classifier's training and another part is for testing the classifier. Approximately 80% of the data is used for the training of the classifier and the rest of the 20% is used for the testing of the classifier. Our significant research has been conducted through the handcrafted preprocessing strategy. The preprocessing is done to standardize and normalize the dataset before approaching the ultimate step. It helped eliminate the unnecessary factors and obsolesce of the brain MRI scans to better train the data for unwanted cases. Four different classifiers have been utilized to detect the presence and severity level of Alzheimer's disease. Data labeling, data splitting phases are also conducted afterward to prepare the data for proper training and testing. The procedure was then completed with a customized 18-layer approach based on Convolutional Neural Networks, a sort of deep neural network most typically used to analyze visual images. CNN (Convolutional Neural Network) has been used in the process, which is a class of deep neural networks most commonly applied to analyzing visual imagery. The deep learning-based CNN algorithms will help us analyze the data to get an in-depth pattern of the brain affected by Alzheimer's disease. This will help us differentiate the brains carrying the signs of Alzheimer's disease from a healthy/non-AD-affected brain. If there is any anomaly found in the proposed multiple classification-based 18-layer CNN model, we can conclude that the particular pattern found in the data refers to an affected brain, and on the other hand, a healthy brain or unaffected brain will be classified separately. Lastly, we conducted a comparative study between the primary multiclass model and the binary class. We evaluated our model's performance through a comparison of the existing renowned architectures in both classifications. Our preprocessing strategy was also thoroughly evaluated by implementing data from the three anatomical planes separately and comparing it with our generalized strategy.

To summarize, the methodology proceeds by these following steps:

1. Dataset Collection

2. Data Pre-processing
3. Data Labeling
4. Data Splitting
5. Proposed 18-layer Convolutional Neural Network (CNN) Model
6. Demonstrating the Performance of Our Proposed Model and Result Analysis (Multiclass (Proposed) and Binary Class)
7. Evaluating the Preprocessing Strategy in Separate Anatomical Planes
8. Loading Pre-trained Convolutional Neural Network (CNN) Models
9. Comparison among the Proposed Model and the Pre-trained Convolutional Neural Network (CNN) Models (Multiclass (Proposed) and Binary Class)

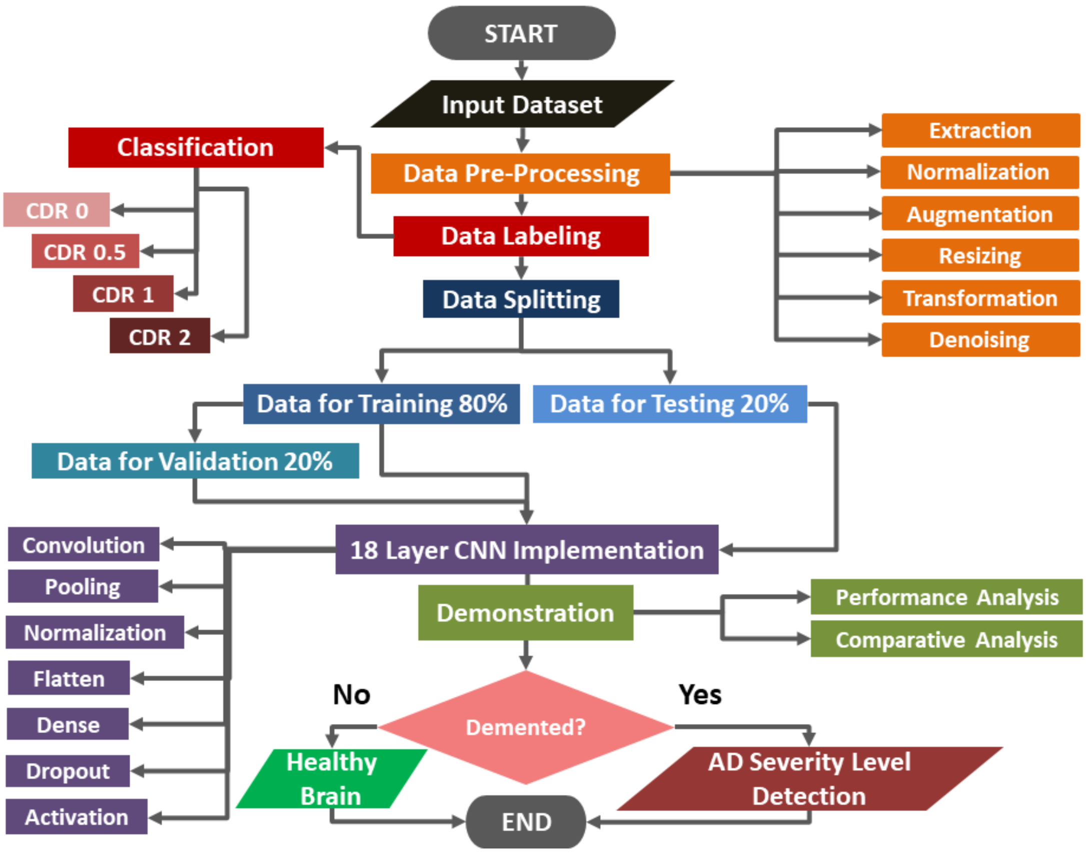

Figure 3.1: Workflow

In this process, whenever new data is tested or fed, this model will provide the tested result of detection of Alzheimer's disease based on its multiple classes that indicate the severity level and provide the scope to further in-depth analysis of the pattern of the effect in the brain.

3.2 Primary Contributions of the Research

Our significant research contribution was implemented through the handcrafted preprocessing strategy and a thorough study of the dataset. We tried to evaluate all the possible cases in different lights to measure the reliability of the proposed 18-layer model.

Till this date, on this research work, our primary contributions have been:

- Implementing our 18-layer architecture based on multiclass (four) classification shows a significant performance for this dataset (approximately over 80% accuracy).
- Evaluating the proposed model through two other existing architectures along with a comparative study on binary classes (Ahead in each case with approximately over 92% accuracy).
- Implementing our signature research contribution working with all three brain anatomical planes collectively through the proposed preprocessing strategy.

Chapter 4

Dataset

4.1 Dataset Collection

OASIS-1, which is the abbreviated form of Open Access Series of Imaging Studies [21], is our primary and only dataset source. The cross-sectional collection of 416 participants is included in this dataset. The participants' ages range from 18 to 96. About 100 of the 416 people in the study are over 60 and have mild to severe dementia. In addition, the dataset includes 20 non-demented people to establish dependability. Images from their visit within 90 days of the initial session are included in the dataset. The participants are all right-handed, male and female.

Each participant has two MRI scan files in the 'nii.gz' format, with many slices in each anatomical plane — sagittal, coronal, and transverse. From each subject's single session, three or four separate T1-weighted MRI scans were collected. T1 weighted image is also shortened as T1WI or even spin-lattice relaxation time. It demonstrates differences in tissue T1 relaxation times which is a simple MRI pulse-sequence.

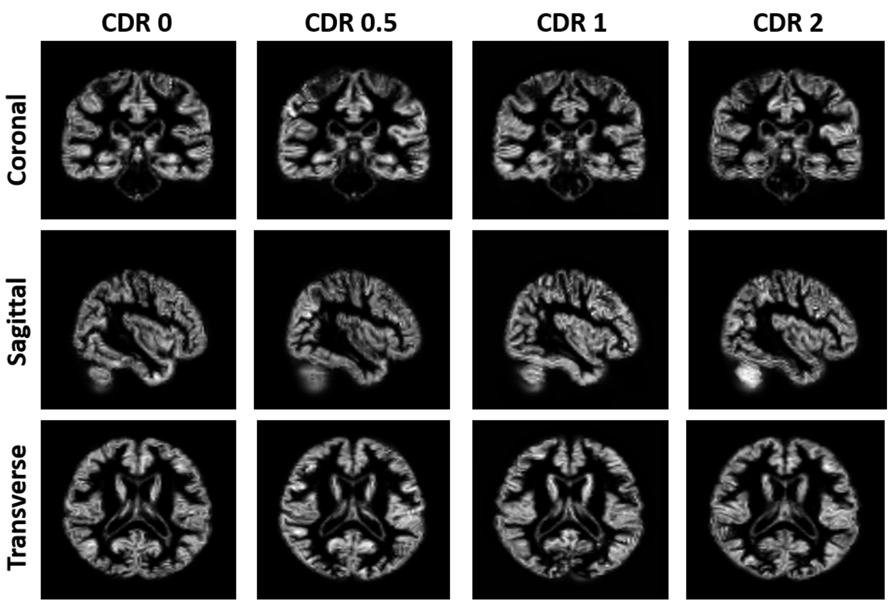

Figure 4.1: Brain Anatomic Planes in Four Classes (OASIS-1)

4.2 Data Preprocessing

The data preprocessing step is the highlight of the proposed methodology. Without proper standardization of the images, the architectures suffer from severe data loss and low accuracy.

First of all, we extracted our image files and implemented preprocessing parameters. The OASIS-1 dataset consists of 2 cross-sectional MRI scan files per patient. The file formats are in 'nii.gz,' which is the default MRI scan format. We dissected the MRI scans using the 'MRIcro' software to better study the anatomical dimensions of the scans. Each scan includes slices based on the three anatomical planes (sagittal, coronal, and transverse). The total extractable slices from each plane are $91, 109, 91$, respectively (sagittal, coronal, and transverse).

Unlike most existing models' approaches, we extracted all three anatomical plane-based slices from each sample. Here, the files were segmented based on their labeling process, which coincides with the preprocessing. Initially, we found 270, 140, 56, and 4 MRI scans from CDR 0, CDR 0.5, CDR 1, and CDR 2, respectively. Then the dataset is normalized for further balancing in the four classes, respectively. The extracted normalized data were augmented by shuffling. To normalize and make the dataset balanced, we considered one slice each from the three planes (CDR 0), two slices from the three planes (CDR 0.5), five slices from the three planes (CDR 1), Sagittal: 58 slices, Coronal: 66 slices, Transverse: 49 slices for per MRI Scan (CDR 2). The dataset was further processed using a programmed approach by eliminating the obsolete images or blank slices within a certain trivially determined range. So, the total slices after normalization were 798 for CDR 0, 822 for CDR 0.5, 785 for CDR 1, 692 for CDR 2.

Lastly, we conducted data augmentation for better and more precise training purposes. The normalization and randomization using augmentation ensure unbiasedness among the data. Moreover, the model gets trained more precisely in all four classes in an equal manner. It ultimately helps to prevent data loss and low accuracy while implementing the algorithm. Ultimately, the total amount stands at – **Class 1:** 9,942 (from 270 files) **Class 2:** 10,003 (from 140 files) **Class 3:** 9,913 (from 56 files) **Class 4:** 9,645 (from 4 files). It ends up in a total of 39,503 image data to work with from 416 MRI samples. The normalization and randomization along with augmentation ensure unbiased training for all classes. It ultimately helps to prevent data loss and low accuracy while implementing the algorithm.

Figure 4.2: Before Pre-processing (Random Samples)

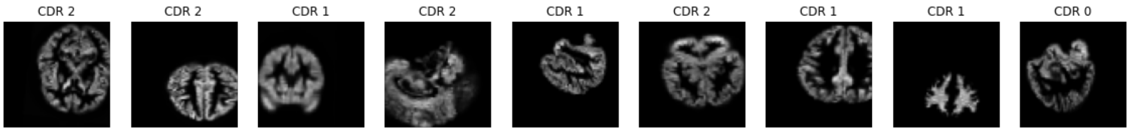

Figure 4.3: After Pre-processing (Random Samples)

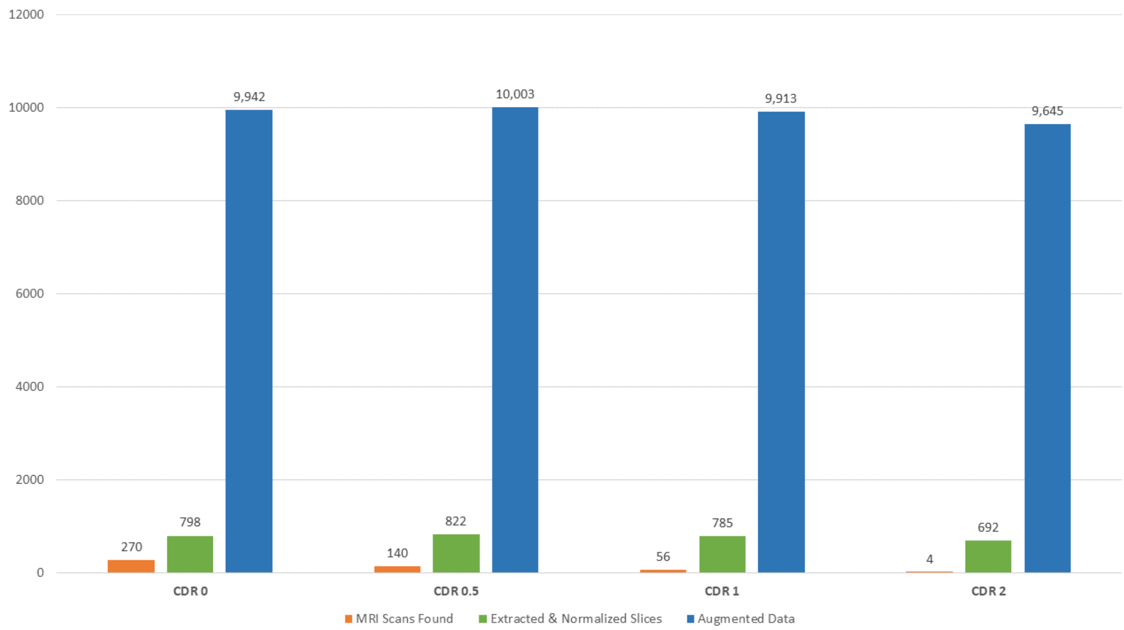

Figure 4.4: Workable Image Samples Before and After Preprocessing

We also attempted the preprocessing of the validation and testing data. This splitting occurs right after preprocessing and labeling. The parameters were mainly used to resize the image and transform each image into a standardized version. This improved the training phase of the dataset significantly. The used parameters helped get a more workable view of the dataset, ensuring noise reduction, a significant challenge while working with particular medical image scan datasets.

Table 4.1: Data Preprocessing

Data Segment	**Parameters**	**Value**
Training Data	Rescale	1/255
	Validation split	0.2
	Rotation range	10
	Width shift range	0.2
	Height shift range	0.2
	Shear range	0.2
	Zoom range	0.2
	Horizontal flip	True
	Vertical flip	True
	Fill mode	nearest
Validation Data	Rescale	1/255
	Validation split	0.2
Testing Data	Rescale	1/255

4.3 Data Labeling

Most existing models are implemented using binary classification or triple class models, and our architecture works with more integrated multiple class approaches. Hence, we have conducted the classification with four classes. This assists in detecting the presence of Alzheimer's Disease and categorizes the disease in a precise way. Our labeling was done after the preprocessing phase of the dataset. The four classes are – **1)** Clinical Dementia Ratio (CDR) 0, **2)** Clinical Dementia Ratio (CDR) 0.5 **3)** Clinical Dementia Ratio (CDR) 1 and, **4)** Clinical Dementia Ratio (CDR) 2. These four classes indicate Non-demented, mild demented, moderately demented, and severely demented samples, respectively. So, CDR0 corresponds to a healthy brain, and CDR 0.5, CDR 1 and, CDR 2 correspond with Alzheimer's Disease and the severity level.

Table 4.2: Data Labeling

Label	**Corresponding State**
Clinical Dementia Ratio (CDR) 0	Non-demented
Clinical Dementia Ratio (CDR) 0.5	Mild Demented
Clinical Dementia Ratio (CDR) 1	Moderately Demented
Clinical Dementia Ratio (CDR) 2	Severely Demented

4.4 Data Splitting

The data segmentation or splitting and the preprocessing were conducted simultaneously. The entire dataset was splitted into 80% and 20% for training and testing data, respectively, based on an 8:2 ratio randomly. Moreover, later on, 20% of the training data was used for validation purposes. The split datasets have a batch size of 32 each, and the target size is (64, 64). 25364, 6340, and 7799 sample images have been used for training, validation, and testing purposes.

The following table shows the summary of the data splitting phase:

Table 4.3: Data Splitting

Data Segmentation	Percentage	Total	Parameters	Value
Training	80% of the total data	25364	Target Size	(64, 64)
			Class Mode	categorical
			Subset	training
			Batch Size	32
Validation	20% of Training data	6340	Target Size	(64, 64)
			Class Mode	categorical
			Subset	validation
			Batch Size	32
Testing	20% of the total data	7799	Target Size	(64, 64)
			Class Mode	categorical
			Subset	testing
			Batch Size	32

Chapter 5

Architecture

5.1 Convolutional Neural Network (CNN) Architecture

The usage of convolutional neural networks is sky-rocketing year by year because of the efficiency of this particular architecture in image processing. A comprehensive study can be conducted due to the architecture's flexibility that significantly helps medical data analysis, especially biomedical anomaly detection. Hence, most of the existing researchers have chosen CNN as their ideal approach, and our proposed 19-layer classification model for Alzheimer's Disease detection extends the CNN-based multiple-class approaches for reasonable accuracy and efficiency.

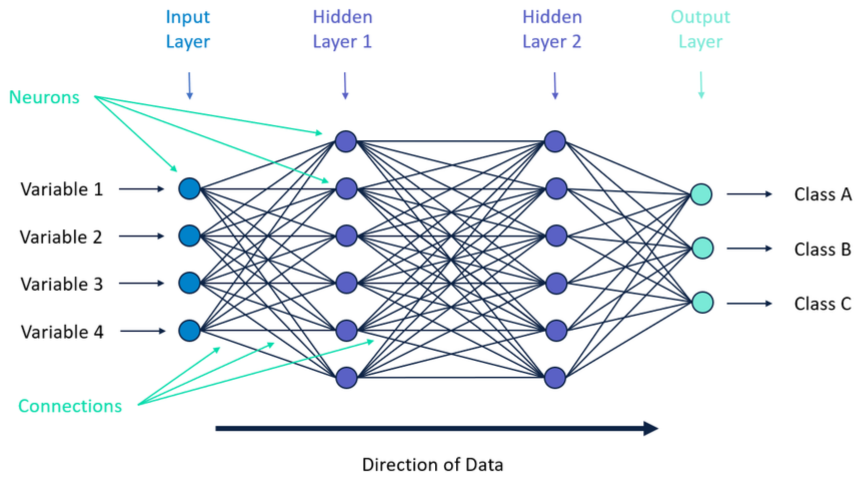

Figure 5.1: Basic Neural Network Structure

In a convolutional neural network, the main building blocks are convolutional layers. Conversion is the basic process of applying a filter to an input to produce an output. When the same filter is entered multiple times, an object map is generated, showing the position and strength of the recognized input object, such as an image. The ability of an accumulative neural network to learn a large number of filters in parallel, especially for a training dataset, within the confines of a particular predictive model problem, such as image classification, is its only indication. As a result, very specific features can be detected almost anywhere.

Convolution: In a convolutional neural network, the main building blocks are convolutional layers. Transformation is the primary procedure of applying a filter to an input in order to generating an output. When the same filter is entered multiple times, an object map is generated, showing the position and strength of the recognized input object (e.g image data). The ability of CNN is to learn a large number of filters in parallel, especially for a training dataset, within the confines of a particular predictive model problem, such as image classification. As a consequence, very detailed and exclusive features can be identified.

The first layer of a CNN is usually a convolutional layer. Accumulate layers of convolutions on the input before passing the output to the next layer. All pixels in the receptive region of convolution are converted to a single value. We can use several types of complexes depending on the problem we are trying to solve and the features we want to learn.

Two Dimensional Convolutional Layer: The 2D convolution layer is the most usual type of convolution and is often abbreviated as 2D convolution or Conv2D. The filter or multiplier 'slides' over the 2D input data in the convection layer, performing element-by-element multiplication. Therefore, the results get aggregated into a single output pixel. The kernel then performs the same procedure for each point it slides through, changing the 2D matrix of the entities to another two dimensional matrix of the entities.

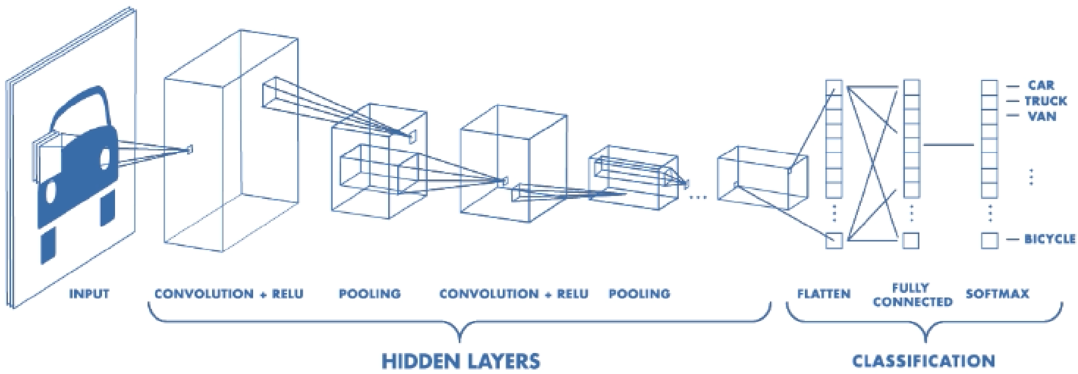

Figure 5.2: Vanilla Convolutional Neural Network Example

5.2 Proposed 18-Layer Model Architecture

Our proposed model works on many preprocessed datasets set to train the architecture for the best possible outcome in terms of multiple classifications. The model consists of several layers compared to the conventional approaches in the existing models. The 18-layer model has seven steps in total all over the architecture. The steps have been explained below:

5.2.1 Convolutional Layer

Our proposed CNN architecture consists of three Conv2D. Conv2D is used as a mandatory parameter that determines how many convolutional layers the network will learn. In this case, the number is three.

5.2.2 Pooling Layer:

In our pooling layer, we have used MaxPooling2D as the pooling parameter. The three MaxPooling2D corresponds with the three Conv2D layers we have implemented. Pooling operation MaxPooling2D is responsible for calculating the largest or maximum value of the feature map. The effect of applying filters to an input image is captured by the feature maps of a CNN. The feature map is the output of each layer, hence the output of each layer is the extracted feature map. The purpose of looking at a feature map for a certain input image is to gain a better grasp of the features that our CNN has recognized.

5.2.3 Batch Normalization

Our architecture has two batch normalization layers. The batch normalization layers help the entire network learn more independently and normalizes the previous layer's output. Using batch normalization also helps to standardize the output and input layers in sequential models.

5.2.4 Flatten Layer

We have two flatten layers in the sequential model that helps to flatten the output by creating a single long feature vector in the convolutional layers. The flatten layer is also connected with the ultimate classification model known as the fully connected layer.

5.2.5 Dense Layer

The architecture has seven dense layers, and a dense layer passes on the outputs from its previous layers to the subsequent neurons, and each neuron provides an output to the next layer.

5.2.6 Dropout

In our proposed architecture, the dropout layer is used to prevent our model from overfitting. This overfitting happens when a model gets trained according to the information and learns it. In that stage the noise of the data also gets learnt by the model which can cause a negative impact on the result. This situation is called overfitting. So, the dropout layer drops some of the redundant error-prone noisy data to reduce overfitting.

5.2.7 Activation Function

The sequential model architecture uses 2 activation functions. The functions are Rectified Linear Unit or ReLU and Softmax. The activation functions are shown as follows:

ReLU: ReLU is a piecewise linear function that gives an immediate result if the input is positive. Otherwise, it will provide a zero output. It seems to be the standard activation function for several neural network models as it's comparatively faster and usually produces improved results.

$$a = max(0, x) \tag{5.1}$$

Softmax: On the other hand, the Softmax function takes the numeric output of the previous layer in a multiclass classification-based neural network and converts it to probabilities using the exponents of each output. Then, one by one, the values are normalized using the total of the exponents. Softmax is the activation function for multiclass classification problems requiring class membership on more than two labels.

$$\sigma(\vec{z}_i) = \frac{e^z i}{\sum_k^{j=1} e^{zj}} \tag{5.2}$$

Here, the visual representation of our custom 18 layer CNN model has been shown below:

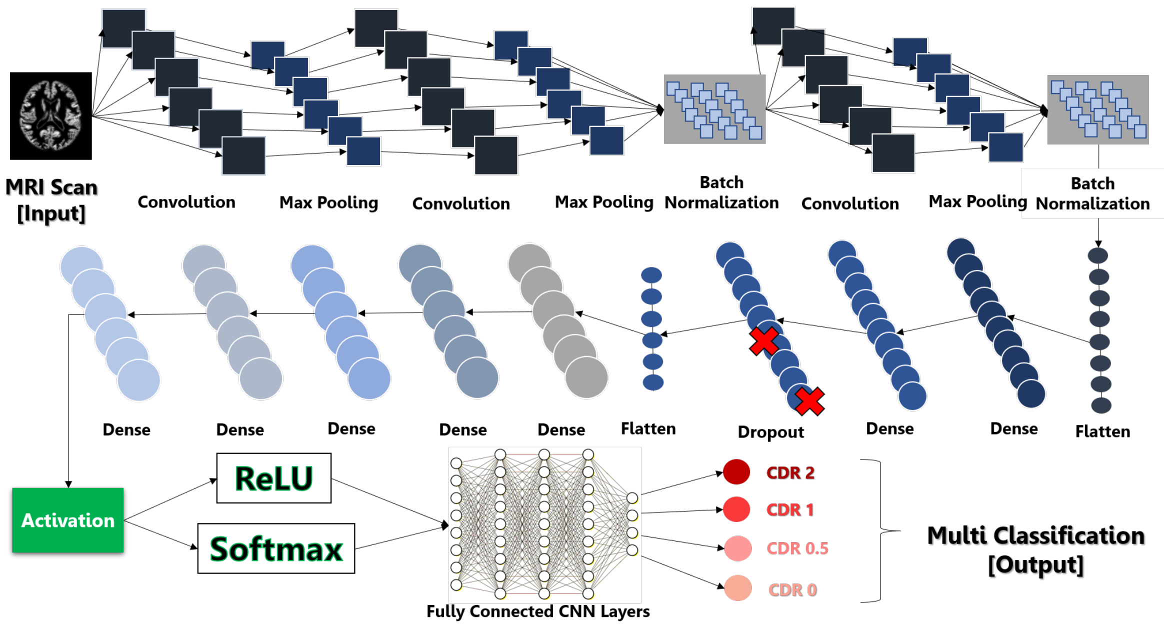

Figure 5.3: The Visual representation of the Proposed 18-layer Architecture (Multiclass)

5.2.8 Summary

Our 18-layer convolutional neural network acts to determine the detailed patterns found in the preprocessed dataset. The following tables represent the summary of the architecture:

Table 5.1: A summary of our proposed 18 Layer CNN Model (I)

Layer (Type)	**Output Shape**	**Param #**
conv2d (Conv2D)	(None, 80, 80, 25)	1900
max_pooling2d (MaxPooling2D)	(None, 32, 32, 25)	0
conv2d_1 (Conv2D)	(None, 40, 40, 25)	31300
max_pooling2d_1 (MaxPooling2D)	(None, 20, 20, 50)	0
batch_normalization (Batch Normalization)	(None, 10, 10, 50)	200
conv2d_2 (Conv2D)	(None, 5, 5, 70)	31570
max_pooling2d_2 (MaxPooling2D)	(None, 2, 2, 70)	0
batch_normalization_1 (Batch Normalization)	(None, 2, 2, 70)	280
flatten (Flatten)	(None, 280)	0
dense (Dense)	(None, 100)	28100
dense_1 (Dense)	(None, 100)	10100
dropout (Dropout)	(None, 100)	0
flatten_1 (Flatten)	(None, 100)	0
dense_2 (Dense)	(None, 32)	3232
dense_3 (Dense)	(None, 32)	1056
dense_4 (Dense)	(None, 32)	1056
dense_5 (Dense)	(None, 32)	1056
dense_6 (Dense)	(None, 4)	132

Table 5.2: A summary of our proposed 18 Layer CNN Model (II)

Summary	**Value**
Total params	109,982
Trainable params	109,742
Non-trainable params	240

Chapter 6

Implementation and Result

6.1 Demonstrating the Performance of Our Proposed Model

The implementation starts from running our proposed sequential model by training and testing the data epoch by epoch. Here we considered 50 epochs while running the model. We have considered five metrics to evaluate our model's performance – accuracy, precision, recall, F1 score, and AUC or Area Under ROC (Receiver operating characteristic) Curve. We have also implemented a separate function to handle the exponential decay.

The equations of accuracy, F1-score, precision, and recall are shown as follows:

Accuracy:

$$\begin{aligned} ACC &= \frac{TP + TN}{P + N} \\ &= \frac{TP + TN}{TP + TN + FP + FN} \end{aligned} \tag{6.1}$$

F1 Score:

$$\begin{aligned} F1 &= 2 \times \frac{PPV \times TPR}{PPV + TPR} \\ &= \frac{2TP}{2TP + FP + FN} \end{aligned} \tag{6.2}$$

Precision:

$$PPV = \frac{TP}{(TP + FP)} = (1 - FDR) \tag{6.3}$$

Recall:

$$PPV = \frac{TP}{(TP + FP)} = (1 - FDR) \tag{6.4}$$

Where (in equations (3) to (6)),

- ACC = Accuracy
- TP = True Positive
- TN = True Negative

- P = Condition Positive
- N = Condition Negative
- PPV = Positive Predictive Value/Precision
- TPR = True Positive Rate
- FDR = False Discovery Rate
- PPV = Positive Predictive Value/Precision
- TP = True Positive
- FN = False Negative

6.2 Result Analysis

6.2.1 Multiclass (Proposed)

Our primary objective was to work with multiple classes on the specific dataset and propose a suitable model. Hence, we have conducted an in-depth analysis of the performance in terms of multi-classification. However, later on, the comparative analysis shows the reliability of the model.

The following table shows the summary of the performance:

Table 6.1: Performance Summary (Multiclass)

Metric	Value
Accuracy	80.09%
Precision	85.80%
Recall	24.72%
AUC	81.89%
F1-Score	37.73%

The following metrics and their graphical representation describe how the model performed under the proposed multi-classification approach.

Accuracy: The history of the accuracy graph depicts the training and validation process accuracy. It indicated the number of correct predictions by the model. The training data accuracy is favorable for this initial version of the proposed model, about 80%. But the validation data lags slightly to match up to the accuracy bar, which is about 79%. So, the difference is around 1% in terms of the accuracy graph.

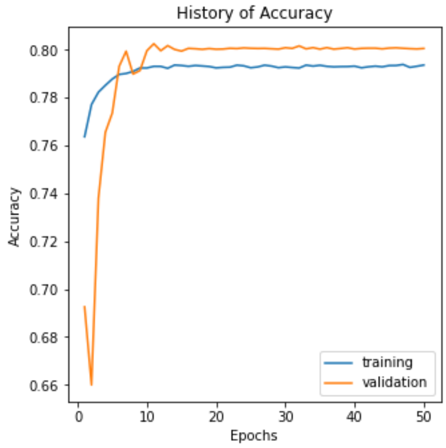

Figure 6.1: History of Accuracy (Multiclass)

Precision: The graph shows that the training phase has ensured precision of about 62-63%, which has scopes for improvement in further studies regarding this data set in multiclass. The validation data is close to the training data, which is also around 62%.

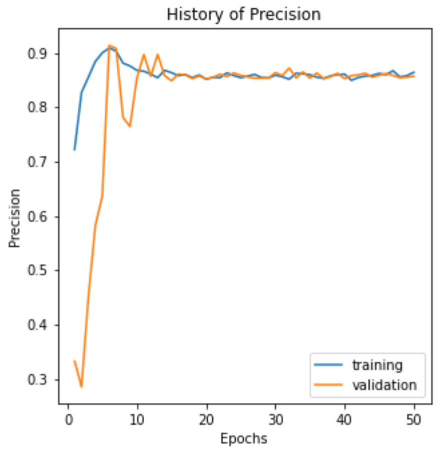

Figure 6.2: History of Precision (Multiclass)

Loss: The loss curve indicates the errors made by the model. It shows how much data is being hampered in terms of precision and decreasing true positive and true negative results in some cases. This can play a role in providing faulty results in testing. The training loss is within the 1 to 1.5 range.

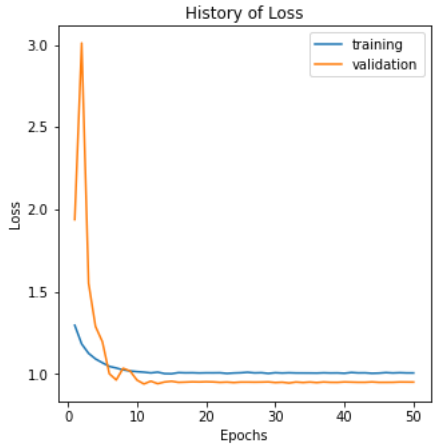

Figure 6.3: History of Loss (Multiclass)

Recall: Higher recall means the model's capability to successfully detect the positive samples. It is the ratio between correctly classified positive samples and the total number of positive samples. The recall of the model is about 24.72%.

AUC: The Area Under the Curve or AUC graph summarizes the Receiver Operating Characteristic or ROC curve in a specific model. It depicts the ability to distinguish between the positive and negative classes. The AUC in our proposed model is about 80-81% in the training data and about 79-80% in the validation data. This ensures that the AUC works well in this model, but it can be spiked up even higher with further improvement.

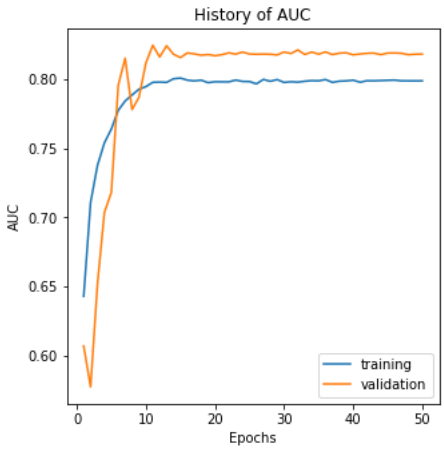

Figure 6.4: History of AUC (Multiclass)

F1-score: The F1-score in our proposed model is about 36-37% in terms of training and around 33-34% mainly in terms of validation data. The F1 score gets low due to low precision and recall value. Our precision and recall values are about 85.80% and 24.72%, respectively. This explains why the F1-score is minor in this regard which is about 37.73%.

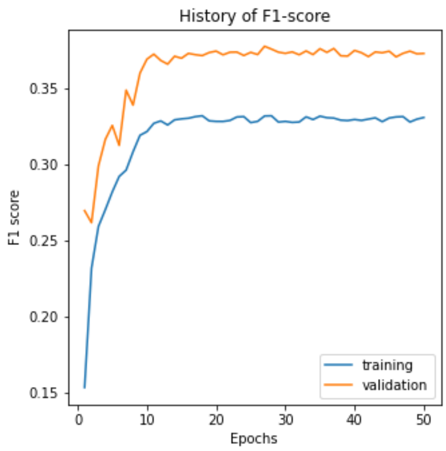

Figure 6.5: History of F1-Score (Multi class)

Confusion Matrix

The confusion matrix helps us to summarize the results predicted from a classification-based model or system. By observing the confusion matrix, we can apprehend the performance of the proposed architecture. Each class represents the correct and incorrect (True Positive/Negative, False Positive Negative) values corresponding to their classes. This also sums up how biased the process has been while training the model and how efficiently the model can detect the goal.

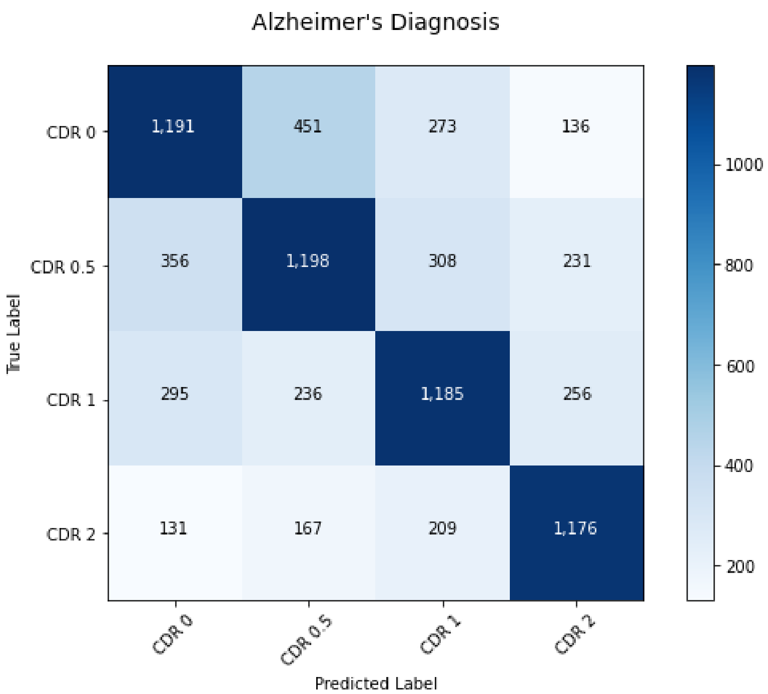

Figure 6.6: Confusion Matrix of the Proposed Model (Multi class)

Here, we can see in the 16 × 16 confusion matrix for four classes correspond to the CDR 0, 0.5, 1, and 2. The x-axis denotes the 'Predicted Label,' and the Y-axis denotes the 'True Label.'

6.2.2 Binary class

We have also implemented our model in terms of binary classes to show the efficiency of the model. The binary classes that were considered were CDR0 and CDR2.
The following table shows the summary of the proposed model in terms of binary class.

Table 6.2: Performance Summary (Binary Class)

Metric	Value
Accuracy	92.78%
Precision	92.78%
Recall	92.78%
AUC	97.88%
F1-Score	92.83%

Accuracy: The accuracy of the training data is almost 92.78%, which is a significant indication of the model's efficiency. The validation dataset stabilizes around 90%.

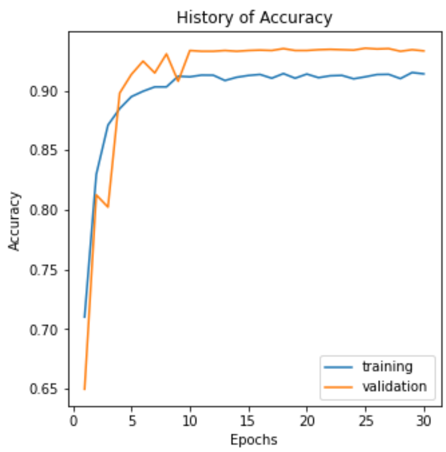

Figure 6.7: History of Accuracy (Binary Class)

Precision: The precision in the binary class is quite significantly high and matches with the accuracy score of the binary class. The precision is around 91-92%.

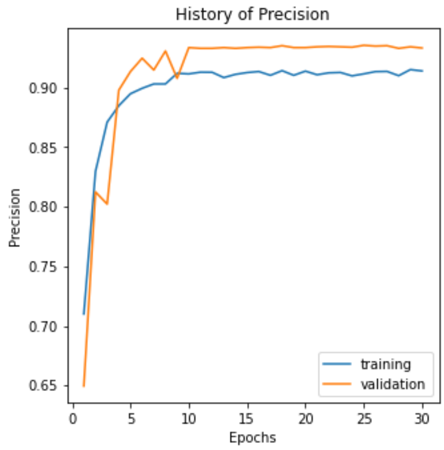

Figure 6.8: History of Precision (Binary Class)

Loss: The binary classification shows the loss value starts and finishes within the range of 0.6 to almost 0.25 gradually in terms of training data. And, the validation data jumps around the range of 0.7 to 0.2.

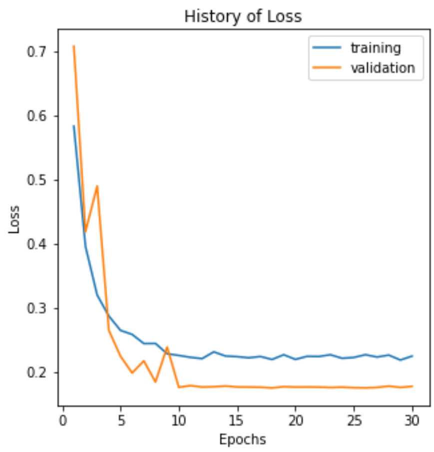

Figure 6.9: History of Loss (Binary Class)

Recall: The recall value is 92.78%, which corresponds with the accuracy and the precision of the model. Higher recall ensures that the performance is quite satisfactory in detecting the true values in the data.

AUC: The AUC or the Area Under the Curve is about 97.88% which is quite impressive. The traning and validation curve stabilizes around 96% and 97-98% respectively.

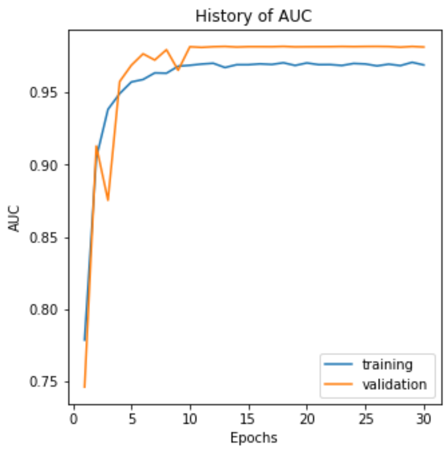

Figure 6.10: History of AUC (Binary Class)

F1-score: The F1-score in the binary classification has reached upto 92.83% and the training and the validation data gets steady around 90-91% and 91-92% respectively.

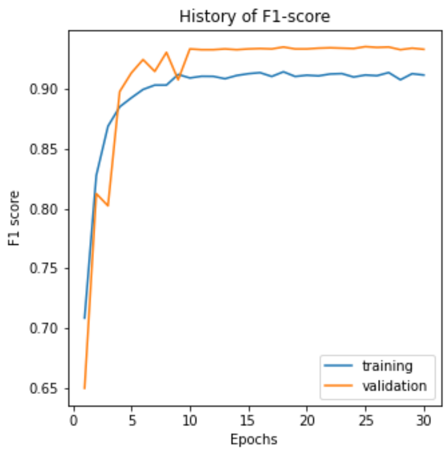

Figure 6.11: History of F1-Score (Binary class)

The binary classification is conducted to represent the reliability and the versatility of the model. Here, we can see a significant spike in all the metrics that correspond with e stronger performance. This also indicates that the model has been able to successfully obtain its goal with a satisfactory performance rate, and the training and validation data difference is almost close to none, which also ensures the reliability of the dataset. Compared to the multiclass, the performance is much higher because of the more balanced data, to begin within both the classes, which resulted in a better training phase.

Chapter 7

Comparative Analysis

7.1 Evaluating the Preprocessing Strategy in Separate Anatomical Planes

As we mentioned that the preprocessing strategy considers all the anatomical planes such as sagittal, coronal, and transverse rather than a lot of the existing approaches where a single anatomical plane is taken under consideration while preprocessing and implementing the model. Hence, we conducted a comparative analysis on all three separate anatomical planes available on the dataset, preprocessed it in the same manner, and compared the result with the proposed preprocessing strategy.

Table 7.1: The comparison among the separate anatomical plane based data and the proposed combined preprocessed data (Multiclass)

Anatomical Plane	**Accuracy**	**Precision**	**Recall**	**AUC**	**F1-Score**
Sagittal	82.35%	78.92%	40.15%	87.23%	52.86%
Coronal	80.66%	86.97%	26.64%	84.00%	40.53%
Transverse	79.22%	67.74%	32.23%	81.74%	43.30%
Combined (Proposed)	80.09%	85.80%	24.72%	81.89%	37.73%

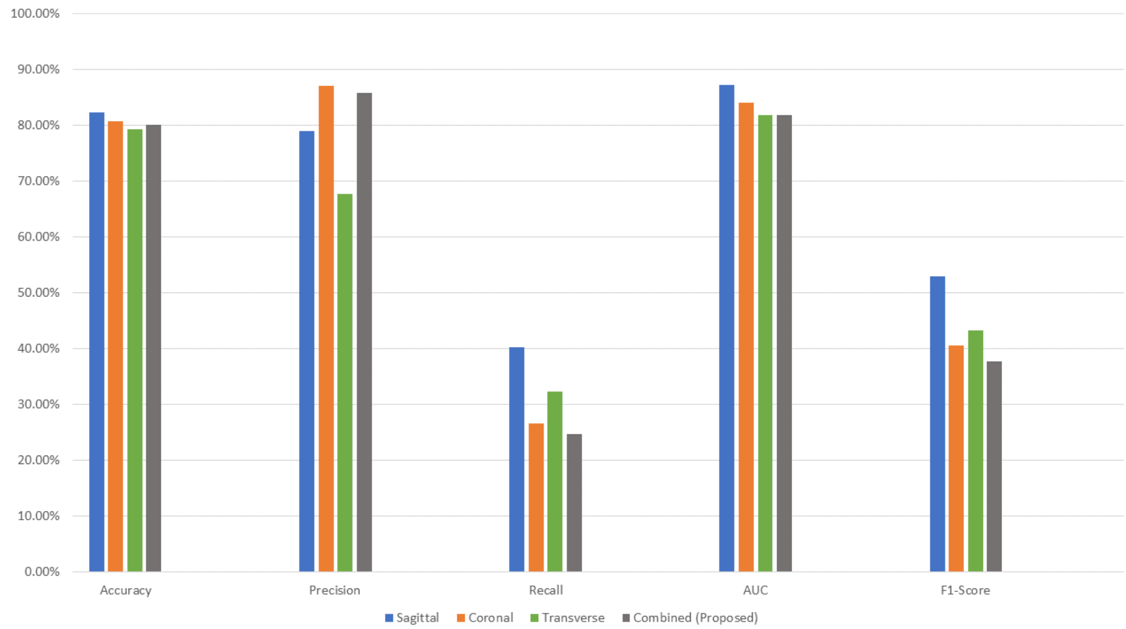

Figure 7.1: Evaluating the Preprocessing Strategy (Multiclass)

The purpose of this specific comparative analysis is to show that the model's training has been done in an equal manner for all four classes. Each cases have almost similar result in terms of accuracy, precision, recall, AUC and F1-score. Here, sagittal has the highest accuracy, about 82.35% and the lowest is of the transverse plane, which is 79.22%. The combined approach has an accuracy of 80.09%. On the other metrics such as recall, AUC, F1-score the sagittal plane slices perform better with 40.15%, 87.23%, 52.86% respectively. But it falls a little short in terms of precision which is about 78.92%. But, the proposed approach performs quite well in terms of precision with about 85.80% but it falls short in recall metric with a score of 24.72% which is closer to the recall of the coronal plane. However, the other metrics such as the AUC are similar to the combined approach and the transverse plane. After the sagittal plane the coronal and transverse planes also have better F1-scores than the combined approach which are about 40.53% and 40.43% respectively.

In usual existing models most of the proposed architectures work with sagittal or coronal planes as per our background research says. However, some can even consider individual slices of the transverse plane also. But, this leaves rooms for the model to be efficient only in a certain setting and certain type of anatomical plane. Hence, we approached a more generalized idea and took a certain number of normalized three anatomical slices from each sample combined in all four classes while preprocessing. The combined result and the separately considered anatomical plane slices evidence that the generalized approach is almost the average of the rest of the three. So, this ensures us that the model is providing more efficient average result in an all around manner. Hence, this is the proposed preprocessing strategy of ours which has been mentioned before.

7.2 Loading Pretrained Convolutional Neural Network (CNN) Models

A model's sustainability and efficiency can only be evaluated through its performance evaluation compared to the predefined and pre-trained existing models. The research works an be extended further by investigating limitations and shortcomings. We considered some of the most well-known existing models to compare our approach with such as InceptionV3, VGG19, VGG16, ResNet, GoogleNet etc.

So far, we have implemented the well known state of the art InceptionV3 architecture and the VGG19 architecture.

7.2.1 InceptionV3

CNN architecture 'InceptionV3' is from the Inception family, and it includes layers like Label Smoothing, factorizing convolutional layers, and auxiliary classifier, which is used to propagate label information further in the network. It also used the batch normalization layer exclusively. The inception architecture has been further analysed in comparison to the other renowned architectures mentioned above in the research work by authors Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016) in their literature [22]. The inceptionV3 architecture is exclusively efficient for processing large amount of dataset without using tons of parameters and has the added benefits of quality feature extraction.

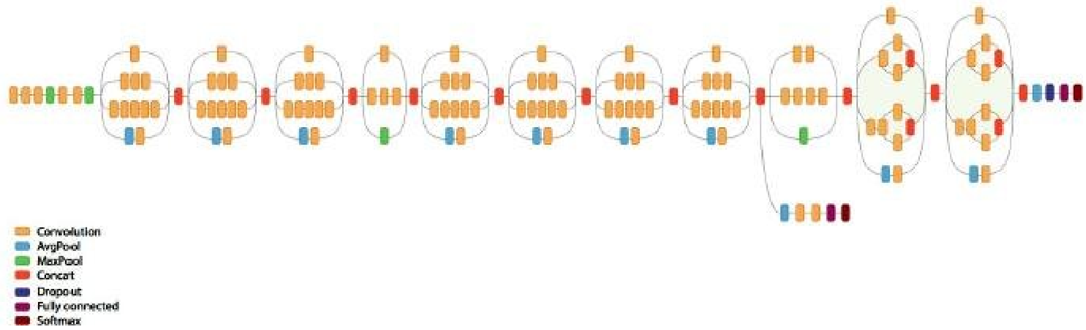

Figure 7.2: The Basic Architecture of InceptionV3

7.2.2 VGG19

VGG or 'Visual Geometry Group' is another classic convolutional neural network architecture that is based on the analytical prospect of incrementing the depth of networks. It uses layers that are small filtering sections and has blocks of max-pooling layers and fully connected layers of 2D convolution. It's well known for its simplicity. The extensive study based on AD detection using VGG architecture has been conducted in the research of authors Adeola Ajagbe, Kamorudeen A. Amuda, Matthew A. Oladipupo, Oluwaseyi F. AFE and Kikelomo I. Okesola (2021) [23]. The VGG19 architecture has specifically 19 layers of detailed CNN structure, hence the name 'VGG19'. Other previous version include VGG16.

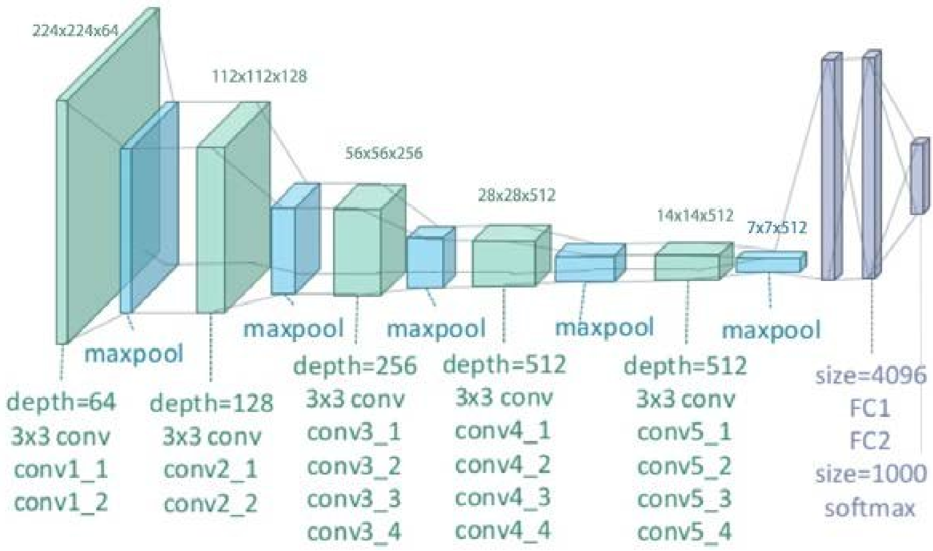

Figure 7.3: The Basic Architecture of VGG19

7.3 Comparison Among the Proposed Model and the Pre-trained Convolutional Neural Network (CNN) Models

7.3.1 Multiclass (Proposed)

The following table and the bar graph show the summary of the implemented pre-defined model(s) and our proposed architecture.

Table 7.2: The Comparison among the Predefined Model(s) and the Proposed 18-layer Model (Multi-class)

Model(s)	Accuracy	Precision	Recall	AUC	F1-Score
Proposed Model	80.09%	85.80%	24.72%	81.89%	37.73%
InceptionV3	78.91%	95.18%	16.47%	77.35%	27.62%
VGG19	77.66%	58.48%	36.67%	81.55%	45.05%

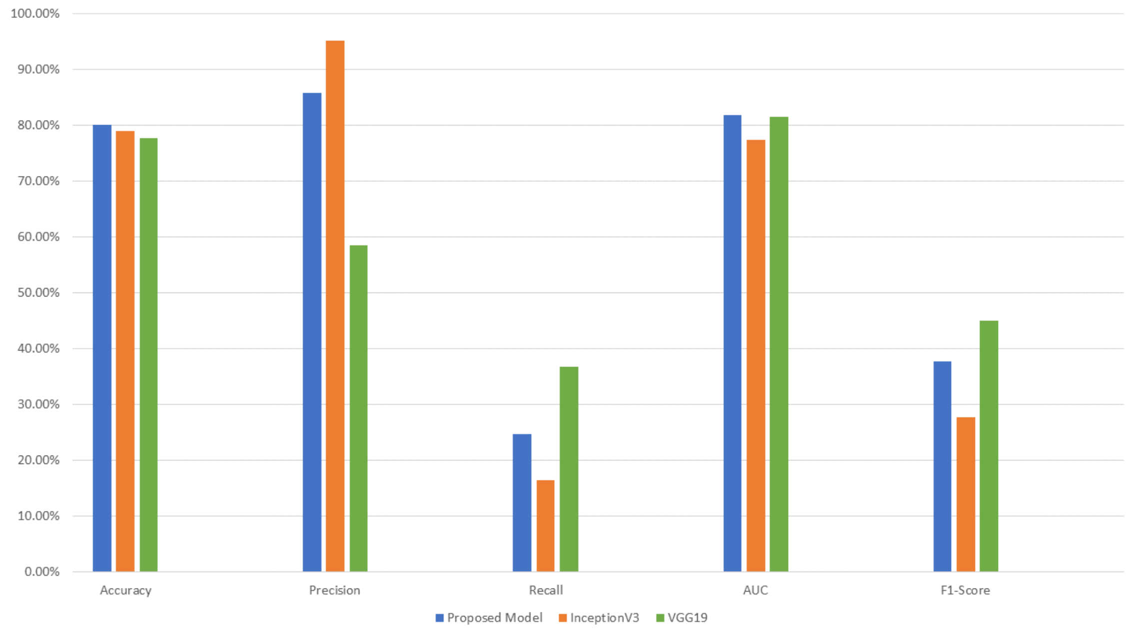

Figure 7.4: Comparitive Analysis (Multiclass)

Here if we further analyze the result, we can observe that the accuracy of our proposed model is about 80.09% whereas the InceptionV3 architecture shows a very similar 78.91% accuracy. The result is 1.18% higher which is very close to the pre-trained model and we can conclude that the result is almost equal in this regard. Then comes the precision which is about 85.80% in our model and 95.18% in the mentioned architecture. In this regard, the InceptionV3 architecture performs slightly better than the proposed model. The recall is higher in our proposed model which is about 24.72% and the InceptionV3 has about 16.47%. So, our model works better in terms of recall metrics. Then comes the AUC. AUC in our proposed model is 81.89% and the pre-trained model is about 77.35% which ensures that our AUC value is higher in our model. The F1-score is also quite similar as both are about 37.73% and 27.62% respectively.

On the other hand, the VGG19 architecture falls short in almost all the metrics except the recall value and F1-score. The accuracy, precision, recall, AUC, F1-score are 77.66%, 58.48%, 36.67%, 81.55%, 45.05% respectively. These existing models in most cases serve higher performance than vanilla CNN. Transfer learning models VGG and InceptionV3 are renowned, but they have shortcomings in some metrics and generalized cases. Whereas, our proposed model ensures a custom-made 18-layer approach that improves the efficiency in terms of computational time and accuracy as it was trained under all possible cases for such datasets. Hence, the validation data and the training data show impressively similar performance on the record.

7.3.2 Binary Class

We also studied the performance of existing InceptionV3 and VGG19 models in terms of binary class and implemented our model for binary classes to comparatively analyse the performance in terms of binary classes [24]. The following table shows the comparative perform in terms of the binary classes.

Table 7.3: The Comparison among the Predefined Model(s) and the Proposed 18-Layer Model (Binary Classes)

Models	Accuracy	Precision	Recall	AUC	F1-Score
Proposed Model	92.78%	92.78%	92.78%	97.88%	92.83%
InceptionV3	91.32%	93.64%	89.93%	96.86%	87.56%
VGG19	50%	50%	25%	50%	33.5%

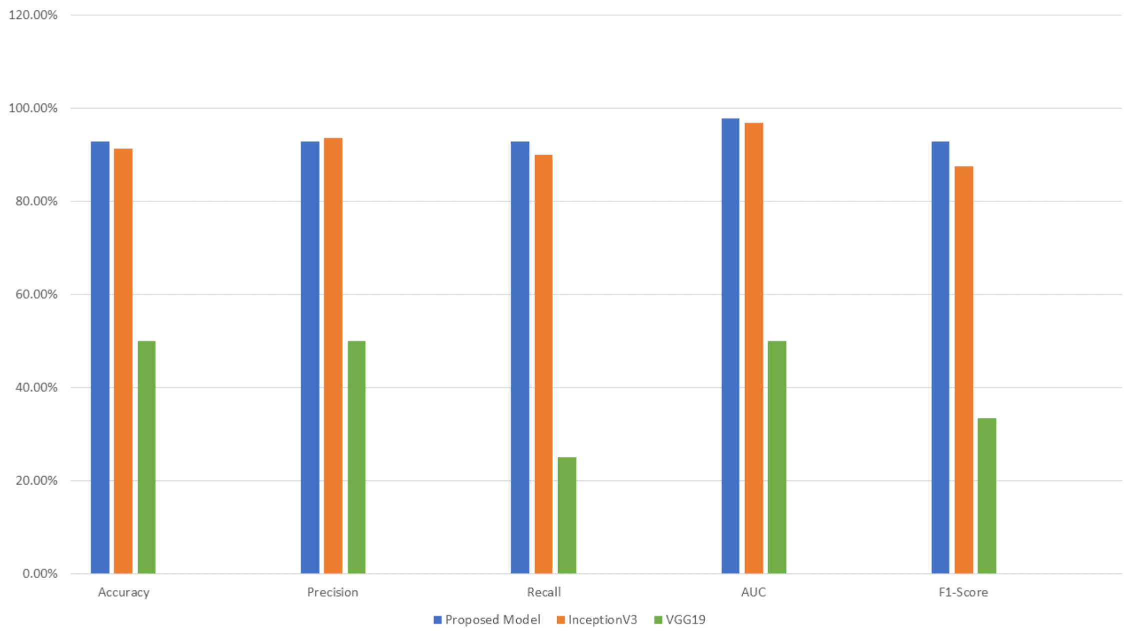

Figure 7.5: Comparitive Analysis (Binary Class)

From the statistical values, we can observe that our model has also shown significant performance strength in binary class. The accuracy of the proposed model and the existing models are the evidence of the our suggested architecture's reliability and efficiency on a comparative ground in this data set. Especially for two well known architecture InceptionV3 and VGG19 the proposed model's performance is strikingly higher. The InceptionV3 has an accuracy of 91.32% in this case which is quite similar to the proposed model's accuracy but VGG-19 falls short drastically when it comes to binary classification. On the other hand, the InceptionV3 is slightly better than the proposed model, but the recall, AUC and F1-score has significantly higher percentage in our proposed model than the rest of the two architectures which are about 92.78%, 97.88% and 92.83%. The InceptionV3 has 89.93%, 96.86% abd 87.56% respectively. On the other hand the existing VGG-19 architecture has some noticeable shortcoming as it has the accuracy, precision, recall, AUC and F1-score of 50%, 50%, 25%, 50% and 33.5% respectively. So, this comparative analysis proves that the proposed model has performed well in a competitive case in both multi classification and binary classification setting.

Chapter 8

Conclusion

8.1 Limitations Discussion and Future Works

Our model has shown impressive strength in terms of a larger dataset. As most of the existing models work with binary or triple classifications, we had considered extending the multiple classification research further. But there are many more areas to improve the performance of the proposed architecture.

In the future, a further extension can be done regarding feeding more training data by ensuring deep unbiased learning. The dataset also has some striking shortcomings in terms of unbalanced data in all four classes, which we tried to overcome with our pre-processing strategy. We are hopeful of receiving much significant performance in any other dataset and plan to work on the different datasets (especially geo-social specific datasets for comparative pattern analysis) in the future to further ensure our model's versatility.

To improve the architecture, we can implement a clustering process, rearrange the existing layers, modify the parameters, and trivially test the epochs through trial and error for the best possible results. K-fold class validation can be considered for better testing validation results and improving the accuracy to more than 90% or more if possible. The pre-processed data can be further augmented and de-noised for more efficient and accurate results. We also believe bringing changes to the loss function and introducing more updated sequential model algorithms can significantly improve accuracy and precision.

In the future, we plan to do a more extensive analysis of Alzheimer's Disease pattern and compare via thorough study among the existing models. So, we look forward to working on these shortcomings and scopes to research further to improve or proposed model. Moreover, if possible, we desire to introduce a medical web application platform for real-time Alzheimer's Detection implementing our architecture.

8.2 Conclusion

The implementation of this proposed 18-layer architecture (based on all three dimensions of the MRI scans) will help us detect at an early stage and get more accurate results compared to other existing models. This will help us detect the disease with efficiency and provide us a suitable classification on severity level. Furthermore, pattern analysis can be conducted for extensive studies and better diagnosis. Using our model, at a premature stage, Alzheimer's disease can be identified without any manual assistance, and fundamental treatment done at these early states will limit the chance of making further complications of the hard to fathom Alzheimer's disease.

Bibliography

[1] N. C. Berchtold and C. W. Cotman, "Evolution in the conceptualization of dementia and alzheimer's disease: Greco-roman period to the 1960s," *Neurobiology of aging*, vol. 19, no. 3, pp. 173–189, 1998.

[2] T. Vos, C. Allen, M. Arora, R. M. Barber, Z. A. Bhutta, A. Brown, A. Carter, D. C. Casey, F. J. Charlson, A. Z. Chen, *et al.*, "Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: A systematic analysis for the global burden of disease study 2015," *The lancet*, vol. 388, no. 10053, pp. 1545–1602, 2016.

[3] W. H. Organization *et al.*, *Dementia fact sheet world health organization*, 2018.

[4] C. V. Balaji, M. B. LR, and M. L. Hingane, "Alzheimer's disease progression & treatment," 2021.

[5] H. Wang, M. Naghavi, C. Allen, R. M. Barber, Z. A. Bhutta, A. Carter, D. C. Casey, F. J. Charlson, A. Z. Chen, M. M. Coates, *et al.*, "Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: A systematic analysis for the global burden of disease study 2015," *The lancet*, vol. 388, no. 10053, pp. 1459–1544, 2016.

[6] S. Bonin-Guillaume, D. Zekry, E. Giacobini, G. Gold, and J.-P. Michel, "The economical impact of dementia," *Presse Medicale (Paris, France: 1983)*, vol. 34, no. 1, pp. 35–41, 2005.

[7] A. Corbett, G. Williams, and C. Ballard, "Drug repositioning in alzheimer's disease," *Front Biosci (Schol Ed)*, vol. 7, pp. 184–188, 2015.

[8] G. Kojima, A. Liljas, S. Iliffe, and K. Walters, "Prevalence of frailty in mild to moderate alzheimer's disease: A systematic review and meta-analysis," *Current Alzheimer Research*, vol. 14, no. 12, pp. 1256–1263, 2017.

[9] H. W. Querfurth and F. M. LaFerla, "Mechanisms of disease," *N Engl J Med*, vol. 362, no. 4, pp. 329–344, 2010.

[10] S. Todd, S. Barr, M. Roberts, and A. P. Passmore, "Survival in dementia and predictors of mortality: A review," *International journal of geriatric psychiatry*, vol. 28, no. 11, pp. 1109–1124, 2013.

[11] N. M. Khan, N. Abraham, and M. Hon, "Transfer learning with intelligent training data selection for prediction of alzheimer's disease," *IEEE Access*, vol. 7, pp. 72 726–72 735, 2019.

[12] R. Sivakani and G. A. Ansari, "Machine learning framework for implementing alzheimer's disease," in *2020 International Conference on Communication and Signal Processing (ICCSP)*, IEEE, 2020, pp. 0588–0592.

[13] S. Sarraf and G. Tofighi, "Classification of alzheimer's disease using fmri data and deep learning convolutional neural networks," *arXiv preprint arXiv:1603.08631*, 2016.

[14] P. Lodha, A. Talele, and K. Degaonkar, "Diagnosis of alzheimer's disease using machine learning," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, IEEE, 2018, pp. 1–4.

[15] R. Mahmood and B. Ghimire, "Automatic detection and classification of alzheimer's disease from mri scans using principal component analysis and artificial neural networks," in *2013 20th International Conference on Systems, Signals and Image Processing (IWSSIP)*, IEEE, 2013, pp. 133–137.

[16] U. R. Acharya, S. L. Fernandes, J. E. WeiKoh, E. J. Ciaccio, M. K. M. Fabell, U. J. Tanik, V. Rajinikanth, and C. H. Yeong, "Automated detection of alzheimer's disease using brain mri images–a study with various feature extraction techniques," *Journal of Medical Systems*, vol. 43, no. 9, pp. 1–14, 2019.

[17] I. Ahmad and K. Pothuganti, "Analysis of different convolution neural network models to diagnose alzheimer's disease," *Materials Today: Proceedings*, 2020.

[18] C. D. Billones, O. J. L. D. Demetria, D. E. D. Hostallero, and P. C. Naval, "Demnet: A convolutional neural network for the detection of alzheimer's disease and mild cognitive impairment," in *2016 IEEE region 10 conference (TENCON)*, IEEE, 2016, pp. 3724–3727.

[19] M. H. Memon, J. Li, A. U. Haq, and M. H. Memon, "Early stage alzheimer's disease diagnosis method," in *2019 16th International Computer Conference on Wavelet Active Media Technology and Information Processing*, IEEE, 2019, pp. 222–225.

[20] J. Islam and Y. Zhang, "Brain mri analysis for alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks," *Brain informatics*, vol. 5, no. 2, pp. 1–14, 2018.

[21] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open access series of imaging studies (oasis): Cross-sectional mri data in young, middle aged, nondemented, and demented older adults," *Journal of cognitive neuroscience*, vol. 19, no. 9, pp. 1498–1507, 2007.

[22] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[23] S. A. Ajagbe, K. A. Amuda, M. A. Oladipupo, O. F. AFE, and K. I. Okesola, "Multi-classification of alzheimer disease on magnetic resonance images (mri) using deep convolutional neural network (dcnn) approaches," *International Journal of Advanced Computer Research*, vol. 11, p. 53, 2021.

[24] E. Hussain, M. Hasan, S. Z. Hassan, T. H. Azmi, M. A. Rahman, and M. Z. Parvez, "Deep learning based binary classification for alzheimer's disease detection using brain mri images," in *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, IEEE, 2020, pp. 1115–1120.