INVESTIGATING THE EFFICIENCY OF ALZHEIMERS DISEASE CLASSIFICATION BY RADIOMIC FEATURES & DEEP LEARNING TECHNIQUES

Dissertation Phase II - Report

submitted by

Mr. MUHAMMED SUFAIL M K

REG NO : TKM22MEAI11

to

the APJ Abdul Kalam Technological University in partial fulfillment for the award of the degree of

MASTER OF TECHNOLOGY
in
Artificial Intelligence

Under the guidance of
Dr. Nissan Kunju

TKM College of Engineering Kollam

JUNE 2024

Thangal Kunju Musaliar College of Engineering
Centre for Artificial Intelligence

C E R T I F I C A T E

This is to certify that, this report titled **Investigating the Efficiency of Alzheimers Disease Classification By Radiomic Features & Deep Learning Techniques** is a bonafide record of the **Dissertation Phase II** done by **MUHAMMED SUFAIL M K (TKM22MEAI11)**, under our guidance and supervision, in partial fulfillment of the requirements for the award of the degree, **M. Tech in Artificial Intelligence** in **APJ Abdul Kalam Technological University**.

Internal Supervisor

Dr. Nissan Kunju
Associate Professor
Dept of ECE
TKMCE

Co Supervisor

Prof. Chinnu Jacob
Assistant Professor
Centre for AI
TKMCE

Project Cordinator

Dr. Sumod Sundar
Associate Professor
Centre for AI
TKMCE

Head of Department

Dr. Imthias Ahamed T P
Professor and HOD
Centre for AI
TKMCE

ACKNOWLEDGEMENT

A successful project is a fruitful culmination of efforts by many people, some directly involved and some others indirectly, by providing support and encouragement. Firstly I would like to thank the almighty for giving me the wisdom and grace for making my project a memorable one. I thank him for steering me to the shore of fulfillment under his protective wings

I express my gratitude to **Dr. T A Shahul Hameed**, Principal of T.K.M College of Engineering for providing me with the best resources for the completion of my dissertation phase II . I would like to thank **Dr. Imthias Ahamed T P**, Professor and Head of the Department, Centre for Artificial Intelligence, TKM College of Engineering, Kollam, for his unwavering support and encouragement throughout the project.

With a profound sense of gratitude, I would like to express my heartfelt thanks to my project coordinator **Dr. Sumod Sundar**,Associate Professor, Centre for Artificial Intelligence, TKM College of Engineering, Kollam, my internal supervisor **Dr. Nissan Kunju**, Associate Professor, Department of ECE, TKM College of Engineering, Kollam , co-supervisor, **Prof. Chinnu Jacob**, Assistant professor, Centre for AI,TKM College of Engineering,Kollam, and **Mr.Ganeshnath R**,Research scholar,Biomedical engineering research lab, Department of ECE, TKM College of Engineering,Kollam and who have encouraged me throughout this work.I am also grateful to the entire faculty and staff of the Centre for AI at TKMCE for their continuous support and encouragement throughout this project.

I also express my thanks to my loving parents and friends, for their support and encouragement in the successful completion of this work.

MUHAMMED SUFAIL M K

Abstract

Alzheimer's Disease (AD) poses a significant global health challenge, underscoring the need for advanced tools facilitating early and precise diagnosis. Previous studies have employed various traditional machine learning techniques, transitioning from image decomposition methods such as principal component analysis to more sophisticated non-linear decomposition algorithms. The advent of computer vision and neural networks has further propelled advancements in the biomedical field. Given the absence of a definitive cure for Alzheimer's in the medical industry and the significance of formulating effective treatment strategies through early diagnosis, the relevance of early detection is paramount. This study centers on Alzheimer's disease diagnosis through MRI data analysis, leveraging radiomic features. It introduces a novel 3D deep learning model integrated with attention modules to tackle the vanishing gradient problem. Additionally, it outlines the development of a machine learning (ML) model for classification using radiomic features extracted from NIFTI files. The project introduces Alz3Dnet, a network that integrates attention modules and addresses vanishing gradient challenges, achieving a validation accuracy of 92.22%. Furthermore, an ML ensemble model utilizing radiomic features attains an accuracy of 96%

Contents

1 Introduction **1**

2 Literature Survey **4**

3 The Deep Learning Approach - Alz3Dnet **7**
3.1 Objective(s) 7
3.2 Proposed Framework 7
3.2.1 ADNI Dataset 8
3.2.2 Data Pre-processing 8
3.2.3 Alz3Dnet 9

4 The Machine Learning Approach - Radiomics Feature Extraction **13**
4.1 Proposed framework 13
4.1.1 PyRadiomics 14
4.1.2 Feature Selection 17
4.1.3 Random forest feature selection 18
4.1.4 The Ensemble model 19

5 Experimental Analysis and Results **24**
5.1 Performance Metrics 24
5.2 Environmental Setup 26
5.3 Results 26
5.3.1 3D Deep learning model 26
5.3.2 ML model 28

6 Conclusion **34**

List of Figures

1.1 Brain images of normal person and person with AD [1] 1
1.2 MRI images of various stages of dementia 2

3.1 Proposed Framework 7
3.2 The MRI image before and after pre processing 8
3.3 The Architecture of Alz3Dnet 9
3.4 The Architecture of CBAM Block 12

4.1 Proposed ML Model 13
4.2 The Architecture of ensemble model 19

5.1 Accuracy graph of Alz3Dnet 26
5.2 Loss graph of Alz3Dnet 27
5.3 Features extracted from MRI image using Random forest feature selection . 29
5.4 Precision Analysis Among Various Models & Feature Extraction Techniques 29
5.5 Recall Analysis Among Various Models & Feature Extraction Techniques . . 29
5.6 F1 Score Analysis Among Various Models & Feature Extraction Techniques . 30
5.7 Accuracy Analysis Among Various Models & Feature Extraction Techniques 30
5.8 Comparison of precision with various models 31
5.9 Comparison of f1 score with various models 31
5.10 Comparison of precision with various models 31
5.11 Comparison of accuracy with various models 31
5.12 Confusion Matrix from proposed ML model 32

List of Tables

5.1 Comparison of results with various models 28
5.2 Comparison of proposed model with existing works 28
5.3 Detailed Comparison with various models 31
5.4 Detailed results of proposed ML model 33

Chapter 1

Introduction

Alzheimer's disease(AD) is one of the deadliest neurodegenerative diseases that spreads across the world. It is being one of the most fearing health condition for the old aged people nowadays. The worst situation of the AD is being affected badly by the daily life and needs. Slowly each of the patients fall into death [1]. It primarily affects cognitive functions including memory, thinking and behaviour parts of the human brain. As per the studies says, The major cause of the diseases are found to be the presence of abnormal proteins, specifically beta-amyloid plaques and tau tangles [2].which aggregates in the brain. The developed plaques will remove the connection between the brain cells.

The plaques typically grows from mild cognitive impairment to severe dementia. As per the conditions, The stages of Alzheimer's are divided into 3 which are CN(Cognitive normal). MCI(Mild Cognitive Impairment) and AD(Alzheimer's Disease) [3]. The disease starts from the early stage of MCI and ends in severe dementia which leads to death.For a normal person who have an age of 60 years, will have a 0.5% of human brain shrinking rate. That is after 60 years, A normal healthy person's may shrink its volume upto 0.5% per year. But coming to a person with alzheimer's, their brain will shrink upto a rate of 2.5% per year. This is about 5 times of a normal person [4].The difference between the normal brain and alzheimer's disease brain is depicted in Fig 1.1. It will cause the brain shrinking in a rapid rate and will loss the functions like memory, language, etc.The person leads to a complete trauma that fails to take decisions on daily needs.The affected parts will fail the person.

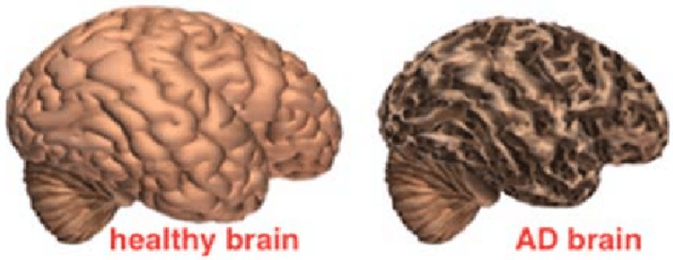

Figure 1.1: Brain images of normal person and person with AD [1]

Currently, the global prevalence of dementia exceeds 55 million individuals, with over

60% residing in low- and middle-income nations. Dementia is diagnosed in someone around the globe every three seconds, with the global population of individuals living with dementia surpassing 55 million in the year 2020 [5]. Annually, there are nearly 10 million new diagnoses. Dementia arises from various brain-affecting diseases and injuries, with Alzheimer's disease comprising 60–70% of cases. It currently ranks as the seventh leading cause of death and a significant contributor to disability and dependency among older populations worldwide. In 2019, dementia incurred a global economic burden of 1.3 trillion US dollars, with roughly 50% of these expenses attributed to informal caregivers, such as family members and close friends, who provide an average of 5 hours of daily care and supervision [1].

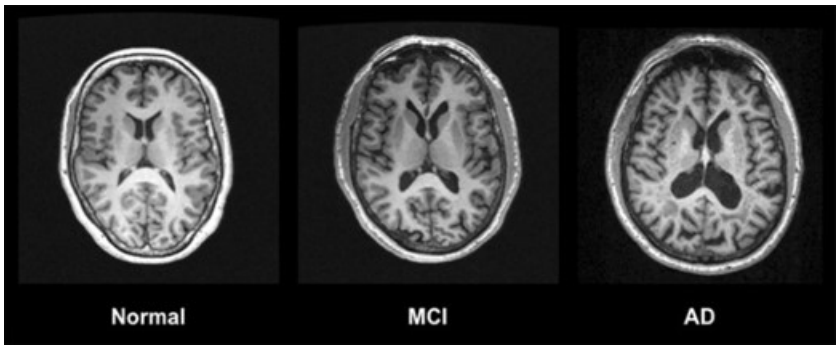

Figure 1.2: MRI images of various stages of dementia

Neuroimaging advances, in particular Magnetic Resonance Imaging (MRI), have created new opportunities for Alzheimer's disease early diagnosis and detection.Fig 1.3 Represents the various stages of Alzheimer's disease in MRI images. MRI provides high-resolution images of brain structures, allowing for detailed examination of anatomical changes associated with AD. However, the complexity and volume of MRI data present significant challenges for manual analysis. This is where radiomics comes into play.

Radiomics is a field of medical study that extracts a large number of quantitative features from medical images using data-characterization algorithms. These features, known as radiomic features, can capture information about the shape, texture, intensity, and spatial relationships of tissues and structures within the brain. When applied to MRI images, radiomics has the potential to reveal subtle changes in brain anatomy that may be indicative of Alzheimer's disease even before clinical symptoms manifest

These features are quantitative descriptors extracted from medical images that capture complex patterns and characteristics of tissues and structures, which are often imperceptible to the human eye. In the context of Alzheimer's disease (AD) classification using MRI images, these features are crucial for early detection and precise diagnosis. Radiomics allows for the identification of subtle changes in brain tissue that precede clinical symptoms, enabling earlier diagnosis and potentially more effective interventions. Unlike traditional diagnostic methods, which can be subjective and variable, radiomic features provide a standardized and objective approach by converting images into numerical data, minimizing human bias and enhancing reproducibility.

These features encompass a wide range of descriptors, including shape, texture, intensity, and wavelet features, which together offer a comprehensive characterization of brain

changes associated with AD. Radiomics can uncover novel biomarkers and differentiate between disease stages, contributing to more targeted and personalized treatment strategies. The integration of radiomic data with machine learning algorithms enhances the predictive power and accuracy of classification models, allowing for robust disease classification and progression tracking in longitudinal studies. By supporting clinical decision-making, radiomic-based tools can corroborate diagnoses, plan treatments, and monitor efficacy, ultimately improving patient care and advancing neuroimaging methodologies in the study and management of Alzheimer's disease.

As Artificial intelligence shows improved innovations in healthcare, using AI algorithms, we need to develop a model which classifies the brain images into the three above-mentioned classes of MCI, NC, AD. The specific research area tried out with a lot of neural networks including pre-trained networks like ResNet50, inception, etc. This project deals mainly focusing on developing two different models for the early detection and classification of Alzheimer's disease. The first model is based on the radiomic features,which are extracted from the brain MRI images, and the second one will clearly classify based on the 3D MRI images.Also the study focuses on the analysis of results in between the ML and DL models and techniques. The newly developed network will address the limitations of the previous networks also classify the brain MRI input into the corresponding category of Alzheimer's disease.

Chapter 2

Literature Survey

Liu et al. proposed a work a network of depthwise separable convolution (DSC) to replace the conventional convolution for the classification of Alzheimer's disease from MRI images [6].The framework is a combination of CNN and DSC and the structure of a standard convolution layer includes convolution module and pooling module. convolution.A typical convolution layer's structure consists of a pooling module and a convolution module. The framework is a hybrid of CNN and DSC.The case where the input feature map is zero-padding and the training step size is one serves as the foundation for this framework.A DSC is composed of a depthwise convolution module, a pointwise convolution module with a stride of one, and a depthwise convolution module with zero padding.A depthwise separable convolution starts with a depthwise convolution. It independently applies one filter to each input channel.This means that no information is shared across channels, which can reduce the number of parameters.Pointwise convolution is the second step in a depthwise separable convolution. It applies a 1x1 convolution to the output of the depthwise convolution. This convolution can be used to combine information across channels, which can increase the model's expressiveness.here, The common training models of AlexNet and GoogLeNet are used for transfer learning to improve the classification accuracy rate of AD detection. The proposed frameworks gives benefits like Reduced computational complexity, Reduced memory requirements and Improved interpretability.

A DL-architecture agnostic ensemble strategy Deep Transfer Ensemble (DTE) for deep neural networks trained through transfer learning [7] .A Deep neural network ensemble (DTE) trained via transfer learning. Additionally, an ensemble of deep neural networks trained from scratch will be built. For large hyperparameter spaces, the time spent tuning hyperparameters is significantly reduced by using random search. With the training time of a single model, ensembles of deep neural networks can be created very effectively using snapshot ensemble. The convolution layers of VGG16 were transferred to DTE, and only the fully connected (FC) layers were trained from scratch.The classification is done by a combination of 2 tasks which are NC vs AD and MCI vs AD. For the large ADNI baseline dataset, the DTE achieved a maximum classification accuracy with best performance.
Another work proposed a multi-stream convolutional neural network used for the classification of Progressive MCI in Alzheimer's Disease Using Structural MRI Images [8] . Using the 3D Slicer software for anterior commissure-posterior commissure correction and N4ITK, an improved nonparametric nonuniform normalisation method, for intensity inhomogeneity

correction. Rigid registration, which entails linearly aligning MRI images to the Colin27 template and resampling them to a size of 155 × 185 × 150 with a resolution of 1 × 1 × 1 mm^3, and skull stripping, which uses a pre-trained U-Net2 to remove both the skull and the dura categorised sMCI and pMCI patients using the same anatomical locations.The proposed model used 3 steps the experiment. Training the proposed multi-stream CNN with MRI images of AD and CN subjects, Transferring weights from the trained model in Step 1 to the identical architecture, fine-tuning it with the data related to sMCI and pMCI patients, and adding biomarkers as an auxiliary modality to examine the reciprocal influence of spectral-spatial features and biomarkers.

Deep Convolutional Autoencoder (CAE) architecture is a tool that can perform an automatic non-linear decomposition of a very large dataset [9].The model is a combination of two analysis from the tasks of classification and regression.The classification part is done by a support vector machine classifier and A neural classifier based on a MLP .prediction of clinical variables and neuropsychological tests using neural regression, a multilayer projection with two hidden layers of 64 neurons.To interface with the Z-layer, typical CAEs use dense layers at the end of the encoder and the beginning of the decoder. Here, a Global Average Pooling (GAP) layer's encoder output takes its place. The data-driven features extracted with this technique have been largely related to other clinical and neuropsychological variables It shows large correlations (greater than 0.63) with clinical variables such as age, tau protein deposits and especially neuropsychological examinations.

3D data can be used without converting it into 2D by dimensionality reduction [10] .The proposed model is defined as a 3D convolutional broad learning system (3D-CBLS).To accomplish end-to-end training, the model is trained in order to extract the features from the images. The MRI images are used to train the model directly, negating the need for further dimension reduction.Inspired by InceptionNet, the proposed feature mapping module is a pipeline with a stacked convolution network block as the mainline.Several parallel convolution modules, with data transmission limited to the bottom convolutional layer, are present in each FMM.Convolution feature pooling is a crucial downsampling technique, and two binary auto diagnostic tasks are used to assess both the suggested and rival approaches.

Another work proposed a 3D Convolution network based on multimodel Alzheimers data [11]. By using MRI and FDG-PET images , designed a Three-dimensional convolutional neural networks (3D CNN) to evaluate the effectiveness of the image fusion approach. A module of Sparse autoencoder added to the network to increase the feature description ability Also Utilizes a multi-modal image fusion approach to combine Magnetic Resonance Images (MRI) with Positron Emission Tomography (PET) images from Alzheimer's disease (AD) patients. The model employs a three-dimensional convolutional neural network (3D CNN) to extract features from the fused images, resulting in richer multi-modal feature information. The work claims that the sparse autoencoder allowed the network to learn the characteristics that best express the sample in a harsh environment, and can effectively reduce the dimensionality of the sample.

Attention based 3D Convolutional networks with multi-sclae integration blocks can perform better in classification tasks with MRI images [12]. The proposed model, AMSNet,

approaches Alzheimer's Disease (AD) classification, by presenting an attention-based 3D multiscale CNN architecture. This model incorporates pivotal components such as the MS Block, Attention Blocks, and a fusion of a global average pooling layer with a single fully connected layer. Given the critical importance of early AD diagnosis in enhancing patient outcomes and quality of life, the AMSNet model is designed to make significant contributions in this domain. Its primary objective is to facilitate early AD diagnosis by delivering exceptional classification accuracy while maintaining efficiency with fewer parameters and a relatively low computational burden.The study claims that, The Central to its efficacy is the MS Block, which integrates multi-scale features utilizing the attention mechanism, thereby augmenting classification performance.

A work that utilized radiomics analysis on multiparametric MRI to identify potential biomarkers for preclinical Alzheimer's disease (AD), showcasing high classification efficiency through SVM and RF models [13]. By identifying three stable high-frequency features from structural MRI, the study offers a promising avenue for early AD diagnosis, providing valuable insights into the preclinical stage of the disease. The advantage of this work lies in its innovative approach of using radiomics analysis to extract imaging biomarkers, which could revolutionize the early detection and intervention strategies for AD. However, the study does have limitations, including a relatively small sample size that may limit the generalizability of the findings. Additionally, the reliance on amyloid-positron emission tomography for participant selection could introduce bias, and the lack of standardization in defining "unstable preclinical AD" poses challenges in interpreting the results. Despite these limitations, the research contributes significantly to the field by highlighting the potential of radiomics analysis in advancing the early diagnosis of AD.

Another work delves into the application of radiomics in neurodegenerative diseases, showcasing the extraction of image features from various modalities like MRI, PET, and CT to analyze physiological and pathological changes [14]. It emphasizes the significance of dataset information, including data collection, feature extraction, and model construction in the radiomics process. The study employs techniques such as feature selection methods and machine learning models to predict disease progression and subtype differentiation. Results indicate the potential of radiomics in enhancing the diagnosis and prediction of neurodegenerative diseases by extracting more data than visible to the naked eye. However, challenges in clinical application, such as the need for large sample sizes, prospective research, and real-world clinical trials, are highlighted to improve model evaluation and acceptance among clinicians and patients.

Chapter 3

The Deep Learning Approach - Alz3Dnet

3.1 Objective(s)

- To design an ML model that classifies the Alzheimer's disease based on its radiomic features.
- To develop a new deep learning model for the AD classification using MRI images, which consists skip connections that might resolve the vanishing gradient issue
- Analyse and study the results from the machine learning and deep learning models.

3.2 Proposed Framework

Figure 3.1 shows the proposed framework.

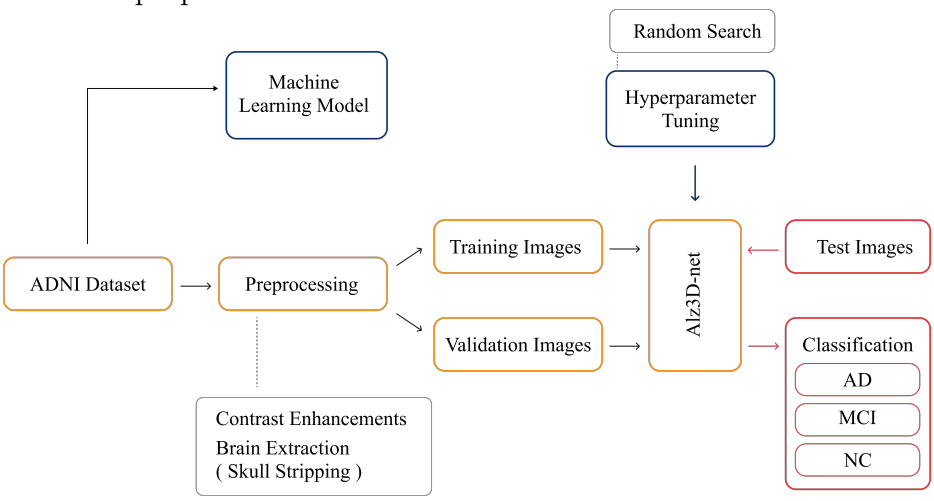

Figure 3.1: Proposed Framework

The project flows with the diagram as mentioned above. The collected dataset from ADNI source is fed on to the pre-processing phase. Here the data flows towards both ML

and 3D network. In the 3D network, The data is divided intop train and validation sets after preprocessing, and flow into the newly developed algorithm Alz3D-Net. After the successful training, and the hyper parameter tuning the results are more precised. The model gives better results when a test image is passed through the network and it will predict as AD/MCI/NC.

3.2.1 ADNI Dataset

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a large-scale, multi-center research initiative launched in 2004 with the goal of advancing research on Alzheimer's disease (AD) through the use of neuroimaging, clinical, cognitive, and genetic data [15]. ADNI is a public-private partnership involving government agencies, nonprofit organizations, and private pharmaceutical companies. The primary objective of ADNI is to accelerate the development of treatments and diagnostics for Alzheimer's disease by providing researchers with a comprehensive dataset.For this project, 1.5T data was utilized for training and validation purposes. The dataset comprised a total of 1123 data points, with 374 belonging to Class AD, 365 to Class CN, and 384 to Class MCI. To facilitate the training process effectively, the dataset was partitioned into train and validation sets, maintaining an 80:20 ratio respectively. This division allowed for robust model training while enabling rigorous validation to ensure the model's generalizability and performance across diverse data samples.

3.2.2 Data Pre-processing

The dataset need to be pre-processed using various techniques for the improved results and better training. In this study, Deepbrain module is used for the contrast enhancements and the skull stripping. The skull stripping or braijn extraction using deepbrain modules removes the irrelevent area from the data. This process will help the model to more focused on the data and the reduce the parameters by training irrelevent area. The difference of images with and without pre-processing is illustrated in Fig 3.2. The successful extraction from the MRI images and the contrast enhancement helped the model for the easy understanding of features and to get the improved output.

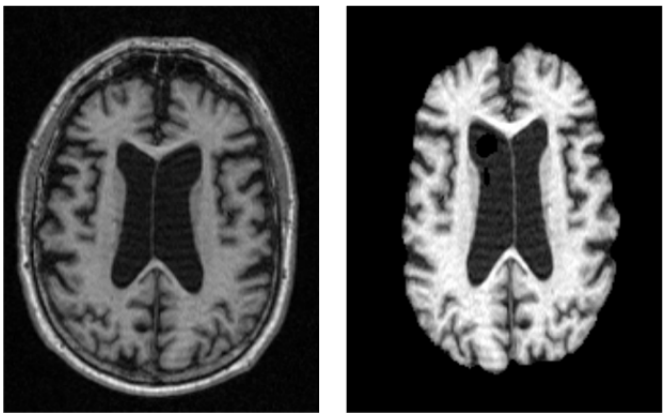

Figure 3.2: The MRI image before and after pre processing

3.2.3 Alz3Dnet

The proposed architecture, Alz3Dnet is an inspiration from a variety of Residual architectures, strategically integrating Residual blocks throughout the model to effectively tackle the gradient descent issue. Additionally, the inclusion of Convolutional Block Attention Module (CBAM) serves to enhance the representational capacity of Convolutional Neural Networks (CNNs), thereby refining the model's ability to discern intricate patterns within the data. Leveraging skip connections at multiple stages further mitigates the challenge of vanishing gradients, ensuring smoother and more efficient training. Notably, CBAM contributes to this by incorporating both channel-wise and spatial-wise attention mechanisms, enabling the model to focus on relevant features while processing the input data. With an input shape requirement of 96x96x96, the architecture is optimized to accommodate datasets of varying complexities, providing a versatile framework for diverse applications.

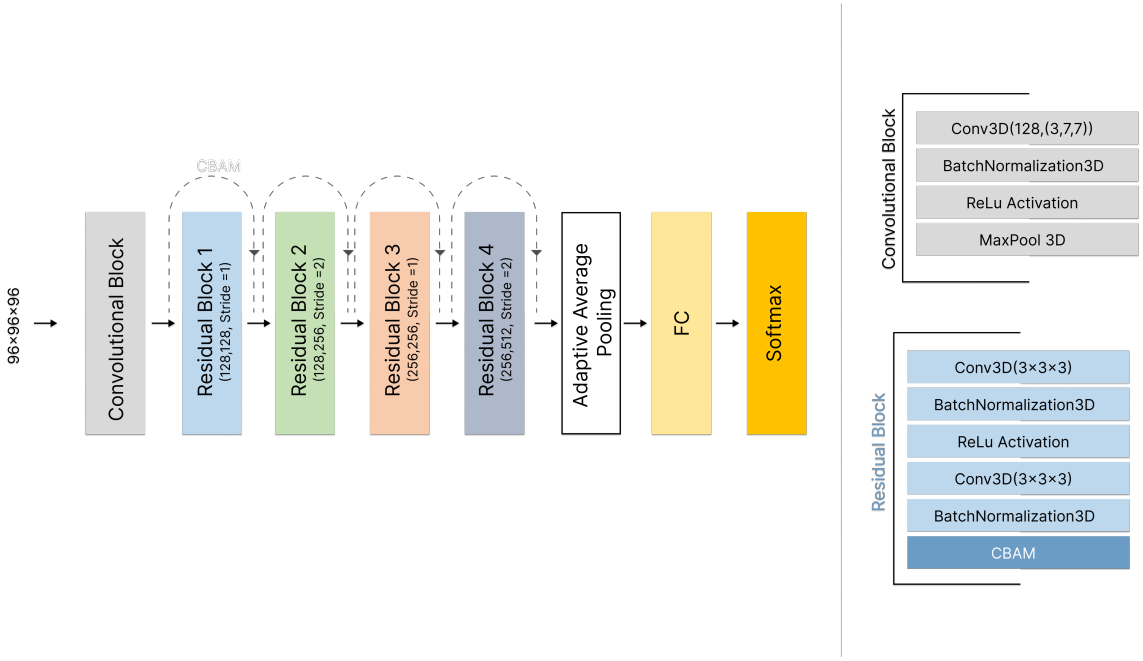

Figure 3.3: The Architecture of Alz3Dnet

The various blocks in the proposed model is explained below:

Initial Convolution and Pooling :

The initial convolutional layer processes the input volume (1 x D x H x W) with a 3D convolutional operation, producing 128 feature maps. This is achieved using a kernel size of (3, 7, 7) and a stride of (1, 2, 2), which reduces the spatial dimensions while maintaining the depth. The convolution is followed by batch normalization to stabilize and accelerate the training process. A ReLU activation function introduces non-linearity, and a subsequent max-pooling layer further reduces the spatial dimensions, down-sampling the feature maps to 128 x 96 x 24 x 24.

Residual Layers :

The network consists of four residual layers, each containing one or more ResidualBlocks. Each block incorporates two sets of 3D convolutional layers, batch normalization, and ReLU activation functions. The structure within these blocks is designed to facilitate the learning of residual mappings, which ease the optimization and improve gradient flow.

Residual Block :

Each Residual Block within the network consists of:

- First Convolutional Sequence: A 3D convolutional layer followed by batch normalization and ReLU activation.
- Second Convolutional Sequence: Another 3D convolutional layer followed by batch normalization.
- CBAMBlock Integration: This block contains both channel and spatial attention mechanisms. The channel attention mechanism utilizes global average pooling followed by two convolutional layers and a sigmoid activation to scale the feature maps adaptively. The spatial attention mechanism applies a convolutional layer followed by a sigmoid function to produce spatial attention maps.
- Identity Connection: If the input and output dimensions differ, a downsampling path consisting of a 1x1 convolution followed by batch normalization is used. The output of the convolutional sequences and the identity connection are added together, followed by a ReLU activation to complete the residual block.

Attention Mechanisms :

The Convolutional Block Attention Module (CBAM) consists of two main components [16]:

Channel Attention : Channel attention mechanisms, integral to convolutional neural networks (CNNs), dynamically modulate the importance of feature channels to augment the network's ability to discern salient features from the input data. Initially, global pooling operations, such as global average pooling or global max pooling, are applied to aggregate information across spatial dimensions, facilitating the computation of channel-wise summaries [17]. These summaries are subsequently processed through fully connected layers, often incorporating non-linear activations like ReLU, culminating in the generation of attention weights. These weights, constrained within the range of 0 to 1 via sigmoid activation, signify the relevance of each channel and are pivotal in determining how much emphasis to allocate to each during feature processing.

Channel attention offers a plethora of advantages, primarily enhancing feature representation within CNNs. By empowering the network to selectively attend to crucial channels, it fosters the learning of discriminative features essential for various tasks such as image classification, object detection, and semantic segmentation. Consequently, this leads to performance improvements across diverse applications, as the network becomes adept at discerning and prioritizing relevant features. Furthermore, channel attention mechanisms, exemplified by

modules like the Squeeze-and-Excitation (SE) block, boast seamless integration into existing CNN architectures, offering a modular and flexible solution for improving model Efficiency.

The fusion of technical nuances with practical Efficiency underscores the significance of channel attention in contemporary deep learning paradigms. By leveraging attention weights derived from global information aggregation, channel attention empowers CNNs to adaptively refine feature representations, amplifying their capacity to discern and leverage informative features. This nuanced approach to feature modulation not only elevates model performance and efficiency but also exemplifies a paradigm shift towards more sophisticated and adaptive deep learning methodologies tailored to tackle the complexities of real-world datasets.

Spatial Attention : Spatial attention mechanisms in deep learning are pivotal for enhancing the discriminative capabilities of convolutional neural networks (CNNs) by selectively focusing on relevant spatial regions within an input feature map. Unlike channel attention, which prioritizes informative feature channels, spatial attention mechanisms aim to highlight important spatial locations or regions [16]. These mechanisms dynamically modulate the importance of different spatial locations, allowing the network to adaptively allocate more resources to regions that are more pertinent to the task at hand, such as object detection, semantic segmentation, or image classification.

The core principle of spatial attention involves computing attention maps that encode the significance of various spatial locations within an input feature map. These attention maps are typically generated through a process of spatial pooling and subsequent processing through convolutional or fully connected layers. Global pooling operations, such as global average pooling or global max pooling, are employed to aggregate spatial information, which is then passed through one or more layers to compute attention weights for each spatial location. These attention weights indicate the relative importance of each location and are used to modulate the feature map, emphasizing informative regions while suppressing less relevant ones.

Spatial attention mechanisms have been shown to improve the performance of CNNs across a range of computer vision tasks. By enabling the network to focus on relevant spatial regions, spatial attention enhances feature representation and promotes better generalization to diverse datasets. Moreover, spatial attention can be integrated into existing CNN architectures seamlessly, offering a flexible and effective means of enhancing model efficacy without significantly increasing computational overhead. Overall, spatial attention mechanisms play a crucial role in advancing the capabilities of deep learning models, contributing to their adaptability and robustness in real-world applications.

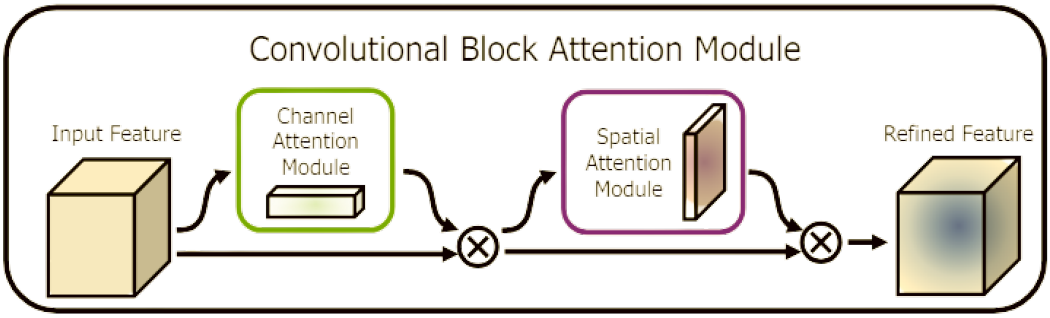

Figure 3.4: The Architecture of CBAM Block

Final pooling and classification :

After the last residual layer, the network applies an adaptive average pooling layer that reduces the feature maps to a fixed size of 1 x 1 x 1, irrespective of the input dimensions. This output is then flattened to a 1D vector, which is passed through a fully connected layer that maps the features to the desired number of classes. Finally, a softmax activation function is applied to generate class probabilities, completing the classification task.

Chapter 4

The Machine Learning Approach - Radiomics Feature Extraction

4.1 Proposed framework

The machine learning approach also utilizes the ADNI dataset that is used in the Deep learning model. But in this case, The pre-processed dataset from the Alzheimer's Disease Neuroimaging Initiative (ADNI) undergoes a thorough analysis using PyRadiomics, a specialized Python package designed for extracting radiomic features from medical images, including CT scans, MRI scans, and PET scans. PyRadiomics processes the NIFTI files to extract a rich array of 130 radiomic features per individual file, encompassing aspects such as texture, shape, and intensity that provide valuable insights into the underlying biological structures.

Figure 4.1 shows the framework of ML model

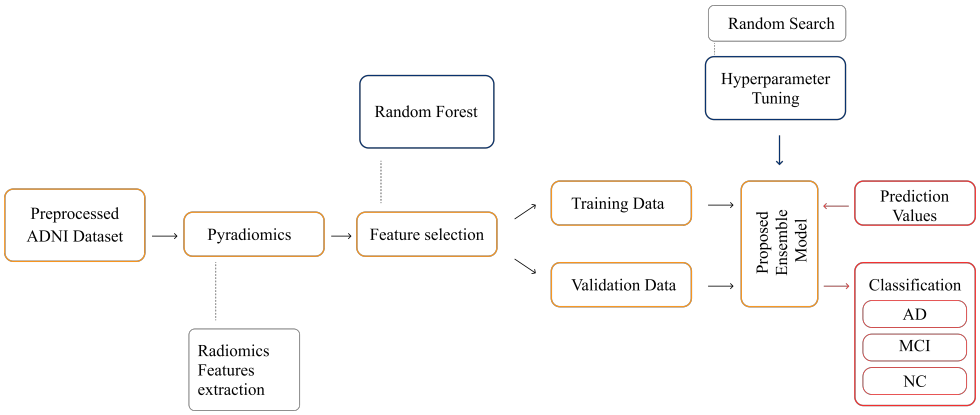

Figure 4.1: Proposed ML Model

As the above mentioned figure, The preprocessed dataset fed up into pyradiomics. The pyradiomics package extracts the radiomics features within the NIFTI images. The features are filtered using Random forest feature selection algorithm and a number of 15 features are selected for the evaluation on the basis of importance. Sequently the data is divided into train and validation at a ratio of 90:10 and trained using an ensemble method. After the

training the model is predicted when give radiomic feature values as input and gives the particular class as output.

To manage the high dimensionality of the dataset and enhance the efficiency of subsequent analysis, Various algorithms are utilized for feature selection. By assessing the contribution of each feature to the overall model accuracy, We identified and retained the 15 most significant features. This reduction not only streamlines the dataset but also ensures that the most relevant features are used for predictive modeling.

Following feature selection, the dataset is partitioned into training and validation sets with a 90:10 ratio. This partitioning strategy is crucial for robust model training and performance evaluation. The larger training set enables the model to learn from a substantial amount of data, improving its generalization capabilities, while the validation set provides an independent measure of the model's performance, ensuring it can accurately predict outcomes on unseen data. This structured framework, leveraging advanced feature extraction and selection techniques, sets the stage for effective predictive modeling in medical image analysis, particularly in the context of Alzheimer's disease diagnosis and progression monitoring.

4.1.1 PyRadiomics

PyRadiomics is a powerful Python package specifically designed for extracting radiomics features from a variety of medical images, including but not limited to CT scans, MRI scans, and PET scans. Leveraging this versatile tool, researchers and practitioners can effortlessly delve into the intricate details of medical images, uncovering a wealth of quantitative features that hold immense diagnostic and prognostic value. With its comprehensive suite of functionalities, PyRadiomics streamlines the process of feature extraction, offering a seamless pathway to analyze radiological data with precision and efficiency. Whether unraveling complex patterns within tumor images or discerning subtle nuances indicative of disease progression, PyRadiomics equips users with the tools needed to navigate the complexities of medical imaging analysis, driving advancements in clinical research and patient care.

In this study the radiomic feature from MRi images are extracted using Pyradiomics. 130 features are extracted from a single MRI images.The features include First-Order Statistics, Shape Descriptors,Gray Level Co-occurrence Matrix (GLCM),Gray Level Size Zone Matrix (GLSZM),Gray Level Run Length Matrix (GLRLM),Gray Level Dependence Matrix (GLDM),Neighboring Gray Tone Difference Matrix (NGTDM) amd Diagnostics Features. The characteristics of the radiomic features are mentioned below .

First-Order Statistics

These features describe the distribution of voxel intensities within the image region of interest (ROI), providing basic statistical properties.

- **Mean:** The average intensity value of the voxels in the ROI.
- **Median:** The middle value of the intensity distribution in the ROI.
- **Standard Deviation:** The measure of intensity variation from the mean.
- **Variance:** The average of the squared differences from the Mean.

- **Skewness:** The asymmetry of the intensity distribution around the Mean.
- **Kurtosis:** The peakedness of the intensity distribution.
- **Energy:** The sum of squared intensity values in the ROI.
- **Entropy:** The measure of randomness in the intensity values.
- **Minimum, Maximum, and Range:** The extreme intensity values and their difference.
- **10th, 90th Percentiles:** The intensity values below which 10% and 90% of the data fall, respectively.

Shape Descriptors

These features describe the geometric properties of the ROI, capturing its size, shape, and spatial orientation.

- **Volume:** The total number of voxels within the ROI.
- **Surface Area:** The area of the ROI boundary.
- **Sphericity:** A measure of how spherical the ROI is (surface area to volume ratio).
- **Compactness:** The degree to which the ROI is compact.
- **Elongation:** The ratio describing how elongated the shape is.
- **Flatness:** The ratio of the smallest to the largest principal component, describing how flat the shape is.
- **Major, Minor, and Least Axis Lengths:** Lengths of the principal axes of the ROI.

Gray Level Co-occurrence Matrix (GLCM)

These features describe the frequency of co-occurring intensity values at a specified spatial relationship (offset).

- **Contrast:** Measures the local intensity variation; high contrast means high variation.
- **Correlation:** Measures the linear dependency of intensity values to their neighbors.
- **Dissimilarity:** Measures the difference between neighboring intensity values.
- **Homogeneity:** Measures the closeness of distribution of elements in the GLCM to the GLCM diagonal.
- **Energy (Angular Second Moment):** The sum of squared elements in the GLCM; measures textural uniformity.
- **Entropy:** Measures the randomness of the intensity distribution.

Gray Level Size Zone Matrix (GLSZM)

These features describe the size of homogeneous zones of connected voxels that share the same intensity.

- **Small Area Emphasis (SAE):** Emphasizes smaller homogeneous zones.
- **Large Area Emphasis (LAE):** Emphasizes larger homogeneous zones.
- **Gray Level Non-uniformity (GLN):** Measures the similarity of intensity values in zones.
- **Zone Size Non-uniformity (ZSN):** Measures the variability of zone sizes.
- **Zone Percentage (ZP):** The ratio of the number of zones to the total number of voxels.

Gray Level Run Length Matrix (GLRLM)

These features describe the length of consecutive voxels having the same intensity value.

- **Short Run Emphasis (SRE):** Measures the distribution of short runs.
- **Long Run Emphasis (LRE):** Measures the distribution of long runs.
- **Gray Level Non-uniformity (GLN):** Measures the similarity of intensity values.
- **Run Length Non-uniformity (RLN):** Measures the similarity of run lengths.
- **Run Percentage (RP):** The ratio of the number of runs to the total number of voxels.

Gray Level Dependence Matrix (GLDM)

These features describe the dependence of voxels on their neighboring voxels' intensity values.

- **Small Dependence Emphasis (SDE):** Emphasizes small dependencies in intensity values.
- **Large Dependence Emphasis (LDE):** Emphasizes large dependencies in intensity values.
- **Gray Level Non-uniformity (GLN):** Measures the variability of intensity values.
- **Dependence Non-uniformity (DN):** Measures the variability of dependencies.
- **Dependence Entropy (DE):** Measures the randomness of dependencies.

Neighboring Gray Tone Difference Matrix (NGTDM)

These features describe the difference between a voxel's intensity and the average intensity of its neighbors.

- **Coarseness:** Measures the texture coarseness; higher values indicate coarser textures.
- **Contrast:** Measures the texture contrast; higher values indicate greater intensity differences.
- **Busyness:** Measures the level of texture busyness; higher values indicate busier textures.
- **Complexity:** Measures the texture complexity.
- **Strength:** Measures the strength of the intensity gradients.

Diagnostics Features

These features provide metadata and quality information about the image and the ROI used in the analysis.

- **Bounding Box:** The coordinates defining the smallest box containing the ROI.
- **Center of Mass:** The center coordinates of the ROI.
- **Spacing:** The physical distance between voxel centers in the image.
- **Volume (of the whole image):** The total volume covered by the image.

4.1.2 Feature Selection

The radiomic data extracted from the NIFTI files encompasses approximately 130 features, presenting a rich but potentially overwhelming dataset. To ensure the optimal performance of our classification model and mitigate the risk of overfitting, it is imperative to conduct feature selection. The inclusion of a large number of features can lead to overfitting and increase the complexity of the model, diminishing its generalizability and interpretability. Thus, we employed various feature selection algorithms to identify and retain the most informative features from the dataset. These techniques aim to distill the wealth of information contained within the radiomic data into a more manageable and relevant subset, enhancing the efficacy and interpretability of our classification model.

Through rigorous experimentation and evaluation, we identified the feature selection algorithms that yielded the most promising results in terms of classification performance. By leveraging these techniques, we were able to extract a subset of features that exhibited the highest discriminatory power in distinguishing between different classes of cognitive health. This refined feature set not only enhances the efficiency of our classification model but also facilitates a deeper understanding of the underlying patterns and relationships within the radiomic data. Furthermore, by reducing the dimensionality of the dataset while preserving its salient features, we strike a balance between model complexity and predictive accuracy, ensuring robust and reliable classification outcomes.

Extra tree Classifer

The Extra Trees classifier is a robust algorithm frequently employed for feature selection tasks in medical imaging, particularly in the diagnosis of Alzheimer's disease using radiomic data from brain images. With its ensemble learning approach, the classifier constructs multiple decision trees using random subsets of features and data samples, effectively reducing the risk of overfitting. By further randomizing the feature selection process through the selection of split points at random, it enhances robustness and provides reliable estimates of feature importance, crucial for identifying the radiomic features most pertinent to disease diagnosis [18].

In this work, accurate feature selection is essential for early detection and intervention, So the Extra Trees classifier excels in identifying the radiomic features that best differentiate between different stages of cognitive health. By analyzing subtle patterns and textures within brain images, it pinpoints the most informative features correlating with disease progression. This capability aids clinicians in making accurate and timely diagnoses, facilitating more effective treatment planning and patient care.

Recursive feature elimination

Recursive Feature Elimination (RFE) is another powerful feature selection technique widely used in medical imaging, especially in diagnosing Alzheimer's disease from radiomic data extracted from brain images. RFE operates by iteratively removing the least significant features from the dataset until the optimal subset of features is identified. This iterative process enables the algorithm to systematically assess the importance of each feature, ranking them based on their contribution to the classification task. By iteratively training the model on increasingly smaller subsets of features, RFE effectively identifies the most discriminative features while minimizing overfitting.

In the diagnosis of Alzheimer's disease, where precise feature selection is crucial for accurate classification, RFE proves invaluable in identifying the subset of radiomic features most relevant to disease diagnosis. By systematically evaluating the importance of each feature and retaining only those that contribute most significantly to the classification task, RFE enhances the interpretability and efficiency of the diagnostic model. This targeted approach not only improves the accuracy of disease detection but also facilitates a deeper understanding of the underlying biomarkers associated with Alzheimer's disease progression [19], ultimately leading to more effective patient management and treatment strategies.

4.1.3 Random forest feature selection

The Random Forest classifier is a versatile and widely-used algorithm renowned for its efficacy in feature selection tasks, particularly in medical imaging applications such as diagnosing Alzheimer's disease from radiomic data extracted from brain images. Unlike traditional decision trees, Random Forest constructs a multitude of decision trees using bootstrapped samples of the dataset and a random subset of features at each node split. This randomized selection of features reduces the risk of overfitting and enhances model robustness by mitigating the dominance of individual features. Additionally, Random Forest calculates the feature importance by measuring the decrease in node impurity, such as Gini impurity or entropy, across all decision trees in the forest.

In this study, precise and interpretable feature selection is critical. Random Forest excels in identifying the subset of radiomic features most pertinent to disease classification [20]. By analyzing the contribution of each feature to reducing impurity across the ensemble of decision trees, Random Forest effectively ranks the features based on their discriminative power. This comprehensive assessment not only facilitates the identification of biomarkers associated with Alzheimer's disease progression but also improves the interpretability of the diagnostic model. Moreover, Random Forest's inherent ability to handle high-dimensional data and nonlinear relationships makes it a robust choice for feature selection in complex medical imaging datasets, ultimately leading to more accurate and reliable disease diagnoses.

4.1.4 The Ensemble model

After the splitting of data into training and validation, various classification models are tried out with the data. By incorporating the various model, we designed a new ensemble method to complete the classification purpose. Fig 4.2 shows the framework of Ensemble model.

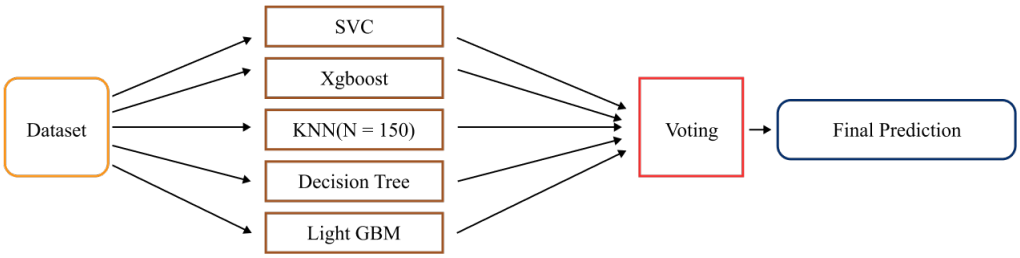

Figure 4.2: The Architecture of ensemble model

The model includes 5 classifiers such as support vector classifier, Xgboost,K-nearest classifier, Decision tree and Light gradient boosting method. The features of these model helped to create a voting based ensemble method and provided better results. The characteristics of the included models are expressed below.

Support Vector Classifier (SVC) :

A Support Vector Classifier (SVC) is a supervised machine learning algorithm used for classification tasks. It is a specific type of Support Vector Machine (SVM) that aims to find the best boundary (or hyperplane) that separates data points of different classes in a high-dimensional space. The key concepts are defined below:
Hyperplane: In the context of SVMs, a hyperplane is a decision boundary that separates data points of different classes. In a two-dimensional space, it's a line; in three dimensions, it's a plane; and in higher dimensions, it's a hyperplane.
Margin: The margin is the distance between the hyperplane and the nearest data points from each class. These nearest data points are called support vectors. SVM aims to maximize this margin to improve the model's generalizability.
Support Vectors: These are the data points that are closest to the hyperplane and are crucial in defining the position and orientation of the hyperplane. The classifier only depends on these support vectors and not on the rest of the data points.
Kernel Trick: When data is not linearly separable in the original feature space, SVM can use

a kernel function to transform the data into a higher-dimensional space where a hyperplane can be used to separate the classes. Common kernels include:
Linear Kernel: No transformation, the data is assumed to be linearly separable. Polynomial Kernel: Maps the data into a higher-dimensional space using polynomial functions. Radial Basis Function (RBF) Kernel: Uses Gaussian functions to map the data into a higher-dimensional space, allowing for non-linear separation.
Regularization Parameter (C): This parameter controls the trade-off between achieving a low error on the training data and minimizing the norm of the weights, which helps in controlling overfitting. A small value of C creates a wider margin but allows more misclassifications, while a large value of C tries to classify all training examples correctly but may lead to a smaller margin and overfitting.

XGBoost :

XGBoost (Extreme Gradient Boosting) is a powerful and efficient implementation of the gradient boosting framework, designed to optimize both performance and speed. It builds an ensemble of decision trees in a sequential manner, where each new tree corrects the errors made by the previous trees. The model starts with an initial prediction and iteratively adds trees that predict the residuals (errors) of the combined predictions of all previous trees. This approach ensures that each subsequent model focuses on the most challenging cases, progressively improving overall accuracy.

The optimization process in XGBoost involves the use of gradient descent to minimize the loss function, which measures the difference between the actual and predicted values. Additionally, XGBoost incorporates regularization techniques, such as L1 (Lasso) and L2 (Ridge) regularization, directly into the objective function to control model complexity and prevent overfitting. This regularization makes XGBoost particularly robust and suitable for handling noisy data and high-dimensional feature spaces. The inclusion of a learning rate parameter further helps in fine-tuning the model by controlling the contribution of each tree to the final prediction.

XGBoost is highly regarded for its scalability and flexibility. It can efficiently process large datasets and is capable of parallel and distributed computing, which significantly speeds up training times. The algorithm also supports a variety of custom loss functions and evaluation metrics, making it adaptable to a wide range of machine learning tasks, including classification, regression, and ranking. While it requires careful tuning of hyperparameters such as the number of trees, maximum depth, and learning rate, XGBoost's ability to deliver high accuracy and its robustness to overfitting make it a preferred choice for many data scientists and machine learning practitioners.

KNN Algorithm :

K-Nearest Neighbors (KNN) is a simple, yet powerful supervised machine learning algorithm used for classification and regression tasks. It operates on the principle that similar data points are often close to each other in a multidimensional space. When making a prediction, KNN identifies the k nearest data points (neighbors) to the input sample and determines the output based on these neighbors. In a classification setting, the algorithm assigns the most common class among the k neighbors, while in regression, it averages the values of the neighbors. For this explanation, we consider k=150, meaning the algorithm will look

at the 150 nearest neighbors to make its prediction.Choosing k =150. k=150 has specific implications. A larger k value, such as 150, can lead to more stable and smooth decision boundaries since it incorporates a broader sample of data points, reducing the impact of noise and outliers. However, it can also lead to more computational overhead as the algorithm must calculate distances for many points, and it may blur distinctions between classes if they are not well-separated. The distance metric (commonly Euclidean) plays a crucial role in determining the neighbors, and feature scaling becomes important to ensure that all features contribute equally to the distance calculation.

Despite its simplicity, KNN has several advantages and disadvantages. It is straightforward to implement and understand, making it a good starting point for many problems. Additionally, it makes no explicit assumptions about the data distribution, making it versatile. However, KNN can be computationally expensive, especially with a large k value like 150, as it requires distance calculations for every data point during prediction. It also demands significant memory to store the entire training dataset. Consequently, while KNN with k=150 can be effective for specific tasks, it is crucial to consider the trade-offs in computational efficiency and prediction accuracy.

Decision Tree :

A Decision Tree is a popular and intuitive supervised machine learning algorithm used for both classification and regression tasks. It works by recursively splitting the dataset into subsets based on the feature that provides the maximum information gain or the minimum Gini impurity. Each internal node of the tree represents a decision based on a feature, each branch represents the outcome of the decision, and each leaf node represents the final output or class. This structure resembles a flowchart and makes decision trees easy to interpret and visualize.

The process of building a decision tree involves selecting the best feature to split the data at each node. For classification tasks, the criteria for selecting the feature often include measures like Information Gain, which is based on entropy, or the Gini Index. For regression tasks, criteria like Mean Squared Error (MSE) are used. The goal is to choose splits that result in the most homogenous subsets possible, thereby reducing the impurity. The tree continues to split the data recursively until it meets a stopping condition, such as a maximum tree depth or a minimum number of samples per leaf node.

While decision trees have several advantages, such as being easy to understand, interpret, and requiring little data preprocessing, they also have notable drawbacks. Decision trees can easily overfit the training data, especially if they are allowed to grow too deep. This overfitting can lead to poor generalization on unseen data. To mitigate this, techniques such as pruning (removing sections of the tree that provide little power) or setting constraints (like maximum depth or minimum samples per leaf) are used. Additionally, decision trees can be unstable because small changes in the data might result in a completely different tree structure. Despite these limitations, decision trees are a fundamental and widely used method in machine learning, and they form the basis for more advanced ensemble methods like Random Forests and Gradient Boosting Machines.

Light GBM :

LightGBM (Light Gradient Boosting Machine) is an advanced gradient boosting framework designed for speed and efficiency, developed by Microsoft. It is particularly well-suited for handling large datasets with high-dimensional features and is widely used for both classification and regression tasks. LightGBM achieves its performance through two main innovations: the Gradient-based One-Side Sampling (GOSS) and the Exclusive Feature Bundling (EFB) techniques. GOSS reduces the number of data instances to be used in each iteration by focusing on instances with larger gradients, while EFB bundles mutually exclusive features to reduce the number of features, thereby speeding up training.

LightGBM builds decision trees sequentially, where each new tree is trained to correct the errors of the previous trees, similar to other gradient boosting methods. However, it uses a leaf-wise growth strategy rather than the level-wise strategy used by many other implementations. This means LightGBM grows trees by expanding the leaf with the maximum loss reduction, leading to deeper and potentially more complex trees. This strategy can result in better accuracy but also requires careful tuning to prevent overfitting. Key parameters include the number of leaves, learning rate, and max depth, among others.

The main advantages of LightGBM include its high efficiency and scalability. It can handle large-scale data with millions of instances and thousands of features with comparatively lower memory usage and faster training times. Additionally, LightGBM supports parallel and distributed computing, further enhancing its scalability. However, like all models, it has its limitations. The leaf-wise growth strategy can sometimes lead to overfitting if not properly managed, and the algorithm's complexity can make it harder to interpret compared to simpler models. Despite these challenges, LightGBM is a powerful tool in the machine learning practitioner's toolkit, offering top-tier performance in many competitive and real-world applications.

Logistic regression

Logistic regression is a statistical method for binary classification that can be extended to handle multiple classes through techniques such as one-vs-rest (OvR) or multinomial logistic regression. In the context of a three-class classification problem, multinomial logistic regression is often used. This approach generalizes logistic regression to predict the probability of each of the three classes. The model uses the softmax function to ensure that the output probabilities sum up to one, with each probability representing the likelihood of the input belonging to one of the classes. The algorithm learns the parameters by maximizing the likelihood of the observed data, typically using a gradient-based optimization method.

In practice, multinomial logistic regression involves fitting a model that assigns a set of coefficients to each class. During training, the model adjusts these coefficients to minimize the difference between the predicted probabilities and the actual class labels in the training data. When predicting, the model calculates the weighted sum of the input features for each class, applies the softmax function to convert these scores into probabilities, and assigns the input to the class with the highest probability. This approach is particularly useful for problems where the classes are mutually exclusive and collectively exhaustive, making it a robust method for multiclass classification tasks.

Multi Layer Perceptron

A multi-layer perceptron (MLP) is a type of artificial neural network used for classification, including multi-class problems such as three-class classification. An MLP consists of an input layer, one or more hidden layers, and an output layer. Each layer is composed of neurons, and each neuron in a layer is connected to every neuron in the previous and next layers. In a three-class classification task, the output layer would typically have three neurons, each representing one of the classes. The network learns to map inputs to the correct class by adjusting the weights and biases of these connections through backpropagation and gradient descent, minimizing the difference between the predicted and actual class labels.

During training, the MLP processes inputs through the network's layers, where each neuron applies an activation function (such as ReLU or sigmoid) to its weighted sum of inputs from the previous layer. For the final output layer in a three-class classification, the softmax activation function is used to convert the raw scores into probabilities that sum to one. The class with the highest probability is then chosen as the model's prediction. The network is trained using a loss function like cross-entropy, which quantifies the difference between the predicted probabilities and the actual labels. This iterative process of adjusting weights continues until the model's performance converges, resulting in a trained MLP capable of accurately classifying inputs into one of the three classes.

Chapter 5

Experimental Analysis and Results

5.1 Performance Metrics

When evaluating the performance of a ResNet-50 model for image classification, several performance metrics are typically used to assess its accuracy and effectiveness. Here are some of the key performance metrics used in the context of ResNet-50 classification.

Accuracy: Accuracy is a fundamental metric that measures the proportion of correctly classified images out of the total number of images in the dataset. It's a simple and widely used metric to gauge overall classification performance.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \tag{5.1}$$

Training accuracy and validation accuracy : These are crucial metrics in evaluating machine learning models. Training accuracy measures how well a model performs on the data it was trained on, indicating the ratio of correctly predicted instances to the total training instances. A high training accuracy suggests effective learning from the training data. On the other hand, validation accuracy assesses a model's ability to generalize to new, unseen data, calculated as the ratio of correctly predicted instances to the total instances in a separate validation dataset. Discrepancies between training and validation accuracy can signal overfitting or underfitting issues. Achieving a balance between high training accuracy and good validation accuracy is essential for building robust models that can make accurate predictions on diverse datasets. Monitoring these metrics during training is a key practice for ensuring model performance in real-world applications.

Precision: Precision quantifies the number of true positive predictions (correctly classified positive instances) divided by the total number of positive predictions. It helps assess the model's ability to avoid false positives. In a classification context, precision is crucial when false positives are costly or undesirable.

Recall (Sensitivity): Recall calculates the number of true positive predictions divided by the total number of actual positive instances in the dataset. It measures the model's ability to identify all relevant instances, which is especially important when false negatives are costly.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5.2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{5.3}$$

The F1 score is a measure of a test's accuracy in binary classification, calculated as the harmonic mean of precision and recall. It ranges from 0 to 1, with 1 indicating perfect precision and recall. The F1 score is particularly useful in scenarios with imbalanced class distributions, balancing the trade-off between false positives and false negatives. It provides a single metric that reflects the test's effectiveness in identifying positive instances accurately, making it invaluable for evaluating models where both precision and recall are important. For example, with 70 true positives, 10 false positives, and 20 false negatives, the F1 score would be approximately 0.821, indicating a good balance between precision and recall.

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5.4}$$

Training loss and validation loss are Another critical metrics in machine learning model evaluation. Training loss gauges the model's fitting performance on the training dataset, aiming to minimise the error between predictions and actual targets through iterative adjustments. Low training loss indicates effective learning. Validation loss assesses the model's generalization to new data, calculated on a separate dataset. The objective is to prevent overfitting, ensuring the model's performance extends beyond training data. Monitoring both losses aids in understanding learning progress and detecting overfitting. Balancing low training loss with effective generalisation is key for robust models. Model selection often involves finding parameters yielding low validation losses, indicating strong generalisation capability.

Confusion Matrix : The confusion matrix is a fundamental tool for evaluating the performance of a classification model, particularly in a multi-class setting. It is a table that summarizes the prediction results by comparing the actual and predicted classes. Each row of the matrix represents the instances of the actual class, while each column represents the instances of the predicted class. For a three-class classification problem, such as differentiating between Alzheimer's Disease (AD), Cognitive Normal (CN), and Mild Cognitive Impairment (MCI), the confusion matrix will be a 3x3 table.

The matrix entries include True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) for each class. Specifically, the diagonal elements represent the number of correctly classified instances for each class (e.g., AD correctly predicted as AD), while the off-diagonal elements indicate misclassifications (e.g., AD misclassified as CN or MCI). This detailed breakdown allows for a comprehensive assessment of model performance, highlighting areas where the model excels and where it needs improvement.

Using the confusion matrix, several key metrics can be derived, such as accuracy, precision, recall, and F1 score for each class. These metrics provide insights into the model's ability to correctly identify each class and handle imbalances in the dataset. For example, in the context of Alzheimer's disease classification, a high recall for the AD class is crucial for ensuring that most patients with Alzheimer's are correctly identified. Overall, the confusion matrix is an invaluable evaluation tool that provides a clear and detailed picture of the classification model's strengths and weaknesses.

5.2 Environmental Setup

The whole works is done using the Python framework. A large storage space and high graphics are required for the loading and training of NIFTI files. The works is done within Kaggle, an open-source notebook that has 16 GB of P100 Graphics and supports 100+ GB of input storage. The python libraries NiBabel,sklearn and basic Libraries are used for deep learning projects. Hyperparameters such as learning rate, Epoch number, and optimizers are altered with every try to achieve the best performance by utilising the most fitted parameters.Performance analysis using the various performance metrics are performed to identify the best model.

5.3 Results

5.3.1 3D Deep learning model

The 3D deep learning model Alz3Dnet undergone training with various parameters. It is shown that the results are promising when comparing with existing pretrained networks like ResNet18. 18, Inception v3, etc. The proposed model gives better results when Adam is selected as optimizer. The epochs are set to be 50 and the learning rate is 0.001. The trainable parameters are 17,938,835, which is less than that of existing networks.The Accuracy curve of the proposed model is shown in fig 5.1 and the loss curve for the same in fig 5.2

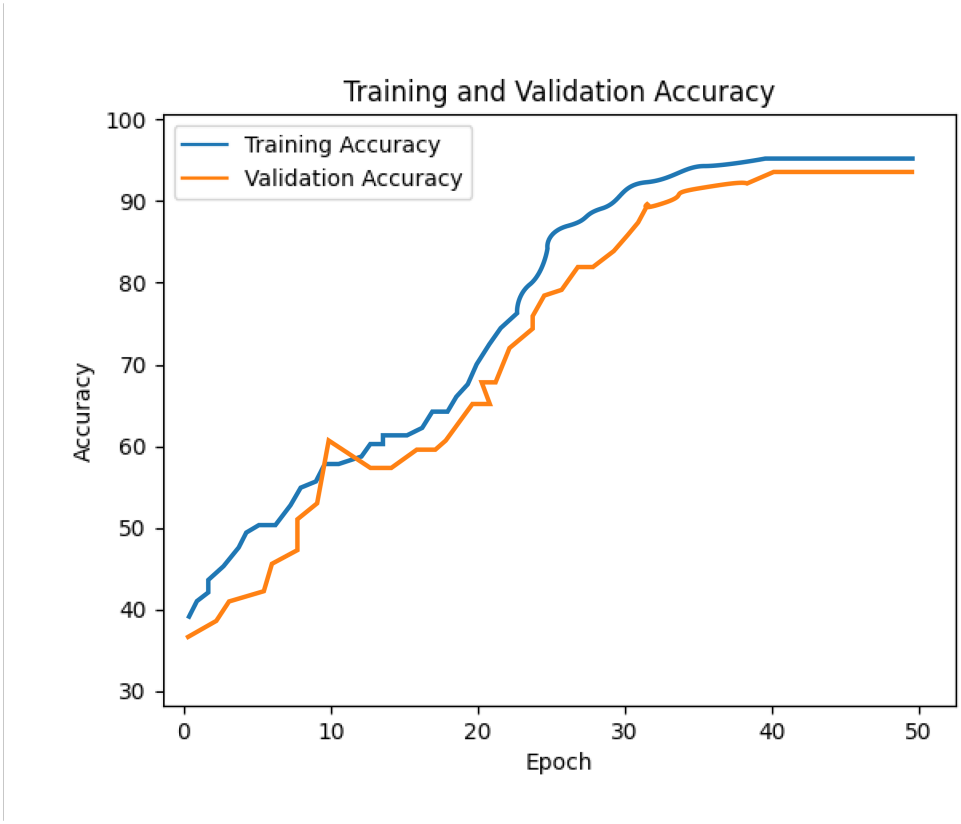

Figure 5.1: Accuracy graph of Alz3Dnet

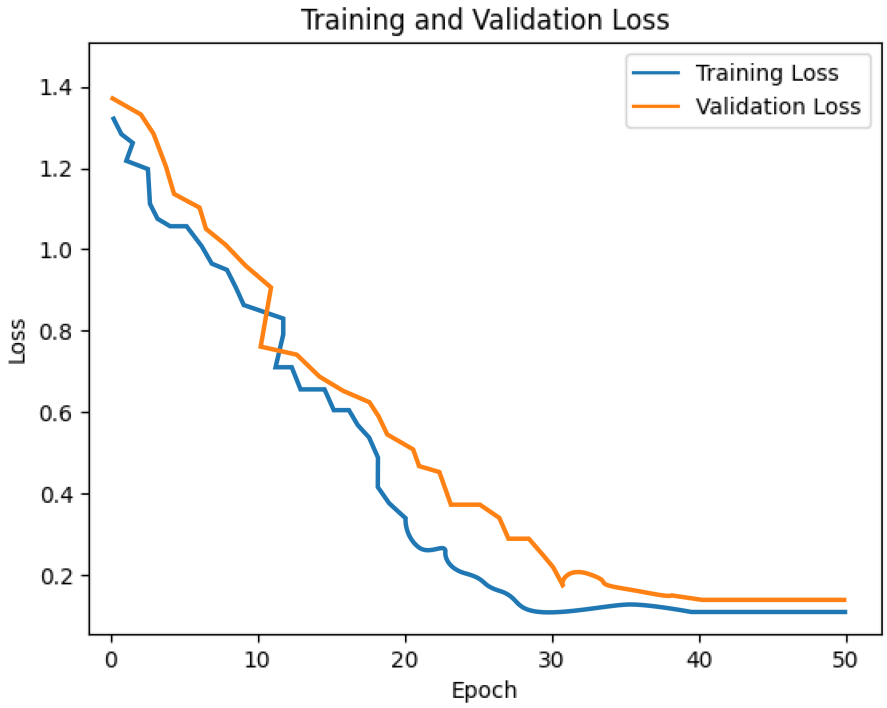

Figure 5.2: Loss graph of Alz3Dnet

From the accuracy and loss curves , The model is trying to become stable in terms of loss and accuracy after around 40 epochs. The model able to give the best performance without overfitting and any other difficulties.

By comparing the results with the existing pretrained networks the proposed model performs better in terms of evaluation metrics. A number of models are trained with the same dataset along with similar conditions and the models which gives better results are depicted in Table 5.1. From the table it is clear that, The proposed model provides better results in accuracy metrics. The learning rate and other parameters are tuned in various ranges and the results are better with a learning rate of 0.001 and Adam as optimizer. The good results may be its adaptive learning rates, momentum incorporation, and bias correction, allowing efficient and robust training across diverse tasks and neural network architectures [21].The incorporation of residual blocks and attention modules helped the proposed model to perform better while comparing with existing pre-trained networks.All the parameters for the proposed model is derived from the results given by random search algorithm, used for hyperparameter tuning.

The proposed model is also compared with the models that are available in existing works and the results are framed in Table 5.2. The proposed model expressed a slight improvement in results comparing with the existing works in terms of validation accuracy. The arrangement of attention modules and the incoporation of skip connections may helped to reduce the vanishing gradient issue and to improve the results.

From the Table it is clear that the proposed model shows improvement in results with

Table 5.1: Comparison of results with various models

Model	Parameters	Trainig Accuracy	Validation Accuracy
ResNet18	Optimizer =Adam Learning rate = 0.001 Epochs = 35	98.75%	83.00
Inception v3	Optimizer =Adam Learning rate = 0.001 Epochs = 35	85.33%	60.25%
Alz3Dnet	Optimizer =Adam Learning rate = 0.001 Epochs = 50	96.56	92.22

Table 5.2: Comparison of proposed model with existing works

Model	Accuracy
Z. Kong et al. [11]	87.67%
Y.Wu et al., [12]	91.3%
R. Han et al., [10]	91.83%
Alz3Dnet	**92.22%**

existing pre-trained networks.

5.3.2 ML model

The first task in this Machine learning part of this work is perform the feature selection from a wide range of features.The wide range of features can make the model too complex and prone to overfitting. Therefore, feature selection methods are used to reduce the number of features, retaining only the most relevant ones for the classification task. In this study, we used several algorithms for feature selection, including the Extra Trees classifier, Random Forest feature extraction, and Recursive Feature Elimination (RFE). These methods help identify and retain the most informative features, improving the efficiency and performance of the classification models.

The Extra Trees classifier is known for its ability to rank features based on their importance by constructing multiple decision trees and averaging the results. Random Forest feature extraction operates similarly, leveraging an ensemble of decision trees to determine the significance of each feature. Recursive Feature Elimination, on the other hand, iteratively removes the least important features and builds the model repeatedly to identify the optimal subset of features. By applying these techniques, we ensure that only the most pertinent radiomic features are used for the classification of NC, AD, and MCI, reducing the dimensionality of the dataset and enhancing the model's robustness.Among the three Random forest feature selection gives the better results and hence the same is used for feature selection. Figure 5.3 shows the selected 15 features that extracted from a single image and sorted based on importance done by the ranking of random forest feature elimination.

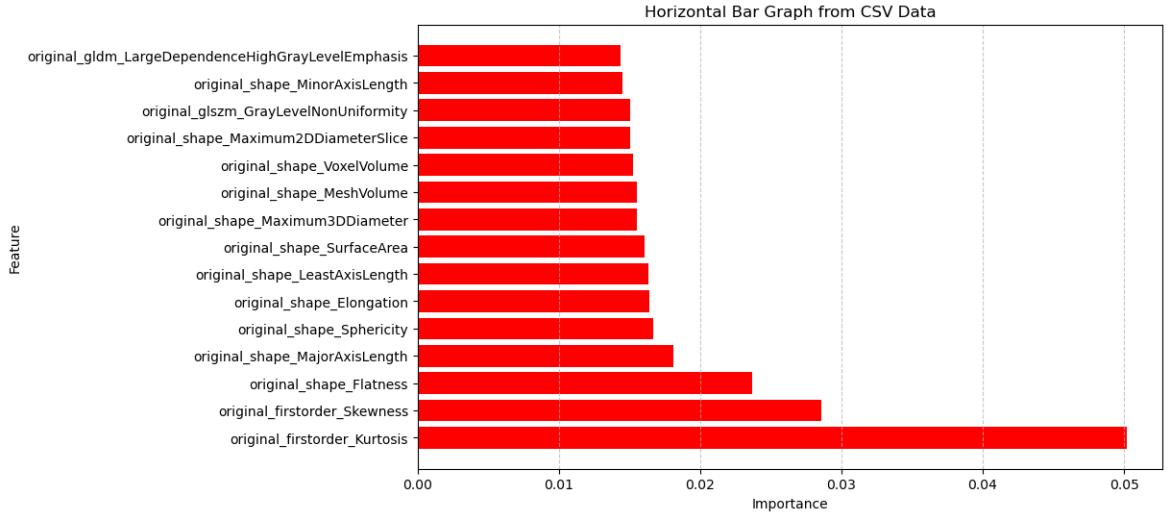

Figure 5.3: Features extracted from MRI image using Random forest feature selection

In the classification of data into categories such as Normal Control (NC), Alzheimer's Disease (AD), and Mild Cognitive Impairment (MCI), various machine learning classifiers have been employed. The results from the various feature selection techniques, combined with different classification models, are expressed in Figure 5.4,5.5,5.6,5.7. The figure illustrates the performance metrics for each combination of feature selection method and classifier. It is evident from the figure that using Random Forest feature extraction in conjunction with ensemble techniques yields superior results compared to other methods. The ensemble approach, which combines the strengths of multiple classifiers, enhances the overall accuracy, precision, recall, and F1 score, demonstrating a clear improvement in classification performance. This highlights the effectiveness of integrating feature selection methods with ensemble learning to achieve more reliable and precise outcomes in the diagnosis of Alzheimer's disease and related cognitive impairments.

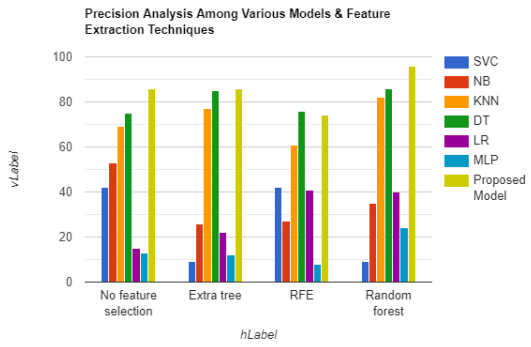

Figure 5.4: Precision Analysis Among Various Models & Feature Extraction Techniques

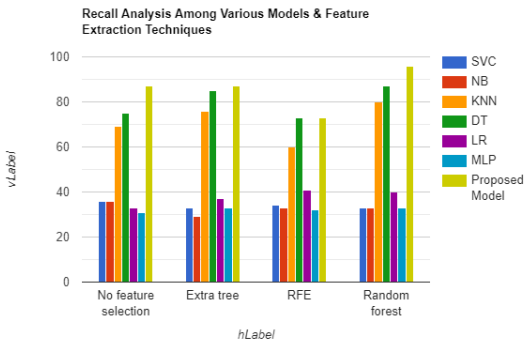

Figure 5.5: Recall Analysis Among Various Models & Feature Extraction Techniques

For the classification of Alzheimer's disease based on the extracted radiomic features after feature selection, traditional classifiers like Support Vector Classifier (SVC), Decision Trees,

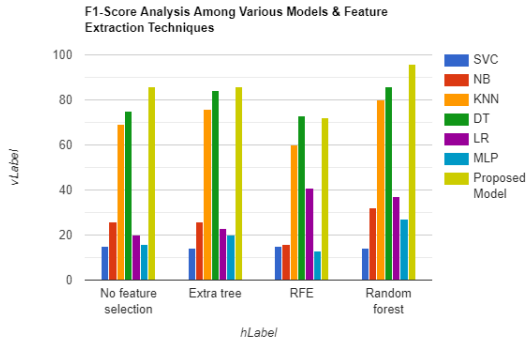

Figure 5.6: F1 Score Analysis Among Various Models & Feature Extraction Techniques

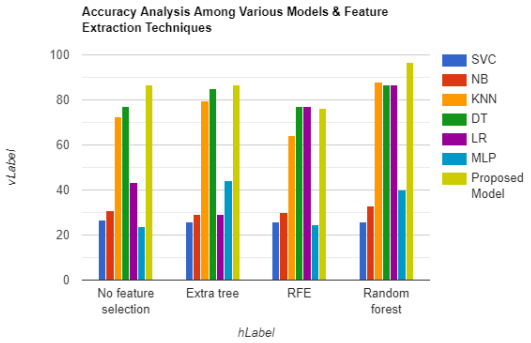

Figure 5.7: Accuracy Analysis Among Various Models & Feature Extraction Techniques

K-Nearest Neighbors (KNN), Naive Bayes classification, Multilayer Perceptron (MLP), and Multinomial Logistic Regression are frequently used due to their effectiveness in handling complex datasets and providing robust predictions. These classifiers are well-regarded for their ability to model intricate relationships within the data and yield reliable results in various medical imaging applications. The results from the various classifiers are expressed in Table 5.3.

SVC is known for its capability to find the optimal hyperplane for separating classes, even in high-dimensional spaces. Decision Trees offer interpretability and can handle both numerical and categorical data. KNN is simple and intuitive, relying on the majority vote of nearest neighbors to classify new instances. Naive Bayes is efficient and works well with large datasets by assuming feature independence. MLP, a type of neural network, is capable of capturing non-linear patterns in the data, and Multinomial Logistic Regression is adept at handling multi-class classification problems with a probabilistic approach.

Despite the strengths of these individual classifiers, the proposed ensemble model demonstrates a slight improvement in evaluation metrics compared to the individual classifiers. Ensemble models, by combining the predictions of multiple classifiers, mitigate the limitations and errors of individual models. This results in a more robust and accurate overall prediction. Techniques such as averaging or voting across the models help to reduce overfitting and improve generalization, especially important in the context of medical diagnoses where accuracy is critical.

The evaluation metrics formed after training the ensemble model with the extracted radiomic features are shown in Table 5.4. These metrics include measures such as accuracy, precision, recall, and F1 score, providing a comprehensive assessment of the model's performance. The slight improvement observed with the ensemble approach underscores the value of combining multiple models to leverage their individual strengths and compensate for their weaknesses, ultimately leading to more reliable and precise classification outcomes for Alzheimer's disease.

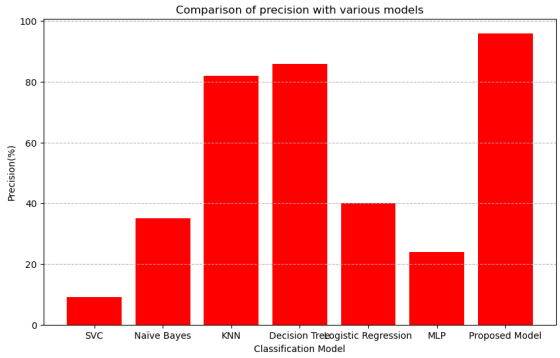

Figure 5.8: Comparison of precision with various models

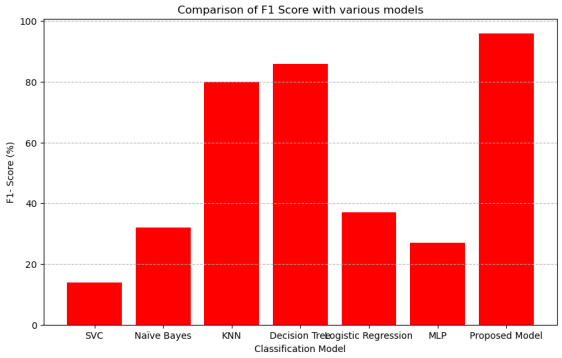

Figure 5.9: Comparison of f1 score with various models

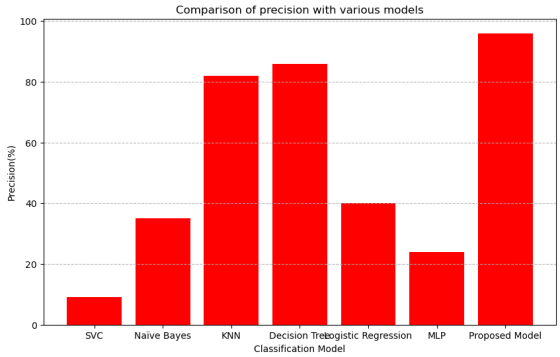

Figure 5.10: Comparison of precision with various models

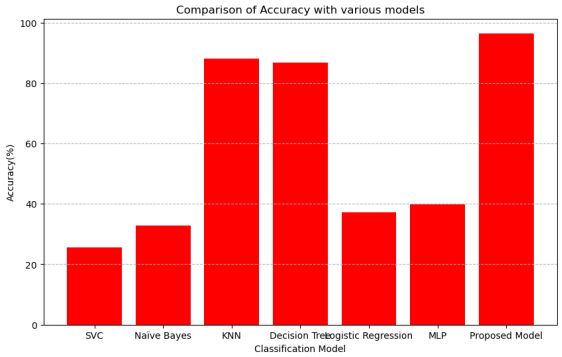

Figure 5.11: Comparison of accuracy with various models

Table 5.3: Detailed Comparison with various models

Model	**Accuracy**
SVC	25.67%
Naive Bayes	32.74%
KNN	88%
Decision Tree	86.72%
Logistic Regression	37.16%
MLP	39.82%
Proposed model	96.46%

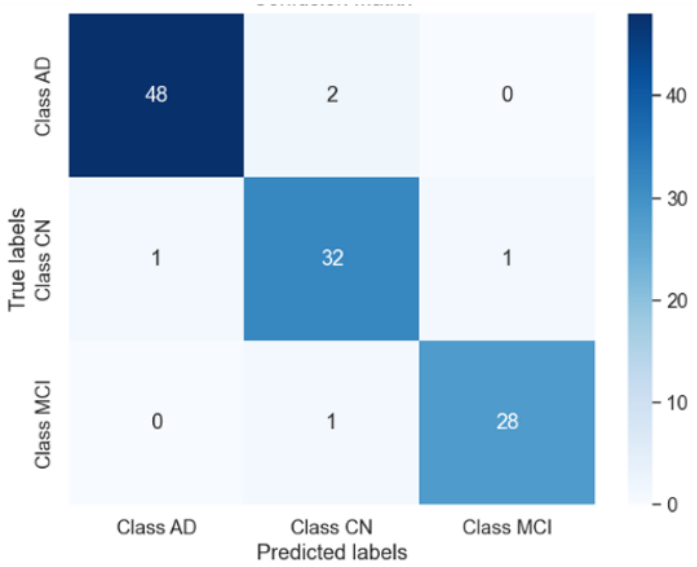

Figure 5.12: Confusion Matrix from proposed ML model

The proposed ensemble model outperforms traditional single models. Ensemble techniques improve classification results by combining multiple models to enhance accuracy and robustness. By averaging or voting across models, they reduce individual errors and overfitting. Methods like boosting train models sequentially, with each new model focusing on the errors made by the previous ones, progressively reducing bias and variance. Stacking combines multiple base models by training a meta-model to aggregate their predictions, leveraging their complementary strengths and compensating for their individual weaknesses.

This collective approach leverages the strengths of each model, leading to better performance than single models. Ensemble methods not only improve predictive accuracy but also increase model stability and generalization by reducing the risk of overfitting to the training data. By integrating diverse models, ensemble techniques can better capture complex patterns and relationships within the data, making them particularly effective for challenging classification tasks.

Ensemble methods operate on the principle that a combination of multiple models can outperform any individual model by aggregating their predictions. This aggregation can be achieved through various techniques such as bagging, boosting, and stacking. Bagging involves training multiple models on different subsets of the data and averaging their predictions, thereby reducing variance and enhancing stability. Boosting sequentially trains models, with each new model focusing on correcting the errors of the previous ones, which increases accuracy by reducing bias. Stacking combines the predictions of multiple models using a meta-model, which learns the best way to integrate these predictions for improved performance.

The advantages of the ensemble approach are multifaceted. Enhanced performance is achieved as the ensemble model leverages the strengths of each individual model, leading to superior predictive power. Robustness is another key advantage; by combining multiple models, the ensemble method mitigates the impact of outliers and noisy data, thus ensuring more reliable predictions. Furthermore, the generalizability of the model is improved as the

Table 5.4: Detailed results of proposed ML model

Accuracy	96.46%
Precision	96.8%
Recall	96.00%
F1 Score	96.00

ensemble technique reduces the likelihood of overfitting, allowing the model to perform well on unseen data.

The increased stability and robustness of ensemble models make them a powerful tool for achieving superior results in various machine learning applications. For instance, in medical diagnosis tasks such as Alzheimer's disease classification, the ensemble approach can significantly enhance diagnostic accuracy by effectively integrating different predictive models. This not only improves the reliability of the diagnostic tool but also provides clinicians with a more comprehensive analysis, thereby facilitating better decision-making.

Overall, the ensemble approach offers a more accurate, stable, and generalizable solution compared to individual models. By leveraging the combined strengths of multiple models, ensemble techniques provide a powerful framework for addressing complex machine learning problems, ensuring robust and high-performing models that are well-suited for practical applications.

From the results provided from the proposed ensemble model , It is clear that the proposed model can perform better with radiomic features. The comparitive analysis expresses the ensemble method can perform better than any single models.

Chapter 6

Conclusion

This study addresses the need for early and accurate diagnosis of Alzheimer's Disease (AD) by applying advanced machine learning techniques to MRI data analysis. Through the utilization of radiomic features and innovative deep learning architectures, significant strides have been made in enhancing classification accuracy and disease recognition.

The development and implementation of the Alz3D net represent a notable achievement in this domain. By incorporating attention modules and effectively handling the vanishing gradient problem via skip connections, the Alz3D net outperforms existing models, achieving an impressive validation accuracy of 92.22% in Alzheimer's dementia classification from MRI images. This underscores the effectiveness of this approach in providing more reliable diagnostic tools for clinicians and researchers.

Furthermore, the integration of the Convolutional Block Attention Module (CBAM) into the Alz3D net has led to further improvements in classification accuracy, demonstrating the importance of incorporating advanced architectural enhancements in deep learning models for enhanced performance.

Additionally, the exploration of ensemble methods alongside radiomic features has yielded promising results, with the ensemble model achieving an accuracy of 96%. This underscores the effectiveness of combining multiple methodologies to capitalize on their respective strengths and achieve superior performance in disease classification tasks.

While the Alz3D net serves as a powerful tool in Alzheimer's diagnosis, the findings also highlight the complementary nature of traditional radiomics-based classification methods. In fact, classification based solely on radiomic features has shown to surpass the accuracy of the Alz3D net, emphasizing the importance of integrating diverse approaches for comprehensive disease analysis. These results indicate that a hybrid approach, leveraging both advanced deep learning architectures and traditional radiomics, can provide the most reliable and accurate diagnostic outcomes for Alzheimer's disease. This study lays the groundwork for future research and development in the field, with the potential to significantly impact clinical practices and improve patient outcomes.

References

[1] World Health Organization. (2020) Dementia. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dementia

[2] Mayo Clinic. (2024) Alzheimer's disease: symptoms and causes. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/symptoms-causes/syc-20350447

[3] NHS. (2024) Alzheimer's disease - causes. [Online]. Available: https://www.nhs.uk/conditions/alzheimers-disease/causes

[4] R. Peters, "Ageing and the brain: This article is part of a series on ageing edited by professor chris bulpitt," *Postgraduate medical journal*, vol. 82, no. 964, pp. 84–88, 2006.

[5] Alzheimer's Disease International. (Year the webpage was last updated or accessed) Dementia statistics. [Online]. Available: https://www.alzint.org/about/dementia-facts-figures/dementia-statistics/

[6] J. Liu, M. Li, Y. Luo, S. Yang, W. Li, and Y. Bi, "Alzheimer's disease detection using depthwise separable convolutional neural networks," *Computer Methods and Programs in Biomedicine*, vol. 203, p. 106032, 2021.

[7] M. Tanveer, A. H. Rashid, M. Ganaie, M. Reza, I. Razzak, and K.-L. Hua, "Classification of alzheimer's disease using ensemble of deep neural networks trained through transfer learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 4, pp. 1453–1463, 2021.

[8] M. Ashtari-Majlan, A. Seifi, and M. M. Dehshibi, "A multi-stream convolutional neural network for classification of progressive mci in alzheimer's disease using structural mri images," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 8, pp. 3918–3926, 2022.

[9] F. J. Martinez-Murcia, A. Ortiz, J.-M. Gorriz, J. Ramirez, and D. Castillo-Barnes, "Studying the manifold structure of alzheimer's disease: a deep learning approach using convolutional autoencoders," *IEEE journal of biomedical and health informatics*, vol. 24, no. 1, pp. 17–26, 2019.

[10] R. Han, Z. Liu, and C. P. Chen, "Multi-scale 3d convolution feature-based broad learning system for alzheimer's disease diagnosis via mri images," *Applied Soft Computing*, vol. 120, p. 108660, 2022.

[11] Z. Kong, M. Zhang, W. Zhu, Y. Yi, T. Wang, and B. Zhang, "Multi-modal data alzheimer's disease detection based on 3d convolution," *Biomedical Signal Processing and Control*, vol. 75, p. 103565, 2022.

[12] Y. Wu, Y. Zhou, W. Zeng, Q. Qian, and M. Song, "An attention-based 3d cnn with multi-scale integration block for alzheimer's disease classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 11, pp. 5665–5673, 2022.

[13] T.-R. Li, Y. Wu, J.-J. Jiang, H. Lin, C.-L. Han, J.-H. Jiang, and Y. Han, "Radiomics analysis of magnetic resonance imaging facilitates the identification of preclinical alzheimer's disease: an exploratory study," *Frontiers in Cell and Developmental Biology*, vol. 8, p. 605734, 2020.

[14] J. Feng, Y. Huang, X. Zhang, Q. Yang, Y. Guo, Y. Xia, C. Peng, and C. Li, "Research and application progress of radiomics in neurodegenerative diseases," *Meta-Radiology*, p. 100068, 2024.

[15] A. D. N. Initiative, "Alzheimer's disease neuroimaging initiative (adni) dataset," 2004, accessed: 2024-05-23. [Online]. Available: http://adni.loni.usc.edu/

[16] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[17] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "Eca-net: Efficient channel attention for deep convolutional neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 534–11 542.

[18] G. Alfian, M. Syafrudin, I. Fahrurrozi, N. L. Fitriyani, F. T. D. Atmaji, T. Widodo, N. Bahiyah, F. Benes, and J. Rhee, "Predicting breast cancer from risk factors using svm and extra-trees-based feature selection method," *Computers*, vol. 11, no. 9, p. 136, 2022.

[19] X. Zeng, Y.-W. Chen, and C. Tao, "Feature selection using recursive feature elimination for handwritten digit recognition," in *2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2009, pp. 1205–1208.

[20] B. H. Menze, B. M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, and F. A. Hamprecht, "A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data," *BMC bioinformatics*, vol. 10, pp. 1–16, 2009.

[21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.