AGHC-NeuroTransformer: An Attention-Guided Hybrid CNN-Transformer Model for Robust Classification of Neurodegenerative Disorders from MRI Scans

by

Saj Saha
21201377
Farzia Akhter
21241019
Mitul Roy Tanny
21201457
Nabil Hossain Chowdhury
21241006

A thesis submitted to the Department of Computer Science and Engineering for the fulfillment of the requirements of the degree of Bachelors of Science in Computer Science.

Department of Computer Science and Engineering
Brac University
June 2025

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Farzia Akhter
21241019

Mitul Roy Tanny
21201457

Saj Saha
21201377

Nabil Hossain Chowdhury
21241006

Approval

The thesis/project titled "AGHC-NeuroTransformer: An Attention-Guided Hybrid CNN-Transformer Model for Robust Classification of Neurodegenerative Disorders from MRI Scans" submitted by

1. Saj Saha (21201377)
2. Farzia Akhter (21241019)
3. Mitul Roy Tanny (21201457)
4. Nabil Hossain Chowdhury (21241006)

Of Spring 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science & Engineering on June 20, 2025.

Examining Committee:

Supervisor:
(Member)

Dr. Md. Ashraful Alam
Associate Professor
Department of Computer Science and Engineering
BRAC University

Co-supervisor:
(Member)

Sahib Kowsar
Lecturer
Department of Computer Science and Engineering
BRAC University

Program Coordinator:
(Member)

Md Golam Rabiul Alam, PhD
Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

Sadia Hamid Kazi
Chairperson
Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

Neurodegenerative diseases like Alzheimer's and Parkinson's gradually damage the brain's structure and function, leading over time to serious and lasting problems with memory, thinking, behavior, and movement. These diseases often manifest subtly in their early stages, making timely diagnosis challenging yet critical for slowing disease progression and improving patient outcomes. In our paper, we proposed an attention-guided hybrid deep learning framework for the early and accurate classification of AD, PD, and Healthy controls using axial PD-T2 weighted Magnetic Resonance Imaging (MRI) slices. Our model leverages a Convolutional Neural Network (CNN) backbone, specifically EfficientNetB0, to capture low- and mid-level spatial features with high computational efficiency.To help the model better understand the bigger picture and relationships across multiple MRI slices, we added a Transformer-based encoder after the CNN feature extractor. A Convolutional Block Attention Module (CBAM) is introduced between the CNN and Transformer components to refine feature maps by adaptively weighting spatial and channel dimensions. Furthermore, the model incorporates an attention-based slice pooling mechanism, allowing it to automatically prioritize and aggregate the most informative MRI slices from each subject without requiring manual slice annotations. The proposed hybrid architecture is trained and evaluated on a curated dataset of 2D axial brain MRI slices, which have been preprocessed and organized at the subject level. Through extensive experiments, we demonstrate that our model achieves competitive classification accuracy with a significantly lower parameter count compared to larger transfer learning baselines, such as ResNet-50 and VGG-19. Our framework effectively combines local feature extraction, long-range dependency modeling, and adaptive attention mechanisms to address the complexities of MRI-based neurodegenerative disease diagnosis. This study states the potential of combining CNNs, Transformers, and attention modules for medical image analysis, offering a scalable, interpretable, and clinically relevant diagnostic tool for the early detection of AD and PD. Our experiments report a decent accuracy of 95% as opposed to a 93% accuracy reported by Yan's (2025) model on our dataset, while using a significantly lower learnable parameter count. As a result, these systems could be integrated into real-life neurological screenings, thereby making it easier to identify diseases early and support early intervention and preventive strategies.

Keywords: Neurodegenerative; Alzheimer's Disease; Parkinson's Disease; Convolutional Neural Network; Deep Learning; Magnetic Resonance Imaging; Non-Diagnostic Migraines;

Nomenclature

The following list describes several symbols and abbreviations that will be used throughout this document.

AD Alzheimer's Disease

ADNI Alzheimer's Disease Neuroimaging Initiative

AUC Area Under the Curve

BN Batch Normalization

CBAM Convolutional Block Attention Module

CNN Convolutional Neural Network

CSF Cerebrospinal Fluid

DL Deep Learning

EffNet EfficientNet

F1 F1 Score

FC Fully Connected

FN False Negative

FP False Positive

FPR False Positive Rate

GM Gray Matter

MCI Mild Cognitive Impairment

MMDNN Multimodal and Multiscale Deep Neural Networks

MRI Magnetic Resonance Imaging

ND Neurodegenerative Diseases

NN Neural Network

PD Proton Density

PD Parkinson's Disease

ReLU Rectified Linear Unit

ResNet Residual Network

ROC Receiver Operating Characteristic

ROI Region of Interest

SCCAN Spatial Convolutional Channel Attention Network

SGD Stochastic Gradient Descent

TP True Positive

TPR True Positive Rate

TN True Negative

ViT Vision Transformer

ViT-CBAM Vision Transformer integrated with Convolutional Block Attention Module

VGG Vision Geometry Group

WM White Matter

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Introduction to Neurodegenerative Disease:
Neurodegenerative diseases are a group of conditions that gradually damage and break down nerve cells in the brain and spinal cord, affecting how the body and mind function over time. These diseases are often driven by a combination of factors, including genetic mutations, environmental exposures, aging, chronic alcohol use and accumulated DNA damage within brain cells. Among the most prevalent neurodegenerative diseases are Alzheimer's Disease (AD) and Parkinson's Disease (PD), both of which affect millions of individuals globally. According to the World Health Organization, neurodegenerative diseases have afflicted more than 50 million people worldwide [1]. Although currently available treatments only provide symptomatic relief and improvement in quality of life, such a definite cure does not exist.

The main challenge in controlling neurodegenerative diseases is because the symptom indicators emerge too late to prevent neuronal damage from happening. Early detection along with precise diagnosis remains essential to start appropriate treatments which will decelerate disease advancement while enhancing results for affected patients. Standard diagnostic tools which include cognitive evaluations together with genetic analysis and blood biomarker tests identify disease conditions only after extensive neural damage has occurred.

MRI used to be the other healthier super tool for early detection of neurodegenerative diseases: a kind of high-resolution view of the brain's structure. MRI scans would ideally reveal certain patterns of atrophy and other abnormalities correlated with earlier stages of AD and PD. Nonetheless, manual analysis of MRI scans has been time-consuming and prone to inter-observer variability, thereby constraining scalability in the clinical arena.

The recent advancements in Artificial Intelligence (AI) and Machine Learning (ML), particularly in deep learning, have opened new avenues toward the automation and the increased accuracy of the medical image analysis. Deep learning models exhibit almost magical instances of extracting hierarchical features starting right from the raw imaging data through to the application versus the older methodology of feature engineering. These models have the ability to spot subtle and complex patterns associated with disease progression that remain hidden to conventional methods of analysis. Building upon this foundation, our work explores a novel hybrid deep learning framework that integrates CNNs with Transformer-based attention mechanisms to improve the early classification of Alzheimer's, Parkinson's, and Healthy subjects from 2D axial PD-T2 MRI slices. By incorporating attention modules and leveraging both local spatial features and global contextual relationships across MRI slices, our approach aims to address the limitations of existing models and contribute to more scalable, interpretable, and effective tools for early-stage neurodegenerative disease diagnosis.

1.1 Motivation

Neurodegenerative diseases have become a growing concern due to their irreversible impact on cognitive and motor functions. With a large number of aging population, these diseases are becoming more prevalent. Traditional diagnostic methods, such as cognitive assessments and clinical evaluations, often detect these diseases at advanced stages when treatment options are limited to managing the symptoms only, rather then a cure. So it is crucial to find ways to detect much earlier, which has a higher potential of cure.

Magnetic Resonance Imaging (MRI) is a powerful way to spot structural changes in the brain. But going through MRI scans by hand takes a lot of time and can lead to mistakes. That's where deep learning comes in it's been very effective in medical imaging by automating the process and boosting accuracy. Using models like Convolutional Neural Networks (CNNs) can help catch neurodegenerative diseases earlier, which could lead to better care and outcomes for patients.

Even with quite significant advancements in AI based diagnostics, challenges remain, such as computational efficiency or dealing with imbalanced data.Our study aims to develop an efficient deep learning framework that can reliably classify MRI scans into three groups: Alzheimer's Disease, Parkinson's Disease, and healthy individuals. By integrating deep learning techniques and optimizing data pre-processing, our research seeks to provide a scalable and reliable solution for early detection, while keeping requirements of computation low. We hope that the findings of this research have that potential to aid in the medical professionals in making more accurate and efficient diagnosis, creating the way for future developments in AI-assisted healthcare.

1.2 Research Problem

The methods of anatomizing various diseases such as Alzheimer's Disease (AD) or Parkinson's Disease (PD) are progressive disorders that lead to the gradual loss of neuron function, significantly impairing cognitive function and damaging motor skills. The damage caused by these disorders are irreversible, resulting in permanent impairments. As discussed, the currently available diagnostic methods, which largely rely on clinical symptoms or conventional imaging methods, often fail to detect these diseases until significant neuronal damage has already occurred. The delay in diagnosis poses a major challenge for effective treatment and management of these diseases. In order to manage these disorders, early identification is essential since it enables prompt action and may even reduce the disease's course.

Magnetic Resonance Imaging offers valuable insights into structural changes in the brain, but the complexity of MRI data often makes it difficult to interpret subtle early-stage abnormalities using traditional analysis methods. Despite advances in medical imaging, the early detection of neurodegenerative diseases remains inadequate. So, there is a need to develop efficient techniques to analyze MRI data to predict these diseases at an earlier stage, when therapeutic interventions could be more effective.

1.3 Research Objective

The main aim of the proposed research is the creation and development of an improved deep learning-based model that can process MRI scans and classify them as containing a neurodegenerative disorder (Alzheimer-Disease and Parkinson-Disease) or a healthy person. The purpose of this system is to contribute to the early detection and diagnosis of these conditions, which is crucial to intervene on time and to have better results and outcomes on the patient. In order to solve such a complicated problem, we are going to concentrate on the following essential objectives:

- First and most obvious is the creation and implementation of an efficient and strong deep learning model combining Convolutional Neural Networks (CNNs), attention model like CBAM, and Vision Transformers (ViTs). The hybrid design is meant to combine the utility of both local feature extraction and global context modeling to provide high-performance classification of neurodegenerative disorders using MRI images. Architectural innovations, proper integration of the model entities, and strict optimization are emphasized in order to guarantee not only precision but also reliability in clinical uses.
- To ensure that the proposed system is viable in practical health care settings, we enhanced the predictive performance as well as the computational performance of the model. Among them are reducing false positives and false negatives, training and inference speeds, and model size that should be lightweight to execute on the typical medical computing hardware. The long-term goal is to provide more reliable and quicker diagnosis results than the manual or semi-automated methods.
- A vital element of our goal comprises the capability of detecting and processing essential structural and functional abnormalities in brain MRI images that signify the presence of neurodegenerative diseases. The model enables the reliable classification of the data since the meaningful features and biomarkers extracted help in understanding the disease progression better. This understanding can assist clinicians to identify early symptoms and make suitable interventions at earlier periods.

Through the realization of these goals, our contributions hope to make much impact in the domain of medical imaging and artificial intelligence, leading to smarter, more accessible and scalable neurodegenerative disease diagnosis tools.

Chapter 2

Literature Review

2.1 Background

Recent Developments in deep learning (DL) have revolutionized the early diagnosis of neurodegenerative disorders, especially when using imaging data such as MRI scans. A growing number of studies are showing that Deep Learning and Machine Learning hold real promise for diagnosing conditions like Parkinson's and Alzheimer's, which often appear as subtle and overlapping changes in brain structure. Important methods and techniques that have been proposed for achieving this objective are presented in the literature that follows.

A study by Rohini and Surendran integrates Gaussian Naive Bayes classifiers, K Nearest Neighbors, and Support Vector Machine (SVM) to present an ensemble machine learning model for early identification of AD [2]. The Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset's behavioral and demographic data are combined with MRI-derived variables, such as changes in brain size and the thickness of the brain's outer layer, to create the model. By reducing over-fitting problems that single classifiers frequently encounter, the ensemble technique improves predicted accuracy across a range of sickness stages, including Normal, Mild Cognitive Impairment (MCI), and Pre MCI stages. The ensemble model combines the Gaussian Naive Bayes, KNN and SVM enhances the prediction accuracy across different AD stages by mitigating over-fitting. However, reliance on handcrafted features and demographic data may limit its generalizability compared to deep learning models.

2.2 Neurodegenerative Disease Detection from MRI using CNN

Convolutional Neural Networks are a type of deep learning model specifically built to handle data organized in a grid-like structure, such as images. These models are effective and widely used in training that involves image or visual data. CNNs are particularly useful in learning the spatial hierarchies of features from images, making them the ideal model for training with MRI scans.

Neurodegenerative disease diagnosis using MRI images has demonstrated great effectiveness when using deep learning, and Convolutional Neural Networks (CNNs) in particular. According to a study, a deep learning pipeline that employed a shallow CNN model on 2D T1-weighted MRI scans was 99.68% accurate in diagnosing AD[3]. Transfer learning is used to increase precision in classification and adaptability for both localized and global disease development stages. While the accuracy rate of detecting AD using a shallow CNN on 2D T1-weighted MRI scans is 99.68% it may indicate potential overfitting or dataset bias. Moreover as transfer

learning refines precision and adaptability, its performance depends on the quality and diversity of pre-trained models and datasets so if pre-trained models and datasets are not well qualified the results may vary.

Additionally, another study highlights the nature of computational developments in detecting various diagnostic patterns in the MRI images. Using deep learning algorithms for neurodegenerative data is reviewed here. CNNs have been deemed encouraging results in the detection of neurodegenerative disorders with high diagnostic accuracy, despite some drawbacks such as the need for more reliable datasets and enhanced model specification[3]. While CNNs show remarkable accuracy in diagnosing neurodegenerative diseases from MRI scans , their performance heavily relies on high-quality datasets and well-tuned models. The use of transfer learning techniques enhances adaptability, but the need for more reliable and comprehensive datasets remains a key challenge in improving diagnostic accuracy.

2.3 CNN-Based Feature Extraction for Brain MRI

The authors Samar Fouad & Duaa AlSaeed[4] presumed a method for detecting Alzheimer's disease from MRI data without the need for human feature engineering by integrating deep learning with convolutional neural networks (CNN) for feature extraction. The CNN model ResNet-50, which extracts features from MRI images, and three classifiers , Softmax function, SVM, and Random Forest were being used by the authors. Their dataset consisted of 708 MRI scans from MIRIAD 741 MRI images from ADNI. These images were resized and are being enhanced before being classified. The Softmax function classifier has performed better than the others, achieving a 99% accuracy on the ADNI dataset and a 96% accuracy on the MIRIAD dataset. After reaching 92% for ADNI and 90% for the MIRIAD, SVM achieved an accuracy of 85.7% for ADNI and 84.8% for MIRIAD in the Random Forest Classifier. The dependence on a restricted dataset (ADNI & MIRIAD) highlights uncertainties about applicability though the study shows high accuracy in Alzheimer's detection using ResNet-50 and different kinds of classifiers. In addition to that, the substantial performance gap between classifiers implies a need for further evaluation of model robustness and real-world extensibility.

2.4 3D Attention-Based Models for Slice Selection

Alzheimer's disease is a progressive neurodegenerative disorder that predominantly affects individuals over 65 years of age and is linked to cognitive decline and structural brain changes observable up to 20 years before the clinical symptoms [5]. Mild cognitive impairment (MCI) , a transitional phase between normal aging and AD, presents a critical window for the early intervention due to its high conversion rate of AD [6][7].

Magnetic resonance imaging is widely used in AD diagnosis due to its high spatial resolution and deep learning models, especially CNNs have become a prominent tools for automated feature extraction from brain images. [8] Prior methods, including 3D CNNs and multi-task models , have achieved promising results but often rely on region of interest (ROI) preselection and fail to integrate complementary features from different slice views.[9][10][11][12]

To address this , the paper proposes an Attention-Based Multi-view Slice Fusion model that fuses global 3D features with the 2D slice features from three anatomical views using attention

mechanisms to weigh slice importance and contexts.[13] This approach outperforms previous methods by 1.6% - 7.1% achieving 94.3% classification accuracy, indicating that effective AD diagnosis depends not only on datasets scale but also on integrating multi-view features with a neural attention [14]

2.5 Attention-Based CNN Architectures for Early Detection of Alzheimer's Disease from MRI Scans

Attention-based Convolutional Neural Networks (CNNs) have greatly improved medical image analysis, especially for the early detection of Alzheimer's Disease using MRI scans. While traditional CNNs are good at extracting features, they often have difficulty highlighting the most important areas in brain images. Attention mechanisms help to overcome this limitation by selectively highlighting regions that show structural changes associated with AD. These mechanisms integrated into CNN frameworks, improve both interpretability and classification accuracy making them ideal for clinical applications where early and accurate diagnosis is crucial.[15][16]

One notable architecture is the CAPBAM model which integrates Capsule Networks with the convolution block attention module. Capsule networks are adept at preserving spatial hierarchies, while CBAM selectively enhances discriminative features through spatial and channel wise attention which achieved an impressive 99.95 % accuracy on ADNI dataset along with 99.8% precision and recall, and a 0.99 AUC. This fusion allows the model to focus on pathological brain regions indicative of AD, significantly improving classification robustness and performance.[16] Similarly, the Stacked Convolutional Channel Attention Network (SCCAN) offers a lightweight yet powerful approach by layering multiple CNN modules with embedded channel attention, resulting in high classification accuracy with reduced computational load reporting and 99.58%,92.22% and 99.70% accuracy on the ADNI1 dataset.[17]

The Dual Attention Network with Multi-Layer Perceptron (DANMLP) approach suggests a Multimodal fusion method, combining MRI data, clinical measurements, and genetic information such as the APOE genotype [18]. The dual attention mechanisms within the model are employed to attend to salient features from all the available information to enhance diagnostic performance as well as decision-making process explainability [15]. This model achieves a 93% accuracy for AD vs MCI and 82.4% for MCI vs NC in five fold cross-validation on ADNI. Furthermore, Adaptive Hybrid Attention Network (AHANet) combines spatial and channel attention mechanisms such that the model can learn to pay attention to implicit patterns across the brain in an effort to promote early-stage detection and enhance model sensitivity which reached 98.53% accuracy on the ADNI dataset[19][20]. Collectively, these attention augmented CNNs deliver markedly enhanced performance and interpretability over traditional architectures.

These attention-based architectures constitute a central development of neuro-imaging-based diagnostics. In their ability to zero in on salient information and filter out distracting noise, they enable more accurate and earlier AD detection, even possibly before symptoms reach clinical status. With future advancement, such models are on the cusp of being in the vanguard of developing scalable, interpretable, and automated diagnostic systems for Alzheimer's and other neurodegenerative diseases [16][15][19][17].

Model	Key Mechanism	Benefit
CAPBAM	Capsule + CBAM	Spatial hierarchy + fine-grained attention
SCCAN	Stacked CNN + Channel Attention	Lightweight, high accuracy
DANMLP	Dual Attention + MLP	Multimodal fusion (MRI + genetics)
AHANet	Adaptive Hybrid Attention	Implicit brain region modeling

Table 2.1: Comparison of Attention-Based CNN Models for Alzheimer's Detection

2.6 Vision Transformers and Hybrid Models in Alzheimer's Disease Classification

In the paper "TransMed: Transformers Advance Multimodal Medical Image Classification," Dai et al. develop a model that makes use of both CNNs and Transformers to categorize parotid gland tumors from T1 and T2 MRI [21] [22]. CNNs focus on extracting simple characteristics, whereas Transformers pay attention to the interactions between the different modalities. Current fusion methods, like combining data at the input, feature, or decision stages; either limit deep interaction between different types of data, require heavy computation, or struggle to effectively capture combined features. They also do not work as well with small medical data. By using the advantages of CNNs and Transformers, TransMed achieves better results and greater efficiency compared to earlier methods[23].

According to the paper, "Introducing Vision Transformer for Alzheimer's Disease classification task with 3D input", it discusses using two approaches to divide Alzheimer's Disease (AD) patients from controls with 3D MRI scans: the Convolutional Voxel Vision Transformer (CVVT) and the ConvNet3D-4 CNN model. The CVVT uses a 3D Version of a Vision Transformer with convolutional patch embedding to notice non-local relationships between voxels. Still, results are affected by insufficient data and a shortage of inductive bias. Alternatively, when there is less data, the ConvNet3D-4, a lightweight CNN with instance normalization and Leaky ReLU, gets state-of-the-art results (98%) on the ADNI dataset. Currently, approaches that rely on radiomics with manually developed features have difficulty adapting to many cases and performing well across all situations, and CNNs do not work well in many scenarios due to their undersized receptive fields and poor explanation capability. Even though ViT models do well at capturing all image features, their need for much data and lack of CNN prejudices make them less suitable for tasks that use little medical data. Results from the study suggest that basic but well-optimized CNN architectures work better than complex designs when data is limited[24].

Model	Approach	Key Result
TransMed	CNN + Transformer	Better modality fusion (MRI T1 + T2)
CVVT	3D ViT	Non-local voxel relation modeling
ConvNet3D-4	Lightweight CNN	98% Accuracy on ADNI

Table 2.2: ViT and Hybrid CNN-ViT Architectures for AD Classification

2.7 Multimodal, Multiscale, and Attention-Based Learning Approaches

The author of this work, use a special deep neural network called Multimodal and Multiscale Deep Neural Networks for the Early diagnosis of Alzheimer's disease (MMDNN) to help identify Alzheimer's Disease at early stages using information gathered from structural MRI and FDG-PET scans[25] [26]. It draws out patch-based features from different parts of the brain at various scales and includes structural and metabolic details, then trains deep neural networks to

identify individuals suffering from any AD stage. The approach applied ensured high accuracy, as it detected MCI conversion in 82.9% of patients and future AD in 86.4%, with sensitive and specific outcomes. Moreover, existing strategies such as SVMs and CNNs designed for single image types are facing tough challenges due to the shortage of features to combine, the lack of using multiple modalities, or poor performance at early disease stages [27] [28]. More importantly, previous deep learning methods usually have fixed-scale feature extraction, which may result in ignoring fine spatial details. It stands out by using both broad views of the data and close-up details, incorporating both anatomic and physiological data, leading to a more exact approach for classifying AD images. This makes it clear that a significant improvement is visible over previous simpler models.

The paper, "Attention-based Deep Multiple Instance Learning" presents a deep learning framework for Multiple Instance Learning(MIL) that expands its flexibility and makes it easier to understand due to its usage of attention-based pooling [29] [30]. Rather than sticking to standard aggregation functions, the model uses a trainable attention layer to assign weights to the objects in each bag. This design allows the model to provide clear and effective predictions by only focusing on important points. Knowing which parts of the data impact decisions is highly important in areas like medical imaging, so this explanation style is very helpful there. This approach stands out from traditional MIL approaches due to its flexibility and the transparent highlighting of key factor instances for each sample [31].

2.8 Application of ResNet-18 in Alzheimer's Disease Detection

ResNet-18 has shown promising results in Alzheimer's disease (AD) detection, particularly when combined with attention mechanisms and transfer learning. Authour Mahmood integrated Squeeze and Ecitation blocks with a pre-trained ResNet-18 architecture and trained the model on 25357 MRI scans. The model achieved a testing accuracy of 88.51% with an ROC of 95 for both AD and cognitively normal classes and 93 for mild cognitive impairment [32]. Their approach demonstrated superior convergence speed and classification performance compared to models trained from scratch. Similarly, fine-tuned a ResNet-18 model using rested state fMRI data from the ADNI2 dataset to classify multiple stages of alzheimer's including early MCI, late MCI , significant and transfer learning in increasing model generalizability across complex classes[33] .

A research uses deep learning, specifically an improved 3D Residual Network, to classify neurodegenerative diseases using MRI images, a significant computer vision field. The authors said deep learning systems cannot grasp high-dimensional features connecting distant brain locations, making AD diagnosis difficult [34] . They extended ResNet-18's receptive field with dilated convolutional layers to extract additional information from MRI data for this challenge. Attention was added to the model to focus on the most important aspects for accurate categorization. Alzheimer's, MCI, and NC patients evaluated the proposed model.

The model classified AD vs. NC 92.12%, AD vs. MCI 74.07%, and NC vs. MCI 87.16% accurately. The improved ResNet model extracted discriminative properties for precise classification better than basic ResNet architectures and 3D CNNs. This study showed that the updated 3D Residual Network easily identifies complex MRI patterns to diagnose and classify neurodegenerative diseases. To further explore volumetric brain data, researchers have extended ResNet-18 into three dimensions. Researchers have developed a 3D ResNet-18 model by adapting 2D filters to handle volumetric MRI data, achieving remarkable performance metrics : 96.88%

accuracy , 100% sensitivity and 93.75% specificity using the ADNI dataset [35]. The 3D architecture proved effective in capturing spatial dependencies that 2D CNNs often miss. These studies validate ResNet18 utility and adaptability in both structural MRI and fMRI modalities for detecting early and advanced stages of neurodegeneration.

2.9 Multi-Class and Multi-Disease Detection Using ResNet-18 Variants

ResNet variants have also been explored for multi-class classification oF Alzeheimer's disease and other neurodegenerative conditions. Author has conducted a comparative study using ResNet 18, ResNet 50 and ResNet 101 to classify MRI scans into three categories AD,MCI and CN. ResNet101 yielded the highest accuracy demonstrating the scalability with limited computational resources [36]. Expanding beyond Alzheimer's disease, evaluated the ability of ResNet-18 in distinguishing between MRI scans for five types of diseases: normal, cerebrovascular, neoplastic, degenerative, and inflammatory [37]. The experiment used extensive data augmentation and showed that even simple CNNs like ResNet-18 were capable of achieving high accuracy on multi-disease classification tasks. The scope of applications from AD binary classification to more general neurological disease classification illustrates the generality of ResNet-18 as a robust backbone in medical image analysis.

2.10 Hybrid CNN-Transformer Architectures for Enhanced Alzheimer's Disease Classification

Recent advances in deep learning have generated much interest in coupling Convolutional Neural Networks (CNNs) and Transformer models to improve the classification of Alzheimer's Disease (AD) from MRI data[38]. Conv-Swinformer, a model that couples CNNs with Swin Transformers through a shift window attention mechanism. While the CNN layer extracts planar features from single slices of an MRI, the Transformer layer uses local attention to scan semantic relations in 3D space for greater fine-grained feature extraction. In the same vein, introduced the 3D Hybrid Compact Convolutional Transformer (3D HCCT), which extracts both spatial and long-range contextual features from 3D MRI scans. This model not only attains state-of-the-art performance on the ADNI dataset but also emphasizes interpretability a critical attribute in a clinical environment [39].

Based on temporal and spatial relationships, a new model is introduced named ViTranZheimer, which recycles video vision transformers to 3D MRI by processing volumetric data as videos rather than video sequences, allowing the model to see complex structural progression between slices through self-attention [40]. In yet another influential paper, proposed LGG-NeXt, a lightweight hybrid model that combines CNNs and Transformers employing the use of special Global and Local NeXt blocks, with added use of a pixel graph neural network and mask attention to allow for effective aggregation of global and local features from 2D structural MRI. Supporting these models, a further model is introduced which consists of multimodal, multiclass AD classification hybrid model that utilized CNNs effectively to extract localized features and Transformers to model global context in various phases of AD. In total, the research indicates the better efficiency and diversity of CNN-Transformer hybrids to enhance the accuracy and trustworthiness of Alzheimer's diagnosis using medical imaging [41] [16].

2.11 Vision Transformers in PET Imaging for Alzheimer's Disease Diagnosis

Current research has explored the application of Vision Transformers (ViTs) in the analysis of 18F-Florbetaben (FBB) PET scans for the classification of Alzheimer's Disease (AD). It has been compared with the performance of ViT models with VGG19, which is a convolutional neural network (CNN), for binary (normal vs. abnormal) and ternary (healthy control, mild cognitive impairment, and AD) classifications. The ViT model fared better than VGG19 in binary classification but showed worse performance when undertaking ternary classification. The study also reported the existence of overfitting and data imbalance resulting from data augmentation and reported that despite the potential of ViTs in some of the diagnostic tasks, they may not perform any better than CNNs under all classification conditions [25].

In general, researcher put forward DiaMond, a multi-modal approach that integrates MRI and PET information with vision transformers equipped with self-attention and bi-attention mechanisms. The approach achieved a highly balanced accuracy of 92.4% in the diagnosis of AD and was also stable for differential diagnosis tasks such as the differentiation of AD from front temporal dementia (FTD). DiaMond's success highlights the promise of integrating more than one imaging modality with high-level transformer architectures to improve diagnostic precision in neurodegenerative illnesses [42].

2.12 Hybrid CNN and Vision Transformer for Alzheimer's Diagnosis

Recent studies have explored the application of Vision Transformers(ViTs) in analyzing neurodegenerative diseases like Alzheimer's Disease , Parkinson's Disease through MRI based analysis. EfficientNet-B0 has become a favoured CNN architecture due to its balanced depth width and resolution scaling, offering strong performance on relatively small and imbalanced medical datasets [43][44]. To amplify its feature extraction capacity, models have been developed attention modules such as the Convolutional Block Attention Module (CBAM), which refines spatial and channel attention to highlight AD relevant regions. The AHANet model, for example fuses CBAM with DenseNet-169 and achieves an impressive 98.53% classification accuracy, illustrating how attention mechanisms enhance focus on disease specific features[45].

Similarly, Vision Transformers (ViTs) have emerged as transformative tools in medical imaging due to their self-attention mechanisms capable of modeling long-range spatial dependencies, something traditional CNNs struggle with. Their integration into hybrid CNN-transformer architectures facilitates a richer understanding of brain morphology. Researcher showed that combining CNNs with ViTs yields over 92 % accuracy in multi-stage Alzeheimer's classification from 3D MRI scans [46]. Similarly, researcher proposed a 3D Hybrid Compact Convolutional Transformer 9HCCT) that merges the local feature sensitivity of CNNs with the global context modeling of transformers [39]. However, despite their impressive accuracy, these models pose challenges in computational overhead and data requirements. For real-world clinical transition , future search must focus on simplifying architectures, improving interoperability and designing robust models that perform reliably with limited annotated data [47].

Model	Design Focus	Outcome
Conv-Swinformer	Shift Window Attention	Enhanced 3D MRI modeling
3D HCCT	CNN + Transformer Compact	State-of-the-art on ADNI
ViTranZheimer	ViT + 3D video-style MRI	Tracks progression in 3D
LGG-NeXt	Local + Global + GNN	Lightweight, interpretable
DiaMond	PET + MRI + Bi-attention	92.4% Balanced Accuracy

Table 2.3: Hybrid CNN-Transformer and ViT-PET Models for AD Diagnosis

2.13 Summary of the Reviewed Literature

In summary, the overall literature review extends about the early and precise diagnosis of neurodegenerative diseases like Alzheimer's and Parkinson's through the application of advanced deep learning models. It critiques traditional methods such as cognitive assessments and manual MRI analysis for their delayed detection capability and proposes a hybrid model combining Convolutional Neural Networks (CNNs), attention modules like CBAM, and Vision Transformers (ViTs) to enhance classification accuracy from 2D PD-T2 MRI scans. This hybrid approach aims to effectively capture both local structural features and global contextual information, ensuring high diagnostic performance, low computational overhead, and clinical applicability. Through a comprehensive literature review, the paper explores a spectrum of approaches from traditional ensemble models to modern attention-enhanced CNNs and hybrid CNN Transformer architectures. Notable models such as AHANet achieved up to 98.53% accuracy on the ADNI dataset, CAPBAM reported 99.95% accuracy, and SCCAN reached 99.58%, demonstrating substantial improvements in early Alzheimer's diagnosis. The review also highlights multimodal strategies like DiaMond, which combines MRI and PET data using bi-attention mechanisms to achieve a balanced accuracy of 92.4%. Overall, the proposed model is positioned as a scalable and interpretable solution aligned with the direction of emerging AI-powered diagnostics in healthcare.

Chapter 3

Methodology

3.1 Objective

The main goal of this work is to create a deep learning pipeline that can accurately and efficiently use MRI data to predict neurodegenerative disorders like Parkinson's disease (PD) and Alzheimer's disease (AD). The project aims to identify significant patterns in MRI scans for early diagnosis by utilizing sophisticated deep learning models, such as EfficientNetB0, CBAM, Vision Transformer , Attention based slice pooling and proposed a Hybrid CNN-Transformer with CBAM with the aim to improve clinical decision-making and patient outcomes.

3.2 Data Source and Description

In our research, we acquired access to data from ADNI and PPMI studies from Image and Data Archive (IDA) of neuroscience data. The samples are distinctly categorized into three classes: Alzheimer's Disease (AD), Parkinson's Disease (PD) and Healthy samples. The images are taken from the Axial Plane, where the slices are horizontal cuts parallel to the ground. The images have combined weighting of T2 and Proton Density (PD). Proton Density contrast weighting that reflects the number of hydrogen protons in tissues. For detection of Alzheimer's, these images can visualize general brain atrophy, particularly in temporal lobes and hippocampus. T2-weighting highlights fluid, such as CSF expansion, whereas PD contrast offers good gray/white matter distinction.

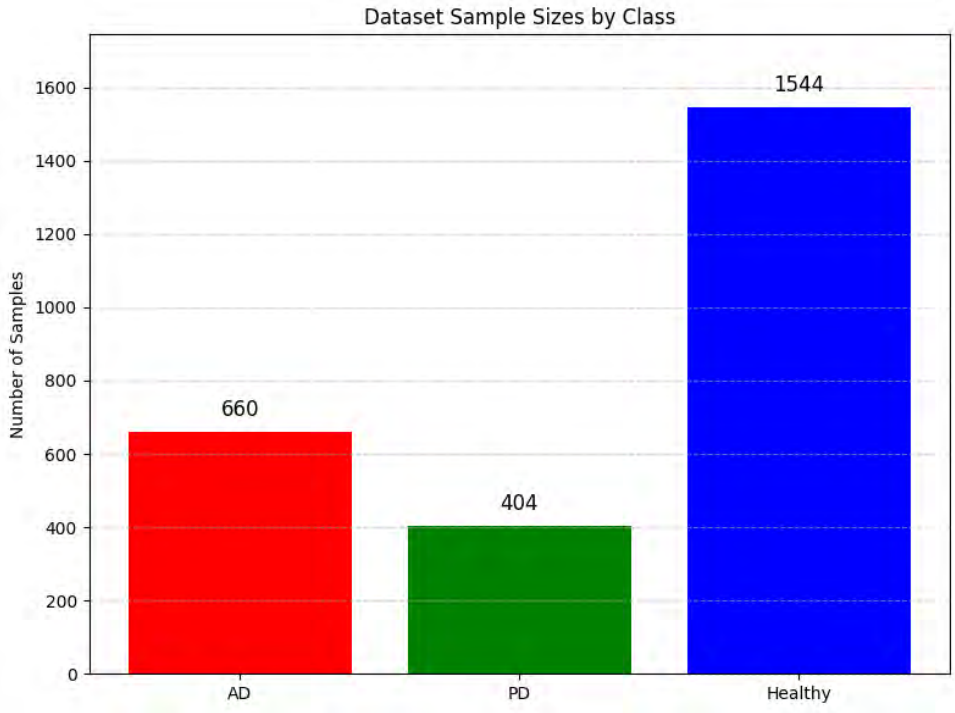

Figure 3.1: Dataset class distribution

The detection of Parkinson's is useful in these images because the images help detect signal changes in the substantia nigra and basal ganglia. Additionally, PD-T2 TSE may show hypointensity or structural changes due to neurodegeneration.
Each MRI sample of every class has 47 slices on average. Each slice has a thickness of 3.0 mm. The manufacturer of the MRI scanner is SIEMENS. Ensuring that the data collected is from the same machine, as well as the same acquisition parameters is crucial. Different scanners exhibit different artifacts, resulting in variations from each other. We have a total of 660 AD samples, 404 PD samples and 1544 Healthy samples, all collected from the same source, and under the same parameters. Figure 3.1 gives a visualize of the size of the dataset. The table 3.1 gives a summary of the dataset characteristics how it has been provided. Figures 3.2, 3.3 and 3.4 shows a 3D view of a sample from each class, and Figure 3.5 shows a particular slice from each class.

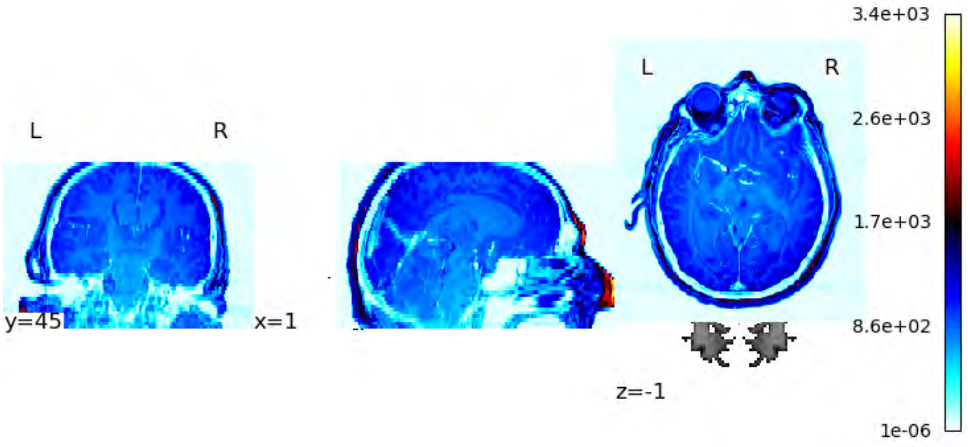

Figure 3.2: A 3D sample of an AD patient

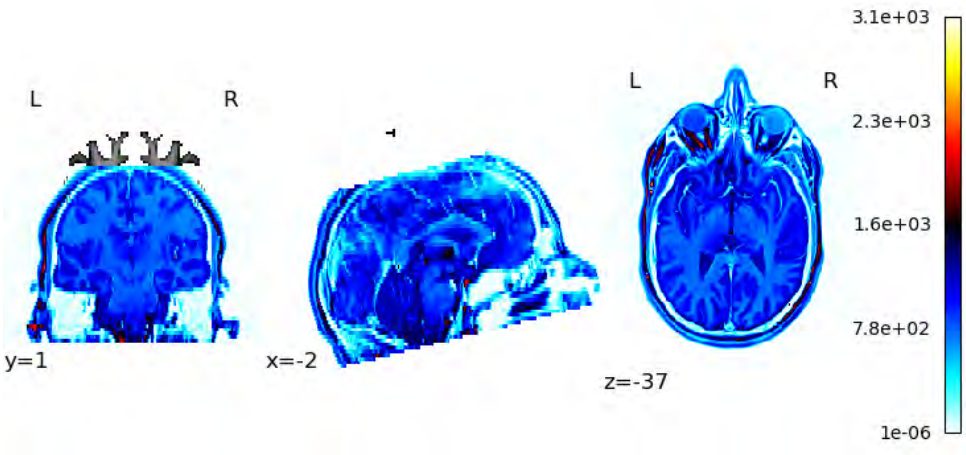

Figure 3.3: A 3D sample of a PD patient

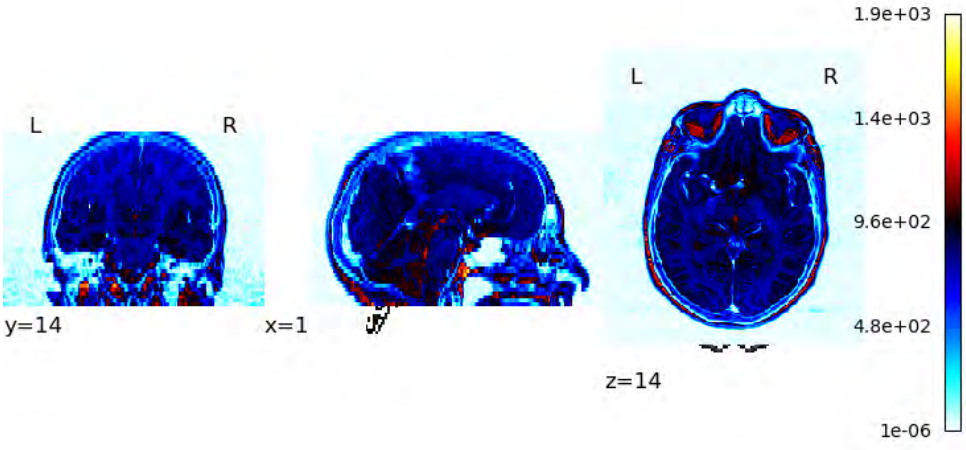

Figure 3.4: A 3D sample of a Healthy patient

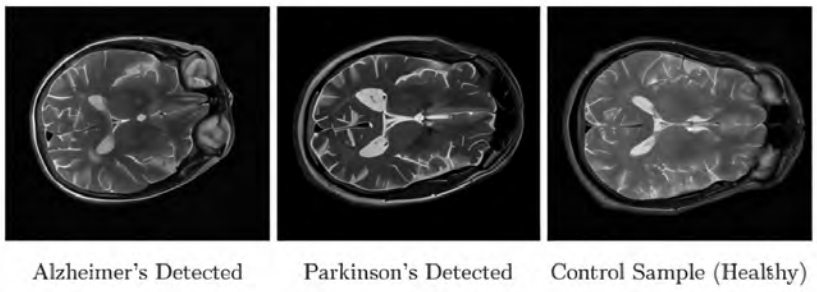

Figure 3.5: Sample 2D slices of AD, PD, and Healthy MRI scans

Characteristic	Value
Source	ADNI, PPMI
Number of AD Subjects	660
Number of PD Subjects	404
Number of Healthy Subjects	1544
MRI Sequence Type	PD T2-Weighted
Scanner Manufacturer	Siemens
MRI Plane	Axial
Gender Distribution	Female: 51%, Male: 49%

Table 3.1: Summary of Dataset Characteristics

3.2.1 Slice Selection

We chose to use the middle 20 slices per subject because these typically capture the most informative brain regions, such as the basal ganglia, hippocampus, and ventricles. The top and bottom slices often contain irrelevant information or non-brain tissue. Additionally, limiting to 20 slices ensures a consistent input size across subjects. Selecting more slices might lead to excessive redundancy, while selecting lesser might cause loss of important information.

3.3 Data Pre-processing

We applied High boost filtering to our dataset. This technique is an image enhancement method in the spatial domain, specifically used to accentuate high-frequency components, such as edges and fine details, while preserving a portion of the image's low-frequency background information.

The mathematical formula for this technique is:

$$H(x, y) = A \cdot f(x, y) - f_L(x, y) \tag{3.1}$$

- $f(x, y)$: original image
- $f_L(x, y)$: blurred image
- A: boost factor
- $H(x, y)$: high-boost filtered image

We used Gaussian Blur for our Blurred Image, and our boosting factor is 1.5. The boosting factor has no universally optimal value, we had to use trial and error to decide which factor is best for our classification.

In the context of medical imaging like MRI, fine structural details such as cortical thickness, sulcal patterns, or subtle lesions are paramount for the accurate diagnosis of neurodegenerative diseases. While high-pass filters would be effective at sharpening these details, they can also remove important low-frequency information that provides overall anatomical context.

High-boost filtering offers a balanced approach, enhancing diagnostically relevant features without sacrificing the broader anatomical context, thus providing a more complete and informative input for our analysis. Deep learning models are inherently designed to learn hierarchical features, commencing with low-level attributes such as edges and textures in their

initial layers. Figure 3.6 shows how the images are being filtered by high boost filtering. By Pre-processing with high-boost filtering, the model is presented with already enhanced edge information. This can potentially accelerate the learning of relevant low-level features or enable the network to allocate its capacity more effectively towards learning higher-level.

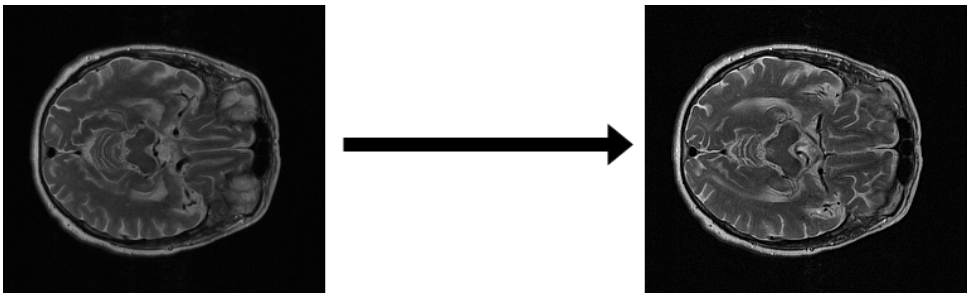

Figure 3.6: High Boost Filtering

To counter the imbalance and scarcity of samples in the dataset, we applied some data augmentation techniques. Medical datasets are frequently constrained in size, and deep learning models, particularly those with complex architectures, demand vast quantities of data to learn robust features and mitigate overfitting. We use augmentation to address this by artificially expanding the dataset, thereby increasing its diversity and rendering the model more resilient to variations encountered in real-world scans. Below are the augmentations we applied:

Augmentation	Description
Random Crop	The cropped area covers 80% to 100% of the original image, introducing variability in zoom and framing.
Random Affine	Applies random translation (up to 5% shift) and scaling (±5%).

We avoided any technique that could significantly alter the anatomical structure of the scan, such as flipping or high levels of rotation. An augmentation strategy that generates biologically implausible variations could lead the model to learn features that do not exist in real patients, resulting in poor generalization to clinical data or, more critically, misdiagnosis.

Besides these we applied Normalization on the images, as well as resized all slices to 224 x 224 pixels.

3.4 Data Splitting

The dataset is split into 80% for training and 20% for testing. We selected the middle 20 slices from each sample, where the significant features that display the diseases lie. The samples are split based on patient-level, rather than image-level.

3.5 Model Architecture

3.5.1 Overview

Convolutional Neural Networks (CNNs) excel in medical imaging because they can efficiently capture and organize local features in a hierarchical way. More recently, Vision Transformers

(ViTs) have emerged as a powerful alternative, utilizing self-attention mechanisms to capture global dependencies by treating images as sequences of fixed-size patches. ViTs have shown impressive performance across various visual tasks, including image classification. However, ViTs initially lack the inherent spatial localization inductive biases of convolutions and are notably data-intensive, necessitating extensive datasets for optimal performance.

This is where the proposed hybrid CNN-Transformer architecture comes in. Hybrid models combine CNNs for robust local feature extraction with ViTs for capturing long-range dependencies across the image. Such hybrid approaches aim to significantly improve performance for tasks that require integrating both localized features and global contextual relationships. Figure 3.7 shows the proposed architecture.

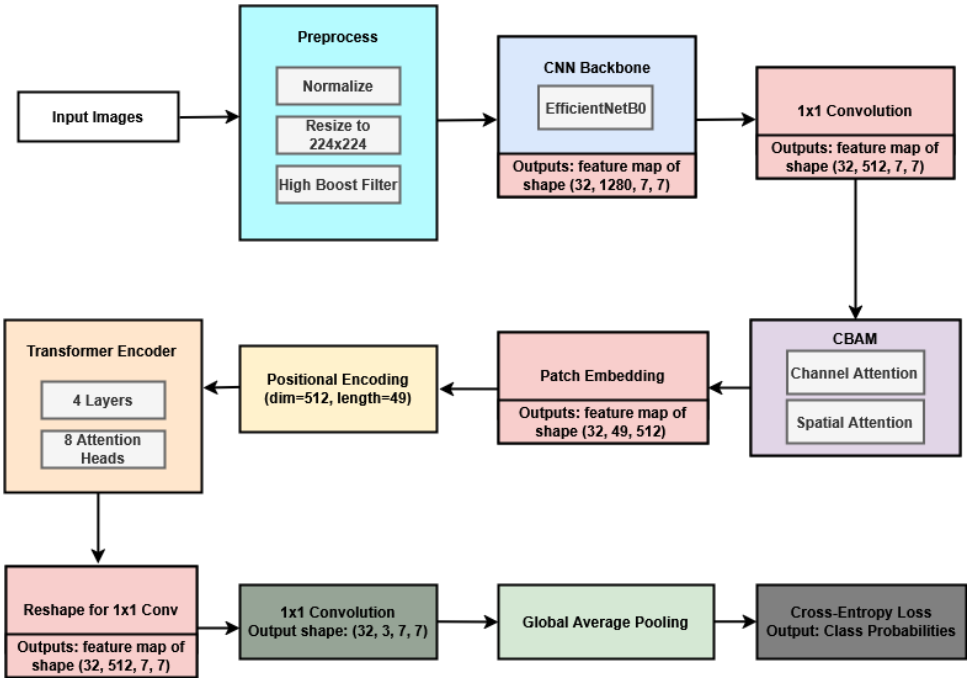

Figure 3.7: Proposed Architecture of AGHC-Neurotransformer

Attention mechanisms are designed to simulate the human visual system's capacity to selectively focus on salient parts of an image. In medical imaging, this allows the model to dynamically weigh the importance of different features. We applied the Convolutional Block Attention Module (CBAM), which is a simple yet highly effective attention module suitable for integration into feedforward CNNs.

3.5.2 EfficientNet-B0

EfficientNetB0 is a convolutional neural network architecture proposed by Tan and Le (2019)[48] which is a part of EfficientNet Family. It achieves decent performance with significantly fewer parameters and lower computational cost compared to conventional models like ResNet (which we also used for comparisons) or VGG. Traditional models scale depth, width, or resolution arbitrarily. EfficientNet introduces a compound coefficient to uniformly scale these three dimensions:

Depth (d): number of layers (how deep the network is),

Width (w): number of channels per layer,

Resolution (r): input image size.

EfficientNetB0 builds upon MBConv blocks from MobileNetV2. Each MBConv block consists of: Depthwise separable convolutions (computationally efficient) Squeeze-and-Excitation (SE) blocks for channel-wise attention Skip connections for better gradient flow.

Total Parameters: $\tilde{5}$.3 million FLOPs: $\tilde{3}$90M Dropout: 20%

EfficientNetB0 is particularly well-suited for medical image analysis for efficient feature extraction from high-resolution images (like 2D MRI slices) and low parameter count allows training on limited datasets without severe overfitting. Figure 3.8 shows the architecture that has been followed in this model.

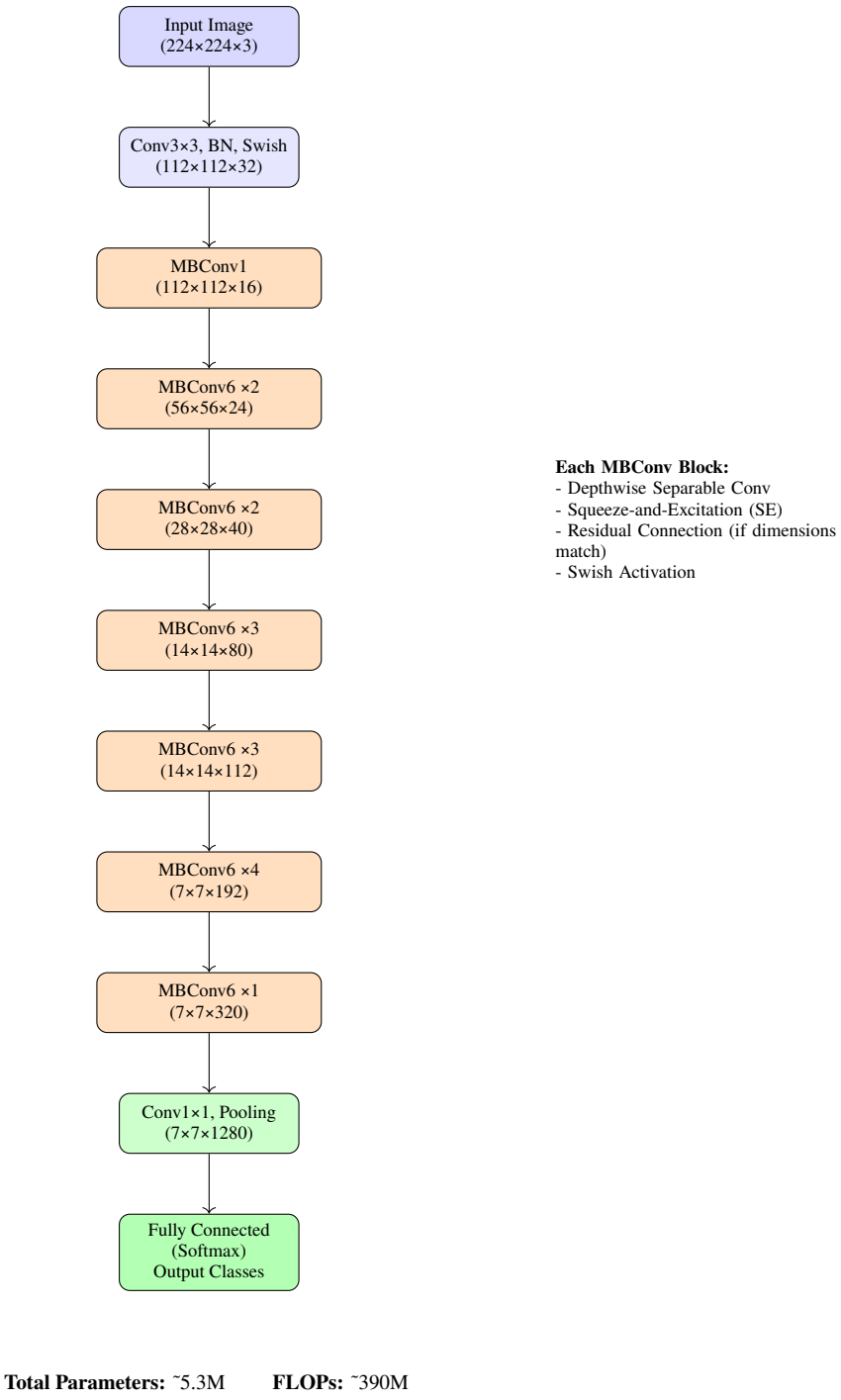

Figure 3.8: EfficientNet-B0 Architecture with MBConv Blocks, Squeeze-and-Excitation, and Dimension Annotations

3.5.3 Convolutional Block Attention Module (CBAM)

CBAM is an attention mechanism that sequentially infers attention maps along the channel and spatial dimensions. This consist of two components:

- Channel Attention
- Spatial Attention

Channel Attention: Channel attention applies average pooling and max pooling across spatial dimensions independently, producing two descriptors. These pass through a shared MLP to

capture inter-channel relationships. The outputs are added and passed through a sigmoid activation to generate the final channel attention map. This map is multiplied element-wise with the input feature map to enhance relevant channels.

Spatial Attention: This component applies channel-wise average and max pooling, resulting in two spatial descriptors. Concatenates them along the channel axis and passes them through a convolutional layer followed by a sigmoid activation to create a spatial attention map. This map is multiplied element-wise with the input to refine spatial details.

In our architecture, CBAM is applied after the CNN backbone to refine the extracted spatial feature maps (B × 512 × 7 × 7) and 20% dropout is performed before feeding them into the transformer. This enhances the representational quality of features by suppressing irrelevant information and emphasizing meaningful parts. This improves the transformer's ability to model long-range dependencies on clean, relevant feature embeddings. By highlighting what the model pays attention to, attention maps can provide a degree of interpretability. If the model consistently focuses on known pathological regions, such as the hippocampus for AD or the substantia nigra for PD, it significantly increases confidence in its decision-making process. Figure 3.9 shows the architecture that has been followed in this model.

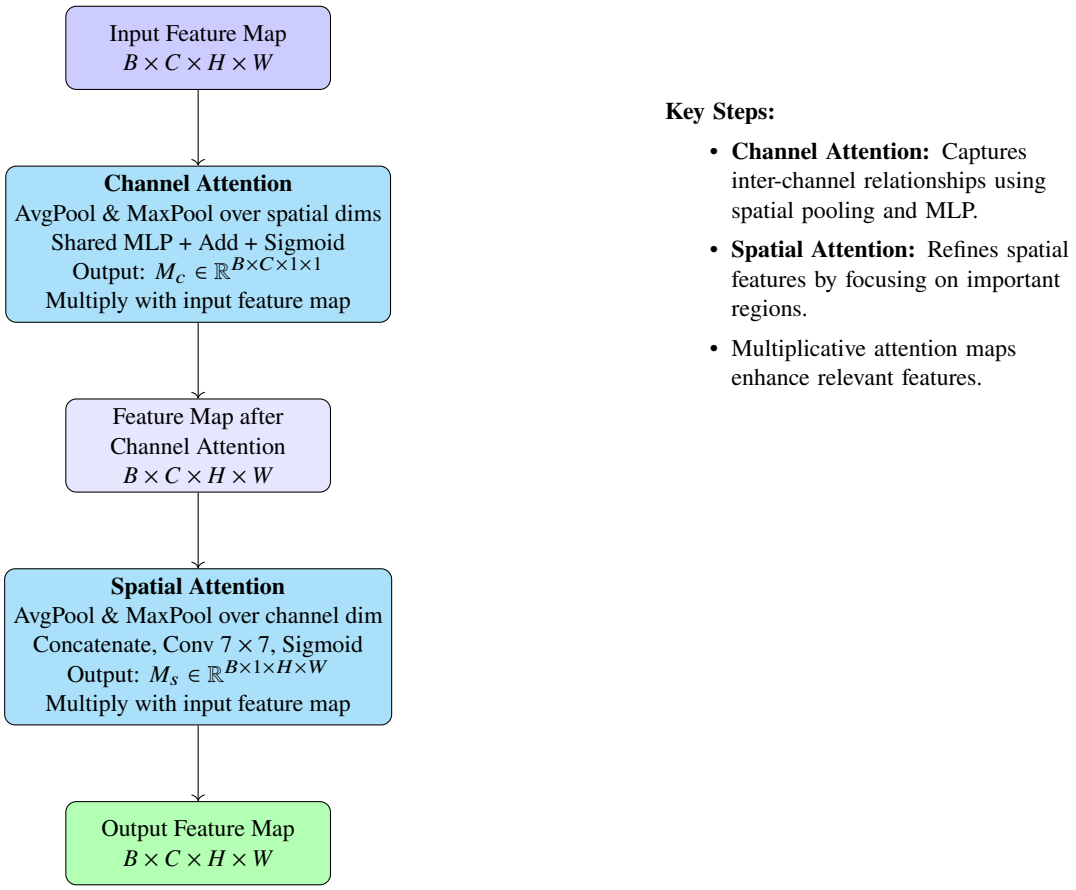

Figure 3.9: Architecture of the Convolutional Block Attention Module (CBAM) showing sequential Channel and Spatial Attention applied to input feature maps.

3.5.4 Vision Transformer

We propose a customized Vision Transformer model which contains the following core components:

- Patch Embedding Layer (1x1 Patches)
- Positional Encoding Module
- A stack of Transformer Blocks (self-attention + MLP layers)
- Tokenized image representation using patch flattening

In our model, the input to the patch embedding layer is a high-level feature map produced by a CNN encoder and enhanced by CBAM. The patch embedding module partitions this feature map into non-overlapping regions and projects each to a fixed-size embedding vector using a 1×1 convolution. This effectively transforms the 2D spatial feature map into a sequence of tokens, enabling the application of transformer blocks for global context modeling. In a standard Vision Transformer, patch dimensions are generally 16x16 or 8x8, but we chose 1x1 because our input is not raw images, but a rich representation of the original image, thanks to the CNN and CBAM. This method of patching is particularly useful when the transformer is applied on top of CNN features rather than directly on the image, as each token carries semantically meaningful information from its corresponding receptive field.

Instead of a learnable positional encoder, we use sinusoidal positional encoding because learnable positional embeddings can lead to overfitting when data is scarce. Additionally, in our case, both techniques give similar results, so we picked sinusoidal due to not having any additional learnable parameters. Given an input of shape (B, H×W, C) where H×W is the number of patches (flattened pixels or tokens) and B is the number of batches, the positional encoding injects spatial order into the token sequence. Figure 3.10 shows the Vision Transformer architecture that has been followed in this scenario.

Each Transformer blocks contains the following:

- Multi-Head Self-Attention
- Feedforward MLP (10% dropout)
- Layer Normalization & Residual Connection

Architectural Flow:

- Input: CBAM enhanced feature map of shape (B, C, H, W), The output from EfficientNetB0 after applying CBAM.
- Patch Embedding: Each 1×1 spatial location is projected into a fixed-dimensional embedding using a 1×1 convolution, treating each location as an individual patch.
- Flattening: The embedded feature map is reshaped into a sequence of shape (B, H×W, C), where each token corresponds to a spatial position.
- Positional Encoding: Fixed sinusoidal positional encodings are added to retain spatial order in the sequence.
- Transformer Encoder: Sequence passes through 4 layers of Transformer blocks.
- Output: A refined sequence of contextualized embeddings of shape (B, H×W, C).

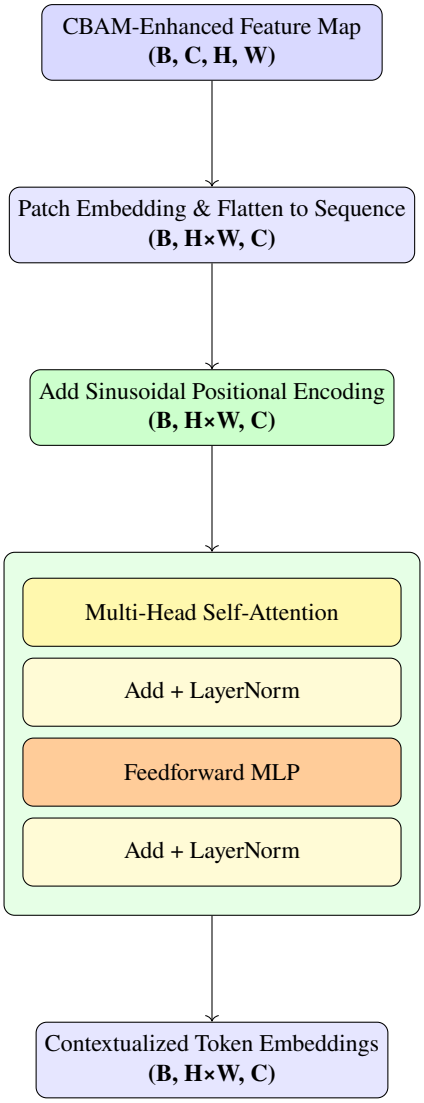

Figure 3.10: Vision Transformer Architecture with Internal Transformer Block Components and Dimensional Annotations

3.5.5 The Proposed Hybrid CNN-Transformer with CBAM

The proposed model begins with a CNN backbone based on EfficientNet-B0 (excluding final layers) to extract spatial features, followed by a 1×1 convolution to reduce channel dimensions. A Convolutional Block Attention Module (CBAM) is applied to refine features by sequentially applying channel and spatial attention. The refined features are reshaped into a sequence of tokens, enriched with fixed sinusoidal positional encodings, and passed through a multi-layer Vision Transformer to capture long-range dependencies. The output tokens are reshaped back into spatial dimensions, passed through a 1×1 convolution to map features to class logits, and aggregated using adaptive average pooling to produce final class predictions.

Component	Description
Input	2D Axial MRI Slice
CNN Backbone	EfficientNet-B0 (pretrained on ImageNet, excluding classification layers) used to extract rich spatial features
Channel Reduction	1×1 Convolution layer to reduce channel dimensions for efficiency and compatibility with attention modules
Attention Module	Convolutional Block Attention Module (CBAM), applied sequentially for channel and spatial attention refinement
Positional Encoding	Fixed sinusoidal encoding added to the token sequence to preserve spatial positional information
Transformer Encoder	Multi-layer Vision Transformer blocks that apply multi-head self-attention to model long-range dependencies
Feature Mapping	1×1 Convolution layer to convert Transformer outputs into class-specific logits
Aggregation	Adaptive Average Pooling used to reduce spatial dimensions and aggregate features globally
Output	Softmax activation for final classification (e.g., Alzheimer's, Parkinson's, Healthy)

Table 3.2: Model Overview of the Proposed Hybrid CNN-Transformer with CBAM

Table 3.3 Displays the shapes of input and output at different components of the model.

Component	Input Shape	Output Shape
CNN Backbone	[1, 3, 224, 224]	[1, 1280, 7, 7]
Channel Reduce	[1, 1280, 7, 7]	[1, 512, 7, 7]
CBAM	[1, 512, 7, 7]	[1, 512, 7, 7]
Patch Embedding	[1, 512, 7, 7]	[1, 49, 512]
Positional Encoder	[1, 49, 512]	[1, 49, 512]
Vision Transformer	[1, 49, 512]	[1, 49, 512]
Reshape	[1, 49, 512]	[1, 512, 7, 7]
1x1 Convolution	[1, 512, 7, 7]	[1, 3, 7, 7]
Global Average Pooling	[1, 3, 7, 7]	[1, 3, 1, 1]

Table 3.3: Input-Output overview of the model

Finally applying softmax on the class logits returned by the Global Average Pooling gives the class probabilities.

Chapter 4

Experimental Setup

4.1 Implementation

To evaluate the effectiveness of the proposed architecture combining CNN, CBAM, and Vision Transformer, multiple model configurations were implemented using the PyTorch deep learning framework. Each configuration was designed to analyze how variation in CNN backbone complexity, attention mechanisms, classifier design, and transformer modules influence the classification performance.

4.1.1 Hyperparameters

Batch Size: 32
Epoch: 15
Optimizer: AdamW (lr = 1e-4, weight_decay = 1e-4)
Scheduler: Cosine Annealing
Loss Function: CrossEntropyLoss (Label Smoothing = 0.1)
Dropouts
CNN Backbone: 20%
CBAM: 20%
Vision Transformer: 10%

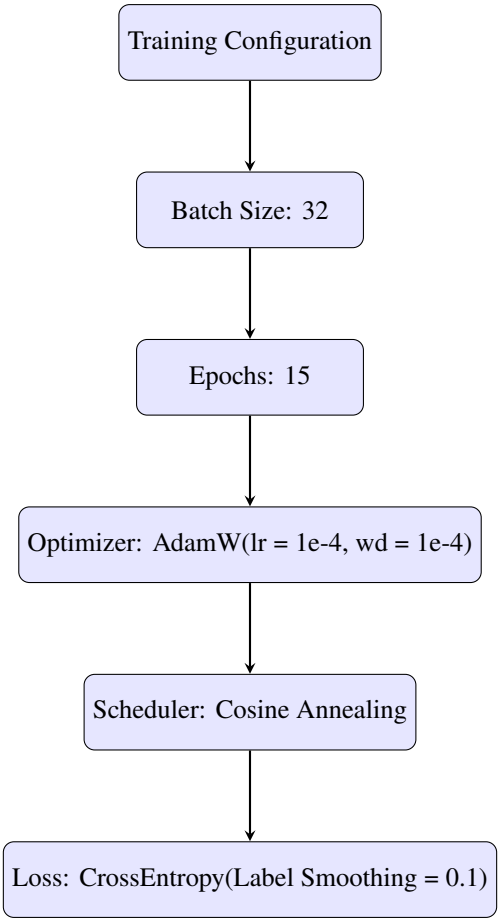

Figure 4.1: Training Configuration Flowchart

4.1.2 Loss Function

Since our task is a multi-class classification, we applied CrossEntropyLoss as our loss function. The standard CrossEntropyLoss inherently assumes equal learning contribution from all classes, which can lead to a bias towards the majority class. Consequently, the model might frequently predict test samples as belonging to the dominant normal class, diminishing its diagnostic utility for the minority disease classes. We mitigate this issue by modifying the optional weight argument of CrossEntropyLoss and assign weights (AD: 2.0, PD: 3.0, Healthy: 1.0) to the different classes due to the imbalance. Additionally we set the label smoothing parameter of CrossEntropyLoss to 0.1. This regularizes the target distribution, reducing overfitting.[49]

$$\mathcal{L} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{4.1}$$

Where:

- C = number of classes
- y_i: true label (1 if the class is correct, 0 otherwise)
- $\hat{y}_i$: predicted probability for class i

4.1.3 Optimizer Configuration

One of the most popular and widely used optimizers is Adam, due to its efficiency and adaptive learning rates. This also requires minimal hyperparameter tuning, making it highly accessible. We used a more advanced version of Adam, the AdamW optimizer. This decouples weight decay from gradient updates and treats it as a separate step in the optimizer, applied directly to the weights after the gradient update. This design results in more precise regularization, significantly aiding models to generalize better.[50]

4.1.4 Learning Rate Scheduling

Our architecture makes use of Cosine Annealing Learning Rate, which is a scheduling technique that starts with a relatively large rate, then gradually decreases to a value close to zero before increasing again. Every time this restart occurs, the good weights from the preceding cycle is used as the starting point. This nature of Cosine Annealing allows the model to periodically jump out of shallow local minima and then converge more precisely in potentially better regions. The strategic reuse of good weights from previous cycles makes this process highly efficient.[51]

4.1.5 Evaluation

Metrics

Overall accuracy alone can be highly misleading, especially in our imbalanced medical dataset. Precision, Recall (Sensitivity), and F1-score offer a far more nuanced perspective on performance, particularly for minority classes such as AD and PD.

Overall Accuracy: The percentage of correctly classified samples across all classes.

Confusion Matrix: To visually represent and analyze per-class performance, specifically identifying patterns of mis-classification between AD, PD, and Healthy groups.

Per-Class Precision, Recall, and F1-score: These are crucial for understanding the model's performance on each specific class (AD, PD, Healthy). High Recall for disease classes is clinically vital as it signifies fewer false negatives. Conversely, high precision is equally important as it indicates fewer false positives. The F1-score, which represents the harmonic mean of Precision and Recall, provides a balanced measure. A high F1-score for AD and PD classes indicates that the model is effectively identifying these conditions without reporting too many false alarms. To assess the performance of the proposed deep learning model for neurodegenerative disease classification, the following evaluation metrics were used: Accuracy, Precision, Recall, and F1-Score. These metrics are computed based on the number of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.2}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.3}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4.4}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.5}$$

For multi-class classification, these metrics are computed for each class separately using a one-vs-rest approach, and then averaged to obtain the final performance scores.

Chapter 5

Results & Analysis

5.1 Discussion

This section presents the experimental results of the proposed hybrid CNN-Transformer model and analyzes its performance in classifying MRI slices into Alzheimer's Disease (AD), Parkinson's Disease (PD), and Healthy categories. Evaluation was conducted using classification metrics and visualizations.

The model was trained over 15 epochs using the AdamW optimizer with cosine annealing learning rate scheduling. Figure 5.1 shows that the training accuracy reaches 98.16%, while validation accuracy reaches 94.70% at the final epoch, indicating good generalization. Additionally, validation loss stabilizes after Epoch 12.

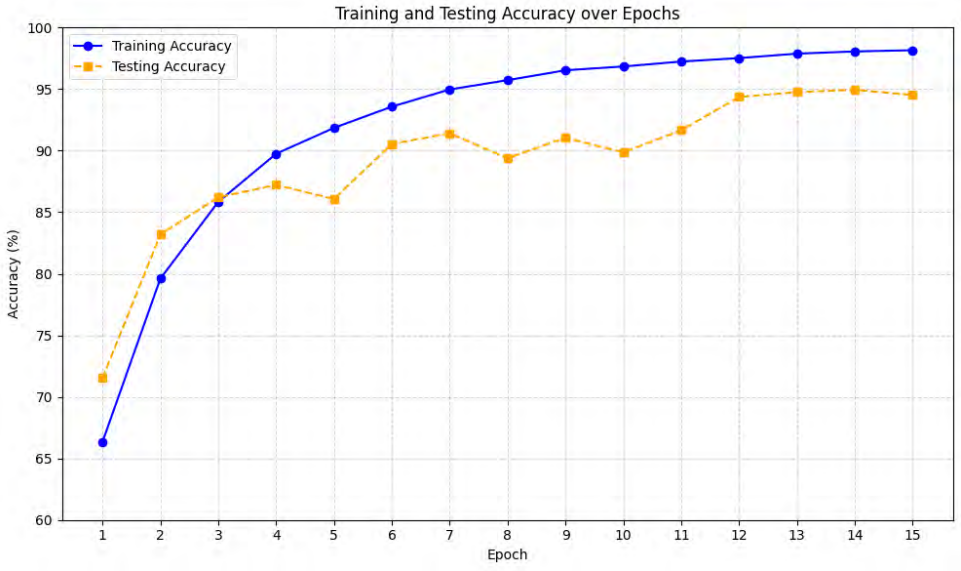

Figure 5.1: Accuracy Over Epochs

The confusion matrix in Figure 5.2 displays the accurate classifications and misclassifications of each class.

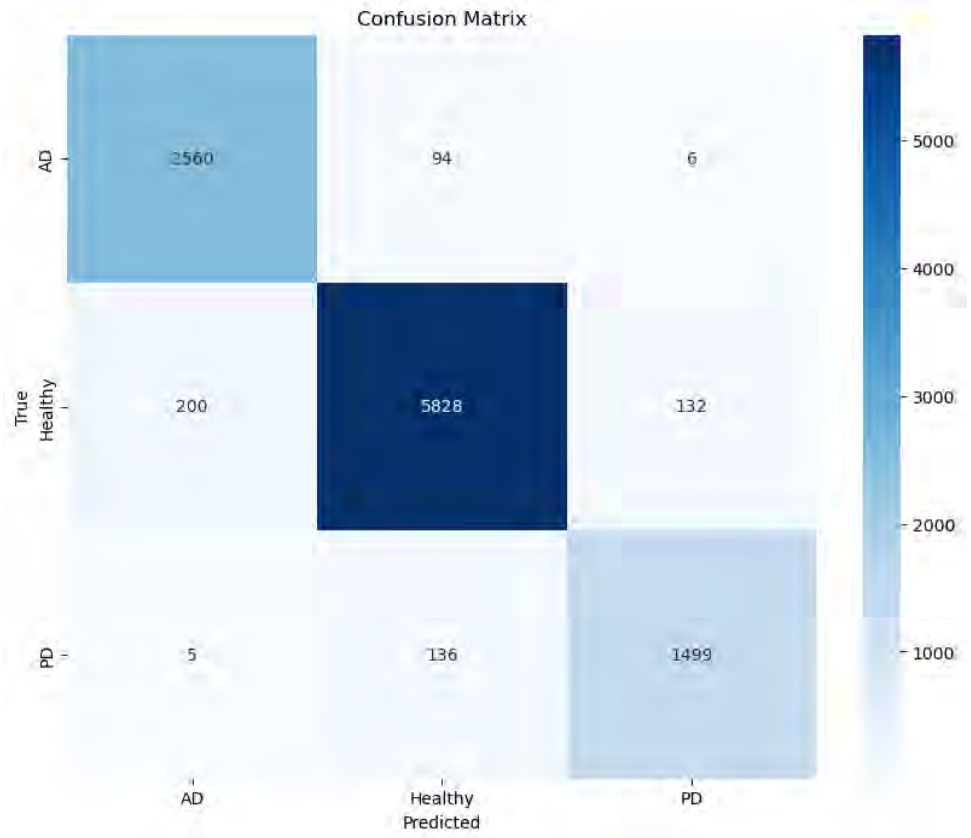

Figure 5.2: Confusion Matrix of Slices

High precision and recall across all three classes makes it suitable for real-world usage. Table 5.1 shows the class-wise Precision, Recall and the F1-scores. Healthy class is very well predicted, primarily due to higher sample count. AD and PD also perform strongly, with slight reduction of Precision and Recall for PD. Macro avg and Weighted avg are both 0.94–0.95.

Class	Precision	Recall	F1-Score
AD	0.93	0.96	0.94
PD	0.92	0.91	0.91
Healthy	0.96	0.95	0.95

Table 5.1: Classification Report

5.2 Comparative Analysis

We have tested several different methods before coming to our final model, such as a different CNN Backbone, or a simplified Vision Transformer, different epochs, etc to evaluate the importance of various components. Table 5.2 gives an overview of the results and values.

Model	ViT	Positional Encoding	Train Accuracy (%)	Test Accuracy (%)	Parameters (Millions)	Epochs
ResNet18	Minimal ViT	None	98.00	91.00	24.0	10
ResNet34	Minimal ViT	Learnable	98.79	93.43	34.1	10
EfficientNetB0	Minimal ViT	Learnable	88.62	86.04	17.2	10
EfficientNetB0	Custom ViT	Sinusoidal	93.80	91.09	17.6	10
EfficientNetB0	Custom ViT	Sinusoidal	97.29	94.27	17.6	15
ResNet18 + Dropouts	Custom ViT	Sinusoidal	98.18	93.46	24.4	15
ResNet34 + Dropouts	Custom ViT	Sinusoidal	98.60	94.60	34.5	15
EfficientNetB0 + Dropouts	Custom ViT	Sinusoidal	98.16	94.70	17.6	15
EfficientNetB0 (No CBAM)	Custom ViT	Sinusoidal	98.09	93.00	17.6	15

Table 5.2: Performance of Different CNN + ViT Architectures

Resnet18 with Simple ViT

Learnable Parameters: 24 million
We used Resnet18 as the CNN backbone and combined it with a simple ViT consisting only of transformer blocks and no positional encodings. This reported a significant overfitting, giving a Training accuracy of 98%, and a testing accuracy of 91%.

Resnet34 with Simple ViT

Learnable Parameters: 34.1 million
Resnet34 combined with a lightweight ViT (with a learnable positional encoder). This also reported overfitting, with a training accuracy of 98.79% and testing accuracy of 93.43%.

EfficientNetB0 with Simple ViT

Learnable Parameters: 17.2 million
EfficientNetB0 combined with the same ViT (containing a learnable positional encoder). This reported a training accuracy of 88.62% and testing accuracy of 86.04%.

Final Model (EfficientNetB0 with Customized ViT) with Augmented Dataset

Learnable Parameters: 17.6 million
EfficientNetB0 combined with the customized ViT (Containing patch embeddings and positional encoding) of the final model. This reported a training accuracy of 93.80% and testing accuracy of 91.09%. Here the following Augmentations were applied:

Augmentation	Description
Random Crop	The cropped area covers 80% to 100% of the original image, introducing variability in zoom and framing.
Random Rotation	Rotate the images randomly within ±15 degrees.
Random Affine	Applies random translation (up to 5% shift) and scaling (±5%).
Color Jitter	Randomly adjusts brightness and contrast by ±10%.

Final Model (EfficientNetB0 with Customized ViT) with Augmented Dataset in 15 epochs

This changes the epoch of the previous task from 10 to 15. This gives a training accuracy of 97.29% and a test accuracy of 94.27%. Additionally Rotation and Color Jitter Augmentations were removed.

Final Model (EfficientNetB0 with Customized ViT) with Augmented Dataset in 15 epochs with added Dropouts

Modified the previous task and added 20% dropout in CNN and CBAM layers. Additionally, two augmentations (Random Rotation & Color Jitter) were removed. This is our best results, giving a train accuracy of 98.16% and a test accuracy of 94.70%, while also maintaining a significantly lower parameter count compared to ResNet architectures, which display similar results as shown in Table 5.2.

Final Model (EfficientNetB0 with Customized ViT) with Augmented Dataset in 15 epochs with added Dropouts and No CBAM

To evaluate the impact of the attention mechanisms, we compared our model with a CNN+ViT model without CBAM. While the results were close, this model demonstrated lower testing accuracy (93%) and a much lower recall for PD class (0.79).

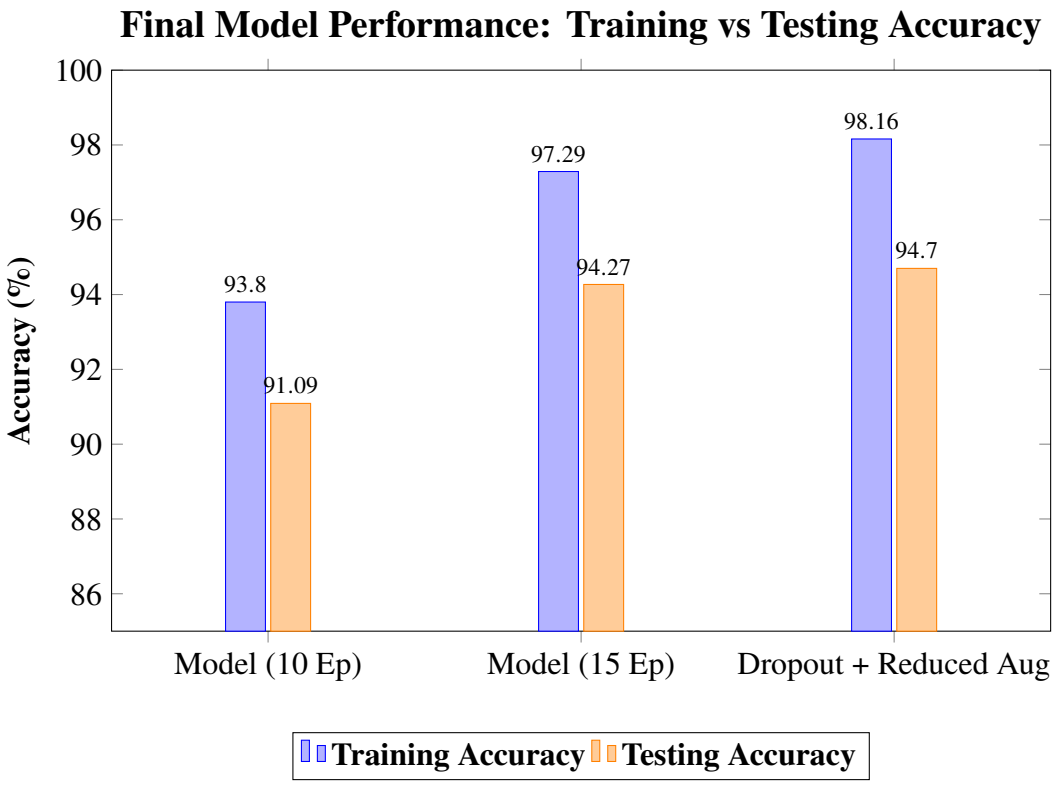

Figure 5.3: Comparison of Training and Testing Accuracy for EfficientNetB0 Variants

5.3 Limitations

Although the dataset used was preprocessed and augmented, it remains relatively limited in size and diversity, especially for Parkinson's Disease samples. This may affect the generalizability of the model to unseen, real-world data from diverse populations or MRI protocols. Our model was trained using 20 MRI slices from each sample. Some slices could be non-informative, which could result in additional noise in the training process. Even though attention mechanisms were used in our architecture, explicit explainability techniques such as Grad-CAM were not integrated to visualize model decisions. This limits clinical trust and interpretability.

Our pipeline did not implement any mechanism to prioritize or weigh the relative importance of individual slices during classification. Incorporating slice selection or attention-based fusion might further improve performance. A weighted slice selection model should do better than uniform input. Another issue is that the study is limited to T2-weighted axial MRI scans. Other modalities like T1, FLAIR, or PET scans might provide complementary information that could enhance diagnosis accuracy.

Chapter 6

Conclusion

The proposed research aims to create an automated system of deep learning to classify and identify neurodegenerative brain diseases (Alzheimer (AD) and Parkinson (PD)) based on brain magnetic resonance imaging (MRI) scans. Timely treatment depends upon early diagnosis, and this is why we aim to create a reliable, accurate and efficient tool that will help the clinicians to treat the patients in time.

We suggest the model combining the efficient net B0 (one kind of CNN), CBAM (Convolutional Block Attention Module), and ViTs (Vision Transformers). EfficientNetB0 handles the detailed local features of MRI slices, CBAM boosts the emphasis on essential areas of the brain through channel and spatial attention, and ViTs achieve global patterns identification as well as long-range dependencies. This combination results in local area actions with a wider perspective of context.

We enhanced the learning and avoided overfitting by high-boost filtering to emphasize features in the images and data-augmentation methods such as rotating and scaling. Our last model performed at a nearly perfect accuracy (95%), but the parameter count (5.3 million) was low, which, in turn, allows us to deploy it even on restricted medical devices.On balance, our blended strategy works much better (and more efficiently) than most conventional ones. It has great promise as a scalable and accessible diagnostic to use in the clinical setting, early and accurate detection of neurodegenerative diseases in better outcomes of the patients.

Bibliography

[1] O. Dimmer. *Uncovering the culprits of neurodegenerative disorders*. Feinberg News Center, April 11. 2024. URL: https://news.feinberg.northwestern.edu/2024/04/11/uncovering-the-culprits-of-neurodegenerative-disorders/.

[2] M Rohini and D Surendran. "Classification of Neurodegenerative Disease Stages using Ensemble Machine Learning Classifiers". In: *Procedia Computer Science* 165 (2020), pp. 66–73. DOI: 10.1016/j.procs.2020.01.071.

[3] MBT Noor et al. "Detecting Neurodegenerative Disease from MRI: A Brief Review on a Deep Learning Perspective". In: *Brain Informatics*. Vol. 11976. 2019, pp. 115–125. DOI: 10.1007/978-3-030-37078-7_12. URL: https://link.springer.com/chapter/10.1007/978-3-030-37078-7_12.

[4] D. AlSaeed and S. F. Omar. "Brain MRI Analysis for Alzheimer's Disease Diagnosis Using CNN-Based Feature Extraction and Machine Learning". In: *Sensors* 22.8 (2022). DOI: 10.3390/s22082911. URL: https://www.mdpi.com/1424-8220/22/8/2911.

[5] EM Reiman et al. "Alzheimer's disease: Early detection and possible prevention through imaging, biomarker and lifestyle-based interventions". In: *Alzheimer's Dementia* 8.4 (2012), pp. 353–362. DOI: 10.1016/j.jalz.2011.11.003.

[6] Rosebud Roberts and David S Knopman. "Classification and epidemiology of MCI". In: *Clinics in Geriatric Medicine* 29.4 (2013), pp. 753–772. DOI: 10.1016/j.cger.2013.07.003.

[7] Modupe Odusami et al. "Analysis of neuroimaging data using deep learning for early detection of Alzheimer's disease". In: *Biomedicines* 10.1 (2022), p. 145. DOI: 10.3390/biomedicines10010145.

[8] Caiyi Lian et al. "Attention-guided multi-modality fusion for Alzheimer's disease diagnosis". In: *Medical Image Analysis* 60 (2020), p. 101622. DOI: 10.1016/j.media.2019.101622.

[9] Sergey Korolev et al. "Residual and plain convolutional neural networks for 3D brain MRI classification". In: *Neurocomputing* 320 (2018), pp. 183–195. DOI: 10.1016/j.neucom.2018.09.082.

[10] J. Cheng, M. Liu, D. Zhang, et al. "Classification of MR brain images by combination of multi-CNNs for AD diagnosis". In: *Medical Imaging 2017: Computer-Aided Diagnosis*. Vol. 10134. 2017, p. 101342V. DOI: 10.1117/12.2281808.

[11] Mingxia Liu et al. "Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis". In: *Neuroinformatics* 16 (2018), pp. 295–308. DOI: 10.1007/s12021-018-9370-4.

[12] Mingxia Liu, Danni Cheng, and Wenya Yan. "A multi-model deep convolutional neural network for automatic hippocampal segmentation and Alzheimer's disease diagnosis". In: *NeuroImage* 208 (2020), p. 116459. DOI: 10.1016/j.neuroimage.2019.116459.

[13] Yuchuan Qiao, Jing Chen, and Xiaojin Zhu. "Multi-view convolutional neural networks for early diagnosis of Alzheimer's disease". In: *IEEE Transactions on Medical Imaging* 40.1 (2021), pp. 1–10. DOI: 10.1109/TMI.2020.3013254.

[14] Caiyi Lian et al. "Hierarchical attention-based multimodal fusion network for Alzheimer's disease prediction". In: *IEEE Transactions on Image Processing* 31 (2022), pp. 145–158. DOI: 10.1109/TIP.2021.3098036.

[15] Y Tang et al. "Dual-attention network and MLP-based Alzheimer's disease classification framework with multimodal data". In: *Frontiers in Neuroinformatics* 18 (2024). DOI: 10.3389/fninf.2024.1507217.

[16] Yan Zhang et al. "CAPCBAM: A capsule network with CBAM for Alzheimer's disease classification". In: *Scientific Reports* 15 (2025), p. 85703. URL: https://www.nature.com/articles/s41598-025-85703-x.

[17] L Zhao, H Wang, and Z Lin. "Lightweight stacked CNN with channel attention for Alzheimer's diagnosis". In: *Artificial Intelligence Review* 58.3 (2025), pp. 679–696. URL: https://link.springer.com/article/10.1007/s10462-025-11146-5.

[18] Binasss. “Alzheimer’s disease prediction using DANMLP model”. In: *ACM Transactions on Computing for Healthcare* 6.1 (2024), pp. 1–18. URL: https://dl.acm.org/doi/abs/10.1145/3665689.3665728.

[19] Fang Yuan et al. “Adaptive hybrid attention network for early Alzheimer’s disease detection using MRI”. In: *Applied Intelligence* 55.1 (2025), pp. 112–124. DOI: 10.1007/s00521-021-06368-x.

[20] T. Akan, S. Alp, and M. A. N. Bhuiyan. “Vision Transformers and Bi-LSTM for Alzheimer’s disease diagnosis from 3D MRI”. In: *arXiv preprint arXiv:2401.03132* (2024). URL: https://arxiv.org/abs/2401.03132.

[21] X. Fan, H. Li, L. Liu, et al. “Early diagnosing and transformation prediction of Alzheimer’s disease using multi-scaled self-attention network on structural MRI images”. In: *Journal of Alzheimer’s Disease* 92.3 (2024), pp. 1235–1250. URL: https://www.researchgate.net/publication/376962808.

[22] Z. Hu et al. “Conv-Swinformer: Integration of CNN and shift window attention for Alzheimer’s disease classification”. In: *Computers in Biology and Medicine* 152 (2023), p. 106458. URL: https://www.sciencedirect.com/science/article/abs/pii/S0010482523007692.

[23] Jing Chen et al. “TransMed: Transformers Advance Multimodal Medical Image Classification”. In: *arXiv preprint arXiv:2103.05940* (2021). URL: https://arxiv.org/abs/2103.05940.

[24] Y Zhang et al. “VGG-TSwinformer: Transformer-based deep learning model for early Alzheimer’s disease prediction”. In: *Computer Methods and Programs in Biomedicine* 229 (2022), p. 107291. DOI: 10.1016/j.cmpb.2022.107291.

[25] H. Shin et al. “Vision Transformer Approach for Classification of Alzheimer’s Disease Using 18F-Florbetaben Brain Images”. In: *Applied Sciences* 13.6 (2023), p. 3453. DOI: 10.3390/app13063453. URL: https://doi.org/10.3390/app13063453.

[26] Donghuan Lu et al. “Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer’s disease using structural MR and FDG-PET images”. In: *Scientific Reports* 8.1 (2018), p. 5697. DOI: 10.1038/s41598-018-22871-z.

[27] Z Zhou, Y Li, and X Wang. “Transformer and convolutional neural network: A hybrid model for multimodal data in multiclass classification of Alzheimer’s disease”. In: *Mathematics* 13.10 (2024), p. 1548. URL: https://www.mdpi.com/2227-7390/13/10/1548.

[28] Hyun Shin et al. “Vision transformer approach for classification of Alzheimer’s disease using 18F-Florbetaben brain images”. In: *Applied Sciences* 13.6 (2023), p. 3453. DOI: 10.3390/app13063453.

[29] PubMed Central (PMC). *Attention-enhanced deep learning for MRI-based Alzheimer’s diagnosis*. 2025. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11815323/.

[30] Sourav Roy et al. *AD-Lite Net: A lightweight and concatenated CNN model for Alzheimer’s detection from MRI images*. 2024. arXiv: 2409.08170 [cs.CV]. URL: https://arxiv.org/abs/2409.08170.

[31] Maximilian Ilse, Jakub M. Tomczak, and Max Welling. “Attention-based Deep Multiple Instance Learning”. In: *Proceedings of the 35th International Conference on Machine Learning*. Vol. 80. Proceedings of Machine Learning Research. PMLR, 2018, pp. 2127–2136. URL: https://proceedings.mlr.press/v80/ilse18a.html.

[32] ASM Mahmood et al. “Transfer learning with ResNet-18 and squeeze-and-excitation for Alzheimer’s disease classification using MRI”. In: *Frontiers in Neuroinformatics* 18 (2024). DOI: 10.3389/fninf.2024.1507217.

[33] MNI Qureshi et al. “Multi-disease classification of brain MRI images using deep convolutional neural network with transfer learning”. In: *Diagnostics* 11.6 (2021), p. 1071. DOI: 10.3390/diagnostics11061071.

[34] Linda Delali Fiasam et al. “Efficient 3D residual network on MRI data for neurodegenerative disease classification”. In: *Thirteenth International Conference on Graphics and Image Processing (ICGIP)*. Vol. 12083. 2022, 120831A. DOI: 10.1117/12.2623238. URL: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/12083/120831A.

[35] M. Z. Alom et al. “A 3D ResNet-18 based transfer learning approach for Alzheimer’s disease classification”. In: *Proceedings of the 17th International Conference on Brain Informatics (BI 2021)*. Springer, 2021.

[36] Mohammad Pashaei et al. “Deep learning models for Alzheimer’s disease classification: ResNet-18 vs. ResNet-50 vs. ResNet-101”. In: *Recent Advances in Computer Science and Communications* 15.1 (2022), pp. 45–57. DOI: 10.2174/2666255815666220208164829.

[37] Muhammad Sajjad et al. “Multi-grade brain tumor classification using deep CNN with extensive data augmentation”. In: *Journal of Neuroscience Methods* 319 (2019), pp. 39–49. DOI: 10.1016/j.jneumeth.2019.02.017.

[38] Y Zhao et al. “Conv-Swinformer: Integration of CNN and shift window attention for Alzheimer’s disease classification”. In: *Computers in Biology and Medicine* 164 (2023), p. 107304. DOI: 10.1016/j.compbiomed.2023.107304.

[39] A. Majee et al. “Enhancing MRI-based classification of Alzheimer’s disease with explainable 3D hybrid compact convolutional transformers”. In: *arXiv preprint arXiv:2403.16175* (2024). URL: https://arxiv.org/abs/2403.16175.

[40] T. Akan et al. “Leveraging Video Vision Transformer for Alzheimer’s disease diagnosis from 3D brain MRI”. In: *arXiv preprint arXiv:2501.15733* (2025). URL: https://arxiv.org/abs/2501.15733.

[41] A. Jaiswal and A. Sadana. “Early detection of Alzheimer’s disease using bottleneck transformers”. In: *arXiv preprint arXiv:2305.00923* (2023). URL: https://arxiv.org/abs/2305.00923.

[42] Yicheng Li et al. *DiaMond: Dementia diagnosis with multi-modal vision transformers using MRI and PET*. 2024. arXiv: 2410.23219 [cs.CV]. URL: https://arxiv.org/abs/2410.23219.

[43] P. C. Wong et al. “Transfer learning for Alzheimer’s disease diagnosis using EfficientNet-B0 convolutional neural network”. In: *Journal of Advanced Research in Applied Sciences and Engineering Technology* 30.1 (2023), pp. 91–100. URL: https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/article/view/1668.

[44] Q.-H. Trinh et al. “EfficientNet for brain-lesion classification”. In: *arXiv preprint arXiv:2208.04616* (2022). URL: https://arxiv.org/abs/2208.04616.

[45] A. Kadri, S. Messaoud, and B. Lyes. “Vision Transformer approach for classification of Alzheimer’s disease using 18F-Florbetaben brain images”. In: *Applied Sciences* 13.6 (2023), p. 3453. DOI: 10.3390/app13063453. URL: https://doi.org/10.3390/app13063453.

[46] Y. Zhao et al. “Vision transformer-equipped convolutional neural networks for automated Alzheimer’s disease diagnosis using 3D MRI scans”. In: *International Journal of Imaging Systems and Technology* (2024). URL: https://pubmed.ncbi.nlm.nih.gov/39737424/.

[47] A. Ebrahimi-Ghahanavieh, S. Luo, and R. Chiong. “Transfer learning for Alzheimer’s disease detection on MRI images”. In: *2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*. 2019, pp. 133–138. DOI: 10.1109/ICIAICT.2019.8784845.

[48] Mingxing Tan and Quoc V. Le. “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”. In: *International Conference on Machine Learning* (2019). URL: https://arxiv.org/abs/1905.11946.

[49] Anqi Mao, Mehryar Mohri, and Yutao Zhong. “Cross-Entropy Loss Functions: Theoretical Analysis and Applications”. In: *ICML 2023* (2023).

[50] Ilya Loshchilov and Frank Hutter. “SGDR: Stochastic Gradient Descent with Warm Restarts”. In: *ICLR 2019 conference paper* (2017). URL: https://arxiv.org/abs/1711.05101v3.

[51] Ilya Loshchilov and Frank Hutter. “Decoupled Weight Decay Regularization”. In: *ICLR 2017 conference paper* (2017). URL: https://arxiv.org/abs/1608.03983v5.

[52] Hongjie Yan et al. “Hybrid-RViT: Hybridizing ResNet-50 and Vision Transformer for Enhanced Alzheimer’s disease detection”. In: *PLOS ONE* (2025). URL: https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0318998.