الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
وزارة التعليم العالي و البحث العلمي
Ministry of Higher Education and Scientific Research
جامعة أبي بكر بلقايد تلمسان
Abou Bekr Belkaid University -Tlemcen- Faculty of Technology

Thesis

Presented to obtain the degree of **DOCTORATE Third Cycle Degree**

Field of Study: **Biomedical Engineering**
Specialization: **Medical Imaging**

Theme

Development of a diagnostic aid system for neurodegenerative pathologies: Application to the detection of Alzheimer's disease

---

Presented by
Saim Meriem

Defended on: **15,May,2025**
In front of the jury composed of

Mr. Brixi Reguig Fethi	Professor	Univ. Tlemcen	President of the Jury
Mm. Feroui Amel	Professor	Univ. Tlemcen	Thesis Supervisor
Mr. Ammar Mohammed	Professor	Univ. Boumerdess	Examiner 1
Mr. Messadi Mahemmed	Professor	Univ. Tlemcen	Examiner 2
Mm. Benchaib Yasmine	MCA	Univ. Tlemcen	Examiner 3
Mr. Bessaid Abedelhafid	Professor	Univ. Tlemcen	Guest
Mr. Mansseri Nabil	Doctor	CHU Tlemcen	Guest

Academic Year: 2024/2025

Dedication

To my dearest Father **Saim Mohamed** and beloved Mother **Zenagui Fatima**,
for your endless love, guidance, and sacrifices.
Your unwavering support has been my greatest strength.

And to my sweet sister **Saim Malek Imane**,
for being my constant companion and source of joy.
Your laughter and kindness light up my world.

With all my love and gratitude,

Saim Meriem

Acknowledgement

First and foremost, I would like to express my deepest gratitude to Almighty ALLAH, whose blessings and guidance have been my source of strength and inspiration throughout this journey. Without His grace, none of this would have been possible.

I am profoundly thankful to the members of the jury for their time, valuable insights, and constructive feedback. Their expertise and dedication have greatly contributed to the completion of this work. In particular, I would like to acknowledge:

Pr. Brixi Reguig Fethi, President of the Jury, for his leadership and guidance during the evaluation process. **Pr. Feroui Amel**, my Thesis Supervisor, for her unwavering support, patience, and invaluable advice throughout this research. **Pr. Ammar Mohammed**, for his thoughtful comments and suggestions. **Pr. Messadi Mahemmed**, for his critical review and encouragement. **Mm. Benchaib Yasmine**, for her insightful feedback and encouragement. My sincer appreciation goes to **Pr. Bessaid Abedelhafid** and **Dr. Mansseri**, for their presence and contributions to this academic endeavor.

My deepest gratitude is extended to my **Parents** and my **sister** for their unwavering support and patience throughout this undertaking. Their affection, supplications, and support have consistently served as a source of inspiration.

I am grateful to all Biomedical Engineering Laboratory members. I'd want to express my profound appreciation to Chaima CHERFI, Abdi Hadjer, Mokhtari aicha, Fatiha Youbi, Zineb Aziza Elaouaber, and Hafida Belfilali, all of whom made significant contributions to the success of our adventure.

Abstract

Alzheimer's disease (AD) is a degenerative disorder and one of the most widespread forms of dementia, with no current cure available. This absence of a cure has led the medical field to focus on managing the symptoms of the disease. However, its progressive nature complicates the identification of disease stages, often requiring years of expertise. Consequently, computer-aided diagnostic systems are essential to assist clinicians in accurately defining disease stages, enabling more targeted and effective treatments. Given the lack of a cure, early detection of AD, particularly at the Mild Cognitive Impairment (MCI) stage, is crucial to slow or halt disease progression. However, distinguishing MCI symptoms from normal aging remains challenging, even with MRI imaging, due to the subtle differences across MCI sub-stages.

This thesis focuses on accurately classifying AD stages, emphasizing early-stage detection. Two MRI databases were utilized, leading to four methodologies addressing specific challenges while fulfilling the research objectives. The first system combines the Histogram of Oriented Gradients (HOG) with the Bias Correction Fuzzy C Means algorithm and machine learning classifiers, achieving an accuracy of 96.8% for the first database and 96% for the second. The second system shifts to the frequency domain, employing the Fast Finite Shearlet Transform (FFST) and Gray Level Co-occurrence Matrix (GLCM) angles, resulting in 72% accuracy for the first database. Building on these approaches, the third system leverages inductive transfer learning with layer-wise fine-tuning of ten pre-trained models. The best results were achieved using the Xception architecture, yielding 85.19% accuracy for the first database and 77.23% for the second using VGG19.

Finally, the fourth methodology integrates machine learning and deep learning by automatically extracting features and refining them using Bayesian optimization. It achieves an accuracy of 98.45% for the first database and 78.54% for the second. These methodologies address critical research questions and highlight the importance of feature quality, hyperparameter optimization, and data augmentation techniques in medical imaging. The findings underscore the potential of advanced computer-aided diagnostic systems to enhance the detection and staging of Alzheimer's disease.

Keywords— Alzheimer's disease (AD), Multiclassification, Histogram of Oriented Gradients (HOG), Fast Finite Shearlet Transform (FFST), Gray Level Co-occurrence Matrix (GLCM), Machine learning algorithm, Inductive transfer learning, Bayesian optimization

Resume

La maladie d'Alzheimer (MA) est un trouble dégénératif et l'une des formes de démence les plus répandues, pour laquelle aucun traitement curatif n'est actuellement disponible. Cette absence de traitement a conduit le domaine médical à se concentrer sur la gestion des symptômes de la maladie. Cependant, sa nature progressive complique l'identification des stades, nécessitant souvent des années d'expertise. Par conséquent, les systèmes de diagnostic assisté par ordinateur sont essentiels pour aider les cliniciens à définir avec précision les stades de la maladie, permettant ainsi des traitements plus ciblés et efficaces. Étant donné l'absence de traitement, le dépistage précoce de la MA, en particulier au stade du déclin cognitif léger (DCL), est crucial pour ralentir ou arrêter sa progression. Toutefois, distinguer les symptômes du DCL de ceux du vieillissement normal reste un défi, même avec l'imagerie par IRM, en raison des différences subtiles entre les sous-stades du DCL.

Cette thèse se concentre sur la classification des stades de la maladie d'Alzheimer, en mettant l'accent sur la détection précoce. Deux bases de données IRM ont été utilisées, aboutissant au développement de quatre méthodologies traitant des défis spécifiques tout en répondant aux objectifs de recherche. Le premier système combine l'histogramme des gradients orientés (HOG) avec l'algorithme BCFCM et des classificateurs d'apprentissage automatique, atteignant une précision de 96,8% pour la première base de données et 84% pour la deuxième. Le deuxième système passe au domaine fréquentiel, utilisant la transformée de Shearlet fini rapide (FFST) et les angles de la matrice de cooccurrence des niveaux de gris (GLCM), obtenant une précision de 72% pour la première base de données. En s'appuyant sur ces approches, le troisième système exploite l'apprentissage inductif par transfert avec un ajustement couche par couche de dix modèles pré-entraînés. Les meilleurs résultats ont été obtenus en utilisant l'architecture Xception, avec une précision de 85,19% pour la première base de données et 77.23% pour la deuxième en utilisant VGG19. Enfin, la quatrième méthodologie intègre l'apprentissage automatique et l'apprentissage profond en extrayant automatiquement des caractéristiques et en les affinant à l'aide de l'optimisation bayésienne, atteignant une précision de 98,45% pour la première base de données et 78,54% pour la deuxième. Ces méthodologies répondent à des questions de recherche cruciales et mettent en évidence l'importance de la qualité des caractéristiques, de l'optimisation des hyperparamètres et des techniques d'augmentation des données en imagerie médicale. Les résultats soulignent le potentiel des systèmes avancés de diagnostic assisté par ordinateur pour améliorer la détection et le stade de la maladie d'Alzheimer.

Mots Clés— Maladie d'Alzheimer (MA), Classification multiclasse, Histogramme des gradients orientés (HOG), Transformation rapide de Shearlet finie (FFST), Matrice de co-occurrence des niveaux de gris (GLCM), Algorithmes d'apprentissage automatique, Apprentissage par transfert inductif, Optimisation bayésienne.

ملخص

مرض الزهايمر (AD) هو اضطراب تنكسي وأحد أكثر أشكال الخرف انتشارًا، ولا يوجد علاج حاليًا متاح له. وقد أدى غياب العلاج إلى تركيز المجال الطبي على إدارة أعراض المرض. ومع ذلك، فإن طبيعته التقدمية تعقد تحديد مراحل المرض، وغالبًا ما يتطلب الأمر سنوات من الخبرة. وبالتالي، فإن أنظمة التشخيص المدعومة بالحاسوب أساسية لمساعدة الأطباء في تحديد مراحل المرض بدقة، مما يمكّن من تطبيق علاجات أكثر استهدافًا وفعالية.

نظرًا لعدم وجود علاج، فإن الكشف المبكر عن مرض الزهايمر، خاصة في مرحلة ضعف الإدراك المعتدل (MCI)، أمر بالغ الأهمية لإبطاء أو إيقاف تقدم المرض. ومع ذلك، فإن تمييز أعراض MCI عن أعراض الشيخوخة الطبيعية لا يزال يمثل تحديًا، حتى مع تصوير الرنين المغناطيسي (IRM)، بسبب الفروق الدقيقة بين المراحل الفرعية لـ MCI.

تركز هذه الأطروحة على تصنيف مراحل مرض الزهايمر بدقة، مع التركيز على الكشف في المراحل المبكرة. تم استخدام قاعدتي بيانات للرنين المغناطيسي، مما أدى إلى تطوير أربع منهجيات تتناول تحديات محددة مع تحقيق أهداف البحث. يجمع النظام الأول بين تدرج التاريخ المدعوم (HOG) مع خوارزمية BCFCMومصنفات التعلم الآلي، محققًا دقة قدرها 96.80 %للقاعدة الأولى و 84% للقاعدة الثانية. أما النظام الثاني، فينتقل إلى مجال التردد، مستخدمًا التحويل السريع للمصفوفة اللانهائية (TSFF) وزوايا مصفوفة التوافق للرمادي (GLCM)، محققًا دقة 72% للقاعدة الأولى.

وبناءً على هذه الأساليب، يستخدم النظام الثالث التعلم التحويلي الاستقرائي مع التعديل التدريجي لطبقات عشرة نماذج مدربة مسبقًا. تم تحقيق أفضل النتائج باستخدام بنية Xception، مما أدى إلى دقة قدرها 91.58% للقاعدة الأولى ووصلت إل77 % للقاعدة الثانية باستخدام VGG19. وأخيرًا، يدمج المنهج الرابع التعلم الآلي والتعلم العميق من خلال استخراج الميزات تلقائيًا وتحسينها باستخدام التحسين البيزي، محققًا دقة قدرها 98.45 % للقاعدة الأولى و78.54 % للقاعدة الثانية.

تتناول هذه المنهجيات أسئلة بحثية حاسمة وتبرز أهمية جودة الميزات، وتحسين المعلمات الفائقة، وتقنيات زيادة البيانات في التصوير الطبي. كما تؤكد النتائج على إمكانات أنظمة التشخيص المدعومة بالحاسوب المتقدمة في تحسين الكشف وتحديد مراحل مرض الزهايمر.

الكلمات المفتاحية--- مرض الزهايمر (AD), التصنيف المتعدد, هيستوجرام الاتجاهات (HOG), التحويل الشيرلي السريع (FFST), مصفوفة التوافق الرمادي (GLCM), خوارزميات التعلم الآلي, التعلم الانتقالي الاستقرائي, تحسين بايزي

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

ACRONYMS

AD Alzheimer's Disease. 6, 10–12, 15, 16, 19–21, 26, 42, 87

ADNI Alzheimer's Disease Neuroimaging Initiative Dataset. 25

BCFCM Bias-Corrected Fuzzy C-Means. 2, 42, 43, 61, 125

BO Baysian Optimization. 86, 89, 90

CNN Convolutional Neural Network. 19, 20

CoT Contourlet Transform. 25

CSF Cerebrospinal Fluid. 20, 25

CuT Curvelet Transform. 25

CWT Continuous Wavelet Transform. 25

DF Deep Features. 88

DFE Deep Features Extractions. 86

DT Decision Tree. 21

DTCWT Dual-Tree Complex Wavelet Transform. 25

DWT Discrete Wavelet Transform. 25

ELM Extreme Learning Machine. 25

EWT Empirical Wavelet Transform. 25

FC Fully Connected Layers. 91, 100

FCM Fuzzy C-Means Clustering. 21, 25

FD Frequency Domain. 20

FFST Fast Finite Shearlet Transform. 51, 62

GGWO Group Grey Wolf Optimization. 21

GLCM Gray Level Co-occurrence Matrix. 20, 21

GLRLM Gray Level Run Length Matrix. 21

GM Gray Matter. 20

GP Gaussian Process. 25, 90

GPR1 Gaussian Process Regression Model 1. 25

GPR2 Gaussian Process Regression Model 2. 25

HC Hippocampus. 26

HOG Histogram of Oriented Gradient. 20, 42

HPO Hyperparameter Tuning Optimization. 88

KNN k-Nearest Neighbors. 21

LBP Local Binary Pattern. 20

LDA Linear Discriminant Analysis. 25

MCI Mild Cognitive Impairment. 20, 25, 26

ML Machine Learning. 88, 90

MRI Magnetic Resonance Imaging. 14, 15, 20

NCSIN Nonsubsampled Contourlet Subband-based Individual Network. 26

NN Neural Network. 21, 26

OASIS Open Access Series of Imaging Studies Dataset. 25

PCA Principal Component Analysis. 25

PCC Posterior Cingulate Cortex. 21, 26

PHT Polar Harmonic Transforms. 27

PLS Partial Least Squares. 25

PNN Probabilistic Neural Network. 21

PSNR Peak Signal-to-Noise Ratio. 59

ROI Regions of Interest. 21, 26

SaDEWN Self-Adaptive Differential Evolution Wavelet Neural Network. 27

SD Spatial Domain. 20

SIFT Scale-Invariant Feature Transform. 20

SPM Statistical Parametric Mapping. 20

ST Shearlet Transform. 25, 26

SVM Support Vector Machine. 25

TPR True Positive Rate. 120

WHO World Health Organization. 1

WM White Matter. 20

WTEF Wavelet Transformation Energy Features. 26

ACRONYMS

INTRODUCTION

Cognitive decline, memory impairment, loss of thinking ability, and difficulties in performing daily activities are hallmark symptoms of dementia. Dementia is a condition characterized by a gradual deterioration in cognitive abilities. It serves as an umbrella term for several neurodegenerative diseases that affect neurons. The number of people living with dementia is expected to increase to 82 million by 2030 and 152 million by 2050, owing to an aging population. While some memory loss is expected as we age, the progressive memory degradation associated with dementia is not.

Alzheimer's disease is the most common type of dementia worldwide, accounting for 60-80% of cases. The World Health Organization (WHO) estimates that 44 million individuals have Alzheimer's disease, rising to 75 million by 2030 and 131.5 million by 2050 [20].

The rising incidence of Alzheimer's disease among elderly individuals has become a substantial burden on healthcare systems and socioeconomic. This condition frequently requires continuous nursing care, which incurs significant costs, in addition to expensive medications needed to manage symptoms and related needs. Furthermore, the psychological impact on family members is considerable. Alzheimer's disease disrupts the daily lives of patients and adversely affects their families, placing immense pressure on healthcare infrastructure.

Despite recent scientific advancements in various medical aspects of Alzheimer's disease treatment, a cure remains elusive, leading to profound despair among both family members and healthcare professionals. Consequently, there has been a shift in how we approach the disease. Current studies are focusing on individuals at risk of developing Alzheimer's disease (AD) due to the irreversible nature of the condition. The preclinical stage of the disease typically begins 15 to 20 years before the first symptoms manifest, highlighting the potential for preventive measures. While this approach is promising, it requires extensive monitoring and careful consideration, as it addresses future risks without providing solutions to the current challenges posed by the disease.

One of the significant difficulties in managing AD is accurately detecting and identifying each stage. As previously mentioned, AD is a progressive condition that is evaluated over time. It begins with a preclinical phase—a silent period—advancing to mild cognitive decline and culminating in clinically diagnosed Alzheimer's disease. The diagnostic process generally involves psychological evaluations, cognitive assessments, and imaging techniques (more details will be found in Chapter 1).

This thesis focuses on imaging methods, particularly MRI, which offers a robust visualization of the brain and facilitates the observation of structural changes associated with the various stages of the disease. In the early stages, an MRI scan of the brain may show

normal or resemble normal aging. As the disease progresses, MRI scans reveal atrophy in different brain regions, indicating a neurodegenerative process. This work addresses the challenge of identifying cognitive stages through MRI, where visual differentiation from normal aging can be difficult. Thus, this thesis aims to tackle this issue by proposing a computer-aided diagnostic system that leverages artificial intelligence technologies, precisely machine and deep learning.

With advancements in medical image acquisition and digitization, computer-aided diagnostic (CAD) systems have become essential as decision-support tools for medical professionals. Various state-of-the-art CAD systems have been proposed for AD detection, primarily based on machine learning algorithms or deep learning techniques.

In ML, handcrafted feature extraction and feature engineering are used to obtain relevant features. ML methods are typically categorized into supervised and unsupervised approaches. In supervised learning, the extracted features are fed into a classifier, where each scan or feature is related to its ground truth during the training process. Several well-established supervised learning models are commonly used, such as K-nearest neighbors, Support Vector Machines, Decision Trees, and Random Forests. Conversely, unsupervised learning algorithms are applied when ground truth labels are unavailable. These algorithms find patterns in the data by using mapping functions. This thesis employs the Bias-Corrected Fuzzy C-Means (BCFCM) algorithm as an unsupervised learning method for segmentation. Deep learning, a subfield of ML, is a powerful technique based on model architectures and large-scale data. However, the success of deep networks, mainly when training models from scratch, depends significantly on the amount and quality of the data samples across all disease stages. This poses a considerable obstacle in the medical domain, where access to data is frequently restricted.

This thesis encompasses machine learning (ML), considered a traditional learning strategy, and deep learning networks, the most recent learning approach. However, this work focuses not solely on the learning methods but also on the feature set, as features significantly impact the outcome of the learning algorithm. It is essential to highlight that these networks rely on discriminative features that facilitate classification through feature selection, feature mapping, dictionary learning, or convolutional neural networks (CNNs). Another important consideration is the combination of hyperparameters in the classifier, which maps samples from the feature space to the class label space. Therefore, within the scope of this thesis, we utilize handcrafted features, referred to as low-level features, instead of traditional methods. Specifically, we explore spatial and frequency domains to obtain these low-level features. Additionally, we expand to recent techniques, such as CNNs, which automatically extract high-level features.

Considering the intricate architecture of brain MRI images, specifically the hippocampal coronal scan, we represent the complexity of the MRI scan by proposing two low-level feature representations. The first approach is based on the spatial domain, where we utilize an unsupervised learning algorithm to obtain regions of interest. We then apply the Histogram of Oriented Gradients (HOG) methods. In this proposed algorithm within the spatial domain, we demonstrate that integrating unsupervised and supervised methods can yield promising results. For the second algorithm, we shift to the frequency domain, where we apply the Shearlet Transform, specifically the Fast Finite Shearlet Transform (FFST), which has not been extensively investigated in the literature for AD feature space representation. This approach summarizes the FFST coefficients using Gray Level Co-occurrence Matrix (GLCM) features. The main idea of this algorithm is to address each direction of the FFST separately and identify the direction that provides the best outcomes when coupled with machine learning algorithms. These algorithms are based on

low-level features; however, our research further explores high-level features by utilizing Convolutional Neural Networks (CNNs), specifically through inductive transfer learning. This approach applies a fine-tuning strategy to investigate ten different pre-trained CNN models. The objective was to identify the most reliable CNN architecture capable of automatically extracting useful high-level features. The primary disadvantage of traditional methods resides in the necessity of handcrafted features, which require domain expertise. To address this limitation, we extend our research by coupling the high-level features extracted by CNNs with machine learning algorithms. Moreover, instead of relying solely on automatically obtained features, we also adapt the hyperparameters of each machine learning algorithm automatically using a Bayesian optimization algorithm.

This research thesis addresses multiple research questions, leading to the proposal of more than one algorithm, each serving as a response to one or more of these questions. However, several challenges were encountered during this work, which can be summarized as follows:

- Addressing AD as a multi-class classification problem rather than a binary classification problem. Given that AD is a progressive disease with multiple stages, binary classification is not suitable for accurately capturing its complexity.
- Managing the significant resemblance between the beginning stages of the illness emerged as one of the primary challenges of this study.
- Identifying an appropriate approach to accurately and effectively classify the various stages of AD, with particular emphasis on the mild cognitive impairment stages.
- Enhancing classification performance by investigating features at different levels, ranging from low-level to high-level features.
- Addressing the limited number of samples in the dataset poses a significant constraint for applying deep learning algorithms.
- Mitigating the impact of an imbalanced dataset, especially for the mild cognitive decline class.
- Developing an optimal fine-tuning strategy to maximize the benefits of transfer learning.
- Employing effective data augmentation strategies to address the challenges posed by data limitations.
- Overcoming hardware limitations constrained the ability to fully implement and evaluate deep learning models.

This thesis addresses these challenges through innovative methodologies, feature exploration, advanced machine algorithms, and deep learning strategies. The challenges above highlight the complexity of tackling AD as a multi-class problem. These difficulties reveal a significant gap in the existing literature, motivating this work and shaping the direction of our contributions. By systematically addressing these challenges and literature gaps, this thesis not only bridges these gaps but also proposes various methodologies to overcome some of the challenges and insights. The key contributions of this research are:

1. Considering the limitations of the available databases, traditional methodologies were identified as the most suitable initial approach. Nonetheless, this research introduces two methodologies based on low-level features, which yield promising outcomes.

2. Low-level features were extracted from the spatial domain by combining an unsupervised learning algorithm with a computer vision algorithm. This resulted in a refined set of features capable of distinguishing all stages of AD.

3. The investigation was extended from the spatial domain to the frequency domain, specifically by applying the Fast Finite Shearlet Transform (FFST). This technique has not yet been explored in the literature for AD feature representation.

4. Various directional components of FFST were analyzed, with approximately 30 textural parameters computed to optimize the classification performance.

5. The most meaningful and impactful FFST directional coefficients for AD classification were identified, enhancing the discriminative power of the model.

6. Deep learning techniques, particularly inductive transfer learning, were utilized to classify multi-class AD.

7. Ten pre-trained CNN models were systematically evaluated to identify the most suitable architecture for achieving the research objectives.

8. The research expanded from low-level to high-level feature extraction, thereby addressing the limitations of traditional methods that require manual feature engineering.

9. Machine learning algorithms were optimized through Bayesian optimization to determine the most appropriate hyperparameters for various feature sets.

10. Two distinct databases with varying levels of complexity and class imbalances were employed to validate the proposed methods.

11. The issue of dataset imbalance was addressed by implementing and assessing the effectiveness of data augmentation techniques in the learning process.

To present the findings coherently, this thesis is organized as follows, with each chapter addressing one of the main contributions of this research:

- **Chapter 1:** This chapter introduces brain anatomy alongside the pathophysiology of Alzheimer's Disease (AD). It explains the changes that occur in the brain during the progression of AD and how these changes manifest in MRI images. Additionally, the chapter covers the MRI modality, including its various sequences and views, and provides an overview of the diagnostic process and stages of AD. The chapter concludes by emphasizing the importance of addressing this topic.

- **Chapter 2:** This chapter is dedicated to the state of the art and is divided into two main parts. The first part reviews research studies that utilize low-level features and machine learning (ML), subdivided into two sections: one focusing on studies that extract low-level features from the spatial domain and the other on studies using the frequency domain. The second part reviews literature employing deep learning techniques. The chapter concludes with a discussion of the gaps in the existing literature and an evaluation of the databases used, as these factors influenced the choice of databases for this research.

- **Chapter 3:** This chapter presents utilizing low-level features, consisting of two main parts. The first part introduces the proposed methodology incorporating BCFCM, HOG, and ML in the spatial domain. The second part discusses the methods based on the frequency domain, where the Fast Finite Shearlet Transform (FFST) is applied alongside GLCM for feature extraction. Implementation details, results, and a comparative discussion of these two methodologies are provided in this chapter.

- **Chapter 4:** This chapter focuses on deep learning algorithms used in the research. It includes a theoretical analysis of Convolutional Neural Networks (CNNs), their main components, and an introduction to the inductive transfer learning approach. Various pre-trained CNN models are evaluated using a layer-wise fine-tuning strategy, and pre-trained transformer models are applied for a deeper investigation. The chapter addresses the constraints posed by data limitations and explores data augmentation techniques utilized in the literature to mitigate overfitting.

- **Chapter 5:** In this chapter, the research extends to applying high-level features with optimized ML algorithms using Bayesian optimization. It investigates the high-level features extracted by the CNN models introduced in Chapter 4 and explores various aspects of using these models to enhance classification outcomes. The focus is on overcoming challenges related to limited data and the labor-intensive process of handcrafted feature extraction.

CHAPTER 1

ALZHEIMER'S DISEASE

1.1 Overview

Alzheimer's Disease (AD) represents the most widespread form of dementia among older people, affecting nearly 44 million individuals worldwide, with 7.7 million new cases each year [20].

The growing healthcare concern of AD extends beyond being a significant social and economic issue; it devastates the affected individuals and their family members and caregivers. The memory loss caused by AD is currently incurable. However, understanding the anatomical and physiological aspects of the brain and the pathology of the disease aids researchers in providing better diagnosis and treatment options.

This chapter delineates the anatomy and physiology of the brain to explain the changes that occur during AD. We will coordinate the disease's progression with the symptoms that appear and illustrate how AD manifests in MRI images. Additionally, we briefly introduce MRI technology and discuss using neuroimaging combined with AI to enhance accurate disease stage identification and early detection.

1.2 Humain Brain

1.2.1 Brain Anatomy

The brain is the most enigmatic component of the human body, as it controls everything. It primarily consists of three parts: the cerebrum comprises gray matter (the cerebral cortex) and white matter at its center, the brainstem (middle of brain) connects the cerebrum with the spinal cord, and the cerebellum ("little brain") is a fist-sized portion of the brain located at the back of the head, below the temporal and occipital lobes and above the brainstem [21]. see figure 1.1.

To discuss the impact of AD on the brain's anatomical structure later, we will delve further into the cerebrum, which forms the significant portion of the brain as shown in figure 1.1,

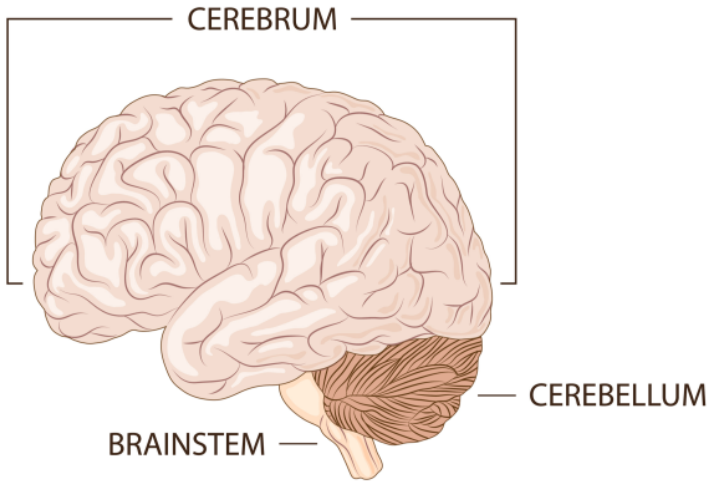


Figure 1.1: The three main part of the brain [1]

The cerebrum is divided into two major parts: the right and left cerebral hemispheres. Those two hemispheres are separated by a fissure (larger grooves) called the tremendous longitudinal fissure; both hemispheres are connected at the bottom by the corpus callosum that serves to deliver the messages between the hemispheres. See figure 1.2. The surface of the cerebrum contains billions of neurons and glia that together form the cerebral cortex, where the cerebral cortex is the outermost layer of the brain, characterized by its grayish-brown appearance. It is responsible for higher brain functions, including sensory perception, cognition, generation of motor commands, spatial reasoning, and language. This region is densely packed with neurons and is divided into several areas by distinct fissures into frontal, temporal, parietal, and occipital lobes, each specialized for different functions, and they interact through complex relationships. [22] Signals in the brain travel along pathways between gyri, lobes, hemispheres, deep brain structures like the thalamus, and other central nervous system regions, with the help of neurons and synapses, which will be covered next.

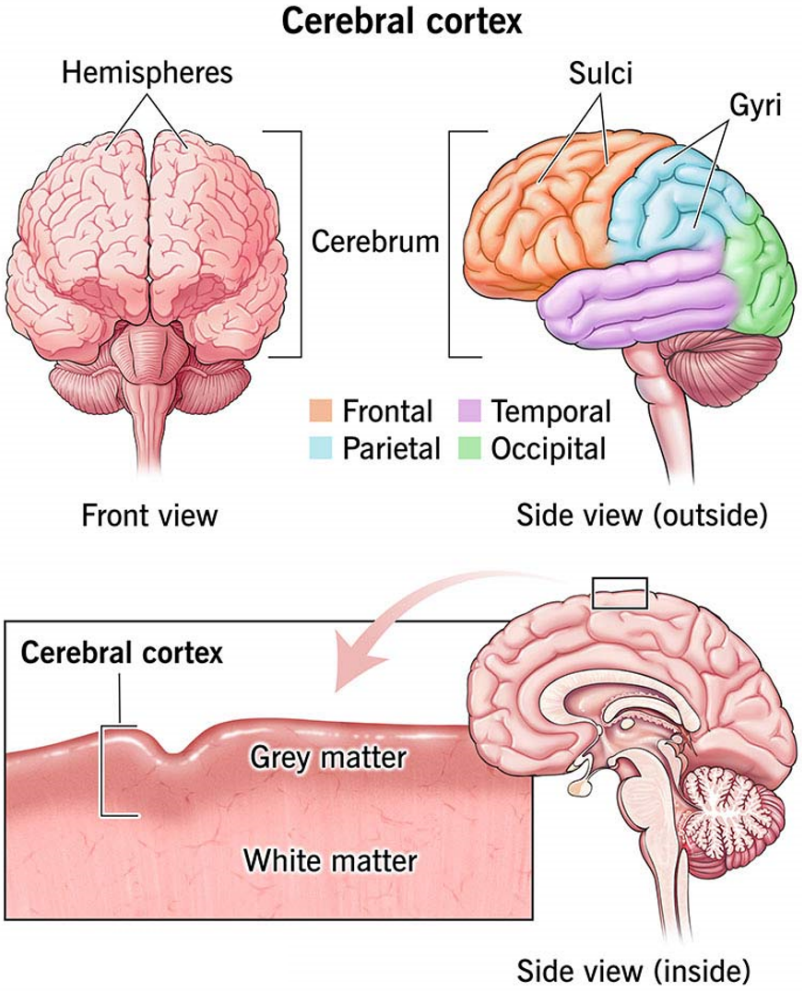


Figure 1.2: Brain cerebral cortex [2]

1.2.2 Neuron and synapses

Neurons are the fundamental units of the brain and nervous system, intricately designed to process and transmit information. Each neuron comprises a cell body (soma), which houses the nucleus and essential organelles, dendrites that branch out to receive signals from other neurons, and a single axon that extends to transmit electrical impulses away from the cell body. At the end of the axon are axon terminals, where neurotransmitters are stored in vesicles. These neurotransmitters are released into the synaptic cleft, a minute gap between neurons, facilitating communication with the adjacent postsynaptic neuron.[22] (see figure 1.3)

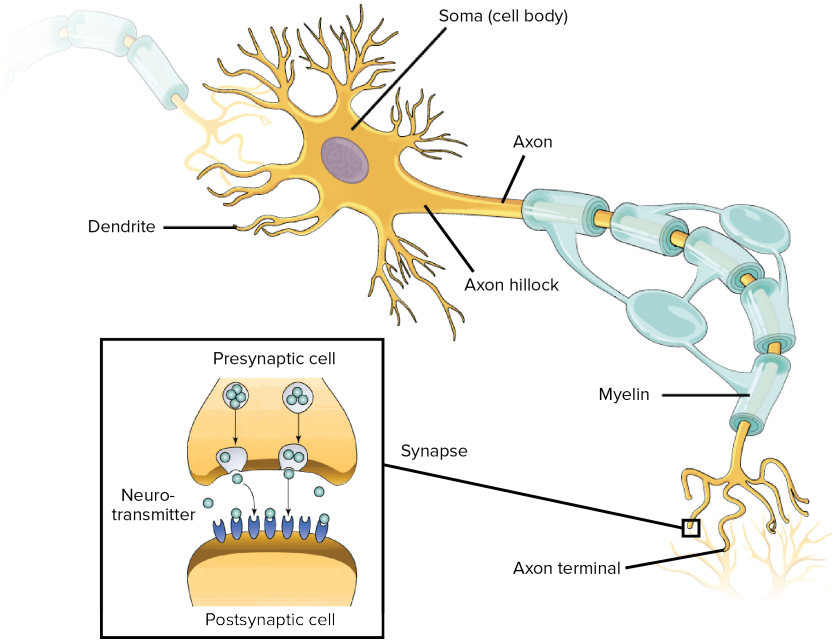


Figure 1.3: Neurons and synapse structure [3]

Neurons can be classified into three types based on their function: sensory neurons that carry information from sensory receptors to the central nervous system, motor neurons that convey commands from the central nervous system to muscles and glands, and interneurons that connect neurons within the central nervous system to integrate and process information.(see figure 1.4)

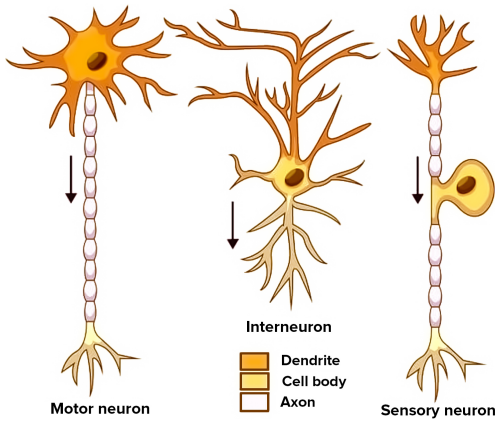


Figure 1.4: Types of neurons [4]

Neuronal communication is a complex electrochemical process. An action potential, or electrical impulse, travels down the axon to the axon terminals. This triggers the release of neurotransmitters into the synaptic cleft. These chemical messengers then bind to specific receptors on the postsynaptic neuron, initiating a new electrical signal in that neuron. This process of synaptic transmission is critical for all brain functions, including memory, learning, and decision-making. Synapses, the junctions where neurons communicate, are

vital for maintaining the brain's intricate neural networks and enabling neuroplasticity, the brain's ability to reorganize itself by forming new neural connections throughout life.

In the context of AD, the normal functioning of neurons and synapses is severely disrupted. The accumulation of amyloid-beta plaques between neurons and the formation of neurofibrillary tangles within neurons impede synaptic communication and lead to the degeneration and death of neurons. These processes will be elaborated upon in the following section, with a more detailed explanation of the amyloid plaques and neurofibrillary tangles.

This loss of synapses and neurons is particularly pronounced in areas of the brain associated with memory and cognition, such as the hippocampus and cortex, leading to the characteristic symptoms of AD.

1.2.3 Brain Function

The brain is an intricately complex organ that performs numerous functions essential to our daily lives. It orchestrates cognitive processes such as memory, learning, attention, and executive functions. The cerebral cortex is crucial for these higher-order functions. Different regions of the cortex are specialized for various tasks: the frontal lobe is involved in executive functions, decision-making, problem-solving, and controlling behavior and emotions; the parietal lobe processes sensory information and is crucial for spatial orientation and navigation; the occipital lobe is dedicated to visual processing; and the temporal lobe is key for auditory processing and memory formation.

The hippocampus, located within the temporal lobe, is central to forming, organizing, and storing new memories. It is also critical for spatial memory and navigation. Damage to the hippocampus, as seen in Alzheimer's disease, results in severe impairments in forming new memories while leaving old memories initially less affected. The amygdala, adjacent to the hippocampus, processes emotions and forms emotionally charged memories.

Neurons and their synapses are fundamental to brain function, facilitating communication within neural networks. The prefrontal cortex, part of the frontal lobe, integrates information from various brain parts to plan, make decisions, and inhibit inappropriate behaviors. It is also involved in working memory, the ability to hold and manipulate information over short periods.

Though not directly involved in higher cognitive functions, the brainstem and cerebellum are crucial for maintaining vital bodily functions and coordinating movement. The brainstem regulates breathing, heart rate, and sleep cycles, while the cerebellum fine-tunes motor activity and is involved in balance and coordination.

The brain's function is a harmonious interplay of its various parts, enabling complex behaviors, thoughts, and emotions. Understanding these functions is essential for appreciating how disruptions, such as those caused by Alzheimer's disease, can profoundly affect an individual's cognitive and physical abilities. Alzheimer's disease particularly impacts areas like the hippocampus and cerebral cortex, leading to memory loss, confusion, and difficulty with reasoning and judgment. This highlights the importance of these regions in maintaining cognitive health.

1.3 Pathophysiology of Alzheimer's disease

1.3.1 Alzheimer's Disease

AD is a progressive neurodegenerative disorder that primarily affects older adults, leading to severe cognitive decline and impairments in daily functioning. First identified by Dr. Alois Alzheimer in 1906 [23], this disease is characterized by the accumulation of amyloid-beta plaques and neurofibrillary tangles in the brain, which disrupt normal neuronal function and ultimately results in neuronal death. A gradual deterioration of memory, reasoning, language, and other cognitive abilities marks AD. The progression of the disease varies among individuals but typically spans several years, culminating in a profound loss of independence and requiring extensive caregiving.[24]

The impact of AD extends beyond the affected individuals to their families, caregivers, and society. It poses a significant emotional, physical, and financial burden on caregivers and healthcare systems worldwide. As the global population ages, the prevalence of AD is expected to rise, making it a critical public health issue. Despite extensive research, the exact cause of Alzheimer's disease remains elusive, with genetic, environmental, and lifestyle factors all playing a role in its development. Current treatments primarily focus on managing symptoms, as no cure is available.

Research efforts are ongoing to understand AD pathophysiology better and develop effective therapies. Advances in neuroimaging, biomarker identification, and genetic studies are shedding light on the underlying mechanisms of the disease, offering hope for early diagnosis and targeted treatments. The significance of early detection and accurate identification of the stages in potentially delaying the onset or progression of AD is also gaining recognition.

1.3.2 Amyloid Plaques and Neurofibrillary Tangles

Amyloid plaques and neurofibrillary tangles are the hallmark pathological features of the disease, playing a crucial role in the neurodegenerative processes that characterize this condition. Amyloid plaques are extracellular deposits primarily composed of amyloid-beta (Aβ) peptides. [25]

These peptides are produced through the cleavage of amyloid precursor protein (APP) by the enzymes beta-secretase and gamma-secretase. In Alzheimer's Disease, an imbalance between the production and clearance of Aβ leads to accumulating these peptides, which aggregate to form insoluble plaques in the brain's extracellular spaces. These plaques disrupt cell-to-cell communication, trigger inflammatory responses, and induce oxidative stress, contributing to neuronal injury and death.[25]

Neurofibrillary tangles, on the other hand, are intracellular aggregates composed of hyperphosphorylated tau protein. Under normal conditions, tau stabilizes microtubules, which are essential for maintaining the structure and function of neurons. In Alzheimer's Disease, abnormal chemical changes cause tau to detach from microtubules and aggregate into twisted threads known as tangles. This process destabilizes the microtubules, impairing nutrient and organelle transport within neurons and leading to cell death [26] [27]. The presence of these tangles correlates with the severity of dementia, reflecting the extent of neuronal loss and cognitive decline. Figure 1.5 shows how the plaque and neurofibrillary tangles appear under the microscope.

The formation and accumulation of amyloid plaques and neurofibrillary tangles are interconnected processes that exacerbate each other. Amyloid-beta deposition can promote

tau pathology, and together, they disrupt neural networks and synaptic function, leading to the hallmark symptoms of Alzheimer's Disease, such as memory loss, confusion, and impaired cognitive abilities. Research into the molecular mechanisms underlying plaque and tangle formation is critical for developing targeted therapies to prevent or slow AD progression. Understanding these pathological features provides valuable insights into the disease's progression and highlights potential therapeutic targets for intervention.

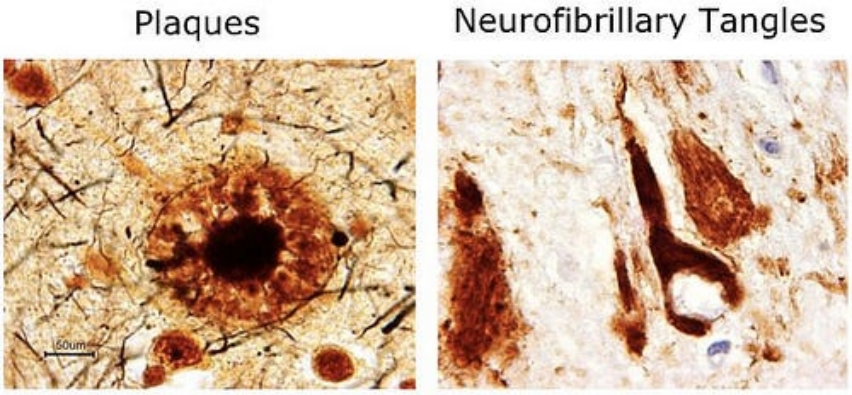


Figure 1.5: A microscope images of amyloid plaques and the neurofibrillary tangles [5]

1.3.3 Neurodegeneration and Brain Atrophy

Neurodegeneration and brain atrophy are central features of AD, resulting from the progressive loss of neurons and synapses, particularly in regions critical for memory and cognition. The degeneration begins at the cellular level, primarily involving the abnormal accumulation of amyloid-beta plaques and neurofibrillary tangles composed of tau protein.

The neurodegenerative process in AD typically begins in the hippocampus, a region critical for memory formation and spatial navigation. Memory impairment and disorientation are among the earliest symptoms observed as neurons in the hippocampus deteriorate due to plaque and tangle accumulation. Patients may experience difficulties forming new memories, often repeating themselves or forgetting recent events. As the pathology progresses, it spreads to the entorhinal cortex and other parts of the cerebral cortex, affecting areas responsible for higher cognitive functions such as language, reasoning, and behavior. This spread corresponds with additional symptoms, including language difficulties, impaired judgment, and changes in personality and behavior.

Neurodegeneration in AD affects both gray matter and white matter. Gray matter, which contains neuronal cell bodies and synapses, undergoes significant atrophy due to the loss of neurons and synaptic connections. This reduces cortical thickness and brain volume, manifesting clinically as more pronounced cognitive deficits and functional impairments. White matter, composed of myelinated axons that facilitate communication between different brain regions, also suffers from the disease's progression. White matter degeneration results from axonal damage and loss, disrupting the brain's connectivity and overall network integrity, leading to further cognitive decline and difficulties in coordination and motor functions.

Anatomically, brain atrophy in AD begins with the shrinkage of the hippocampus and the medial temporal lobe. As the disease progresses, atrophy spreads to the parietal and frontal lobes, eventually affecting almost the entire cortex.

The ventricles, fluid-filled cavities within the brain, enlarge as surrounding brain tissue atrophies. This overall reduction in brain volume is evident in advanced stages of AD and profoundly impacts cognitive and physical functions. Patients in advanced stages often require full-time care as they lose the ability to perform basic activities of daily living, experience severe memory loss, and have significant communication difficulties.

The anatomical and functional deterioration associated with Alzheimer's Disease (AD) arises from a sequence of pathological processes triggered by the accumulation of amyloid-beta and tau pathology, culminating in extensive neurodegeneration within the brain atrophy. The clinical symptoms align closely with these brain changes, starting with memory issues and progressing to global cognitive and functional impairments.

1.4 Diagnosis of Alzheimer's Disease

1.4.1 Clinical Assessment

Cognitive and neuropsychological testing are critical components of the clinical assessment for Alzheimer's Disease (AD). These tests systematically evaluate various cognitive domains to identify patterns of impairment consistent with AD [28]. Below are some of the commonly used tests and what they measure:

- **Mini-Mental State Examination (MMSE):** The MMSE is a brief 30-point questionnaire widely used for screening cognitive function. It assesses orientation to time and place, immediate and delayed recall, attention and calculation, language abilities, and visuospatial skills. Scores range from 0 to 30, with lower scores indicating greater cognitive impairment. A score below 24 often suggests cognitive impairment.[29]

- **Montreal Cognitive Assessment (MoCA):** The MoCA is a 30-point test designed to detect mild cognitive impairment (MCI) and early AD. This assessment encompasses tasks related to attention and concentration, executive functions, memory, language, visuospatial skills, conceptual thinking, mathematical calculations, and orientation. A score of 26 or higher is regarded as normal. The MoCA demonstrates greater sensitivity than the MMSE in identifying early cognitive impairment.[30]

- **Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog):** The ADAS-Cog is a more comprehensive tool specifically used to measure the severity of cognitive symptoms in AD. It evaluates memory, language, praxis, and orientation through word recall, naming objects, following commands, and copying geometric figures. The total score ranges from 0 to 70, with higher scores indicating more severe cognitive impairment. [31][32]

- **Neuropsychological Battery:** A detailed battery of tests that provides an in-depth assessment that covers a wide range of cognitive abilities like memory, language, executive function, attention and processing speed, and visuospatial skills.[33] [34]

- **Clock Drawing Test:** A quick screening tool to assess visuospatial abilities, executive function, and attention. Performance is evaluated based on the accuracy and completeness of the clock drawing.[35][36]

- **Verbal Fluency Tests:** These tests measure language and executive function. Patients are asked to name as many words as possible in one minute that start with a specific letter (phonemic fluency) or belong to a certain category (semantic fluency). The number of correct words produced is counted, with lower scores indicating potential impairment.[37][38]

- **Trail Making Test (TMT)**: TMT assesses cognitive flexibility, visual attention, and task switching. Has two parts first part Connect numbered circles sequentially the second Alternate between numbers and letters Time taken to complete the tasks is measured, with longer times indicating impairment.[39]

- **Stroop Test:** Evaluates processing speed, attention, and cognitive control. Patients must name the ink color in which words are printed, which can match or conflict with the color name (e.g., the word "RED" printed in blue ink). The number of correct responses and the time taken are recorded.[40] [41]

These tests help clinicians identify specific cognitive deficits and differentiate AD from other forms of dementia or mental impairments. The pattern of results across these tests can provide a detailed cognitive profile, guiding further diagnostic steps and interventions.

1.4.2 Neuroimaging Techniques

Although clinical assessment is typically the initial step in evaluating neurodegenerative diseases, it is essential to emphasize the significance of neuroimaging techniques in visualizing internal structures within the human body. Neuroimaging offers a more comprehensive perspective, confirming the presence of diseases, assessing disease progression, and understanding anatomical and physiological changes in the brain. Two categories of imaging techniques exist: structural imaging (MRI, CT, and DTI) and functional imaging (such as PET, fMRI, and SPECT). Alongside imaging, additional biomarkers utilized in AD diagnosis include Cerebrospinal Fluid (CSF) Markers and Blood-Based Markers. Since we focus on using MRI images for AD disease detection and classification in this work, we will discuss the MRI image in the following subsection.

1.4.2.1 MRI in Alzheimer's Disease: Brief Overview

Magnetic Resonance Imaging (MRI) is a sophisticated imaging modality that employs powerful magnetic fields, radiofrequency waves, and gradient fields to produce highly detailed representations of the body's internal structures, including exceptionally soft tissues like the brain. This non-invasive method operates on the principles of nuclear magnetic resonance. When a patient is positioned within a powerful magnetic field, the protons within their body, primarily found in hydrogen atoms, align themselves with the field. A radiofrequency pulse is then administered, temporarily disrupting this alignment. As the protons return to their original position, they emit radiofrequency signals. These signals are detected by the MRI scanner and processed by a computer to produce comprehensive images of the body's internal structures.

Two crucial parameters in MRI are the Echo Time (TE) and the Repetition Time (TR). TE denotes the time elapsed between the delivery of the radiofrequency pulse and the peak of the induced signal in the coil. This parameter influences the image contrast. On the other hand, TR represents the time interval between successive radiofrequency pulses administered to the same slice, which affects the relaxation of protons and, consequently,

the image contrast. By adjusting TE and TR, MRI is capable of generating images with varying levels of contrast, including:

1. **T1-Weighted Imaging:** Provides high-resolution images of brain anatomy produced using short TE and TR times where the White matter appears bright, gray matter appears darker, and cerebrospinal fluid (CSF) is dark.

2. **T2-Weighted Imaging:** Highlights differences between gray matter, white matter, and CSF produced using longer TE and TR times. CSF appears bright, white matter is dark, and gray matter is lighter than white matter.

3. **Fluid-Attenuated Inversion Recovery (FLAIR):** Suppresses the CSF signal to better visualize lesions or abnormalities near CSF. CSF appears dark; lesions appear bright.

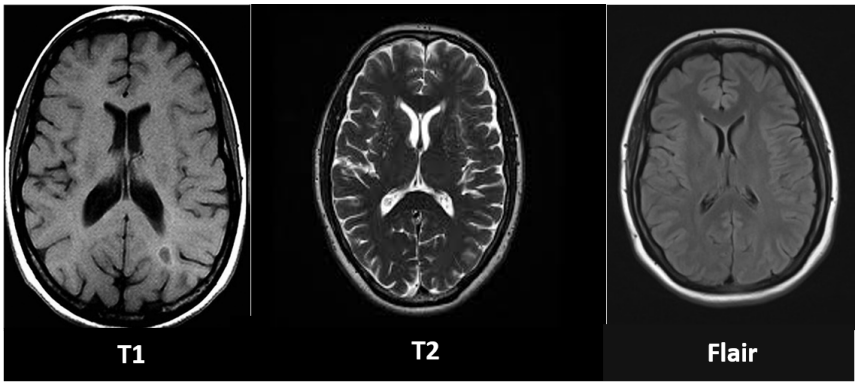


Figure 1.6: Sequential brain [6]

After discussing the various MRI sequences and their specific applications in visualizing brain structures, it is essential to consider the orientation of these images. MRI scans are acquired in different planes to provide a thorough and detailed view of the brain. These planes include axial, coronal, and sagittal.

The axial scan represents an XY plane that slices parallel to the ground, dividing the brain into upper and lower halves. It is commonly employed to assess overall brain structure and detect cortical and hippocampal atrophy.

Conversely, the coronal scan represents an XZ plane that slices from the front to the back. It separates the brain into anterior and posterior halves and is particularly valuable for visualizing the hippocampus and medial temporal lobes.

Lastly, the sagittal plane represents a YZ plane that slices parallel to the midline, dividing the brain into left and right halves. It helps examine the corpus callosum and other midline structures.

Each plane offers unique perspectives indispensable for a comprehensive brain anatomy and pathology assessment.

In this study, we utilized axial and coronal views. While both coronal and axial scans provide valuable information for detecting AD, coronal scans are often considered more suitable for this purpose.

Coronal scans provide a more detailed and clear view of the medial temporal lobes, including the hippocampus and entorhinal cortex, among the first regions to exhibit atrophy in AD. This orientation enables precise measurement of hippocampal volume and assessment of early neurodegenerative changes critical for early diagnosis.

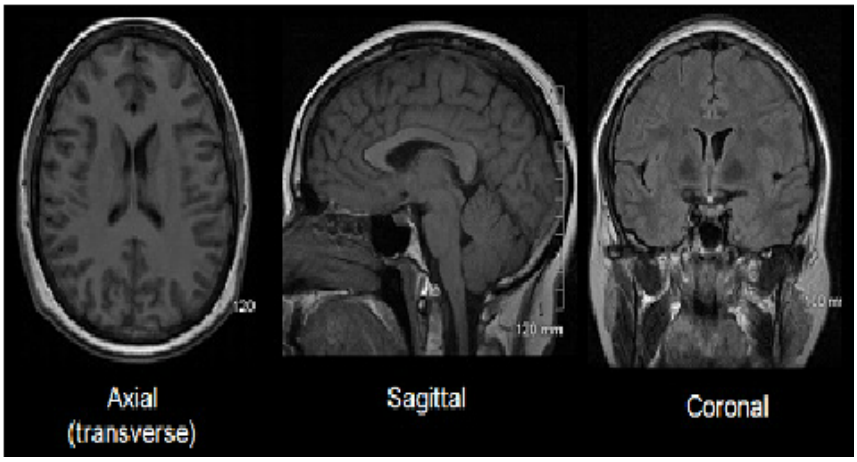


Figure 1.7: Different planes of MRI brain image [7]

1.5 Alzheimer's disease stages

AD progresses through several stages, each characterized by distinct cognitive impairments and corresponding changes in the brain that can be visualized using MRI scans.

Early Stage (Preclinical AD and Mild Cognitive Impairment - MCI): In the early stages, individuals may experience subtle memory lapses and difficulty in word-finding or problem-solving. These symptoms are often attributed to normal aging but represent the onset of Mild Cognitive Impairment (MCI). On MRI, this stage is marked by mild atrophy in the hippocampus and entorhinal cortex, which is critical for memory formation. The hippocampus begins to shrink, but the overall brain structure remains relatively intact.

Early MCI (EMCI): As MCI progresses to Early MCI, cognitive impairments become more noticeable, especially in memory and executive functions. MRI scans reveal more pronounced hippocampal atrophy and early signs of cortical thinning in the temporal lobes. The lateral ventricles may start to enlarge slightly due to the loss of surrounding brain tissue.

Late MCI (LMCI): Late MCI is characterized by more significant memory loss, confusion, and difficulty in performing complex tasks. MRI scans show increased atrophy in the hippocampus, parahippocampal gyrus, and medial temporal lobes at this stage. There is also noticeable cortical thinning in the parietal and frontal lobes, indicating the spread of neurodegenerative changes. The enlargement of the lateral and third ventricles becomes more evident, reflecting the ongoing loss of brain volume.

Moderate Stage: As AD progresses to the moderate stage, individuals experience more severe memory impairment, confusion, and difficulty recognizing familiar faces and places. Behavioral and personality changes, such as increased irritability and anxiety, may also occur. MRI scans reveal widespread atrophy across the cortex, with significant shrinkage in the temporal and parietal lobes. The hippocampus and surrounding medial temporal structures show extensive atrophy. Cortical thinning is prominent, and the ventricles are markedly enlarged.

Severe Stage: In the severe stage of AD, cognitive and functional abilities continue to decline. Individuals become unable to communicate effectively, require assistance with daily activities, and may lose the ability to walk or sit independently. MRI scans at this stage show severe global brain atrophy. The cortex is extensively thinned, especially

in the temporal, parietal, and frontal lobes. The hippocampus is severely atrophied, significantly reducing the brain's overall volume. The ventricles are greatly enlarged, and sulcal widening is pronounced, reflecting the extensive loss of brain tissue.

The MRI findings correlate closely with the clinical symptoms throughout these stages, visually representing the disease's progression. By tracking changes in specific brain regions over time, MRI scans play a crucial role in diagnosing AD, monitoring its progression, and evaluating the effectiveness of potential treatments.

1.6 Artificial Intelligence and Computer-Aided Diagnosis

Alzheimer's Disease (AD) profoundly impacts individuals, families, and society. Personally, AD leads to progressive memory loss, impaired judgment, behavioral changes, and loss of independence. Families face emotional strain, caregiving burdens, and significant financial pressures. Societally, AD places a heavy burden on healthcare systems and presents an important public health challenge as the aging population increases. Economically, the direct costs of medical care, long-term support, and indirect costs of lost productivity and caregiving are substantial. These multifaceted impacts highlight the urgent need for continued research, improved care strategies, and comprehensive support for those affected by AD.

AD's increasing prevalence and impact have spurred researchers to explore AI and computer-aided diagnosis systems, leveraging machine learning and deep learning technologies. These systems analyze neuroimaging data, such as MRI scans, to detect early signs of AD, classify disease stages, and monitor progression. AI algorithms excel in identifying minute structural changes in the brain that are often missed by traditional methods, enhancing the accuracy of early detection and stage classification. This capability is crucial for initiating timely interventions that can slow disease progression.

Searching on Google Scholar under the term "Alzheimer's disease" revealed 2050000 research papers and books since the discovery of the disease. However, numerous and extensive research has been conducted over the past few years, as shown in figure 1.8. Early studies focused on using machine learning models to analyze structural MRI data, revealing the potential of automated approaches in identifying different stages of AD depending on the brain structure atrophy and other AD markers. As deep learning techniques evolved, researchers developed convolutional neural networks (CNNs) capable of processing vast amounts of imaging data with remarkable precision.

Despite their promise, these technologies face challenges, including the need for extensive investigation, diverse datasets for training, and ensuring the reliability and interpretability of AI-driven insights. Nonetheless, AI's potential benefits in improving diagnostic precision and patient outcomes make it a compelling area of ongoing research and development in the fight against Alzheimer's Disease.

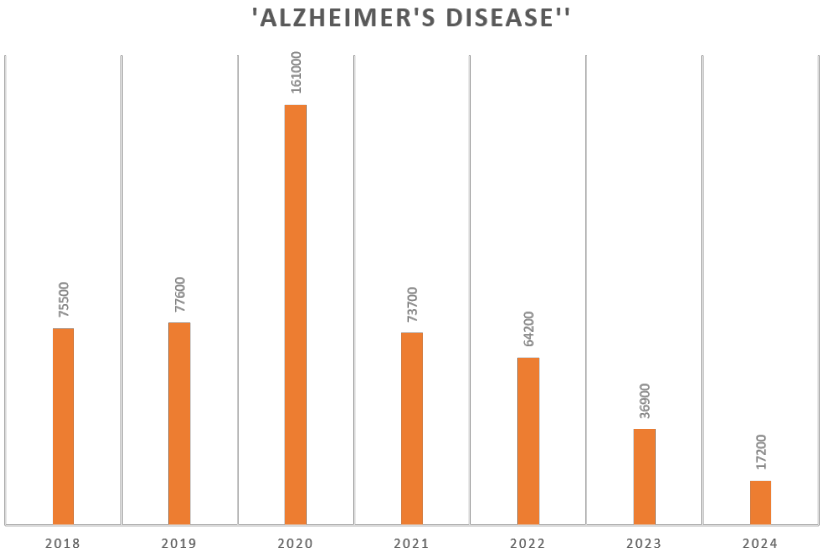


Figure 1.8: Number of the research that includes the word Alzheimer's disease over the years

1.7 Summary

In this chapter, we delved into the anatomy and physiology of the brain, emphasizing the cerebral cortex and neurons.

Our exploration begins by examining AD's pathophysiology and its consequential effects on the brain. We then discuss the stages and diagnosis of AD in depth. Furthermore, we thoroughly analyze the capabilities and limitations of AI and CAD systems. In the subsequent chapter, we will present an updated overview, referencing influential studies in AD stage classification and identification that have significantly influenced our research.

CHAPTER 2

LITTERATURE REVIEW

2.1 Overview

This chapter systematically reviews the current state-of-the-art classification in AD. We aim to provide an unbiased comparison of methods that discriminate and classify AD stages. We analyze retrieval approaches that utilize low-level and high-level features, specifically emphasizing commonly used methods. Furthermore, we investigate using spatial and frequency domains to extract distinguishing features. We also explore machine learning algorithms and deep learning techniques for extracting high-level features. Additionally, we delve into transfer learning applications. Lastly, we give special attention to the deep feature works that form the foundation of this thesis and briefly address existing research gaps to introduce novel methods for enhancing performance.

Highlight

- We present a review of work that used low and high-level features for Alzheimer's Disease stage classification
- We provide the state-of-the-art classification of (AD) based on spatial and frequency domain
- we include a literature review of machine learning, deep learning, and transfer learning

This chapter will present an overview of various methodologies employed in the field, which utilize either localized brain regions or the entire brain to address classification challenges. Additionally, we present a comparative analysis of traditional approaches relying on engineered features with classifiers alongside advanced methods leveraging Convolutional Neural Network (CNN), offering a comprehensive benchmark of their performance.

2.2 Literature review

Medical image classification is crucial for various medical applications, including prognostics, diagnostics, and identification. The development of automatic and efficient classi-

fication algorithms has been a focal point in AD stage classification for several decades. However, the efficacy of these algorithms heavily depends on the learned features and the quantity and quality of the data. Before the recent emergence of deep learning methods, medical image classification underwent significant developments, which can be broadly categorized.

In this work, we categorize classification methodologies into two groups based on the nature of features fed into the algorithm. The first group comprises low-level feature-extraction techniques that rely on texture, color, intensity, shape, gradient, histogram, and frequency-based features. Many popular works have used these features to classify Alzheimer's disease, achieving remarkable results.

In recent years, there has been a shift towards using high-level features learned automatically using deep learning techniques by applying CNN trained on large datasets to extract those relevant features, which will be considered as the second methodological group.

Due to the broad spectrum of AD in medical imaging classification, this section presents an exhaustive survey of the current literature on AD classification using 2D MRI images for both categories.

2.2.1 Traditional-Based methods: Low-level

Low-level feature extraction is crucial in image processing when analyzing and interpreting visual data, such as MRI images. These features contain essential information about the structure and content of the images, particularly in cases of AD. By extracting and quantifying these characteristics, researchers can obtain valuable insights into image content, enabling disease stage classification.

Low-level features can be obtained from two distinct domains, the Spatial Domain (SD) and the Frequency Domain (FD).

2.2.1.1 Spacial Domain Features

The spatial domain SD aims to examine pixel relationships within the image. Various researchers extracted textural features from this domain to perform binary and multi-classification. Altaf et al. [42] developed a hybrid feature-based classification method for accurately detecting and classifying AD and Mild Cognitive Impairment (MCI) using a combination of texture-based features extracted from MR images and clinical data. The bag of visual words methodology was used to augment the efficacy of texture-based features in Alzheimer's disease classification by clustering important points derived from pictures with methods such as Gray Level Co-occurrence Matrix (GLCM), Scale-Invariant Feature Transform (SIFT), Histogram of Oriented Gradient (HOG), and Local Binary Pattern (LBP). The authors include preprocessingng by utilizing the gradwarp Correction to improve image appearance and B1 Non-uniformity Correction to correct image color and intensity information. They also apply N3 Bias Field Correction to eliminate the dielectric effects during image acquisition. All those steps are essential for improving the non-uniform gradient in the image. Next, brain texture Segmentation was performed using the segmentation module of Statistical Parametric Mapping (SPM) to extract Gray Matter (GM), White Matter (WM), and Cerebrospinal Fluid (CSF) regions. They achieved the highest accuracy, equal to 79.8% when using the Whole image and the clinical information. Sampath et al. [43] also utilized textural features however, they utilized the Lucy-Richardson approach to eliminate Gaussian noise and deblurring the image, and

then they applied the prolonged adaptive exclusive analytical Atlas (paea) with the Gaussian mixture model (gmm) to segment the brain before extracting GLCM features and selected the important one by mutual information, pointwise mutual information, and Pearson product-moment correlation coefficient values. They achieved an accuracy of 99.26%.

Baskar et al. [44] propose a multimodal classification framework to discriminate patients with AD. combining supervised Fuzzy C-Means Clustering (FCM) and unsupervised learning techniques (Probabilistic Neural Network (PNN)). Additionally, the study focuses on normalizing brain images, extracting Regions of Interest (ROI) such as the hippocampus and Posterior Cingulate Cortex (PCC), then analyzing a comprehensive set of features here in the feature extracting process, the authors extract 457 features, including 447 texture measures and 10 shape measures in three brain planes (axial, sagittal, and coronal). A multiple-criterion feature selection approach was employed to select about 19 highly relevant AD-related features from the extracted feature set, addressing dimensionality issues and improving the discrimination of subjects with AD, achieving an accuracy of (98.63%) for AD vs NC, (95.4%)for MCI vs NC, and (96.4%) for AD vs MCI. Shankar et al.[45] aims to develop an effective model for AD detection using brain MRI data. They extracted the GLCM and the Gray Level Run Length Matrix (GLRLM) features and the histogram features. With a total of 35 features, the authors utilize Group Grey Wolf Optimization (GGWO) for feature selection, followed by classification using diverse classifiers such as Decision Tree (DT), Neural Network (NN), and k-Nearest Neighbors (KNN) to accurately classify MR brain images as AD, MCI, or NC. This study reported an accuracy of 84.5% using the KNN and an accuracy of 88.52% when using DT. Vaithinathan et al.[46] aims to develop a technique for extracting 2D textures from 3D MRI scans, particularly T1 weighted MRI images, to aid AD classification—the proposed method analyses textures at multiple axes from individual slices with different neighbor region sizes. Extracting and refining these textures using various feature selection techniques aims to address the challenge of high-dimensional data and improve the accuracy of AD classification. In their work, the authors select specific image blocks to compute textures for these blocks on individual slices and refine the features using various feature selection techniques. Classification algorithms are then applied, aiming to improve accuracy and efficiency in disease classification. The authors in this work perform a 2-way classification, achieving an accuracy of 87.39% for AD vs NC and 63.16% for AD vs MCI, and using linear SVM for NC vs MCI, the best results were 63.73% and for cMCI vs MCInc the best score was achieved by RF with an accuracy of 59.48%. Uysal et al. [47] analyze the hippocampal region for the early detection of Alzheimer's Disease (AD). They employ machine learning techniques and assess hippocampal volume, measured using a semi-automatic segmentation software called ITK-SNAP. This software allows for the calculation of the volumes of the left and right hippocampus separately, facilitating the creation of three-dimensional models of the hippocampus. This step aims to detect hippocampal atrophy. In addition to using hippocampal volume as the primary biomarker in their study, the authors also incorporate demographic information such as age and gender, enhancing their findings' overall accuracy. The authors employ several ML methods, with the Gaussian Naive Bayes (GNB) model as the best classifier, achieving an accuracy of 82%. Zaina et al. [48] propose using exemplar pyramid techniques combined with bi-linear interpolation to enhance the quality of MRI scans. Feature extraction is subsequently performed using Gray-Level Co-occurrence Matrix (glcm) and Local Binary Patterns (lbp) from the gray matter (GM) and white matter (WM) regions of interest. The extracted features are concatenated, resulting in a feature set of over 109,628 dimensions. These features are then input to a

Multilayer Perceptron (MLP) for classification. The classification process employs binary (2-way) and ternary (3-way) strategies. The dataset used in this study includes 31preprocesseded T1-weighted structural MRI scans obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The proposed method achieved an accuracy of **89.80%** for multiclass classification across four categories: Cognitively Normal (CN), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), and Alzheimer's Disease (AD).

Salunkhe et al.[49] utilize texture analysis to examine subtle changes in the hippocampus caused by Alzheimer's disease. This involves extracting statistical properties from MRI images to identify neurodegenerative changes that may not be visible through traditional methods. The study identifies and utilizes important features derived from MRI images' Gray Level Co-occurrence Matrix (GLCM). Not all features are necessary for classification, and a subset of eight crucial features was determined to contribute significantly to the classification accuracy. The top-performing algorithms mentioned in the study include Ensemble methods and Decision Trees, achieving classification accuracies of 90.2% and 88.5%, respectively, along with Support Vector Machines (SVM), which achieved an accuracy of 87.2%. The research focuses on extracting textural features from the MRI images using the Gray Level Co-occurrence Matrix (GLCM). A total of 20 features were initially retrieved from the GLCM, but it was determined that only eight crucial features were necessary for effective classification

Nicolas et al.[50] utilize four feature descriptors—Local Binary Patterns (lbp), Histogram of Oriented Gradients (HOG), Local Wavelet Patterns (lwp), and Local Bit Plane Decoded Patterns (ldbp)—to extract features from MRI scan images. The methodology proposed employs a database (6400 images). The images are categorized into four classes: non-demented, very mild demented, mildly demented, and moderately demented. The K-Nearest Neighbour (K-NN) algorithm is used to classify the images into their respective classes based on the features extracted from the images. Another work also addresses the textural features of Amrutha et al.[51], which aims to separate the lateral ventricles of brain MR images to study textural changes in those areas. The bidimensional multiscale entropy (MSE) approach is applied to the segmented lateral ventricles. This proposed method quantifies textural variations at multiple spatial scales, allowing for extracting entropy features and complexity indices. The entropy feature vector is extracted from the segmented LV images, which contain multiscale e—specifically reflecting the textural characteristics of the brain images. These characteristics help classify patients by differentiating between Normal/MCI and Normal/AD.

The extracted features are used for classification purposes, distinguishing between Normal vs. MCI and Normal vs. AD cases. The study employs classifiers such as Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) to evaluate the classification performance. Moving to a study proposed by Dong et al. where they perform different analyses and combine multiple modalities[52], this study employs several key methods for Alzheimer's disease classification, primarily focusing on latent feature representation learning involves learning a specific projection matrix for each imaging modality. This is achieved by introducing a binary label matrix and local geometry constraints, which help protect the original features of each modality in a low-dimensional target space. The proposed method outperformed traditional methods across various classification tasks. Specifically, the results for classifying AD vs. normal controls (NC). For AD classification, various studies, such as the previous one, address the impact of the region of interests as well as different populations; a work by Chan et al.[53] included a cohort of 65 patients with mild cognitive impairment (MCI), 21 patients with AD, 44 patients with subcortical

vascular MCI (scVMCI), 22 patients with Mixed etiology, and 48 healthy elderly controls. This diverse group allowed for a comprehensive analysis of different dementia types. This work defined specific regions from which to extract biomarkers, including Normal Appearing Brain Matter (NABM), White Matter Lesion (WML) Penumbra, Blood Supply Territory (BST), and White Matter Tracts. The study used textural features extracted from FLAIR MRI images to analyze and classify dementia cohorts. The utility of the textural featFeature proven its impact on the classification outcome as mentioned by Lucas et al.[54] where they employed a dataset from the Alzheimer's Disease Neuroimaging Initiative (ADNI), which includes structural MRI images from 474 subjects categorized into three groups: cognitively normal (CN), mild cognitive impairment (MCI), and Alzheimer's disease (AD). Multiple texture features are extracted from 92 segmented brain regions using three algorithms: Gray Level Co-occurrence Matrix (GLCM), Run Length Matrix (RLM), and Local Binary Patterns (LBP). These features help capture the texture characteristics of the brain regions, which are essential for distinguishing between different cognitive states. The extracted features are used to construct graphs where nodes represent the brain regions, characterized by attributes such as texture feature distributions and brain region volumes. Edges are defined based on the distance between node attributes, with different distance metrics being investigated to determine the best representation of relationships between regions. The study employs graph kernels to measure the similarity between the constructed graphs. This allows SVM classifiers to differentiate between the cognitive states (CN vs. MCI, CN vs. AD, and MCI vs. AD) based on the graph representations.

Table 2.1: Literature review focusing on the spacial domain

Authors	**Database**	**Methods**	**Results**	**Advantages**	**challenges**
Altaf et al. [42]	ADNI	texture-based features BWO	3 way classification=79%	Integration of texture-based features with clinical data. The use of Bag-of-Word optimization	Exploration of a broader range of feature eHandcraftedthods. Handcrafted features are time-consuming
Sampath et al. [43]	ADNI	GLCM features binary classification	99.26%	Imagpreprocessingng, Improved Accuracy, use of ROI	Data quality and quantity, Generalization problem, lack of feature selection
Shankar et al.[45]	ADNI	GLCM GLRLM LBP SIFT Histogram	3-way classification KNN=84.5% DT=88.52 NN=98.20%	the use of various rang of the textural features feature selection process	Dataset Size(287 subjects) Lack of in-depth analysis of the modModel'srformance and the ROI used

Vaithinathan et al. [46]	ADNI	2-way classification RROI texture of GM+WM	RF (AD vs. NC)=87.39% RF(AD vs. MCI)=63.16% LSVM(NC vs. MCI)=63.73% RF(cMCI vs. MCInc)=59.48%	Efficient Texture Extraction Feature Selection Techniques	Imbalanced Datasets. Feature Selection Challenges Generalization
Baskar et al.[44]	ADNI	FCM+ WPNN	AD vs NC: 98.63% MCI vs NC: 95.4% AD vs MCI: 96.4%	FeatFeatureection, High Accuracy, Reduced Training Data, Multimodal Classification	Manual Intervention, Complexity of Feature Extraction, Dependency on Imaging Quality Generalizability Interpretability
Salunkhe et al.[49]	ADNI	GLCM+Ensemble	AD vs NC: **90.2%**.	High accuracy, Texture Analysis, Identification of Key Features	Limited Generalizability, binary classification, lack of experiment, limited features
Nicolas et al.[50]	Kaggle	LBP, HOG, LBDP, LWP	LBP:91.21%, HOG: 86.73%, LBDP:83,56%, LWP:81%	Multiple Feature Descriptors,	Class Imbalance,Computational Complexity,
[51]	OASIS	Entropy Values	NC vs. MCI: 80.10%, AD vs NC:87.60%	Multiscale Analysis, Robust Feature Extraction	Sample Size, Focus on Lateral Ventricles, Computational Requirements
Dong et al. [52]	ADNI	Latent Feature Fusion + RBF-SVM	AD vs.NC:90.74%, MCI vs. NC:74.73%, AD vs. MCI:75.54%	Enhanced Classification Performance, Latent Feature Representation, Local Geometry Constraints, Robustness to Small Sample Sizes	Complexity of Implementation, Potential Overfitting
Chan et al.[53]	CCNA	Textural features	Various results	Comprehensive Analysis	Cross-Sectional Design

Lucas et al.[54]	ADNI	GLCM, RLM, LBP +SVM	CN vs. MCI: 74.1%, CN vs. AD:83.8%, and MCI vs. A:75.4%	Comprehensive Feature Extraction, Graph-Based Approach, Threshold-Based Edge Removal	Complexity of Graph Construction, Limited Exploration of Other Features, Sensitivity and Specificity Concerns

2.2.1.2 Frequency Domain Features

The frequency domain examines the distribution of intensity variations across different frequencies within an image. Various wavelet transforms have demonstrated their effectiveness, and numerous researchers find the frequency domain intriguing for uncovering different perspectives and variations of AD.

Jha et al.[55] authors used the Dual-Tree Complex Wavelet Transform (DTCWT) for feature extraction from MRI images to capture both local and global image characteristics crucial for distinguishing between AD patients and healthy controls; they also applied the Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) for dimensionality reduction they achieved an accuracy of 90.26% when using Alzheimer's Disease Neuroimaging Initiative Dataset (ADNI) and 95.72% for Open Access Series of Imaging Studies Dataset (OASIS) databases. The authors applied the classification process using the Extreme Learning Machine (ELM) with 10 cross-validation. Gao et al.[56] research aims to assess the effectiveness of incorporating contourlet-based hippocampal MRI texture features into multivariant models for improving the classification of AD and the prediction of MCI conversion. Additionally, the study evaluates the feasibility of using Gaussian Process (GP) and Partial Least Squares (PLS) algorithms in developing these multivariant models. the authors achieved an AUC of 92% with PLS, 88% with Gaussian Process Regression Model 1 (GPR1), 78% Gaussian Process Regression Model 2 (GPR2) ,and 83 % with Support Vector Machine (SVM).

Another proposed method by [57] aims to provide a reliable and efficient approach for the early identification of AD based on the analysis of MRI data using Discrete Wavelet Transform (DWT) and SVM techniques. The authors use the FCM clustering technique to segment the brain MRI data into different tissue types, such as GM, WM, and CSF, and extract the following features: the mean values, standard deviation values, and entropy values. They achieved high accuracy rates using the 'Haar' wavelet transformation with the approximation sub-band at the third level, demonstrating the best solution with an accuracy value of 97.37%; those results were based on the GM. Acharya et al. [58] proposed a system that incorporates various feature extraction techniques such as DTCWT, DWT, Empirical Wavelet Transform (EWT), Shearlet Transform (ST), Contourlet Transform (CoT), Curvelet Transform (CuT), and Continuous Wavelet Transform (CWT) Feature selection is performed using Student's t-test. Classification is carried out using the k-nearest

Neighbor (KNN) classifier. The system was tested on a clinical dataset and benchmark images using a 3-fold stratified cross-validation (CV) procedure to evaluate its performance in terms of accuracy, precision, sensitivity, specificity, and F1 score. With the textural features, the authors obtain the best results when using the ST with an accuracy of 98.48%. In addition, Feng et al.[59] propose a methodology that leverages wavelet transformation energy features to improve the identification of AD and MCI patients from Hippocampus (HC) based on structural MRI images. The authors utilize the Wavelet Transformation Energy Features (WTEF) to capture subtle differences in the energy distribution of brain atrophies, which are difficult to detect through spatial analysis. With 90 ROI from GM, the author obtains various energy features and applies the NN for the classification. Feng et al. [59] obtain an accuracy of 93.93% for AD/HC, 86.40% for AD/MCI, and 82.12% for MCI/HC. Expanding on this research, Feng et al.[60] published a follow-up study in 2021 that used the nonsubsampled contourlet subband Nonsubsampled Contourlet Subband-based Individual Network (NCSIN) rather than WTEF. This second work is based on the directional subbands at different scales. Then, three energy features are extracted from each column of the directional subband to capture abnormal energy distribution patterns associated with AD. These energy features are used to create a column energy feature vector. Next, the column energy feature vectors from each directional subband at one scale are considered nodes in the individual network. These nodes represent the directional subbands and their energy characteristics. The edges between nodes in the NCSIN are weighted with connection strength, measured by PCC. This quantifies the relationship between nodes in the network. With this proposed methodology, they achieved an accuracy of 94.21% for AD vs HC, 90.03% for AD vs MCI, and 84.64% for MCI vs. HC. With this second proposed methodology, the results exhibited an improvement. The authors investigated another frequency domain method for their methodological approach as their work progressed by suggesting a third work based on the ST [61]. The author in this work maintains a similar procedure as their previous work, and they achieved an accuracy of 94.78% for AD vs HC, 90.00% AD vs MCI, and 89.71% for MCI vs HC. While the authors of the previous studies [61] have made significant contributions in this area, other researchers have also explored the shearlet transform domain. For instance, in a study published by Alinsaf et al.[62], their proposed method involves combining shearlet-based descriptors with deep features to enhance the classification performance. Specifically, they utilize the 3D shearlet transform to extract features from MRI volumes, which helps capture anisotropic features efficiently. These shearlet-based descriptors are combined with deep features obtained from CNN to create a comprehensive feature representation for each MRI sample.

Furthermore, the authors suggest two different pipelines for their proposed system. In the first pipeline, the MRI samples undergo preprocessing and transformation using the 3D shearlet transform, then summarizing the shearlet coefficients using various texture-based methods such as GLCM, LBP, LOSIB, and SFTA. This pipeline focuses on extracting features using shearlet-based descriptors.

Several CNN models are fine-tuned in the second pipeline based on the MRI samples to extract deep features. These fine-tuned CNN models are then used as feature extractors to complement the shearlet-based descriptors. Unlike the shearlet-based approach, CNN models have the flexibility to learn relevant features directly from the MRI volumes. The authors obtained accuracies ranging from 70% to 80% using the ADNI database and accuracies between 62% and 68% using the OASIS database These results were obtained using 3D ST and GLCM textural features with an SVM classifier. For the second proposition, the authors perform two fusions: Fusion handcrafted all handcrafted descriptors and

deep features, leading to 11,018 descriptors for the ADNI dataset and 8,288 descriptors for the OASIS dataset. The second fusion, which combined shearlet-based descriptors and deep features, yielded the best classification results, demonstrating the effectiveness of feature fusion. Urooj et al.[63] also propose work that is based on the transformation domain; however, they did not use the frequency domain transform they apply the Polar Harmonic Transforms (PHT) for extracting features from MRI data, specifically from the GM tissues of the brain after selecting the most relevant features they perform a binary classification using Self-Adaptive Differential Evolution Wavelet Neural Network (SaDEWN) and achieving 94.4% for AD vs. HC, 92.9%, and 93.7% for AD vs. MCI. Raju et al. [64] propose a novel architecture for the ternary classification of AD, MCI, and . The proposed architecture incorporates a multiscale and multi-path design with wavelet pooling to enhance feature extraction capabilities, prevent vanishing gradient issues, and reduce computational load. The main path focuses on multiscale feature extraction, while the parallel path combines early features with multiscale features to effectively learn local and global features. Furthermore, the wavelet-pooling technique is utilized to extract texture features of the MCI and AD atrophy regions, with the Daubechies-4 wavelet chosen for its efficiency due to its orthogonal and compact nature. This choice of wavelet helps capture subtle atrophy regions of the disease, contributing to the overall effectiveness of the architecture in early disease prediction. With this methodology, the authors achieve 96.52% for 3-way classification.

A study by Silva et al.[65] uses structural MRI texture analysis to differentiate between AD, MCI, and CN by extracting a set of co-occurrence matrix and texture statistical measures from a two-level discrete wavelet transform decomposition of MRI images. These measures include contrast, correlation, energy, homogeneity, entropy, variance, and standard deviation. By analyzing these texture features extracted from the MRI images, the study aims to capture subtle differences in brain tissue characteristics among individuals with CN, MCI, and AD. Furthermore, the study employs machine learning algorithms such as Decision Trees, Discriminant Analysis, Logistic Regression, Naive Bayes, Support Vector Machines (SVM), and k-nearest Neighbors (KNN) to classify individuals into the three different stages based on the extracted texture features. Each classifier is optimized with specific parameters to enhance its performance in distinguishing between CN, MCI, and AD. The authors performed a 2-way and 3-way classification strategy and they achieved an accuracy of 93.3% accuracy for AD vs CN, 87.7% accuracy for AD vs MCI,88.2% accuracy for CN vs MCI for the multiclassification they obtain 75.3% for All vs All.

Table 2.2: Literature review focusing on the Frequency domain

Authors	**Database**	**Methods**	**Results**	**Advantages**	**challenges**
Jha et al.[55]	ADNI and OASIS	2-way classification Wavelet and ELM	90.26% (ADNI) and 95.72% (OASIS)	Potential for Early Detection, Multiscale Analysis, Multiscale Analysis, enhanced Feature Extraction	Complexity of Methodology, Training Data Bias, Limited Comparison Metrics

Gao et al.[56]	ADNI (MRI/ PET)	2-way classification	/	Improved Classification and Prediction - Feasibility of different- Models Comparative Performance- Potential Clinical Impact -Methodological Advancement	Sample Size - the focus on the hippocampal MRI texture features-sensitive mod-Model high-dimensional data.
Wulandari et al.[57]	ADNI	2-way classification DWT and SVM	AD vs NC=97.37%	High Accuracy - Interdisciplinary Approach - Reproducibility - ROI	Sample Size - Dataset Variability - Lack of the number of features
Acharya et al. [58]	University of Malaya Medical Centre Database	2-way classification ST +KNN	98.48%	High Accuracy- Efficiency- Feature Extraction Techniques - Comparative Analysis - Fewer Features Required	Limited Dataset - Feature Selection Bias - Classifier Dependency - Limited literature Comparison - Future Scalability
Feng et al.[59]	ADNI	2-way classification WTEF +	93.93% for AD vs HC, 86.40% for AD vs MCI, and 82.12% for MCI vs HC.	Integration of Spatial and Transformation Domain - Effective Feature Extraction - Segmentation with Brain Mask - Multiscale Analysis - Multiscale Analysis	the information redundancy caused by the non-downsampling process of the wavelet transform -noise or irrelevant information into the feature representation
Feng et al.[60]	ADNI	2-way classification NCSIN + NN	94.21% for AD vs HC, 90.03% for AD vs MCI, and 84.64% for MCI vs HC	Innovative Approach - Effective Feature Extraction - Low Dimensionality	lack of clear biological interpretation

Feng et al.[61]	ADNI	2-way classification SSBIN + SVM	94.78% for AD vs HC, 90.00% AD vs MCI, and 89.71% for MCI vs HC	Improved Detection Performance - Comprehensive Representation - Frequency Domain Analysis	Complexity and Interpretability- possible overfitting - Computational Resources - Complexity and Interpretability
Alinsaf et al.[62]	ADNI and OASIS	2-way classification 3D ST and GLCM and Deep features and CNN	Various results were reported	Effective Feature Fusion - Generalization to Unseen Data - Novel Application of 3D Shearlet Transform	Dimensionality of Feature Vectors - Model Overfitting - Data leakage problem - Computational Intensity
Urooj et al.[63]	ADNI	2 way classification PHT + SaDEWN	94.4% for AD vs. HC, 92.9%, and 93.7% for AD vs. MCI	Innovative Hybrid Approach - Statistical Significance -	highly dependent on the quality and size of the dataset - Model Complexity
Raju et al.[64]	ADNI	3-way classification wavelet pooling	AD vs MCI vs NC =96.52%	Efficient Classification - Improved Feature Extraction: - Reduced Computational Complexity	Data Imbalance - External Validation - Lack of the modModelterpretability
Silva et al.[65]	ADNI	2-way and 3-way classification - Wavelet-Texture features bagged trees classifier	AD vs CN=93.3% AD vs MCI=87.7% CN vs MCI=88.2% AD vs MCI vs CN=75.3%	High 2-way Classification Accuracies - texture statistical provides objective biomarkers	Low multi-Class Classification results

2.2.2 Deep learning-based methods: High-level

High-level features are abstract representations of input data that capture complex and semantic information. They are typically derived through multiple layers of processing in deep learning models. These features are more abstract and particularly useful for classification, detection, and recognition tasks, as they represent data in terms of meaningful concepts rather than low-level details. In deep learning, high-level features are usually obtained from the deeper layers of a neural network. The network extracts increasingly complex patterns and relationships as data passes through each layer. For instance, in a Convolutional Neural Network (CNN), early layers may detect edges and textures (low-level features). In contrast, deeper layers identify parts of objects or entire objects (high-level features).

High-level features are crucial for making accurate predictions and decisions. In image classification, for example, identifying an entire object is more informative than detecting individual pixels or edges. Additionally, high-level features learned by pre-trained models on large datasets can be transferred to new tasks. This transferability is why pre-trained models are often fine-tuned for specific applications, as the high-level features they have learned can generalize well across different tasks.

In this section, we will present works that use deep learning in the classification AD stage; hence, the following review will include works that apply various deep learning strategies, including transfer learning.

Xiaowang et al. [66] introduce a deep Metric learning algorithm to diagnose Alzheimer's Disease by learning a mapping from the original MRI image features to a low-dimensional vector space. This process allows the algorithm to minimize the distance between objects (MRI images) of the same class (e.g., AD patients) while maximizing the distance between objects of different classes (e.g., AD patients vs. healthy controls). As a result, the DML algorithm enhances MRI images' classification accuracy and stability, improving the clarity and quality of the images used for diagnosis. Additionally, the DML model demonstrates faster convergence during training compared to traditional methods, which aids in the efficient classification of early AD patients. The database used in this work was obtained from ADNI. The authors perform a binary classification, obtaining an accuracy of 83.00% for AD vs HC and an accuracy of 65.00% for MCI vs HC.

Tufail et al. [67] present a distinctive two-dimensional CNN constructed with separate convolutional layers. This architecture is designed to automatically learn features from local brain images, combined to make final classifications for AD diagnosis. In addition to the custom CNN, the methodology incorporates two well-known transfer learning architectures: Inception version 3 and Xception. These models are pre-trained on the ImageNet dataset. They are utilized to leverage their learned features for the classification task on the MRI images from 416 subjects, including those diagnosed with mild to moderate AD, allowing for a comprehensive classification problem analysis. The study emphasizes the importance of addressing data imbalance and data leakage issues, which can lead to biased classification performance. The authors explore these challenges to improve the robustness of their models. The best results were obtained by the transfer learning approach with an accuracy of 64.87% M.Fang et al. [68] introduce an a-way classification CNN by applying DTI images as the input to their CNN. This multiclassification is more valuable for clinical diagnosis than existing methods focusing on binary or limited classifications. The approach incorporates a re-transfer learning method that combines transfer learning and multimodal learning theories. This method is particularly beneficial in addressing the challenges of class imbalance and data scarcity, which are common in medical imaging

datasets. This study performs a binary classification.

Al-Adhaileh et al. [69] The study primarily utilized advanced deep learning techniques. Among the models evaluated, the AlexNet transfer model was identified as particularly effective for analyzing brain MRI scans. The research also compared deep learning architectures, including ResNet50, to determine their performance in diagnosing AD from MRI images. The utilized CNN algorithm, particularly the AlexNet model, demonstrated promising results in classifying Alzheimer's disease compared to existing diagnostic systems. The study reported high accuracy rates, achieving 92.85% for multiclass classification using the transfer learning method with fine-tuned parameters. This performance was evaluated against the physician's decisions, indicating that the CNN model is a supportive tool rather than a replacement for clinical judgment. The performance of the classification models in the study was evaluated using a dataset of brain MRI scans collected from the Kaggle AAlzheimer'sclassification dataset. A total of 1,279 MRI scans were utilized, which were categorized into different groups: Mild Demented (179 scans), Moderate Demented (12 scans), Non-Demented (640 scans), and Very Mild Demented (448 scans). The images had a resolution of 208x176 pixels, and the dataset was used to train, validate, and test the deep learning models.

Fareed et al. [70] propose a new CNN architecture specifically designed for detecting Alzheimer's Disease with relatively few parameters, making it suitable for training on smaller datasets. The research highlights the issue of dataset imbalance in Alzheimer's classification and employs the SMOTETOMEK oversampling algorithm to generate synthetic samples, thereby balancing the number of samples across different classes. The dataset utilized in the research is sourced from Kaggle, consisting of 6400 MRI scan images of anonymous patients, categorized into four classes: Normal (NOD), Very Mild Demented (VMD), Mild Demented (MD), and Moderate Demented (MOD). The data undergoes normalization and one-hot encoding to prepare it for training. The ADD-Net model is built from scratch, featuring a custom-designed architecture with multiple layers known as ADD blocks. This design aims to reduce the number of parameters and computational costs while effectively classifying the stages of Alzheimer's Disease, The dataset is split into training, validation, and testing sets (60%, 20%, and 20%, respectively). The model is then trained using standard CNN techniques, and its performance is evaluated against other state-of-the-art models, achieving a test accuracy of 97.89%.

Taliah et al. [71] The study utilized various deep learning models, including Custom CNN, to classify fMRI images of patients into different stages of Alzheimer's Disease. The research focused on binary classification (distinguishing between MCI and AD) and multiclass classification (categorizing into six stages of AD). The proposed method achieved an overall average accuracy of 98.8% for classifying the stages of Alzheimer's Disease. This high accuracy indicates the effectiveness of the deep learning models and the ensembling techniques used in the study. The research employed a max-voting ensembling technique to improve the results of different deep-learning models. This approach allowed for better performance in classifying the various stages of AD compared to using individual models alone. Multiple deep-learning models are trained on functional MRI images of patients. The study employs various architectures, including Custom CNN, VGG-16, ResNet-18, and others, to capture different features from the data. Each model was trained to classify the images into various stages of Alzheimer's Disease (AD). The study primarily utilizes the max-voting technique for ensembling. In this method, each model votes for a predicted class, and the class that receives the most votes across all models is selected as the final prediction. This approach helps mitigate the impact of individual model errors and enhances overall classification accuracy. While the study mentions other ensembling methods, such

as stacking, blending, and averaging, it finds that the max-voting technique yields the best results for their specific application in classifying the stages of Alzheimer's Disease. The proposed methodology achieved a high overall accuracy of **98.8%** using the max-voting ensembling technique for multiclass classification of the six stages of Alzheimer's Disease. This indicates a significant improvement in classification performance compared to previous methods.

Nawaz et al. [72] developed a deep learning model using DenseNet with transfer learning, achieving an impressive accuracy of 97.84% in the multiclass classification of Alzheimer's disease stages based on MRI scans of gray matter (GM). The research involved segmenting MRI scans into three parts: gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). The focus was primarily on the GM, which is significantly affected by Alzheimer's disease. The modeModellized a pre-trained DenseNet model, retraining only the last two blocks for classifying Alzheimer's stages. This approach allowed for effective feature extraction from the MRI images, enhancing the Model's performance. The proposed model performs several other well-known convolutional neural network (CNN) architectures, such as Inception V4, AlexNet, GoogLeNet, and ResNet-18, which had lower accuracy rates ranging from 92.80% to 96.88%.

Faizal et al. [73] proposed a deep Siamese convolutional neural network (SCNN). The SCNN consists of two or more identical sub-networks that share the same weights and parameters. This design allows the network to process two input images simultaneously, generating feature vectors (embeddings) for each image. The goal is to learn a similarity model that can effectively differentiate between images of the same class and those of different classes. The model employs a pre-trained VGG16 architecture as the encoder to extract features from the input MRI images. This pre-trained model the images into lower-dimensional embeddings, which capture the essential characteristics of the images while reducing their complexity; the SCNN utilizes a triplet-loss function, which is designed to optimize the embeddings such that the distance between embeddings of similar images (anchor and positive) is minimized. In contrast, the distance between embeddings of dissimilar images (anchor and negative) is maximized. This approach helps create a more discriminative feature space, enhancing the model's ability to classify different stages of Alzheimer's disease effectively.

Pushpandra et al. [74] discuss several key deep transfer learning models for the class-wise prediction of Alzheimer's disease, including **Xception**, **EfficientNetB7 ResNet50**, **VGG19**, **DenseNet201**. The study proposes to improve the early detection of Alzheimer's disease through **Utilization of MRI Images Classifying Multiple Stages** and using transfer learning. The DenseNet201 model achieved the highest validation accuracy of **96.59%**, making it the most effective model for those tested.

Assami et al.[75] The authors reviewed existing works on Alzheimer's disease classification, highlighting the importance of early detection and the role of deep learning in improving diagnostic accuracy. They implemented various deep learning architectures, including Convolutional Neural Networks (CNNs) and transfer learning models such as AlexNet, ResNet-18, and ResNet-101, to classify different stages of Alzheimer's disease based on MRI scans. The authors conducted extensive experiments to assess the performance of their approach. This analysis allowed them to compare the effectiveness of different architectures and strategies. The study evaluated six pre-trained deep learning models—VGG-19, VGG-16, ResNet-50, Inception-V3, Xception, and DenseNet-169. The results indicated that VGG-16 and VGG-19 outperformed the other models regarding classification accuracy for Alzheimer's disease stages. The models were trained and tested on a dataset of 6,400 MRI images, which included four classes: Mild Demented, Mod-

erate Demented, Non-Demented, and Very Mild Demented. The deep learning models demonstrated a high level of accuracy in distinguishing between these classes, effectively categorizing patients based on the severity of their condition. The use of transfer learning significantly improved the performance of the models, allowing them to adapt pre-trained architectures to the specific task of Alzheimer's classification. This approach helped mitigate issues related to overfitting and enhanced the models' ability to generalize from the training data. The authors noted that the findings suggest a promising avenue for future research, focusing on developing specialized deep-learning networks tailored for Alzheimer's disease classification.

Akbar et al. [76] By combining features from the original brain MR images and their transformations (using the Orthogonal Ripplet II Transform and the Two-Dimensional Discrete Orthonormal Stockwell Transform), the method increases the amount of information available for classification. This multi-faceted approach allows for a more comprehensive analysis of the brain's condition, which is crucial for detecting early signs of Alzheimer's disease. Convolutional Neural Networks (CNNs) for deep feature extraction enable the model of complex patterns and representations from the MR images. This deep learning approach is particularly effective in capturing subtle changes in brain structure that may indicate the early stages of Alzheimer's, which traditional methods might miss. Neighborhood Component Analysis (NCA) helps select the most informative features from the concatenated feature vector. By focusing on the most relevant features, the model enhances its classification accuracy and reduces the impact of redundant or irrelevant data, essential for reliable early diagnosis. The study reports that the feature-level fusion approach, mainly when using EfficientNet-B7 with an Artificial Neural Network (ANN), achieves the highest accuracy in classifying brain MR images. This high level of accuracy is critical for early detection, as it allows for timely intervention and management of the disease. They extract deep features from the original brain MR images and their transformations. The study uses the Orthogonal Ripplet II Transform (ORT-II) and the Two-Dimensional Discrete Orthonormal Stockwell Transform (2D DOST) to obtain different representations of the images. The CNNs are designed to learn spatial hierarchies of features, capturing both low-level and high-level patterns in the images. In this approach, multiple CNNs classify each transformed image separately. The final classification decision is based on a majority vote among the CNNs, which helps integrate the decisions from different feature sets. Features extracted from the original and transformed images are concatenated to form a comprehensive feature vector. Neighborhood Component Analysis (NCA) is applied to enhance classification performance and select the most relevant features from this vector before classification. The study evaluates different classifiers, including Artificial Neural Networks (ANN), k-nearest Neighbors (kNN), and Support Vector Machines (SVM), to determine the best-performing model classifying the brain MR images into the respective categories (CN, MCI, AD).

Shukla et al.[77] The principal methodology of the study involves the use of various convolutional neural network (CNN) models to classify different stages of Alzheimer's disease (AD) based on medical imaging data. The study utilized a dataset from the Kaggle repository, which includes images classified into different stages of Alzheimer's disease: non-demented (nD), mildly demented (mD), moderately demented (moD), and very mildly demented (vmD). The data was preprocessed to ensure compatibility with the chosen models. Five convolutional network models were selected. Each mode model is trained separately on the preprocessed data. Techniques such as batch normalization and residual connections were employed to optimize performance.

Daniel et al.[78] The study explored and compared two approaches for MRI image

segmentation in predicting Alzheimer's disease. The first approach utilized the K-means clustering algorithm, and the second employed Principal Component Analysis (PCA) for dimensionality reduction. The results indicated that the PCA method provided superior performance in feature extraction compared to the K-means method. This suggests that factorial methods like PCA can enhance the extraction of relevant features from MRI images, which is crucial for accurate diagnosis. The primary methodology of the research involves two distinct approaches for diagnosing Alzheimer's disease using MRI images; the first approach uses The K-means clustering algorithm to segment the images into three clusters representing white matter, grey matter, and cerebrospinal fluid. A Convolutional Neural Network (CNN) was then trained on the segmented images to classify the presence of Alzheimer's disease based on the identified features. In the second approach, PCA was utilized to reduce the dimensionality of the MRI images, extracting relevant features that represent white matter, grey matter, and cerebrospinal fluid. The reduced feature set was then fed into a Multilayer Perceptron (MLP) to classify Alzheimer's disease cases. Both approaches were evaluated using a dataset of 602 MRI images from the Alzheimer's Disease Neuroimaging Initiative (ADNI), which included images from cognitively normal individuals and Alzheimer's disease patients. The performance of each method was compared based on accuracy and processing time, with the PCA + MLP approach demonstrating superior results in accuracy and efficiency.

El-Assy et al.[79] proposed CNN architecture for Alzheimer's disease classification, including several key features. The architecture employs two convolutional neural network models designed with distinct filter sizes and pooling layers. This allows the model to capture various features from the MRI data effectively. The outputs of the two CNN models are concatenated at the classification layer. This novel approach enhances the model to learn task-specific features and improves overall classification accuracy. The architecture is designed to handle multiclass classification tasks, addressing challenges across three, four, and five categories of Alzheimer's disease. This is a departure from many studies that focus solely on binary classification. The network leverages the hierarchical nature of convolutional, pooling, and fully connected layers to extract both local and global patterns from the MRI images, facilitating accurate discrimination between different AD categories. The CNN architecture is designed to automatically extract intricate features from the MRI images without manual feature extraction. This capability allows the identification of relevant patterns associated with Alzheimer's disease stages, such as changes in grey and white matter volumes, cortical thickness, and other structural characteristics. The proposed CNN architecture for classifying Alzheimer's disease is designed to effectively process MRI data and achieve high accuracy in distinguishing between different stages of the disease. The proposed CNN Architecture consists of two distinct CNN models, each designed with different filter sizes and pooling layers. This dual-model approach allows the network to capture various features from the MRI images. The input to the CNN models is a tensor of size $224 \times 224 \times 3$, representing the MRI images. This size is chosen to standardize the input dimensions for the network. Each CNN model begins with multiple convolutional layers. For instance, the first model has two convolutional layers, each with 16 filters of size 3×3. The second model utilizes larger filters (e.g., 5×5) to capture different spatial features. Following the convolutional layers, max-pooling layers are employed to reduce the spatial dimensions of the feature maps, which helps retain the most significant features while reducing computational complexity. The outputs from the two CNN models are concatenated in a classification layer. This innovative approach allows the models to learn task-specific features and complement each other's capabilities, enhancing overall classification performance. After concatenation, fully connected layers

are used to process the combined features and produce the final classification output, which categorizes the MRI images into multiple classes.

The training begins with preparing the MRI images from the Alzheimer's Disease Neuroimaging Initiative database. This includes processing steps such as normalization, resizing, and data augmentation to enhance the dataset's quality and variability. The dataset is divided into training, validation, and test sets. The training set is used to train the model, and the validation set is used to tune hyperparameters and prevent overfitting. The test set is reserved for evaluating the performance after training. To prevent overfitting, the training process may include early stopping, where training is halted if the performance on the validation set does not improve for a specified number of epochs.

Zhang et al.[80] the author proposed a CNN model called ADNet model classifying Alzheimer's disease involves several key components and techniques aimed at enhancing the performance of the classification task. DNet is built upon the VGG16 architecture, a well-known convolutional neural network (CNN) model. The architecture is modified to improve its efficiency and effectiveness in classifying MRI images. This technique replaces traditional convolution layers with depthwise separable convolutions. This approach involves two steps: applying a single convolutional filter to each input channel (depthwise convolution) followed by a pointwise convolution that combines the outputs. This significantly reduces the number of parameters and computational costs while maintaining performance. The model was trained using multiple datasets, including the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, which contains a wealth of MRI images and clinical information from Alzheimer's disease patients and cognitively normal individuals.

The summary of this literature review will be found in the following table:

Table 2.3: Deep learning-based method literature review

Authors	Database	Methods	Results	Multi-classification	Advantages	Limitations
[67]	OASIS	TL Xception	AD vs. NC:64.87	no	Effective Use of Transfer Learning, Addressing Data Challenges, use of cross-validation	Data Imbalance overfitting
[66]	ADNI	DML	AD vs. NC:83% MCI vs. NC: 65%	no	Convergence Speed,	Limited Comparison with Other Methods, Small Sample Size, focus on binary classification
[68]	ADNI	CNN4AD	AD vs. NC: 94.6% LMCI vs. EMCI: 80.8% EMCI vs. NC: 93.5% LMCI vs. NC: 90.9%	No	Re-Transfer Learning,Robustness Against Overfitting	Complexity of Model Training,Interpretability of the Model,
[69]	Kaggle	TL	ResNet50:92.67% AlexNet:94%	yes	using transfer learning, Fine-tuning strategy	Dataset Size and Diversity, Class Imbalance, Lack of Compariso n, Potential Overfitting, minimum epoch
[70]	Kaggle	ADD-Net	97.89%	yes	Reduced overfitting,Efficient Architecture	Imbalanced Dataset Challenges, Limited External Validation,Computational Resources
[71]	ADNI	Max-voting	98.80%	yes	Ensembling Techniques, Comprehensive Evaluation Metrics, Focus on MMulticlassClassification	Focus on Specific Age Group, Dataset Size and Diversity, Potential Overfitting, Limited Comparison with Other Techniques
[73]	ADNI/ OASIS	SCNN	OASIS:93.85% ADNI:91.83%	yes	Triplet-Loss Function, Improved Classification Accuracy, utilization of two databases, Discriminative Feature Learning, Robustness to Class Imbalance,pre-trained VGG16 model as an encoder	Complexity of the ModeModeld for Extensive Tuning, Potential Overfitting
[75]	Kaggle	CNN+TL	VGG-19: 92.86% VGG-16: 92.83 Inception-V3: 91.04% Xception: 90.57% ResNet-50: 85.99% DenseNet169: 88.64%	Yes	Transfer learning, Hyperparameter optimization,	Unbalanced database, Overfitting problem,
[76]	ADNI	feature-level fusion	CN vs. MCI: .83.9% CN vs. AD:90.3%, MCI v.s AD:85.4%	no	Feature Selection for Improved Performance, fusion strategy, Potential for Early Intervention	Lack of comprehensive evaluation
[77]	Kaggle	CNN+TL	Alz-DenseConvNet: 92%. Alz-ResConvNet: 69% Alz-VGGConvNet: 93%. Alz-MobileConvNet: 94% Alz-XceptionConvNet: 92%	yes	High accuracy, Transfer learning,	Unbalanced database, Overfittingproblemm,
[79]	ADNI	Dual CNN Models	99.30%	yes	Hierarchical Feature Extraction, High accuracy, Concatenation Technique,	Unbalanced database, potential overfitting, Complexity of Data, Model complexity
[80]	ADNI	ADNet	MCI vsCN:79.62%,AD vs. CN:87.12%		Enhanced Stability, Multi-task Learning Benefits	Limited Data Sources, Binary classification
[78]	ADNI	PCA+MLP	(AD vs NC)K-means + CNN : 82.0% PCA + MLP:91.0%	No	Improved Accuracy,Efficient Processing Time (PCA + MLP for 106.70 seconds for training over 500 epochs, K-means + CNN:11.50 hours for 60 epochs.	Sensitivity to Hyperparameters, Loss of Spatial Information, Dataset Limitations
[74]	Kaggle	transfer learning models	ResNet50: 93.52% VGG19: 95.08% Xception: 89.77% EfficientNetB7: 83.20% DenseNet201:96.59%	yes	Transfer learning, Data augmentation, comparative analysis	Very imbalanced database, Adding noise to the images, lack of information about augmented data

2.3 Research Gaps

Identifying research gaps is crucial because it helps us recognize areas that require further investigation. Although there has been progress in AD stage categorization, some areas still need more investigation and have hurdles to overcome.

Therefore, we are interested in exploring low-level and high-level features because we believe each methodology has its own utility and applicable environment. In our thematic

review, we highlight both sections and identify a significant challenge in Alzheimer's Disease (AD) research: the database and sample size. This challenge persists in many studies, where traditional methods are often considered the best solution, especially when dealing with limited data samples. Based on our literature review, we identified several key areas requiring further investigation:

- **Investigation of Computer Vision Feature Extraction Techniques:** There is a need to investigate computer vision techniques, such as HOG, in AD stage classification. While algorithms like HOG have proven efficacy in the field, there is limited research on their utility for AD classification.
- **Textural Features for AD Classification:** Many studies have used textural features. The choice of textural features is logical, as the brain exhibits textural properties. Some studies have shifted from the spatial domain to the frequency domain, as calculating textural features in the spatial domain does not always yield supportive results, especially in multi-class classification. However, none of the studies have utilized the Fast Finite Shearlet Transform, which is well known for its efficiency, particularly in fusion tasks. In this research, we will investigate its impact on the outcome of AD stage classification.
- **Overfitting in Deep Learning Models:** In the high-level feature domain, particularly with Convolutional Neural Networks (CNNs) and deep learning, most studies suffer from the overfitting problem. As a solution, many researchers have framed AD classification as a binary classification problem rather than a multiclass classification. Additionally, there is a lack of investigation into the efficiency of data augmentation techniques and the comparative analysis of CNN models for AD classification. Overfitting persists in many studies that apply deep learning or transfer learning despite using data augmentation. In this thesis, we will investigate the utility of data augmentation techniques and aim to identify the best transfer learning model for AD. We will also explore layer-wise fine-tuning strategies, rarely used in AD classification.
- **Multiclass Classification:** Several studies have addressed AD as a binary classification problem, and even those that have used multiclass classification typically exclude the early stage of the disease. In contrast, this thesis will address AD as a 4-way and 5-way classification problem rather than the conventional 2-way or 3-way classification.

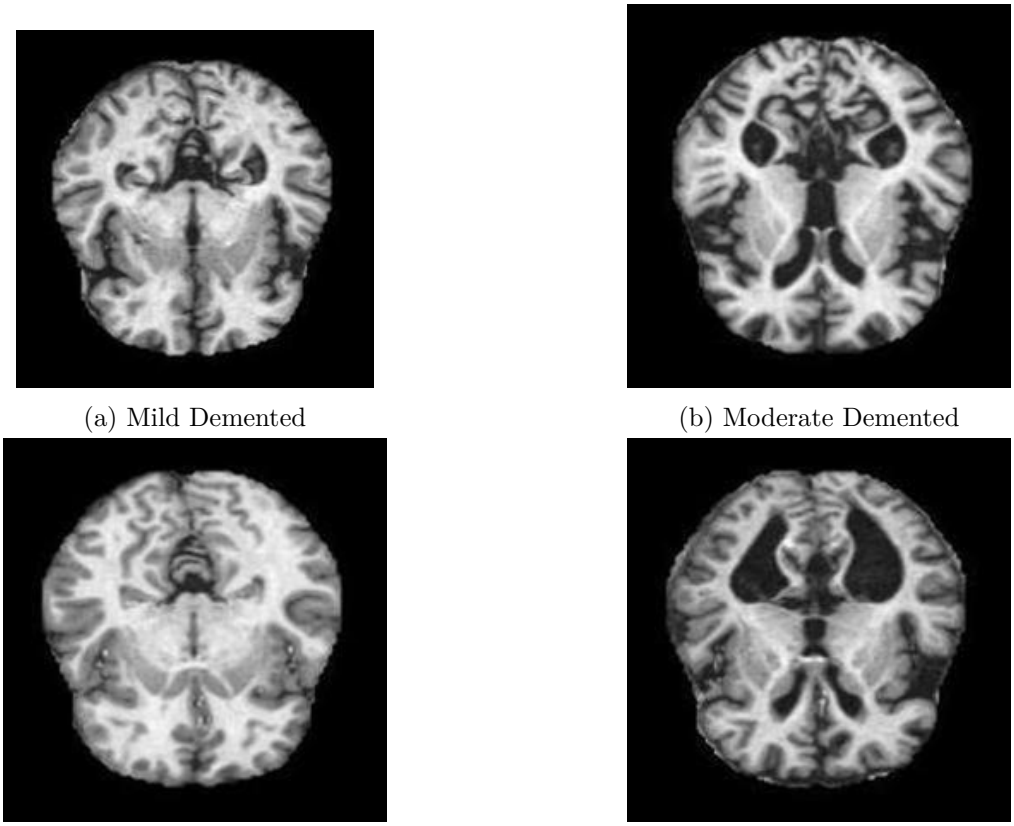
The literature has focused on feature extraction, which is crucial for model performance. However, the choice of machine learning algorithms and their parameters is equally essential. Therefore, this thesis will also address the choice of machine learning algorithms and their parameters in AD classification.

2.4 Databases

Based on the preceding state-of-the-art section, we conclude that the ADNI database is the most commonly used dataset in most studies. We also observed that a dataset obtained from Kaggle is frequently employed, particularly in studies utilizing transfer learning. Therefore, we will utilize these two databases in this research because they are the most commonly used and fulfill our objective of applying 4- and 5-way classification strategies. In the following section, we will introduce both databases.

2.4.1 Kaggle database

The Kaggle database is sourced from the Kaggle website and consists of data collected from various websites, hospitals, and public repositories. This dataset includes four categories of preprocessed brain MRI images. (`https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images/data`). Consulted in September 2021. Those categories are mild demented (MD) see figure 2.1a, moderate demented (MOD) see figure 2.1b, non-demented (ND) see figure 2.1c, and very mild demented (VMD) see figure 2.1d



(a) Mild Demented

(b) Moderate Demented

(c) Non Demented

(d) Very Mild Demented

Figure 2.1: Kaggle database samples.

The dataset is organized into two main folders: the training and test folders. Each of these folders contains four subfolders corresponding to the four classes. The distribution of data samples across the classes and subfolders is illustrated in Figure2.2.

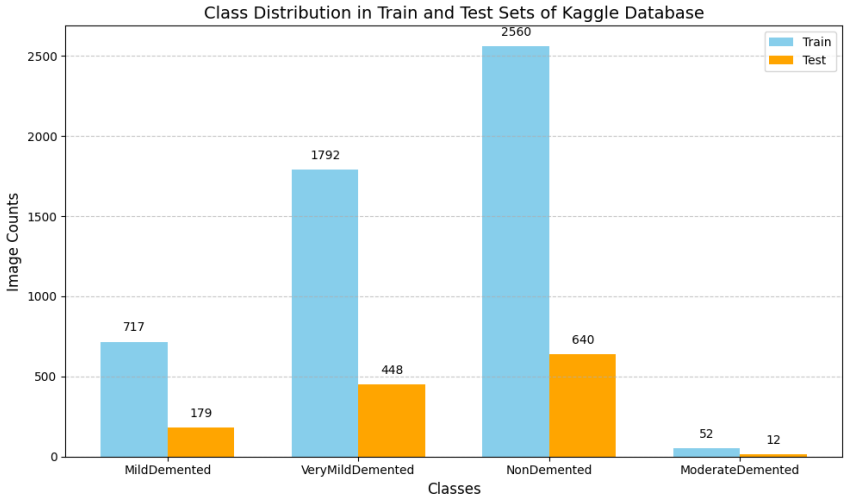


Figure 2.2: Kaggle database class distribution

2.4.2 ADNI3 database

The ADNI3 database (`https://adni.loni.usc.edu/`) is a project initiated by the National Institute on Aging (NIA) in 2016. It includes several categories of cognitive impairment: early mild cognitive impairment (EMCI), mild cognitive impairment (MCI), late mild cognitive impairment (LMCI), Alzheimer's disease (AD), and cognitively normal (CN) individuals.

This dataset features T1-weighted structural MRI scans of the brains of participants of both genders, aged 55 to 90, acquired using GE medical system scanners. The images in the ADNI3 dataset encompass those from ADNI-1, ADNI-GO, and ADNI-2, along with additional scans for EMCI, MCI, and AD.

In this thesis, we will utilize both ADNI2 and ADNI3 datasets. ADNI2 contains 200 elderly controls, 600 cases of MCI, and 200 cases of AD, with data collected from 2011 to 2016. In contrast, ADNI3 includes 371 individuals, comprising 133 elderly controls, 151 MCI cases, and 87 AD cases, with data collected from 2016 to 2022. The distribution of sample sizes across the ADNI database is illustrated in the figure. 2.3

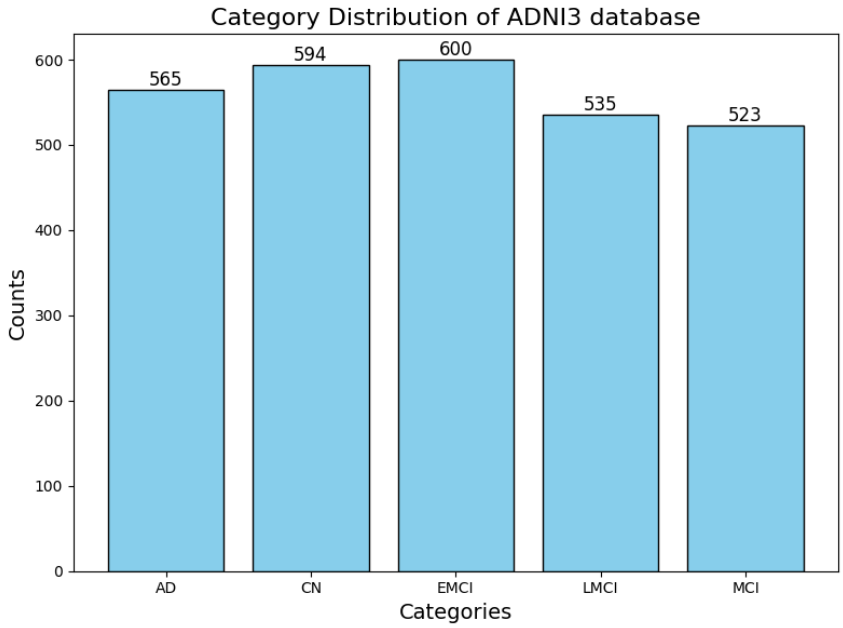
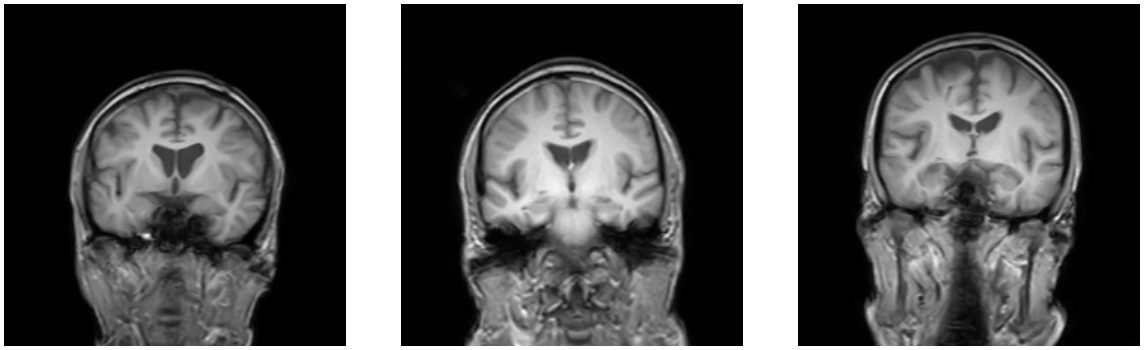


Figure 2.3: ADNI3 database class distribution

With the assistance of radiologists, we selected 4 to 5 two-dimensional coronal hippocampal slices from the three-dimensional MRI images of each patient. A sample example of this database is shown in the figure2.4.



(a) EMCI (b) MCI (c) LMCI

(d) AD (e) NC

Figure 2.4: ADNI3 Database samples.

2.5 Summary

This chapter presents a thematic literature review that categorizes studies focusing on low-level and high-level features. We extend our literature synthesis to highlight the strengths, challenges, and outcomes of various studies in the context of improving diagnostic accuracy for Alzheimer's stages. Additionally, we emphasize the methodologies, datasets, and outcomes utilized in these studies. Based on the findings presented in this section, we provide a rationale for selecting databases. Furthermore, a description of the datasets is included, along with an examination of identified challenges and research gaps that will be addressed further in this dissertation.

CHAPTER 3

ALZHEIMER'S DISEASE STAGE CLASSIFICATION BASED ON LOW-LEVEL FEATURES

3.1 Overview

Accurately differentiating between AD stages is crucial for diagnosing AD. This chapter introduces two methods for classifying AD stages using low-level features from MRI scans.

The first method focuses on the shape and gradient details in MRI brain images by utilizing the HOG as a feature descriptor. HOG captures how edges and gradients are distributed across the image, revealing the shapes of the brain's structures. This approach efficiently analyzes anatomical geometry and provides shape-based solid features that can help differentiate between the stages of Alzheimer's disease.

The second method highlights the textural features in the frequency domain, which captures delicate patterns in brain tissue. Transforming MRI data into the frequency domain allows for a deeper texture analysis, which is especially useful for detecting subtle changes in tissue properties that might signal different stages of the disease. Both methods rely on low-level features, focusing on fundamental image details like edges, gradients, and textures.

This chapter aims to demonstrate the effectiveness of these shape and texture features, instilling confidence in their potential to analyze and classify the stages of Alzheimer's disease.

3.2 Spatial Domain Case Study

Addressing AD images in a spatial domain involves a comprehensive analysis of all pixels within MRI images. In this context, each image element corresponds to a pixel intensity, allowing for direct manipulation of pixel values, including intensity, texture, and shape. Consequently, this section emphasizes using the spatial domain to extract pixel information. Furthermore, we aim to investigate the distribution of gradient orientation by employing HOG descriptor methods on localized versions of the images. We apply the BCFCM technique to enhance the three brain tissues' localization.

We will first present the BCFCM method utilized for improving brain tissue localization, followed by an introduction to the HOG methodology. These two methods constitute

the core of the proposed approach, which will be further detailed in the subsequent sections.

3.2.1 Segmentation of the region of interest: Bias-corrected Fuzzy C-means

Ahmed et al. [81] proposed the bias-corrected fuzzy c-means (BCFCM), an unsupervised clustering technique. To minimize the objective function J (equation 3.1), the BCFCM approach is based on the same idea as the conventional FCM suggested by Bezdek [82]. Nevertheless, a new objective function, J' (equation 3.2), was presented by Ahmed et al. [82], which incorporates adjacent pixels as a regularization factor.

$$J = \sum_{i=1}^{c} \sum_{k=1}^{n} U_{ik}^{m} \left\| (X_k - V_i) \right\|^2 \tag{3.1}$$

With:
U_{ik} : Degree of belonging and V_i the center of a cluster.

$$J' = \sum_{i=1}^{c} \sum_{k=1}^{n} U_{ik}^{m} \left\| (X_k - V_i) \right\|^2 + \frac{\alpha}{N_R} \sum_{i=1}^{c} \sum_{k=1}^{n} U_{ik}^{p} \left(\sum_{x_r \in N_j} \left\| (X_k - V_i) \right\|^2 \right) \tag{3.2}$$

With:
N_i : The neighboring pixels of pixel i.
N_R : is the total number of neighboring pixels.
α: controls the effect of the neighbor term.

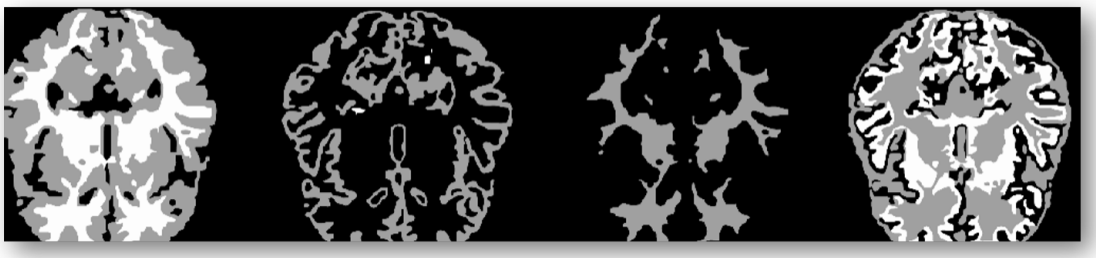


Figure 3.1: The Kaggle database's BCFCM clustering outcomes [8]

Figure 3.2: The ADNI database's BCFCM clustering outcomes [8]

3.2.2 Histogram of oriented gradient (HOG)

The Histograms of Oriented Gradients (HOG) is a widely used feature descriptor in various applications. It is designed to capture an object's shape by measuring the occurrences of gradient orientation in each local region. [83]

The HOG algorithm divides the image into patches, simplifying the computation. The gradient orientation and magnitude are then determined for each patch. The center pixel is considered to calculate the gradient for each pixel, and the differences in values of the surrounding pixels (right/left and above/below) are calculated, resulting in gradients in the horizontal and vertical directions. The HOG feature values are computed based on these calculations and represented as a single vector. In the next step, the magnitude and orientation of each pixel value are mathematically computed using the previously calculated gradient. A histogram is created using the computed magnitude and orientation. Finally, the gradients from image patches of 8 x 8 blocks are normalized to 16 x 16 blocks to reduce lighting variation in the image. [84]

The HOG features for the entire image are generated by combining all the 16 x 16 patches at once. One of the HOG method's key advantages is its resilience to illumination fluctuations. It further emphasizes the delineation and contours of an item, offering supplementary insights on the orientation of the edges [85]. In summary, the HOG algorithm is a powerful and effective approach for describing the features of an image. The HOG algorithm follows these steps:

Algorithm 1 Oriented Gradient Histogram Procedure

1: **Input:** Image I.
2: **Step 1: Resize the image.**
3: **Step 2: Subdivide the image into cells of size $N \times N$ pixels.**
4: **Step 3: Compute gradients for each pixel using filters:**

$$G_x = I * K_x, \quad G_y = I * K_y \tag{3.3}$$

where K_x and K_y are gradient kernels (e.g., Sobel filters).
5: **Step 4: Calculate the gradient magnitude and orientation:**

$$\text{Magnitude: } M = \sqrt{G_x^2 + G_y^2}, \quad \text{Orientation: } \theta = \arctan \frac{G_y}{G_x}. \tag{3.4}$$

6: **Step 5: Construct a histogram of orientations for each cell, weighted by the gradient magnitudes.**
7: **Step 6: Normalize the histograms across overlapping blocks for illumination invariance.**
8: **Step 7: Flatten the normalized histograms into a one-dimensional feature vector.**
9: **Output:** HOG feature vector F.

3.2.3 Machine learning classification

This study included three different machine-learning models. The descriptions of each model are as follows:

- **Support Vector Machine** is a well-known algorithm in ML techniques that can be applied to classification and regression tasks. It's a supervised learning model that aims to find a hyperplane that maximizes the separation of the data points from their potential classes in an N-dimensional space [86]. Different kernel methods such as linear, polynomial, RBF, and sigmoid can be applied to obtain a complete separation between two classes [87], [88] [89]. SVM is a binary classification problem, and since we are addressing a multiclassification problem, we adopt the One Versus One strategy to break down our multiclassification problem into a sub-binary classification problem. Here, we obtain a binary classification for each pair of classes.

- **The K Nearest Neighbors** method is a simple, intuitive, supervised machine-learning technique for classification and regression. It is a non-parametric approach with no trainable parameters. Instead, it uses the distances between the data points to make predictions [90]. It aims to find the K closest data points to a new, unlabeled data point. These K data points are called the "nearest neighbors." Then, it assigns the majority class label among the K nearest neighbors to the new data point. The choice of the K value depends on the experimental performance of the model. Therefore, choosing the appropriate k value is critical to avoiding overfitting caused by a small k value and underfitting caused by a considerable k value [91].

- **Random Forest** (RF) is a popular ensemble learning technique in non-linear and non-parametric analysis. This method is an extension of the decision tree (DT) model, where multiple trees are constructed from subsets of the training dataset. Unlike DT, which constructs a hierarchical structure by iteratively partitioning the dataset according to input features—where internal nodes denote decision rules, branches indicate the outcomes of these rules, and leaf nodes represent final predictions (either class labels for classification tasks or continuous values for regression tasks)—RF builds multiple trees to reduce overfitting. The algorithm seeks to enhance the purity of the resulting subsets through metrics such as Gini impurity, information gain, or variance reduction. Decision trees are susceptible to overfitting despite their simplicity and interpretability, especially when they grow excessively deep or complex. In contrast, RF models are less prone to overfitting and can effectively handle data grouping and regression. However, data overfitting and high error variance are still common issues with RF models. In RF, input variables are evaluated to determine the principal dataset subdivision, and the datasets are progressively segregated into correlative output variable subgroups based on improved residual sum of squares. The final output of the RF model is computed by taking the mode of classes predicted by individual decision trees in the case of classification or the mean prediction of individual trees in regression.[92].

3.2.4 Spatial Domain-Based Low-Level Feature Extraction

For this approach, we propose a methodology for analyzing the shapes and structures present in MRI images. To accomplish this, we employed the Histogram of Oriented Gradients (HOG) technique to accurately capture the brain's shape and appearance. We extracted meaningful information by examining the distribution of intensity gradients or edge directions within the brain MRI image.

To initiate the process, we conducted a pre-processing step that involved filtering and skull stripping. This step was of the utmost importance in preparing the image for

segmentation, which employed the BCFCM algorithm.

Once the segmentation process was complete, we utilized HOG as a feature descriptor. This step enabled us to acquire a feature vector containing valuable information about the image. Armed with this feature vector, we employed machine learning techniques for classification. The figure 3.3 shows the diagram of the this approach

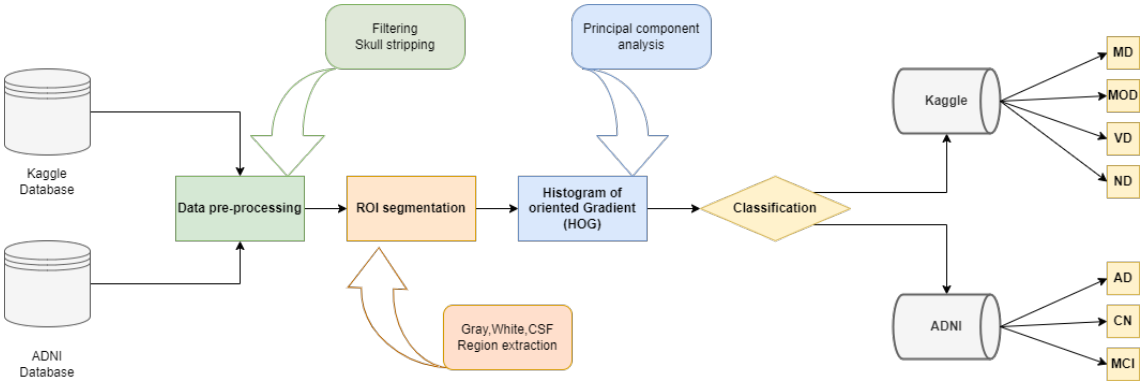


Figure 3.3: Flow diagram of the proposed approach 1

3.2.4.1 Pre-processing

In computer vision and image processing, the preprocessing stage is vital for preparing the data for subsequent analysis. Our study utilized filtering and skull removal to reduce background noise and remove extraneous data. Additionally, the two datasets are structured differently: We use a median filter for both data sets. On the other hand, we remove any non-brain tissue from the ADNI database.

1. **Filtering: Median filter**
   The median filter is a commonly used method for reducing noise, especially when dealing with noise that has large amplitude probability density tails and periodic patterns. It is one of the most effective order-statistic filters and performs well against specific types of noise, including Gaussian, random, and salt-and-pepper noise.[93]

   The median filtering process involves sliding a window over the image and placing the median value of the pixels within the window at the center of the output image. This method effectively removes single-pixel noise artifacts while maintaining the image's overall sharpness [94]. The figure 3.4 below demonstrates how it works.

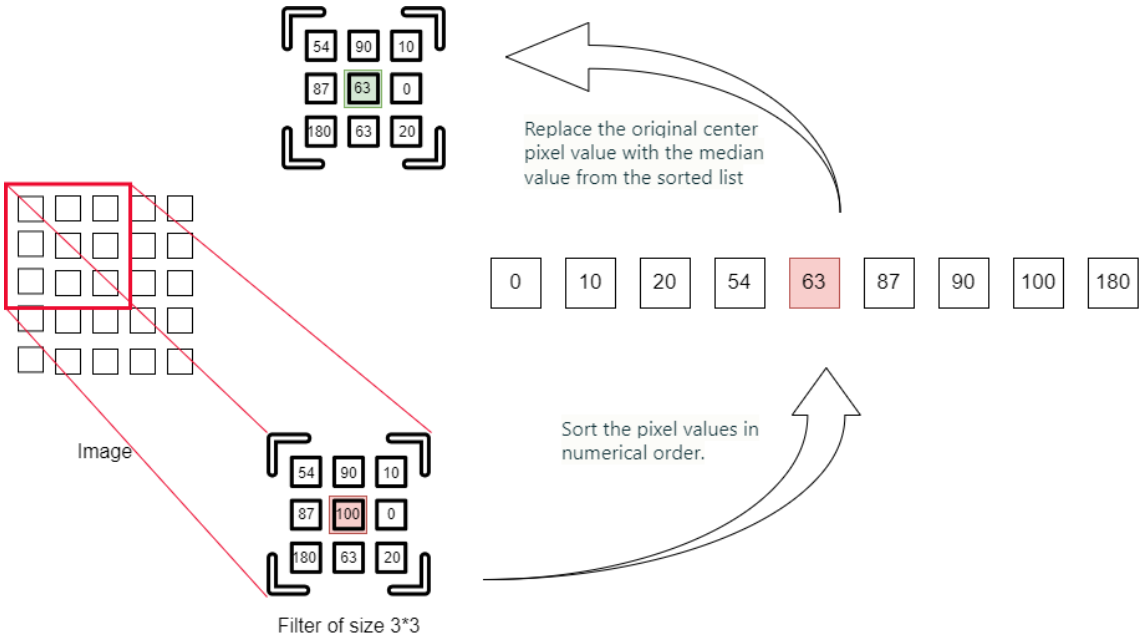


Figure 3.4: Median filter

2. **Skull striping:**
   Skull stripping is a widely employed technique in MRI brain scans to improve the analysis of brain structures. It entails the removal of the skull and other superfluous tissues, which proves particularly valuable in tasks like detecting neurological disorders, monitoring brain volume, or measuring specific brain regions. In this work, we employed morphological operators to extract the brain by applying the erosion and dilatation operators to the structural element [8].

3.2.4.2 HOG feature Descriptor siting

Various parameters within the HOG algorithm determine the size of the HOG feature vectors. Different block sizes and cell counts are used to extract HOG characteristics. Consequently, the number of features varies depending on the size settings. Having too few features indicates a loss of sufficient information, while excessive characteristics are unnecessary, time-consuming [95], and can lead to overfitting machine interpretation algorithms. Therefore, selecting a single block and cell size that remains consistent throughout the entire procedure is challenging, especially when working with two different databases. For this study, we selected a block size of 1×1 and cell size of 8×8 as HOG parameters, with a 50% overlap, based on our experiment. This choice allowed us to determine the appropriate HOG size.

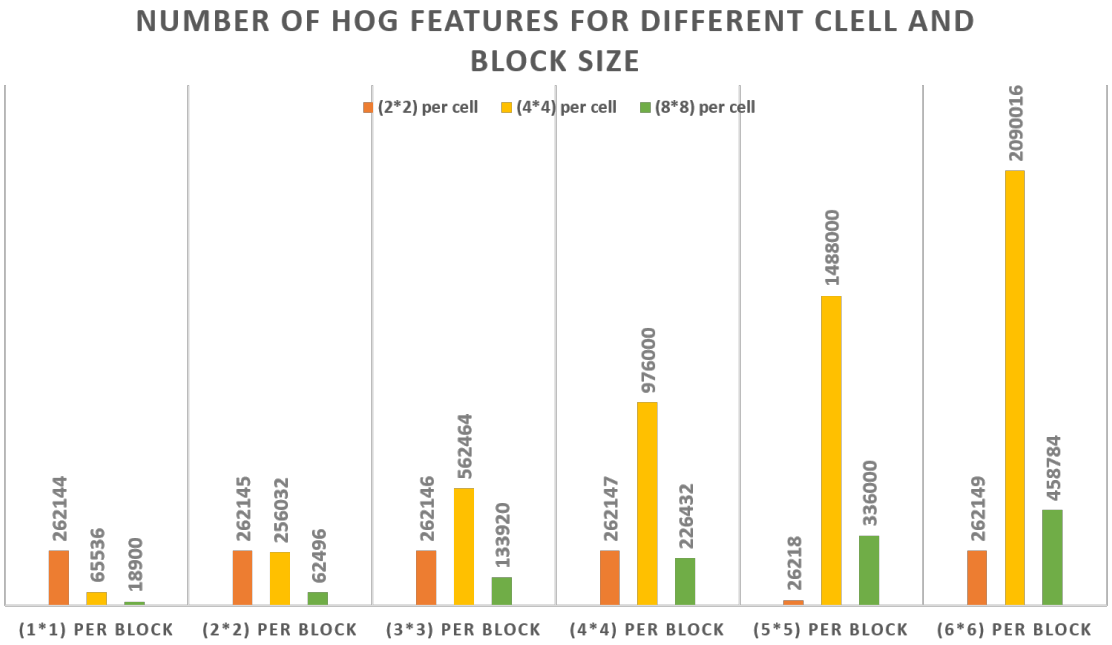


Figure 3.5: The number of HOG features associated with each dimension of the blocks and cells.[8]

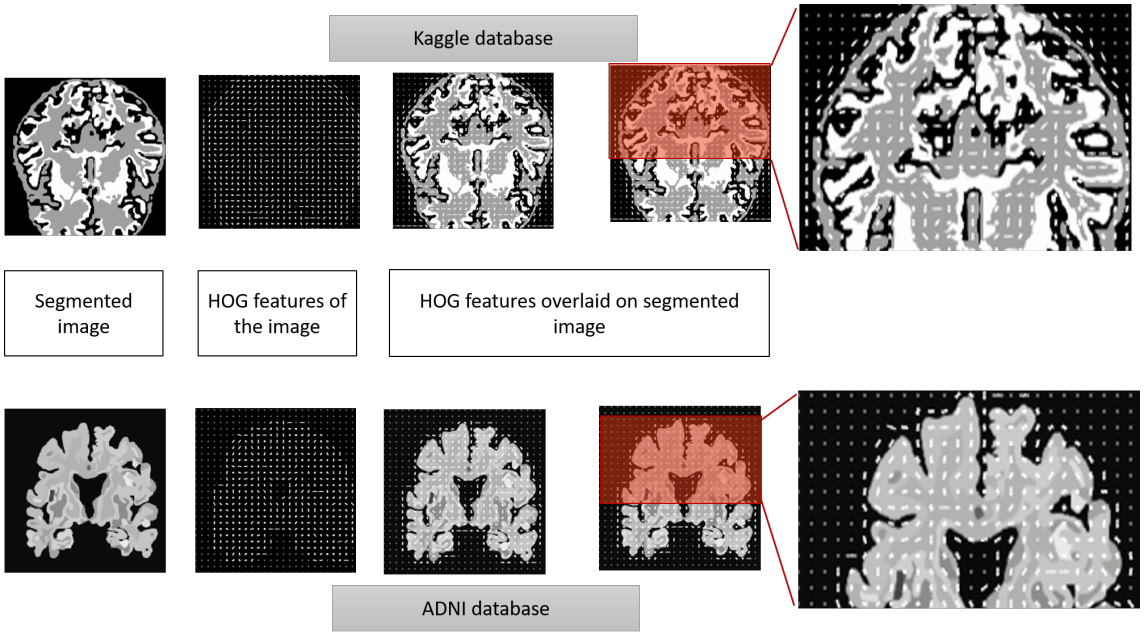


Figure 3.6: Results of the HOG feature representation on the segmented MRI brain images using BCFCM.

3.2.4.3 Principal Component Analysis PCA

PCA is widely used in ML techniques to reduce the dimensions of multivariate data. It provides a strategic solution for managing high-dimensional datasets. While big data can have its benefits, it also presents increased computational demands, potential overfitting,

and decreased interpretability of models. PCA tackles these challenges by eliminating redundant or less informative features, transforming a large dataset into a smaller one while retaining significant patterns and information in the principal components.[96]

Mathematically, PCA is represented by performing eigenvalue decomposition on the covariance matrix of the data. The coefficients for the principal components are obtained from the eigenvector associated with the largest eigenvalue. By selecting only the first few principal components, which contain the majority of the information in the data, dimensionality reduction is achieved.[97] The steps involved in PCA are as follows:

- Standardization: Ensure that all variables are on the same scale.
- Compute the Covariance Matrix: Calculate the covariance between each pair of variables.
- Calculate Eigenvectors and Eigenvalues: Determine the eigenvectors and eigenvalues of the covariance matrix.
- Sort Eigenvalues: Arrange the eigenvalues in descending order.
- Select Principal Components: Choose the first few eigenvectors corresponding to the largest eigenvalues as the principal components.
- Transform the Data: Project the original dataset onto the selected principal components.

3.2.5 Result & Discussion : Approach one

To evaluate the efficacy of the proposed methodology [1], two distinct experiments were conducted:

- ***Experiment 1:*** This experiment examined the influence of brain clustering on classification performance.
- ***Experiment 2:*** This experiment implemented PCA to select features to enhance classification outcomes.

The findings from these experiments, each involving 10-fold cross-validation, are detailed in Tables 3.1 and 3.2. Table 3.1 presents the first experiment's results, which involved a comparative analysis. This analysis contrasted the Histogram of Oriented Gradients (HOG) feature vectors derived from segmented images with those obtained from preprocessed images for classification purposes.

The results demonstrated an approximate 8% enhancement in accuracy across all three machine learning algorithms when clustered brain images were utilized. Specifically, Support Vector Machine (SVM) models exhibited the highest accuracy rates, achieving 89.0% and 88.3% accuracy for the Kaggle and ADNI databases, respectively.

[1] Saim, M., Feroui, A. (2022, November). A New Hybrid Method Based On Bias-Correction Fuzzy C Means And Histogram Of Oriented Gradient For Alzheimer Disease Detection. In 2022 First International Conference on Computer Communications and Intelligent Systems (I3CIS) (pp. 31-36). IEEE.

Table 3.1: The outcome of the first experiment

	Kaggle database				ADNI2 database			
	With BCFCM		Without BCFCM		With BCFCM		Without BCFCM	
	Accuracy	F1 score	Accuracy	F1 score	Accuracy	F1 score	Accuracy	F1 score
SVM	0.89	0.79	0.8	82	0.88	0.83	0.83	0.81
KNN	0.88	0.82	0.72	0.79	0.86	0.79	0.74	0.78
RF	0.79	0.8	0.69	0.7	75.9	0.71	0.7	0.75

The observed improvement in classification performance is likely due to the increased gradient values computed by the HOG method on the clustered images. These elevated gradient values enhance key features' prominence, facilitating more accurate classification results. This effect underscores the importance of effective feature extraction in improving the performance of machine learning models for complex tasks such as Alzheimer's disease detection.

Table 3.2: The outcome of the second experiment

	Kaggle database		ADNI2 database	
	Accuracy	F1 score	Accuracy	F1 score
SVM	0.96	0.58	0.96	0.84
KNN	0.94	0.89	0.92	0.77
RF	0.84	0.81	0.79	0.71

Conversely, the application of the HOG descriptor algorithm with 50% overlapping adds unnecessary data duplication, which adversely affects classification performance. To address this issue, Experiment 2 incorporates PCA for decreasing dimensionality. The results of this experiment, which utilize segmented images combined with PCA, are detailed in Table 3.2. According to the findings, PCA contributes to an approximate 2% improvement in accuracy. As a result, the proposed approach achieves accuracies of 96.8%, 94.4%, and 84.0% for the respective models.

Comparing these results with existing literature, as summarized in Table 3.3, provides valuable insights into the performance of the proposed approach relative to other methodologies.

For instance, the proposed approach achieves an accuracy of 96.8% on the Kaggle database and 96.0% on the ADNI database, outperforming several existing methods. This improvement can be attributed to the effective combination of Bias-Correction Fuzzy C-Means clustering, HOG feature extraction, and PCA-based dimensionality reduction.

The comparative analysis underscores the significance of integrating clustering and feature selection techniques to enhance classification performance in Alzheimer's disease detection. These findings support the proposed methodology's promise as a reliable tool for AD analysis and contribute to the results of previous investigations.

Table 3.3: Approach one comparison with other studies.

Methodology	Database	Accuracy
[50]	Kaggle	86.73%
[98]	ADNI	95.0%
HOG_BCFCM_SVM [8]	Kaggle	96.8%
HOG_BCFCM_SVM [8]	ADNI	96.0%

3.3 Frequency Domain Case Study

Transforming images from the spatial domain to the frequency domain enables the analysis of MRI image information as constituent frequencies or patterns. This approach allows for examining MRI images from various perspectives, emphasizing frequency content rather than pixel values. The transformation is performed using the Fast Finite Shearlet Transform (FFST), a type of wavelet transformation. We have selected this method due to wavelets' significant advancements in signal and image processing, primarily through sparse approximations and improved handling of singularities compared to Fourier methods. Their success is also attributed to the existence of fast algorithms that accurately digitize continuous transforms. However, wavelets struggle with multivariate data due to their isotropic nature, mainly when dealing with distributed singularities, such as edges along curves. Recognizing the need for directional sensitivity, researchers have developed several versions of directional wavelets, including steerable pyramids [99], directional filter banks [100], and 2D directional wavelets[101]. Although these approaches have been beneficial regarding anisotropic features in multivariate data, they fail to provide optimally sparse approximations.

Curvelets [102] were introduced as a significant improvement over wavelets for handling anisotropic features in bivariate functions. Curvelets are defined at multiple scales, locations, and orientations, with supports that become more elongated at finer scales, making them practical for images with edges. However, curvelets are not singly generated and involve rotations that do not preserve the digital lattice, complicating the transition from continuous to digital settings. To overcome the rotation problem, contourlets [103] was introduced as a discrete, filter-bank version of curvelets, allowing efficient numerical algorithms through tree-structured filter-bank implementation. However, contourlets need a proper continuum theory.

The field has continued to evolve, leading to the introduction of shearlets [104],[105] to overcome the limitations of previous methods. Shearlets extend wavelets within a larger class of affine-like systems. They use shearing instead of rotation for directional selectivity, which allows them to be derived from a finite set of generators and preserves the digital lattice. This offers a unified approach to continuous and digital domains due to the shear matrix's preservation of the integer lattice.

Shearlets provide a combination of features that make them highly effective for approximating anisotropic features in multivariate data, supporting fast algorithmic implementations, and maintaining a unified treatment of continuous and digital realms. However, despite their potential, there is a notable gap in the literature concerning exploring AD stage classification within the frequency domain, especially with the FFST used specifically for classification rather than feature fusion. Addressing this gap is essential to advancing the field. In response, the second approach proposed in this chapter focuses

on extracting low-level features, such as texture, in the frequency domain to enhance the performance of machine learning models and further investigate their application in AD classification. Before discussing the proposed approach, we will first delve into the FFST method, explaining it mathematically in the following section.

3.3.1 Shearlet transform

3.3.1.1 Continuous Shearlet Systems

The shearlet transform deals with the anisotropic singularities feature in images where the properties vary significantly in different directions. For example, edges in images are typical anisotropic singularities, as they are sharp changes in intensity that occur primarily in one direction by achieving optimally sparse approximations of signals by using an element consisting of waveforms to analyze and represent the signal that must adapt to its features. For anisotropic singularities, these elements must vary in:

- **Scale**: To capture details at different levels of resolution to generate elements at different scales
- **Orientation**: To detect features in various directions using an orthogonal operator.
- **Location**: Position the elements accurately over the entire image using a translation operator.

Therefore, the continuous shearlet transform is defined by using the scaling matrix A_a and a sear matrix S_s

$$A_a = \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}, a \in \mathbb{R}^+, \quad S_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}, s \in \mathbb{R}. \tag{3.5}$$

The translation operator is noted by T_t. Now for $\psi \in L^2(\mathbb{R}^2)$, the continuous shearlet system $SH(\psi)$ is defined by:

$$SH(\psi) = \left\{ \psi_{a,s,t} = T_t D_{A_a} D_{S_s} \psi : a > 0, s \in \mathbb{R}, t \in \mathbb{R}^2 \right\}. \tag{3.6}$$

Where: ψ is a generation function, and choosing the suitable ψ based on your specific problem is essential. Hence, the system $SH(\psi)$ satisfies a reproducing formula for $L^2(\mathbb{R}^2)$.

3.3.1.2 Fast Finite shearlet transform FFST

The shearlet transform employs a shearlet function that addresses anisotropic singularities by integrating scaling, orientation (via a shearing operator), and localization (using a translating operator).

When converting an image into the frequency domain, the process begins with decomposing it into low- and high-frequency subbands. Subsequently, a shear filter is applied to subdivide these subbands directionally at various scales. This methodology facilitates multiscale and multidirectional decomposition while maintaining the dimensions of the decomposed subbands in alignment with those of the original image.

To apply the FFST to our study, we used a computed spectrum as detailed in the [106] using the following equation: (Equation 3.6).

Since the number of parameters is important, the indexing is used to reduce the number of parameters needed to compute the continuous shearlet transform; therefore, for the low

pass part of the image, the index i is equal to 1 which is considered as one index to the low pass however there are two indices of each frequency band with the diagonals bands the $k = +-2^j$and for the cone bands the index is equal to $2^j + 1$. At the same time, the scale is equal to $j = 2^j + 2$. based on this information with a scale $j = 0$ we get 4 band and with $j = 1$ we obtain 8 bands and with $j = 2$ we get 16 bands. additionally, with the maximum scale $j_0 - 1$ we obtain a number of indices $n = 2^{j0^{+2}} - 3$.

3.3.2 Co-occurrence Matrix Methods (GLCM)

GLCM is a statistical texture-based method employing a second-order conditional probability density combination. In a grayscale image, two pixels are correlated, leading to the definition of the cooccurrence matrix based on the joint probability density of pixels at two positions. The cooccurrence value represents the distribution of these values at a specific distance from a given pixel.

Mathematically, For an image $I(n, m)$, the co-occurrence matrix C_m of a distance $D(x, y)$ from the pixel central $I\,(n_c, m_c)$ and pixel k is :

$$C_M = C_{(D_x, D_y)}(k, k)$$

$$C_M = \sum_{n=1}^{k} \sum_{m=1}^{k} \begin{cases} 1 \text{ if } I(n, m) = k \text{ and } I\,(n + D_x, m + D_y) = k, \\ 0 \quad \text{otherwise} \end{cases} \tag{3.7}$$

Where:

$$D_x = D \cdot \cos(\theta)$$

$$D_y = D \cdot \sin(\theta)$$

Where:
θ : defines the direction of the matrix from the central pixel $I\,(n_c, m_c)$.

After computing the GLCM, statistical characteristics can be extracted to describe the image's texture.

In this study, 30 distinct features were derived from the GLCM, capturing various aspects of the spatial relationships between pixel intensities. These features include widely used metrics such as energy, entropy, contrast, correlation, and additional measures that provide a more comprehensive understanding of the image's texture. Each characteristic contributes to a representation of texture patterns, encompassing uniformity, randomness, and more complex statistical dependencies.

The mathematical equations for these features are presented in the following equations, illustrating the computation process from the GLCM matrix.

1. **Energy:** The energy measures the uniformity of the image's gray levels distribution (see equation 3.8).

$$\textbf{Energy} = \sqrt{\sum_{i,j=0}^{N_g - 1} (p(i, j))} \tag{3.8}$$

2. **Entropy:** measures the complexity of the texture within an image; the high entropy indicates a complex or heterogeneous texture, while a low entropy value indicates a less complex or more homogeneous texture. The following is the equation of the entropy.

$$\textbf{Entropy} \ = -\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} P_\delta(i,j)\cdot\log\left\{P_\delta(i,j)\right\}. \tag{3.9}$$

3. **Mean :** is defined as the average of all elements in the matrix, it indicate that the image is bright when the mean is a high value while when it a low value indicate that the image is more subdued. Mathematically, it can be expressed as:

$$\boldsymbol{Mean} = \frac{1}{N_p}\sum_{i=1}^{N_p} p(i) \tag{3.10}$$

4. **Root mean square (RMS)**: RMS calculates the magnitude of each pixel value, regardless of its sign. Hence, it indicates the brightness of the image.

$$\boldsymbol{RMS} = \sqrt{\frac{1}{N_p}\sum_{i=1}^{N_p}(X(i)+c)^2} \tag{3.11}$$

5. **standard derivation (SD):** SD measures the variability of pixel values in an image. A high standard deviation might indicate a textured image with varying pixel values, while a low standard deviation might suggest a smooth, uniform image.

$$\boldsymbol{SD} = \sqrt{\frac{1}{N_p}\sum_{i=1}^{N_p}(X(i)-\bar{X})^2} \tag{3.12}$$

6. **Skewness:** quantifies the asymmetry in a probability distribution, providing insight into its shape and deviation from a symmetrical bell curve.

$$\textbf{Skewness} \ = \frac{\frac{1}{N_p}\sum_{i=1}^{N_p}(X(i)-\bar{X})^3}{\left(\sqrt{\frac{1}{N_p}\sum_{i=1}^{N_p}(X(i)-\bar{X})^2}\right)^3} \tag{3.13}$$

7. **kurtosis:** This describes the shape of a probability distribution, emphasizing "peakedness" and "tailedness." A high kurtosis value indicates a distribution with a sharp peak and heavy tails, suggesting the presence of a few extreme pixel values far from the mean. In contrast, a low kurtosis value signifies a flatter distribution with lighter tails, indicating that the pixel values are more evenly distributed.

$$\boldsymbol{kurtosis} = \frac{\frac{1}{N_p}\sum_{i=1}^{N_p}(X(i)-\bar{X})^4}{\left(\frac{1}{N_p}\sum_{i=1}^{N_p}(X(i)-\bar{X})^2\right)^2} \tag{3.14}$$

8. **Variance:** Quantifies the spread or dispersion of data points around the mean. In images, high variance signifies a wide range of pixel values within a local area, indicating a textured image with varying intensities. In contrast, low variance suggests a more uniform region with similar pixel values, indicating a smoother texture.

$$\boldsymbol{variance} = \frac{1}{N_p}\sum_{i=1}^{N_p}(X(i)-\bar{X})^2 \tag{3.15}$$

9. **uniformity:** Uniformity describes the consistency or evenness of pixel values within a specific region or throughout the entire image. An image with high uniformity features similar pixel values across its area, whereas a non-uniform image displays significant variations in pixel values.

$$\boldsymbol{uniformity} = \sum_{i=1}^{N_g} p(i)^2 \tag{3.16}$$

With:
$\boldsymbol{X}$: a set of N_p Pixels included in the ROI. $P(i)$: The first order histogram. N_g: The number of non-zero bins. $p(i)$ the normalized first-order histogram.

10. **Autocorrelation:** analyzes the spatial relationships between pixels in an image. High autocorrelation values indicate a strong correlation between neighboring pixels, suggesting a smooth or regular texture, whereas low autocorrelation values signify a more random or textured pattern.

$$\boldsymbol{Autocorrelation} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)ij \tag{3.17}$$

11. **Cluster Prominence (CP):** This metric assesses the asymmetry and flatness of the pixel intensity distribution. It emphasizes the presence of outliers or notable variations in intensity levels, assigning greater weight to pixel pairs that diverge from the mean intensity. In more straightforward terms, it quantifies the degree to which gray-level values cluster around the mean and the extent of that clustering.

$$\boldsymbol{CP} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - \mu_x \ - \mu_y)^4 \, p(i,j) \tag{3.18}$$

12. **Cluster Shade:** quantifies the skewness or asymmetry of the pixel intensity distribution within an image. It highlights the degree to which the gray levels of neighboring pixels deviate from the mean, mainly focusing on regions characterized by skewed or non-uniform textures. Essentially, it aids in determining whether the texture exhibits a bias toward specific intensity values, such as brighter or darker areas.

$$\textbf{Cluster Shade } = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \Big(\boldsymbol{i} + \boldsymbol{j} - \boldsymbol{\mu_x} - \boldsymbol{\mu_y}\Big)^3 \, p(\boldsymbol{i},\boldsymbol{j}) \tag{3.19}$$

13. **cluster tendency:** Quantifies the degree to which gray levels in an image cluster around specific values. This metric indicates the similarity of pixel intensities within a local neighborhood, reflecting the image's overall texture smoothness and regularity.

$$\textbf{Cluster tendency } = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \Big(\boldsymbol{i} + \boldsymbol{j} - \boldsymbol{\mu}_x - \boldsymbol{\mu}_y\Big)^2 \, p(i,\boldsymbol{j}) \tag{3.20}$$

14. **Contrast:** is the difference in brightness or color between an image's lightest and darkest areas. A high-contrast image displays a broad range of tones, whereas a low-contrast image exhibits a more restricted range.

$$\boldsymbol{Contrast} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (\boldsymbol{i} - \boldsymbol{j})^2 p(i,j) \tag{3.21}$$

15. **Correlation:** It illustrates the relationship between two variables and quantifies the degree to which their values change in tandem.

$$\boldsymbol{Correlation} = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{\boldsymbol{N}_g}\boldsymbol{p(i,j)ij} - \boldsymbol{\mu_x\mu_y}}{\boldsymbol{\sigma}_x(i)\boldsymbol{\sigma}_y(j)} \tag{3.22}$$

16. **Difference average (DA):** measure that calculates the average difference between pairs of values in an image

$$\boldsymbol{DA} = \sum_{k=0}^{Ng-1} kp_{x-y}(k) \tag{3.23}$$

17. **Difference variance (DV):** It measures the variation in differences between gray levels and quantifies the texture's roughness or smoothness in an image.

$$DV = \sum_{k=0}^{N_g-1} (k - DA)^2 p_{x-y}(k) \tag{3.24}$$

18. **difference entropy (DE):**Calculates the entropy, a metric that quantifies the randomness of differences in gray levels between neighboring pixels in an image. This value reflects the degree of unpredictability in the variations of these gray-level differences.

$$DE = \sum_{k=0}^{N_g+1} p_{x-y}(k)\log_2\left(p_{x-y}(k) + \epsilon\right) \tag{3.25}$$

19. **Dissimilarity :** Dissimilarity is a metric that quantifies the differences in gray levels between pairs of pixels. It measures the variation in neighboring pixel values,

$$\textbf{dissimilarity} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}|i-j|p(i,j) \tag{3.26}$$

20. **Joint energy:** Joint Energy, also called Angular Second Moment (ASM), is the sum of squared elements in the Gray Level Co-occurrence Matrix (GLCM). A higher Joint Energy value signifies a more uniform texture, where specific gray-level pixel pairs are more frequent. Conversely, a lower Joint Energy value indicates a more significant variation in gray-level relationships, reflecting a more complex texture.

$$\boldsymbol{jointenergy} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(p(i,j))^2 \tag{3.27}$$

21. **joint entropy :** Joint entropy is a statistical measure that quantifies the randomness or uncertainty in the relationships between pixel intensities within an image. It is calculated as the negative sum of the product of the joint probability of each pixel intensity pair and the logarithm of that probability. A higher joint entropy indicates greater randomness, suggesting a lack of strong patterns or correlations between pixels. Conversely, a lower joint entropy signifies more structure or predictability in the distribution of pixel intensities, which may reveal regular patterns or textures in the image.

$$\boldsymbol{jointentropy} = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}p(i,j)\log_2(p(i,j)+\epsilon) \tag{3.28}$$

22. **Information measured of correlation 1 (IMC1):** It evaluates the relationship between the information content of an image's gray levels and the probability distribution of co-occurring pixel pairs. IMC1 quantifies how much an image's texture can be predicted from the gray-level distribution of its neighboring pixels.

$$\boldsymbol{IMC1} = \frac{HXY - HXY1}{\text{mzx}\{HX, HY\}} \tag{3.29}$$

23. **information measured of correlation 2 (IMC2):** Similar to IMC1, it evaluates the degree of correlation between the gray levels of pixel pairs; however, it utilizes a slightly different formulation that highlights the relationship between entropy and the joint probability distribution of these pixel pairs.

$$\boldsymbol{IMC2} = \sqrt{1 - e^{-2(HXY2 - HXY)}} \tag{3.30}$$

24. **inverse difference moment IDM:** It evaluates the similarity or uniformity of pixel values in an image by analyzing the gray levels of neighboring pixels. The Intensity Difference Metric (IDM) emphasizes pixel pairs with more minor intensity differences.

$$\boldsymbol{IDM} = \sum_{k=0}^{N_g - 1} \frac{p_{x-y}(k)}{1 + k^2} \tag{3.31}$$

25. **inverse difference (ID):** It evaluates the degree of similarity between the gray levels of neighboring pixels. The primary distinction between Inverse Difference (ID) and Inverse Difference Moment (IDM) lies in the fact that ID is less sensitive to more considerable intensity differences between pixel pairs, as it does not square the intensity differences, unlike IDM.

$$\boldsymbol{ID} = \sum_{k=n}^{N_g - 1} \frac{p_{x-y}(k)}{1 + k} \tag{3.32}$$

26. **inverse difference normalized IDN** It is a technique used in image processing. The method calculates the difference between each pixel's intensity and the intensities of its neighboring pixels, inverts these differences, and adds them back to the original pixel intensities. This process amplifies the contrast between dark and bright areas of the image, making details easier to distinguish.

$$\boldsymbol{IDN} = \sum_{k=0}^{N_g - 1} \frac{p_{x-y}(k)}{1 + \left(\frac{k^2}{N_g^2}\right)} \tag{3.33}$$

27. **inverse variance :** is a statistical measure used to assess the uniformity of pixel intensities within an image. It helps distinguish between different types of textures based on their level of uniformity. Higher IV values are associated with more uniform textures, while lower values indicate more textured or patterned regions.

$$\boldsymbol{inverseVariance} = \sum_{k=0}^{N_g - 1} \frac{p_{x-y}(k)}{k^2} \tag{3.34}$$

28. **sum average SA:** is a statistical measure employed to quantify the overall intensity level of pixel pairs within an image. A high SA indicates a greater overall intensity level, suggesting a higher concentration of bright or intense pixels. In contrast, a low SA signifies a lower overall intensity level, indicating a predominance of dark or less intense pixels.

$$\boldsymbol{SA} = \sum_{k=2}^{2N_g} \boldsymbol{p}_{x+y}(k)k \tag{3.35}$$

29. **sum entropy SE:**is a statistical measure used in Gray Level Co-occurrence Matrices (GLCMs) to quantify the randomness of intensities within an image.

$$\boldsymbol{SE} = \sum_{k}^{2N_g} p_{x+y}(k) \log_2 \Big(\boldsymbol{p}_{x+y}(\boldsymbol{k}) + \boldsymbol{\epsilon}\Big) \tag{3.36}$$

30. **sum variance SV**: is a statistical measure utilized to quantify the variability of intensities within an image.

$$\boldsymbol{SV} = \sum_{k=2}^{2N_g} (\boldsymbol{k} - \boldsymbol{SA})^2 p_{x+y}(\boldsymbol{k}) \tag{3.37}$$

WITH:

$$\epsilon \approx 2.2 \times 10^{-16}$$

$\boldsymbol{P}(i,j)$: The co-occurrence matrix.
$p(i,j) = \frac{P(i,j)}{\sum P(i,j)}$ The normalized co-occurrence matrix.
N_g The number of discrete intensity levels in image
$p_x(i) = \sum_{i=1}^{N_g} p(i,j)$ The marginal row probabilities.
$p_y(i) = \sum_{i=1}^{N_g} p(i,j)$ The marginal column probabilities.
$\mu_x = \sum_{i=1}^{N_g} p(i)i$ The mean gray level of the marginal row probabilities.
$\mu_y = \sum_{j=1}^{N_g} p(j)j$ The mean gray level of the marginal column probabilities.
σ_y Standard derivation of the marginal row probabilities.
σ_x Standard derivation of the marginal column probabilities.
$p_{x+y}(k) = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} p(i,j)$, **where** $i + j = k$, and $k = 2, 3, \ldots, 2N_g$
$p_{x+y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)$, **where** $|i - j| = k$, and $k = 0, 1, \ldots, N_g - 1$
$HX = -\sum_{i=1}^{N_g} p_x(i) \log_2 (p_x(i) + \epsilon)$ the entropy of the marginal row probabilities.
$Hy = -\sum_{j=1}^{N_g} p_y(j) \log_2 (p_x(j) + \epsilon)$ the entropy of the marginal column probabilities.

$$HXY = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \log_2(p(i,j) + \epsilon)$$

$$HXY1 = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \log_2 (p_x(i)p_y(j) + \epsilon)$$

$$HXY2 = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p_x(i)p_y(j) \log_2 (p_x(i)p_y(j) + \epsilon)$$

3.3.3 Frequency Domain-Based Low-Level Feature Extraction: Approach two

This chapter section presents the second approach[2], which emphasizes low-level features in the frequency domain.

This approach aims to investigate the utility of the frequency domain for extracting textural features and to analyze the impact and distribution of frequency information across eight directions of the FFST. We employed the Gray Level Co-Occurrence Matrix (GLCM) to extract 30 features, as outlined in the GLCM definition section, to obtain the textural features.

This approach is structured into five steps, beginning with preprocessing and culminating in classification. A visual representation of this methodology is provided in the following diagram.(Figure 3.7)

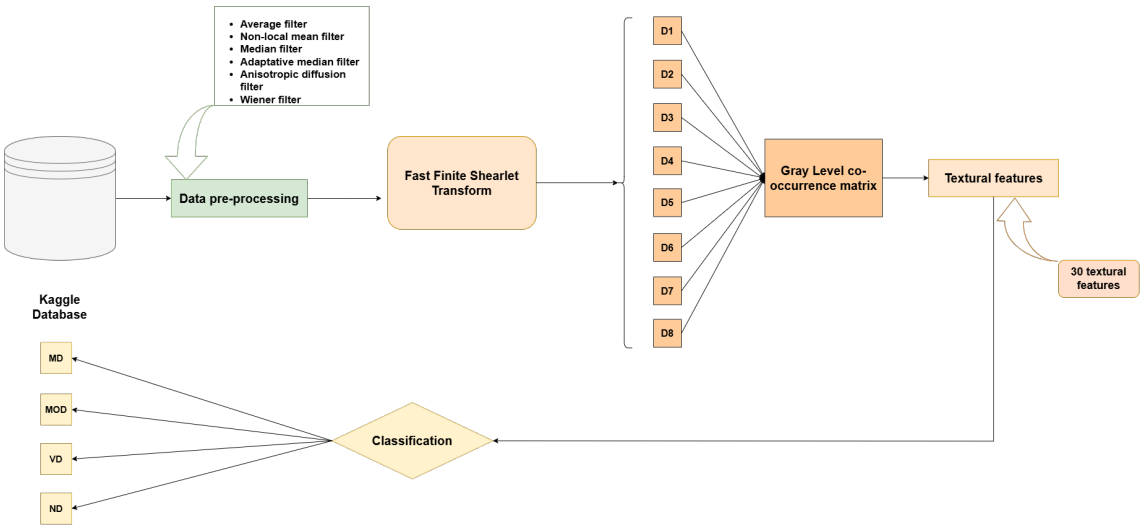


Figure 3.7: The flow diagram of the approach two:FFST-GLCM [9].

3.3.3.1 Data pre-processing

The primary objective of this phase is to enhance the image features essential for subsequent processing.

To accomplish this, we evaluated several filters, including the average filter, non-local mean filter, median filter, adaptive median filter, anisotropic diffusion filter, and Wiener filter. The performance of these filters was assessed using the Peak Signal-to-Noise Ratio (PSNR), resulting in a PSNR value of 46.3 when using the median filter. As a result, the median filter was chosen for noise removal, as it yielded the most favorable outcomes in our experiments.

[2]Saim, M., & Feroui, A. (2022, May). An Efficient Computer System for Alzheimer's Diseases Classification Using Fast Finite Shearlet Transform Domain and Support Vector Machine Classifier. In 2022 IEEE 9th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT) (pp. 337-342). IEEE.

3.3.3.2 The FFST and GLCM implementation

In this second proposed methodology (Figure 3.7), Alzheimer's disease is approached as both a 2-way and 4-way classification problem to explore the varying directions of the FFST and identify which direction provides the most informative features for each stage. To achieve this, six distinct sets—MD/MOD, ND/VD, MD/ND, MD/VD, MOD/ND, and MOD/VD—were examined. For each set, the dataset was partitioned, with 20% of the images randomly allocated for testing and the remaining 80% reserved for training. Following this, a 4-way classification was conducted to evaluate the performance across these configurations.

During the experimental phase, the k-fold cross-validation method was employed to assess the reliability and robustness of the classifier's results. This technique is particularly advantageous for small datasets, as it mitigates the risk of overfitting and ensures a more generalized evaluation. The process involves dividing the dataset into K equally sized, non-overlapping subsets. This research, a 10-fold cross-validation was applied to each of the six sets, ensuring a comprehensive evaluation of the model's performance. The outcomes of this analysis are detailed in Table 3.4, providing a clear comparison of the results across the different configurations.

Table 3.4: The 2-way and 4-way AD stage classification results for the eight directions of the FFST transform

Direction / acc	MD vs MOD	ND vs VD	MD vs ND	MD vs VD	MOD vs ND	MOD vs VD	4 way classification
D1	**0.7**	**0.8**	0.68	0.62	0 .67	0.72	0.7
D2	0.61	0.65	0.58	0.58	0.61	0.66	0.63
D3	0.67	0.69	0.61	0.63	0.69	0.65	0.66
D4	**0.72**	**0.82**	0.7	0.65	0.76	0.63	**0.72**
D5	0.69	0.79	0.6	0.66	0.78	0.7	0.7
D6	**0.72**	**0.82**	0.64	0.6	0.72	0.67	0.71
D7	0.6	0.78	0.53	0.59	0.68	0.6	0.64
D8	0.7	0.77	0.6	0.66	0.7	0.75	0.7

Table 3.4 illustrates the accuracy values for the classification processes conducted using the SVM classifier with a one-against-one strategy.

For the first set (MD vs. MOD), the highest accuracy was achieved with direction four (D4) of the FFST. In the second set, the best results were likewise obtained from D4 and direction six (D6). This indicates that both directions provide similar information that facilitates the distinction between ND and VD. However, D4 yielded notable results for the third set, which analyzes MD and ND exclusively, although the overall accuracy was lower than in the previous sets. This suggests that information from all directions is valuable for discriminating between these classes.

The next set, which examines the MD vs. VD stages of the disease, shows a similar trend: lower accuracy results compared to the first and second sets, with the highest accuracy occurring in direction five(D5). In analyzing the fourth set, direction eight (D8) also demonstrates accuracy comparable to D5, with only a slight difference, indicating that they possess similar discriminative information.

In the subsequent set, focusing on MOD vs. ND, the results again reveal that D5 yields the best outcomes. Finally, direction eight (D8) produced the highest accuracy in the last set.

In conclusion, directions two, three, and seven were not more accurate than the other directions for all sets. Direction four recorded the highest accuracy when addressing AD as a multiclass classification, while directions one, five, six, and eight also demonstrated commendable results. This confirms that the remaining directions do not contain highly discriminative information.

For further evaluation of the proposed methodology, the table 3.5 below provides a comparative summary of the results achieved by Approach 02 alongside those reported in the existing literature. This comparison underscores the promising potential of the proposed method, as it demonstrates competitive or superior performance in terms of accuracy across various classification tasks. The table highlights the effectiveness of the proposed approach, particularly in achieving higher accuracy rates in specific scenarios, such as distinguishing between ND and VD, where it attains an accuracy of 82%. These results suggest that the proposed methodology aligns with and advances the current state-of-the-art in Alzheimer's disease detection, offering promising results.

Table 3.5: Comparison of Approach 02 with previous studies.

References	**Method**	**Accuracy**
[107]	Ensemble SVM	55.6%
[46]	(WM & GM) Feature Extraction + Linear SVM	(AD/ND)=75.65% (AD/MCI)=61.65%
[108]	3D shearlet+GLCM+SVM ADNI DATASET OASIS DATASET	68% 68.8%
[9]	Approach 02	(MD vs MOD)=72% (ND vs VD)=82% (MD vs ND)=70.8% (MD vs VD)=65.3% (MOD vs ND)=76.2% (MOD vs VD)=63.7% **(Multi-SVM)=72.80%**

3.4 Approaches comparison

This chapter employed two distinct methodologies to tackle the challenge of extracting low-level features for AD stage classification. The first methodology utilized BCFCM segmentation, followed by extracting HOG features in the spatial domain. The second methodology leveraged the FFST combined with GLCM based texture analysis in the frequency domain.

The first methodology involved segmenting brain images into three fundamental tissue types: gray matter, white matter, and cerebrospinal fluid. This was achieved using BCFCM, which provides a biologically informed foundation for feature extraction. Such segmentation ensures that features are derived from well-defined brain regions, enhancing the relevance of the subsequent HOG features, which are particularly effective in capturing edge and shape information. HOG operates within the spatial domain, analyzing pixel intensities and gradients directly, thereby offering computational efficiency and ease of

implementation. This presents significant advantages in speed and simplicity, particularly when detecting structural patterns such as edges and boundaries.

However, this approach's notable limitation is its inability to capture complex textural details, as it primarily focuses on local changes in intensity. Consequently, classification performance may be constrained in tasks requiring fine-grained texture analysis, which can limit effectiveness in scenarios where detailed variations in tissue texture are essential for differentiating between pathological and healthy tissues.

The second methodology, which operates in the frequency domain, leverages the FFST to capture features across multiple scales and directions. This is followed by the extraction of 30 GLCM-based texture features per direction.

The frequency domain approach facilitates a more global image analysis, transforming it from the spatial domain into a representation that captures underlying patterns at varying scales and orientations. FFST is particularly adept at extracting directional and multi-scale features, enabling the identification of complex textural patterns that may not be evident in the spatial domain.

Integrating FFST with GLCM provides the benefit of capturing a comprehensive range of textural information across eight directional orientations, including contrast, correlation, homogeneity, and energy. This method is particularly advantageous in scenarios where subtle textural differences may signal early-stage pathological changes in AD. However, frequency domain techniques such as utilizing FFST tend to be more complex than the straightforward methodologies employed in spatial domain approaches (approach 1). While the extensive features extracted can enhance classification performance, analyzing information from multiple directions individually complicates the extraction of direct insights and decision-making regarding the optimal direction for final assessments. Furthermore, to fully utilize the potential of all eight directions, this approach requires further refinement, which involves additional post-processing and dimensionality reduction techniques to address the redundancy of information across directions, especially when examined separately. In addition to the reasons above, addressing the eight directions when dealing with big data makes the process arduous.

Regarding domain-specific advantages, the spatial domain approach (HOG) stands out for its efficiency and ease of interpretation. It is particularly suitable for tasks focused on edge detection and shape recognition. However, it is limited in capturing intricate texture details, essential for distinguishing between delicate tissue structures.

Conversely, the frequency domain approach (FFST-GLCM) provides a more detailed and comprehensive feature set, resulting in superior performance in texture-sensitive tasks. Each of these methodologies addresses specific needs and research questions within their respective domains. However, it is crucial to employ a method that prioritizes shape and textural features and extends beyond them, especially if the objective is to classify the early stages of AD effectively. Furthermore, the methodology should be suitable for big data applications and automatically generate features, as a significant drawback of these two approaches is the manual feature extraction process, which necessitates expertise in both contextual and technical aspects and is also time-consuming.

3.5 Summary

In this chapter, we propose two approaches to addressing low-level features in two different domains: the spatial and frequency domains. The first approach, which focuses on the spatial domain, investigates the impact of HOG features on classification results and

compares the effects of the segmentation step. This helps us address one of the research gaps in the second chapter.

In the second approach, we transformed to the frequency domain, using the FFST and investigating each direction coupled with GLCM. At the end of this chapter, we introduce a comparative analysis of both approaches.

With these two approaches, we have fulfilled our objective and addressed one of the gaps listed in the chapter 2, leading us to another research question discussed in the next chapter.

CHAPTER 4

ALZHEIMER'S DISEASE STAGE CLASSIFICATION BASED ON HIGH-LEVEL FEATURES: TRANSFER LEARNING APPLICATION

4.1 Overview

It is crucial to focus on discriminative features to develop a sophisticated model for classification tasks. These features can be obtained through various methods, including feature extraction and selection (as discussed in Chapter 03) and/or convolutional neural networks (CNNs).

In many applications and studies, CNNs are particularly valued for their ability to learn high-level feature representations that enable straightforward classification, even with simple classifiers like softmax. However, training CNNs to achieve this level of performance often requires extensive, well-balanced datasets. This poses a significant challenge in the medical field, where data is typically scarce and imbalanced, hindering the ability to train models effectively.

Transfer learning has emerged as a widely adopted approach to address these limitations. By leveraging pre-trained CNN models trained on extensive datasets, transfer learning facilitates the adaptation of these models to new tasks by fine-tuning them on smaller, domain-specific datasets.

This chapter investigates the application of transfer learning for multiclass classification in AD, with a primary focus on inductive transfer learning. It explores the effectiveness of this approach in addressing data scarcity and class imbalance. Furthermore, this chapter examines the layer-wise fine-tuning strategy, highlighting the lack of application of this transfer learning method, particularly in adapting transfer learning architectures for multiclass classification in various stages of AD.

Highlight

- We utilize ten pre-trained CNN models through inductive transfer learning to adapt knowledge from ImageNet datasets to our specific database.
- We employ a layer-wise fine-tuning approach to adapt pre-trained CNN models.

- We extensively evaluate and compare the ten CNN models.

4.2 Convolution Neural Network

Convolutional Neural Networks CNNs represent an advanced evolution of traditional artificial neural networks. These networks are designed to process and analyze data with a grid-like structure, such as images comprising interconnected processing layers that independently analyze images to extract hierarchical representations. The lower layers focus on identifying low-level features. Meanwhile, the deeper layers synthesize these elements to discern complex patterns (high-level features). Its powerful technique comprises the use of backpropagation algorithms, which iteratively adjust internal parameters, allowing the network to refine the output of each layer based on the preceding one.

Structurally, CNNs are constructed using a sequence of layers, primarily convolutional layers, activation functions, pooling, and fully connected layers.

4.2.1 Convolutional Layer

The convolutional layer [109] is a fundamental building block of CNNs employing small filters (generally 3∗3) to extract meaningful features from the input data. Each filter slides across the input, performing an operation known as convolution, defined mathematically as:

$$y(i,j) = \sum_{m} \sum_{n} x(i+m, j+n) \cdot w(m,n) + b \tag{4.1}$$

Where:

- $x(i,j)$ represents the input.
- $w(m,n)$ denotes the filter weights.
- b is the bias term, and $y(i,j)$ is the resulting feature map.

The accompanying figure 4.1 visually illustrates this process. The leftmost grid represents the input matrix, the middle section displays a 3×3 convolution filter, and the output grid illustrates the feature map. Each destination pixel in the feature map is calculated by performing a dot product between the filter's weights and the corresponding input pixels within the receptive field, followed by adding the bias.

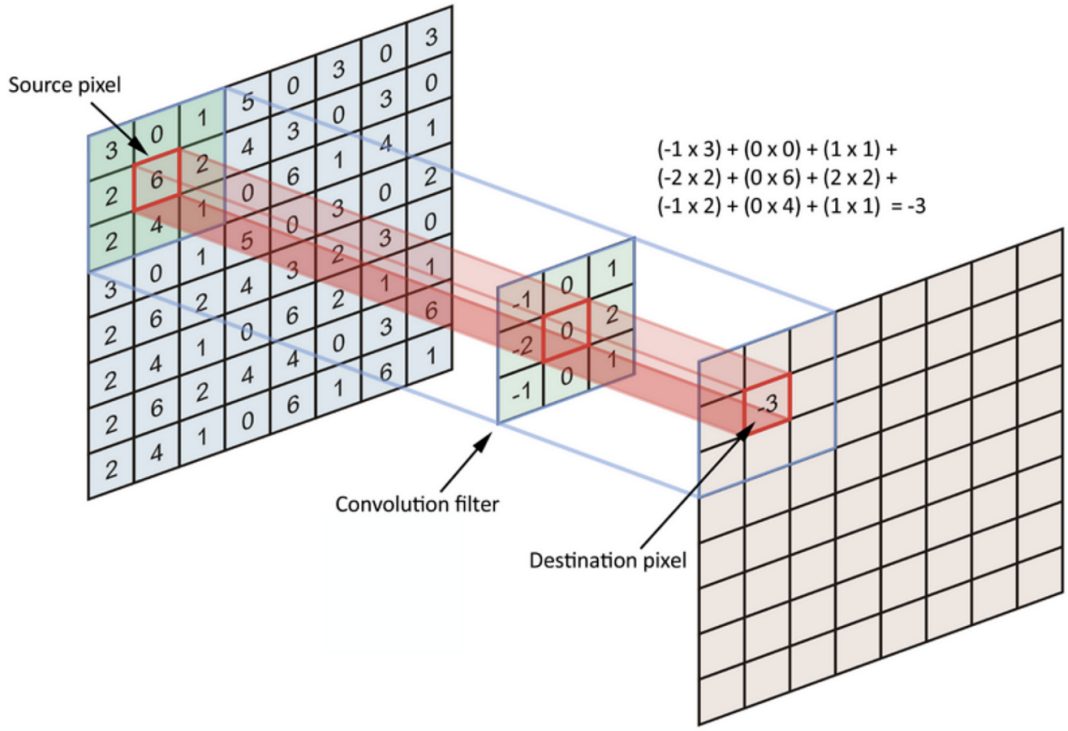


Figure 4.1: the operation of convolution [10]

This operation is repeated across the entire input matrix, moving the filter by the defined stride (step). By detecting spatial features, convolutional layers enable CNNs to learn hierarchical representations of the data. The filter values used in the CNN models are randomly specified to capture features like edge and sharpen, for example. Those filters are updated automatically and continuously as the network is trained.

4.2.2 Activation functions

Activation functions are essential for introducing non-linearity into CNNs, enabling the network to learn complex mappings between input and output. Without them, the entire model would behave as a linear system, regardless of depth.There are several activation functions, and choosing the right one depends on each CNN model's specific needs and objectives. More detailed information about activation functions can be found in[110]

4.2.3 Pooling layers

Pooling layers reduce the spatial dimensions of feature maps while retaining their most essential features, decreasing the computational load and mitigating the risk of overfitting. Max pooling, the most common pooling operation, selects the maximum value in a sliding window, while average pooling computes the average value. [109] The operation is defined as fellow respectively:

$$\text{Max pooling: } y(i,j,k) = \max_{m,n}(x(i+m, j+n, k)) \tag{4.2}$$

$$\text{Average pooling: } y_{i,j} = \frac{1}{K^2} \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} x_{s_i+m, s_j+n} \tag{4.3}$$

Where:K is the pooling kernel size, s is the stride, and $x(i+m, j+n, k)$ represents the feature values in the pooling window, pooling layers introduce translational invariance, ensuring that small spatial shifts in the input do not significantly affect the extracted features.

4.2.4 Fully connected (FC)

Fully connected (FC) layers are typically placed at the end of the network, and the extracted features are combined to produce the final prediction. Unlike convolutional layers, FC layers connect every neuron in one layer to every neuron in the subsequent layer, capturing global information about the input data.[109] Mathematically, the operation of a fully connected layer is expressed as:

$$z = W.x + b \tag{4.4}$$

Here, W represents the weight matrix, x is the input feature vector, b is the bias vector, and z is the output vector. In classification tasks, the output of the last FC layer is typically fed into a softmax function, producing class probabilities.

4.3 Transfer learning

Transfer learning is a machine learning approach that leverages knowledge gained from one task or dataset to enhance a model's performance on a different task or dataset, even when the two are unrelated. This technique improves a model's ability to generalize across various contexts, making it particularly valuable in deep learning applications[111].

By addressing challenges like limited data availability, high costs, and restricted accessibility, transfer learning establishes pre-trained CNN models as powerful tools for medical applications. There are various types of transfer learning. However, this thesis is interested in Inductive transfer learning.

Inductive transfer learning is a comprehensive concept in which a model is initially trained on a source task characterized by a large amount of labelled data, subsequently adapting to a target task that conventionally presents a scarcity of labelled instances.

The core principle of inductive transfer learning asserts a correlation between the source and target tasks. This enables the application of information gained from the source work to improve performance on the target task.

The purpose is to transfer the information included in the model parameters, representations, or features obtained during the training of the source task to the target task [112].

- $\mathcal{T}_S = \{\mathcal{X}_S, \mathcal{Y}_S\}$: Source task with input space $\mathcal{X}_S$ and label space $\mathcal{Y}_S$.
- $\mathcal{T}_T = \{\mathcal{X}_T, \mathcal{Y}_T\}$: Target task with input space $\mathcal{X}_T$ and label space $\mathcal{Y}_T$.

In inductive transfer learning, a model $f_S(x; \theta_S)$ is trained on the source task $\mathcal{T}_S$ to minimize:

$$\mathcal{L}_S = \frac{1}{n_S} \sum_{i=1}^{n_S} \ell\left(f_S\left(x_i; \theta_S\right), y_i\right) \tag{4.5}$$

where:
$x_i \in \mathcal{X}_S, y_i \in \mathcal{Y}_S$, and $\ell(\cdot, \cdot)$ is the loss function.

The learned parameters θ_S or representations $\phi(x)$ are transferred to the target task $\mathcal{T}_T$, where a new model $f_T(x;\theta_T)$ is fine-tuned or built to minimize:

$$\mathcal{L}_T = \frac{1}{n_T}\sum_{j=1}^{n_T} \ell\left(f_T\left(x_j;\theta_T\right), y_j\right) \tag{4.6}$$

With:
$x_j \in \mathcal{X}_T, y_j \in \mathcal{Y}_T$, and θ_T being initialized based on θ_S or the features $\phi(x)$.

4.3.1 Pretrained convolution neural network models

4.3.1.1 AlexNet:

The AlexNet architecture, created by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton [12], is a groundbreaking CNN model. It includes eight layers, five convolutional layers, and three fully connected layers.

The first convolutional layer features 96 filters with an 11x11 size and a 4-stride. The second convolutional layer contains 256 filters with a 5x5 size and a 1-stride, followed by the third and fourth convolutional layers with 384 filters of size 3x3. The fifth convolutional layer comprises 256 filters of size 3x3. Each convolutional layer is followed by a rectified linear unit (ReLU) activation function that introduces non-linearity into the model and improves its ability to learn intricate features.

AlexNet employs two max-pooling layers between the convolutional and fully connected layers, which diminish the spatial resolution of the feature maps and facilitate the model's invariance to small rotations and translations in the input images. The architecture also includes local response normalization, which standardizes the responses of neighboring neurons across various feature maps and enhances the model's generalization ability.[11]

The three fully connected layers each comprise 4096 neurons, followed by a softmax layer that produces the final classification output. AlexNet's design incorporates dropout regularization, which randomly omits neurons during training to prevent overfitting.

This architecture is considered a turning point in computer vision, inspiring a new generation of CNN models. The design principles of AlexNet, such as ReLU activations, max-pooling layers, and dropout regularization, have become conventional practices in CNN architectures, which led to the development of more intricate and advanced CNN models, such as VGG, GoogleLeNet, and ResNet, which exhibited even better performance on visual recognition tasks.

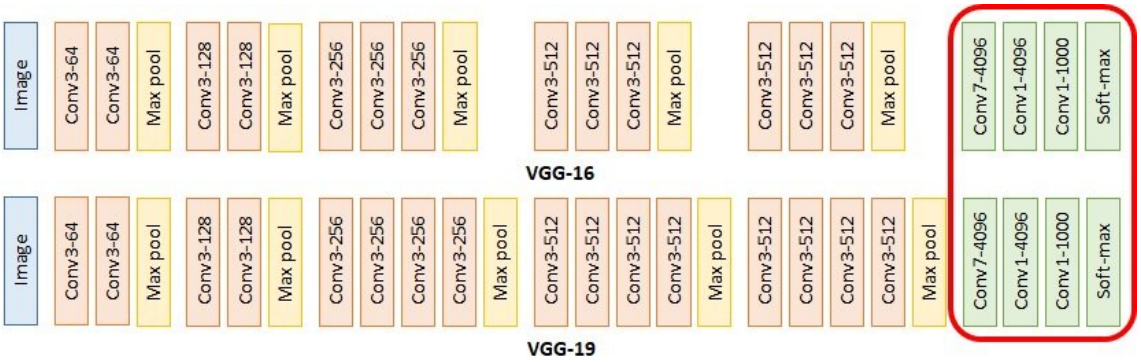


Figure 4.3: VGG16 and VGG19 architecture. [11]

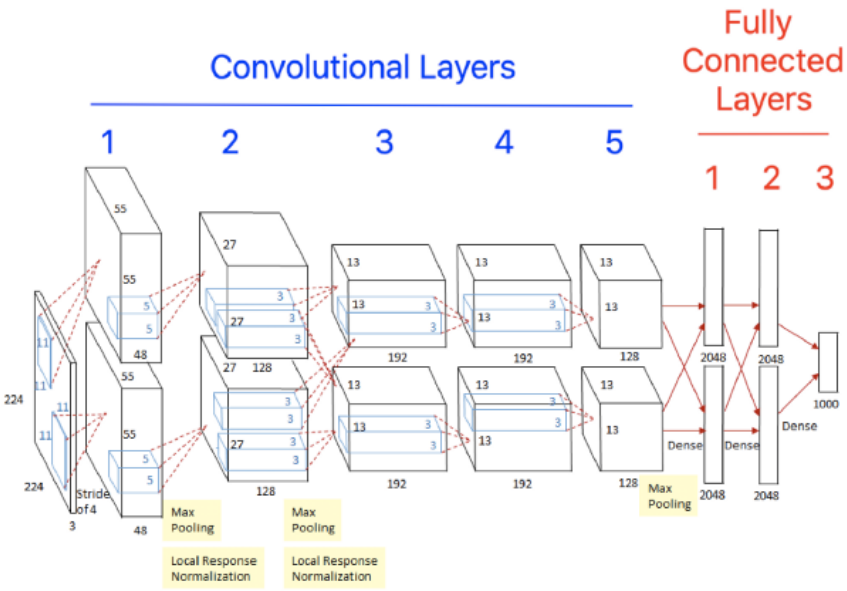


Figure 4.2: AlexNet architecture [12]

4.3.1.2 VGG16 and VGG19:

The Visual Geometry Group developed the VGG16 and the VGG19 at the University of Oxford.

These models have achieved widespread recognition. They comprise 16 and 19 layers, rendering them substantially deeper than prior CNN models such as AlexNet. The acronym VGG stands for Visual Geometry Group, while the numeric values denote the number of layers in each network.

These architectures are sequential convolutional layers followed by max pooling layers to extract features and reduce the number of parameters in the network while retaining the informational content from the input image. Specifically, these models utilize small 3x3 filters with a stride of 1 and zero padding to preserve the spatial dimensions of the input image. Subsequently, the networks deploy a sequence of fully connected (FC) nodes responsible for mapping the extracted features to a final output vector to represent the likelihood of the input image belonging to each classification class.[11]

4.3.1.3 Google Net:

GoogLeNet is a non-sequential convolutional neural network (CNN) that gained recognition at the ILSVRC-2014 contest for its performance in classification and detection tracks.

The network's innovative feature is the Inception module, consisting of 22 layers that utilize convolutional and pooling operations. This CNN architecture includes several auxiliary classifiers and processes input in a parallel workflow while using filters of different sizes in each layer. This enables it to capture patterns of varying scales and generate more comprehensive information. To minimize the number of parameters, 1x1 convolutional layers precede larger kernel convolutional filters (3x3 and 5x5). The ReLU activation function is used after 1x1 convolutional layers to enhance non-linearity and deepen the network. Unlike other CNNs, Google Net does not include fully connected layers and instead uses average pooling. Additionally, the bottleneck layer, which performs 1x1 convolutions on the input, aims to reduce the number of channels in the feature maps, thereby reducing the computational cost of the CNN. By decreasing the number of channels, CNN can extract more meaningful features with fewer parameters, further improving the efficiency of the network.[13]

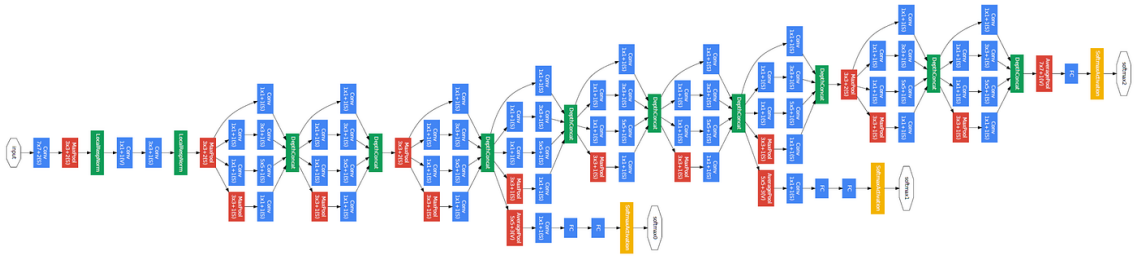


Figure 4.4: GoogleNet architecture.[13]

4.3.1.4 Inception V3

Inception-v3 was introduced by Google in 2015. It is a developed version of the Inception-v1 and Inception-v2 architectures.

The key features of Inception-v1 (GoogLeNet) are the "inception modules" and the auxiliary classifiers. However, the primary difference is that Inception V3 employs a more stringent factorization technique for convolutional layers than Inception V1, which means that the convolutional layers in Inception V3 have smaller filter sizes. Consequently, the number of parameters decreases, and computational efficiency improves. Secondly, Inception V3 uses batch normalization after each convolutional layer, which speeds up the training and boosts the accuracy of the network. [14]

The third difference is that Inception V1 includes two auxiliary classifiers, while Inception V3 uses a single auxiliary classifier. Regarding the stem, Inception V3 contains two 3 x 3 convolutions followed by a max pooling layer, whereas Inception V1 has a single 7 x 7 convolution.

Finally, Inception V3 uses average pooling instead of max pooling in the final layers of the network, which helps reduce overfitting and improves the accuracy of the network.[14]

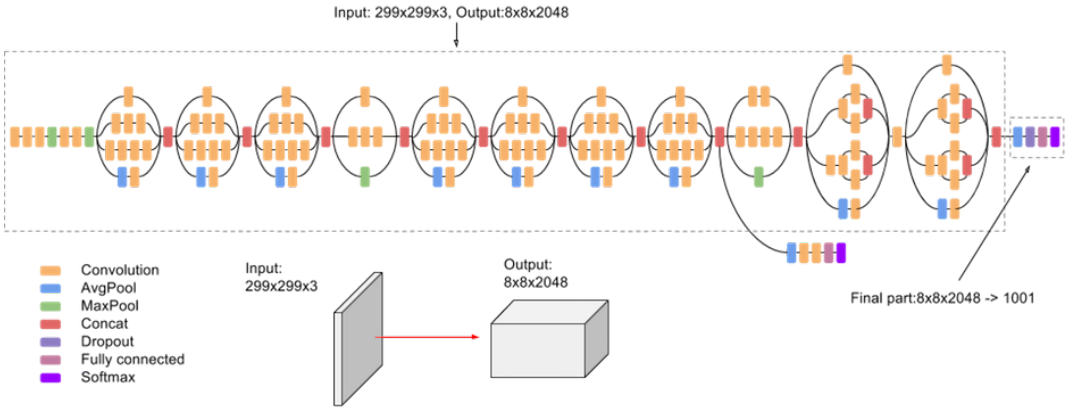


Figure 4.5: Inception V3 Architecture. [14]

4.3.1.5 Residual Network

Microsoft introduced the Residual Network in 2015. Generally, training a deep CNN model has two issues: a high training error and the vanishing gradient problem. This problem occurs when gradients become increasingly small during backward propagation in deep neural networks, limiting the network's effectiveness.

The Resnet architecture solves the vanishing gradient problems by using residual connections. It consists of multiple layers of convolutional and pooling operations, followed by fully connected layers for classification.

The residual block in ResNet adds the input to the output of one or more layers stacked on top of it, enabling faster training by allowing the gradient to flow directly. It has mainly 3x3 convolutional filters, which makes it similar to the VGG model but simpler.

This architecture has many variations, from depthless networks with only a few residual blocks to profound networks with hundreds of layers. ResNet 50 and ResNet 101 are variants of the ResNet architecture, with 50 and 101 layers, respectively. They both consist of multiple residual blocks with 16 residual blocks for ResNet 50. While ResNet 101 has 33 blocks.

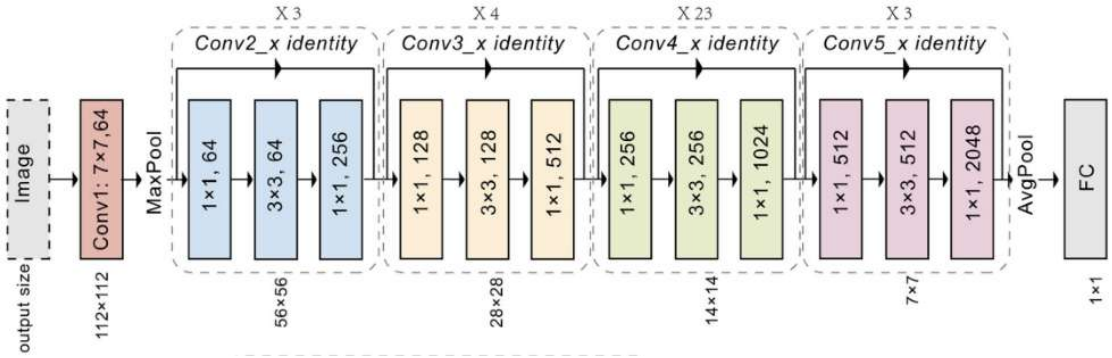


Figure 4.6: ResNet Architecture [15]

4.3.1.6 Xception:

François Chollet introduced Xception in 2016. It is designed to improve the efficiency of the Inception module by using depth-wise separable convolutional layers. Depth-wise

separable convolutions are a type of convolutional layer that can decompose into two separate operations: a depth-wise convolution, which applies a particular convolutional filter to each input channel, and a pointwise convolution, which combines the outputs of the depth-wise convolution into a new set of feature maps.

It's an architecture that combines the strengths of both Inception and ResNet architecture. Therefore, it's named Xception, which stands for Extreme Inception and is structured into two tiers.

The initial tier consists of a single convolutional layer using a filter with 1x1 dimensions.

The second tier encompasses three convolutional layers employing a filter with a 3x3 size. Within the two, an intermediate layer divides the output into three parts and transfers them to the next set of filters. Parting the convolutional operation can lower the network's computational complexity while preserving or enhancing its accuracy.

Since Xception employs residual connections, it allows learning residual functions to alleviate the problem of vanishing gradients and improve the flow of information through the network [86].

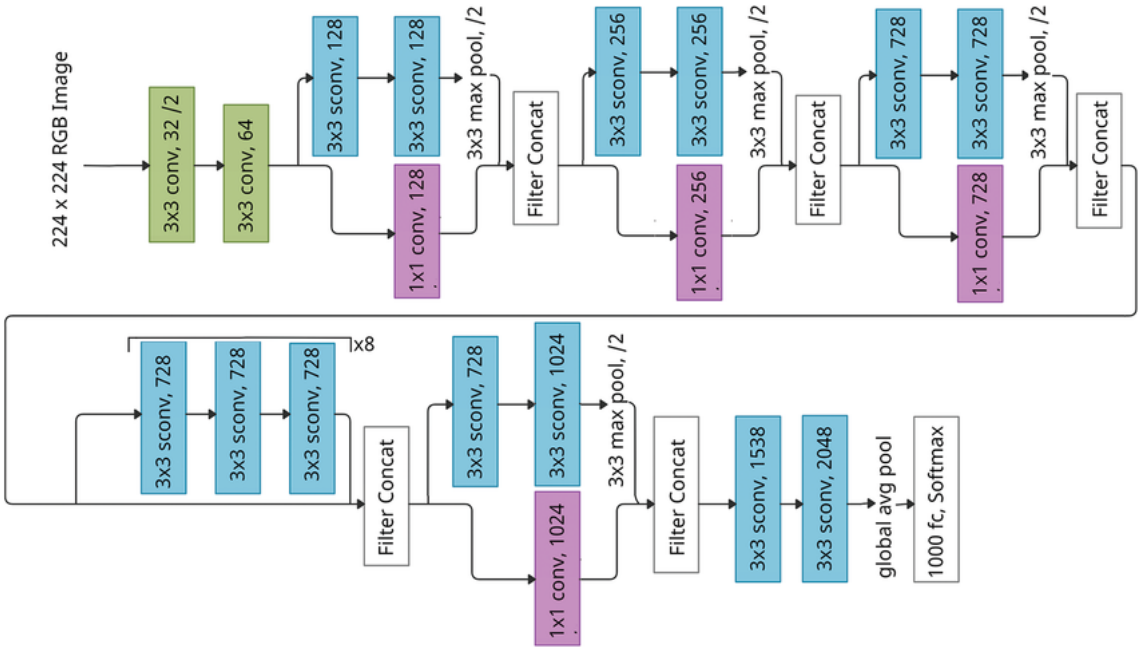


Figure 4.7: Xception architecture [16]

4.3.1.7 Efficient B0

Efficient B0 is a neural network architecture that has attracted considerable attention in computer vision and medical imaging. It is one of the EfficientNet models family, designed to achieve a prominent performance while utilizing minimal computational resources.

Efficient B0's primary objective is to provide an efficient and highly accurate model for image classification tasks. Its architecture employs a compound scaling method that optimizes the network's depth, width, and resolution to achieve uniform scaling and higher accuracy with fewer computational resources. Moreover, the model utilizes depthwise separable convolutions and mobile-friendly MBConv blocks, which enable exceptional accuracy with a single forward pass of 5.3 million parameters and 380 million FLOPs, which is a measure of the computational complexity of a neural network model, referring to the number of arithmetic operations performed per second, and is used to compare the

efficiency of different models and optimize their computational cost while maintaining accuracy.FLOPs stand for floating-point operations per second.

The depthwise separable convolution is used in Efficient B0's architecture, which decomposes the convolutional layer into distinct depthwise and pointwise convolutions, decreasing the model's computational cost while maintaining crucial spatial and channel-wise connections. The MBConv (Mobile Inverted Residual Bottleneck Convolution) blocks are a type of building block designed to capture spatial and channel-wise dependencies efficiently.

The depth-wise convolution applies a separate convolutional filter to each input channel, while the pointwise convolution reduces the number of channels in the output. The nonlinear activation function introduces nonlinearity to the network, enabling it to learn complex representations of the input data. The MBConv block's bottleneck structure helps reduce the model's computational cost by minimizing the number of parameters required for each layer. This is accomplished by using a 1x1 convolutional layer to reduce the number of input channels, followed by the depthwise convolution, and another 1x1 convolutional layer to increase the number of output channels.

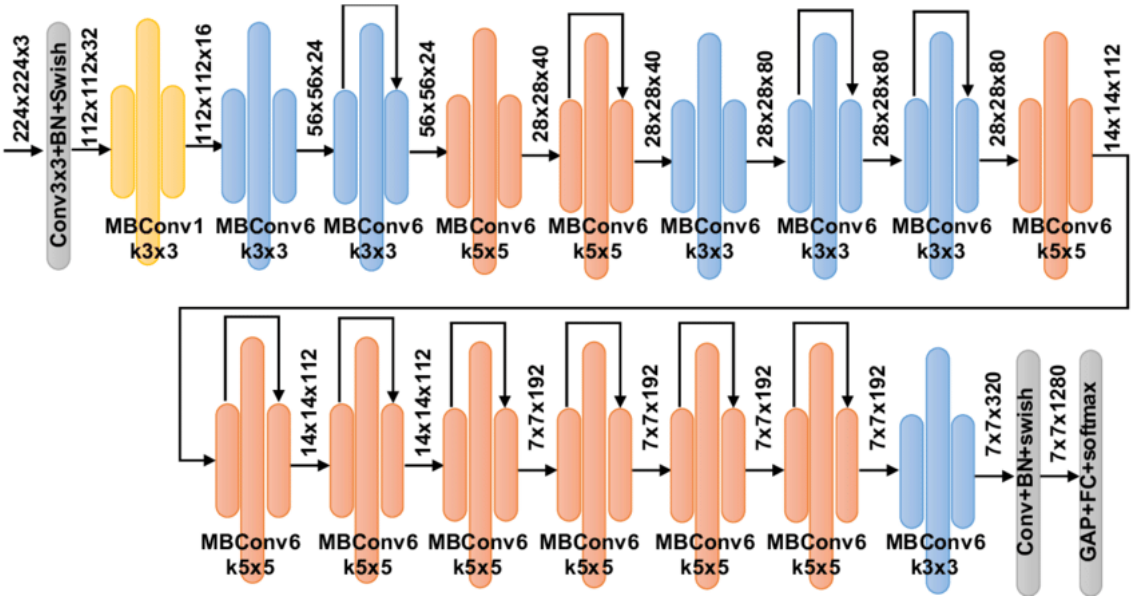


Figure 4.8: EfficientNet B0 architecture [17]

4.3.1.8 MobileNet

MobileNet is a CNN that uses depthwise separable convolutions, which reduce the number of parameters and computational cost. Depthwise separable convolutions consist of two steps: a depthwise convolution that applies a filter to each input channel, followed by a pointwise convolution that combines the outputs using a 1x1 convolution. MobileNet also introduces two multipliers, α and ρ, to adjust the number of channels and the input image resolution.[113]

MobileNetV2 builds upon MobileNetV1 by incorporating inverted residuals with linear bottlenecks. It includes shortcuts between thin bottleneck layers to create more efficient network structures. The linear bottlenecks in MobileNetV2 also ensure higher-dimensional internal representations, reducing memory usage while preserving information. The typical

architecture of MobileNetV2 includes an initial convolution layer, a series of inverted residual blocks, a final convolution layer, global average pooling, and an output layer for classification tasks.[18]

The key advantages of both versions of MobileNet are their efficiency, scalability, and competitive performance. This makes MobileNet highly suitable for applications that require deployment on devices with limited computational resources.

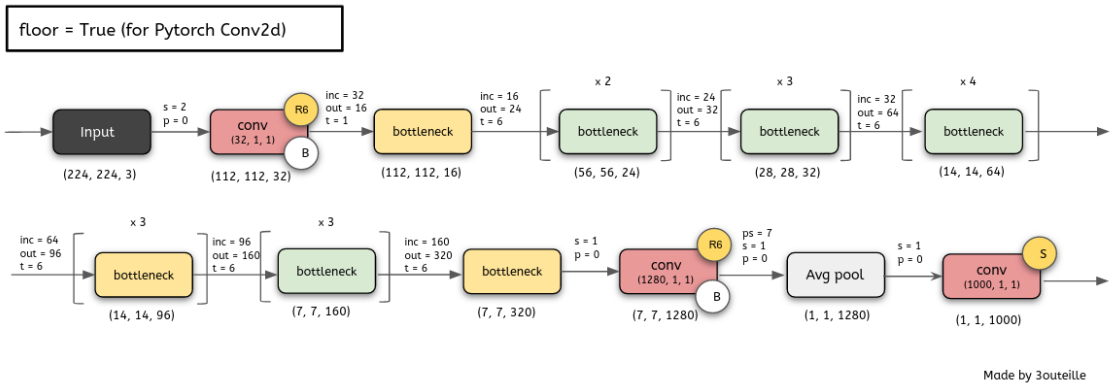


Figure 4.9: MobileNet V2 architecture [18]

4.4 Transfer learning: Alzheimer's disease case study

Inductive transfer learning leverages pre-trained models to address domain-specific tasks like ours. A key aspect of this approach is fine-tuning, which adapts pre-trained models to the target problem. The following section will present how we applied the fine-tuning strategy to classify the AD stage alongside the obtained results.

4.4.1 Fine-tuning pre-trained models

Model fine-tuning is adapting a pre-trained model to a specific task. It involves modifying particular layers of the pre-trained model while retaining the knowledge it gained during initial training on a large dataset. In practice, fine-tuning includes freezing layers to preserve the learned features and adding additional layers for the new task.

In this study, we employed the 10 previously defined pre-trained architectures and fine-tuned them consistently across two databases to investigate their performance for multiclass classification tasks.

Initially, we employ a transfer learning model by integrating a pre-trained base model with additional layers, including a pooling layer to reduce spatial dimensions, a regularization layer, and an output layer configured for multiclass classification. The model was compiled using the Adam optimizer, sparse categorical cross-entropy loss, and accuracy as the evaluation metric. The training was conducted over 50 epochs utilizing the training and validation datasets.

We unfroze specific layers of the pre-trained base model for fine-tuning while maintaining the earlier layers in a frozen state to preserve their learned features. The model was then recompiled with a lower learning rate to mitigate significant weight updates during fine-tuning. This was followed by an additional 50 training epochs, resulting in 100 epochs.

The uniform fine-tuning strategy ensured consistency when comparing the performance of different architectures across the two datasets. The following diagram (Figure 4.10) summarizes the model training and fine-tuning.

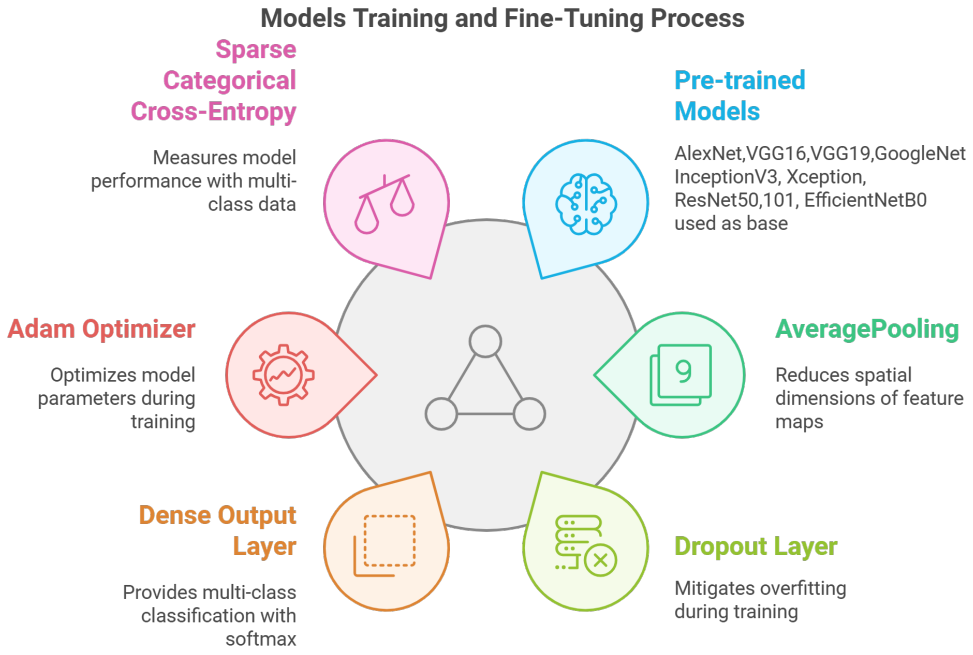


Figure 4.10: Process Pre-trained CNN architecture training and fine-tuning

4.4.1.1 Experiment and results

Experimental setup

In this chapter, we implemented transfer learning techniques utilizing the following CNN architectures: AlexNet, VGG16, VGG19, Inception V1 (GoogleNet), Inception V3, Xception, ResNet50, ResNet101, EfficientNet B0, and MobileNet V2. Each architecture was fine-tuned using a layer-wise unfreezing strategy.

1. **Preprocessing and Dataset Preparation:**
   The input images were resized to match the required dimensions of each pre-trained model. The majority of the architectures necessitate an input image size of $224 \times 224 \times 3$; however, Inception V1 and V3 require an image size of $299 \times 299 \times 3$, and AlexNet requires an input size of $227 \times 227 \times 3$. Furthermore, a data augmentation process was applied to the Kaggle database to address its significant class imbalance, as shown in Figure 4.11a.

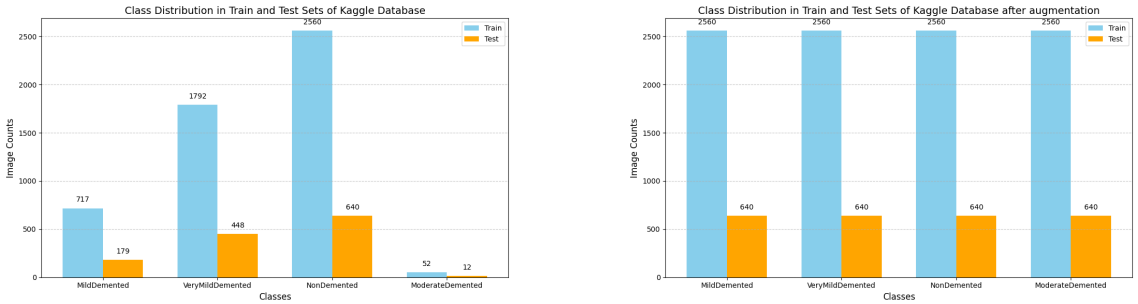
   We first identified the class with the highest number of samples in both the training and testing sets; in our case, the non-demented class in the training folder contained 2,560samplese. Consequently, we oversampled the other classes to equal 2,560 samples. Similarly, in the testing folder, which included 640 samples of the same class, we oversampled the remaining classes to reach 640 samples. All other classes were

augmented accordingly to ensure balance in both the training and testing sets, as depicted in Figure 4.11b. The augmentation process was inspired by Abderghal et al. [114]; however, we did not apply the blur effect to our images. Instead, we implemented rotation, flips (horizontal and vertical), and shear operations. Additionally, we adapted to apply an augmentation process based on the automatic choice of number by finding which class has the most samples. The data augmentation algorithm is shown below (Algorithm 2).

Algorithm 2 Dataset Balancing and Augmentation with Factor F

$\mathcal{D}$: Directories for train and test datasets, $\mathcal{C}$: Classes, F: Augmentation factor, S: Max-Shift. Balanced datasets with augmented images
AugmentImageAugmentImage BalanceClassBalanceClass CountImagesCountImagesInClasses
Procedure (image, S): Randomly shift the image by (dx, dy) within $[-S, S]$ Apply random horizontal flip Rotate the image randomly within $[-30°, 30°]$ Scale the image randomly between 80% and 120% Augmented image
Procedure ($\mathcal{D}_{\text{cls}}$, N, F): Compute $N_{\text{aug}} = N \cdot F -$ current size of $\mathcal{D}_{\text{cls}}$ $N_{\text{aug}} > 0$ Select a random image i from $\mathcal{D}_{\text{cls}}$ Apply (i, S) Save the augmented image with a new name $N_{\text{aug}} \leftarrow N_{\text{aug}} - 1$
Compute $N = \max\{\text{card}(\text{class\_1}), \text{card}(\text{class\_2}), \ldots, \text{card}(\text{class\_n})\}$ across all classes
class $\mathcal{C}_{\text{cls}}$ in $\mathcal{C}$ ($\mathcal{D}_{\text{cls}}$, N, F)
Procedure ($\mathcal{D}$): For each class, $\mathcal{C}_{\text{cls}}$ in $\mathcal{D}$, count images Dictionary of class names and image counts

Figure 4.11 compares how the dataset was distributed before and after augmentation, where figure 4.11b is the augmentation results.



(a) Kaggle database distribution

(b) Augmented database distribution

Figure 4.11: Kaggle Database distribution across the classes before and after augmentation

2. **Transfer Learning and Fine Tuning Process:**
The CNN model was pre-trained on ImageNet, and the models were imported without their top (fully connected) layers. Then, a layer-wise unfreezing fine-tuning strategy was applied, freezing the model's layers to preserve the pre-trained weights from the image. Only specific layers were unfrozen. This allows these layers to adapt to the particular characteristics of our dataset while the earlier layers (which

capture more generic features) remain frozen. By unfreezing only a subset of layers, we focus the fine-tuning process on the deeper layers. These layers are more specific to the ImageNet dataset and are more likely to benefit from adaptation to our task (AD stage classification).

3. **Training Protocol:**
   For all the architectures, both databases were split in the 80% train, and the 20% test, and the models were compiled using the Adam optimizer with a sparse categorical cross-entropy loss function. The primary evaluation metric was accuracy, while validation loss was monitored to track performance. The batch size was 32 for balanced memory usage and computational efficiency. The learning rate for the fine-tuning was reduced to 1×10^{-5}. The experiments were conducted on Google Colab, using TensorFlow as the primary deep learning framework.

Inductive transfer learning Results and discussion

This section presents experimental results from two databases, utilizing an inductive transfer learning approach with a layer-wise unfreezing fine-tuning strategy.

1. **Fine-tuning inductive transfer learning using the Kaggle database**

The following table 4.1 presents the results of ten pre-trained CNN architectures in seven metrics

Table 4.1: layer-wise fine-tuning pre-trained CNN models results using Kaggle database

Model	**Accuracy**	**Precision**	**Recall**	**F1score**	**Kappa**	**MCC**	**AUC**
AlexNet	55.31	53.83	55.32	53.89	40.42	40.76	81.43
VGG16	75.94	77.87	75.94	75.15	67.92	68.88	94.98
VGG19	**77.23**	**79.43**	**77.23**	**76.88**	**69.64**	**70.34**	**94.24**
GoogleNet	65.40	69.80	65.50	55.70	56.30	55.75	81.50
InceptionV3	69.49	71.35	69.49	69.70	59.32	59.74	90.29
Xception	71.99	73.00	71.99	71.84	62.66	63.02	91.35
ResNet50	69.02	74.81	69.02	69.66	58.70	60.55	91.53
ResNet101	74.22	76.24	74.22	74.42	65.62	66.17	92.82
EfficientNet	74.34	77.83	74.34	74.71	65.78	66.77	93.22
MobileNet V2	75.35	75.75	75.35	75.03	67.14	67.41	92.73

The results of the layer-wise fine-tuning of pre-trained CNNs models on an augmented Kaggle dataset are summarized in table 4.1. Among the models tested, VGG19 demonstrated the highest overall performance, achieving an accuracy of 77.23%, followed closely by VGG16 with 75.94% accuracy. These models excelled in accuracy and other key metrics, such as precision and recall. Specifically, VGG16 achieved the highest precision (77.87%) and recall (75.94%), indicating its strong ability to identify positive instances while minimizing false positives correctly. Moreover, VGG19 led with an F1 score of

76.88%, reflecting a good balance between precision and recall, confirming its robustness. On the other hand, AlexNet consistently underperformed across all metrics, with the lowest accuracy (55.31%), precision (53.83%), recall (55.32%), and F1 score (53.89%), suggesting its limited effectiveness for this task.

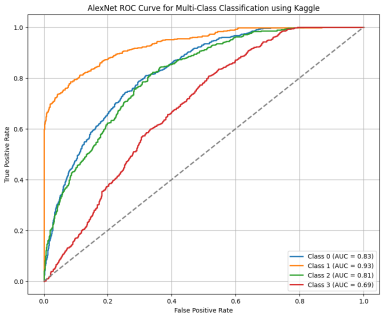
Regarding the Kappa statistic, which measures agreement between predicted and actual values, VGG19 again outperformed all other models with a Kappa value of 69.64, indicating substantial agreement between predictions and true class labels. Similarly, VGG16 achieved a respectable Kappa score of 67.92, while AlexNet showed the weakest Kappa score of 40.42, highlighting its poor agreement. The Matthews Correlation Coefficient (MCC), which evaluates binary classification quality, followed a similar pattern, with VGG19 and VGG16 achieving MCC values of 70.34 and 68.88, respectively, while AlexNet had the lowest MCC of 40.76.

Additionally, the Area Under the Curve (AUC) metric, which reflects the model's ability to discriminate between classes, reinforced the superior performance of VGG16 and VGG19, with AUC scores of 94.98% and 94.24%, respectively. In contrast, AlexNet lagged behind with an AUC of 81.43%, further supporting its underwhelming performance. Models such as ResNet101, EfficientNet, and MobileNet V2 also demonstrated commendable results, with ResNet101 achieving an AUC of 92.82%, indicating their competitive performance, particularly in distinguishing between classes.

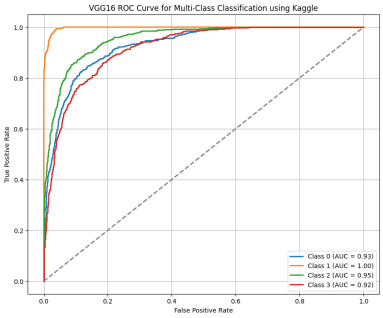
In conclusion, the fine-tuning results suggest that VGG16 and VGG19 are the most effective pre-trained CNN models for this classification task, exhibiting superior performance across all evaluated metrics. Conversely, AlexNet is unsuited for this dataset, showing the lowest values for accuracy, precision, recall, and other metrics. The findings highlight the importance of selecting an appropriate model for the specific characteristics of the dataset and classification task at hand.

In addition to the previously discussed metrics, the Roc curves are essential to further analyses of the discrimination of the model between each class presented in Figure 4.12, which illustrate the performance of various pre-trained CNN models when fine-tuned on the Kaggle database. These curves, which plot the True Positive Rate (sensitivity) against the False Positive Rate at different classification thresholds, provide a comprehensive assessment of the models' discriminative power. To further understand these results, the Area Under the Curve (AUC) values serve as a quantitative measure of each model's ability to differentiate between the four balanced classes.

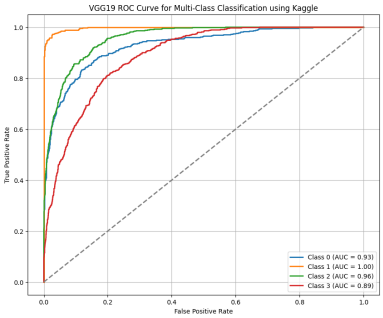
Starting with **AlexNet** (Figure 4.12a), the ROC curve indicates moderate classification performance. The relatively lower trajectory of the curves highlights the model's limited ability to maintain high sensitivity without a corresponding increase in false positives. Consequently, the AUC values are among the lowest, reflecting suboptimal separability across the classes.



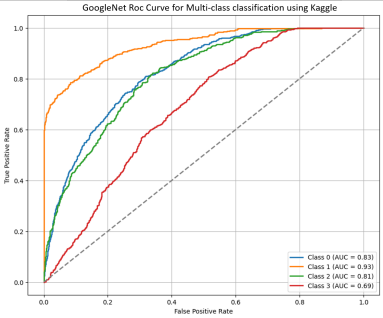
(a) AlexNet ROC curve



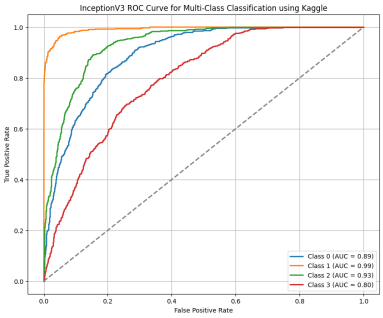
(b) VGG16 ROC curve



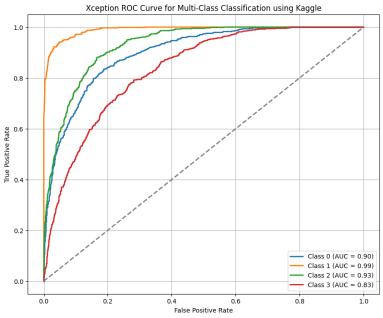
(c) VGG19 ROC curve



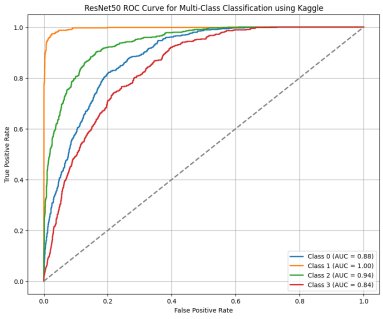
(d) GoogleNet ROC curve



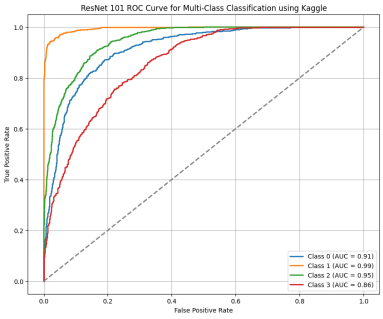
(e) Inception V3 ROC curve



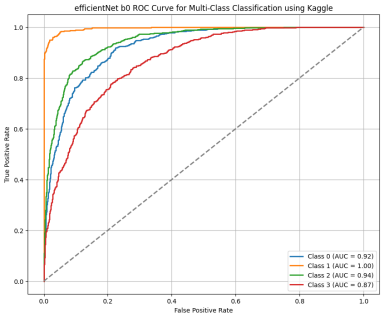
(f) Xception ROC curve



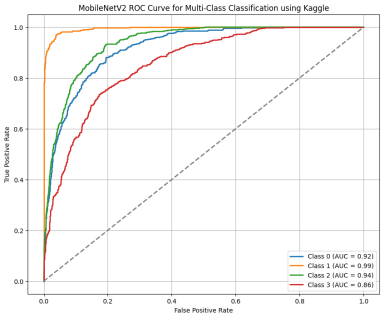
(g) ResNet 50 ROC curve



(h) ResNet 101 ROC curve



(i) EfficientNet B0 ROC curve



(j) MobileNet V2 ROC curve

Figure 4.12: layer-wise fine-tuning pre-trained CNN models ROC curves results using Kaggle database

In contrast, the VGG16 and VGG19 models (Figures4.12b and 4.12c) exhibit outstanding performance. Their ROC curves remain consistently close to the top-left corner, signalling a strong balance between sensitivity and specificity. Notably, these models achieve the highest AUC values, which align well with their superior accuracy and F1 scores observed in earlier analyses. These results emphasize the effectiveness of the deeper architectures in accurately classifying the balanced classes.

When evaluating GoogleNet (Figure 4.12d), a noticeable decline in performance becomes evident compared to the VGG models. The ROC curves for GoogleNet show a less pronounced ability to discriminate between classes, as reflected in its relatively lower AUC values. This suggests that the model struggles to achieve the same level of reliability as its deeper counterparts.

Similarly, the InceptionV3 and Xception models (Figures 4.12e and 4.12f demonstrate competitive performance, with Xception slightly outperforming InceptionV3. The ROC curves for these models are closer to the top-left corner compared to AlexNet and GoogleNet, indicating a more reliable classification performance. However, their AUC values remain slightly lower than those of the VGG architectures, signalling room for improvement.

Moving to the ResNet architectures (Figures 4.12g and 4.12h), both ResNet50 and ResNet101 exhibit solid performance, with ResNet101 showing superior results. The ROC curve of ResNet101 is closer to the ideal top-left corner, and its higher AUC values highlight its improved discriminative power compared to ResNet50.

Turning to EfficientNet B0 (Figure 4.12i), the model demonstrates commendable classification performance. Its ROC curve closely approaches the top-left corner, and the corresponding AUC values are among the higher-performing models. This result underscores EfficientNet's capability to balance sensitivity and specificity effectively.

Finally, the MobileNet V2 model (Figure 4.12j) delivers robust performance, with its ROC curve reflecting a good balance between sensitivity and specificity. While its AUC values are slightly below those of VGG19 and VGG16, MobileNet V2 remains a competitive lightweight model, offering reliable results with relatively less computational complexity.

The ROC analysis reveals significant variability in the performance of the tested models. VGG16 and VGG19 consistently outperform their counterparts, as indicated by their superior ROC curves and highest AUC values. Models such as EfficientNet B0, MobileNet V2, and ResNet101 also perform well, showing strong classification capabilities. Conversely, AlexNet and GoogleNet demonstrate limited discriminative power, as evidenced by their lower ROC trajectories and AUC values. These findings underscore the importance of employing deeper architectures, such as VGG and ResNet, for fine-tuning tasks in multi-class classification problems.

2. **Fine-Tunning inductive transfer learning using the ADNI database**

The results presented in Table 4.2 illustrate the performance of various deep learning architectures that have been fine-tuned using a layer-wise strategy for classifying stages of Alzheimer's Disease, utilizing the ADNI3 database. This database encompasses five classes, three particularly challenging to differentiate.Table 4.2 presents metrics that collectively evaluate the classification performance of the models in identifying AD stages.

Table 4.2: layer-wise fine-tuning pre-trained CNN models results using ADNI database

Model	Accuracy	Precision	Recall	F1score	Kappa	MCC	AUC
AlexNet	78.55	80.2	79	78.78	73.11	73.56	96.87
VGG16	47.66	52.89	46.19	45.97	33.43	34.21	77.14
VGG19	78.2	80.31	78.97	78.64	72.66	73.21	95.34
GoogleNet	70.02	72.20	71.50	70.50	67.70	68.10	95.80
InceptionV3	84.99	85.39	85.08	85.18	81.15	81.18	98.07
Xception	**85.19**	**85.92**	**85.12**	**85.43**	**81.38**	**81.42**	**97.67**
ResNet50	80.51	81.59	80.91	80.99	75.55	75.68	96.69
ResNet101	84.02	85.21	84.25	84.55	79.92	80	97.35
EfficientNet	75.63	76.93	75.55	75.94	69.42	69.53	95.21
MobileNetV2	79.27	80.77	79.25	79.5	74.02	74.27	95.01

The **AlexNet** model achieves moderate performance, with an accuracy of 78.55%, a precision of 80.20%, and an F1 score of 78.78%. Its Kappa (73.11) and MCC (73.56) values indicate reasonably strong agreement and correlation between predicted and actual labels. Furthermore, its AUC value of 96.87 reflects a high discriminative capability.

In stark contrast, **VGG16** performs poorly, with an accuracy of only 47.66% and a corresponding F1 score of 45.97%. Its low Kappa (33.43) and MCC (34.21) values suggest a weak agreement and correlation. Additionally, the AUC of 77.14 indicates a limited ability to distinguish between classes, making it the least effective model in this evaluation.

The **VGG19** model improves substantially over VGG16, achieving an accuracy of 78.20%, precision of 80.31%, and an F1 score of 78.64%. Its Kappa (72.66) and MCC (73.21) values are comparable to those of AlexNet, while its AUC of 95.34 is slightly lower but still indicative of good performance.

GoogleNet delivers an accuracy of 70.02%, which is lower than AlexNet and VGG19, but its AUC of 95.80 suggests strong discriminative power. However, its F1 score (70.50), Kappa (67.70), and MCC (68.10) point to a moderate overall performance.

The **InceptionV3** model demonstrates superior results, achieving the highest accuracy (84.99%) among all models. It also excels in precision (85.39%), recall (85.08%), and F1 score (85.18%). Its Kappa (81.15) and MCC (81.18) values indicate strong agreement and correlation, while its AUC of 98.07 underscores its exceptional classification performance.

Similarly, **Xception** performs exceptionally well, achieving the highest accuracy (85.19%) in the table. It also scores slightly higher than InceptionV3 in precision (85.92%) and F1 score (85.43%), while maintaining comparable Kappa (81.38) and MCC (81.42) values. Its AUC of 97.67 further highlights its strong discriminative ability.

The ResNet architectures also perform well, with **ResNet50** achieving an accuracy of 80.51% and **ResNet101** reaching 84.02%. ResNet101 exhibits slightly superior results compared to ResNet50 in all metrics, including precision (85.21%), F1 score (84.55%), and AUC (97.35%). These results suggest that deeper architectures within the ResNet family provide better classification capabilities.

EfficientNet achieves an accuracy of 75.63%, which is lower than ResNet and Xception. Its F1 score (75.94) and AUC (95.21) indicate moderately strong performance,

though it does not match the effectiveness of deeper models.

Lastly, **MobileNetV2** achieves an accuracy of 79.27%, with an F1 score of 79.50%. While its Kappa (74.02) and MCC (74.27) values reflect strong performance, its AUC (95.01) is slightly lower than other high-performing models such as Xception and InceptionV3.

In summary, Xception and InceptionV3 emerge as the top-performing models, demonstrating consistently high performance across all metrics. ResNet101 also delivers strong results, though slightly below Xception and InceptionV3. By contrast, VGG16 performs the worst, underscoring the limitations of this architecture for the given classification task. These findings highlight the importance of model architecture depth and complexity in achieving superior classification performance in Alzheimer's disease diagnosis using the ADNI database.

To thoroughly evaluate the performance of the models and given the focus on multi-class classification, Receiver Operating Characteristic (ROC) curves were incorporated into the analysis.

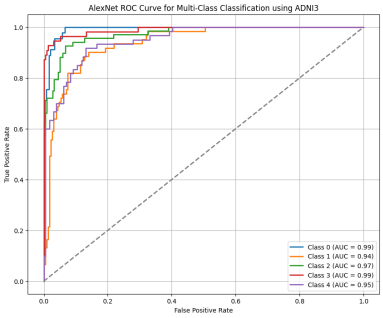
Figure 4.13 presents the ROC curves generated from the layer-wise fine-tuning of various pre-trained CNNs using the ADNI3 database. Each subfigure corresponds to a specific CNN architecture: AlexNet (Figure 4.13a), VGG16 (Figure 4.13b), VGG19 (Figure 4.13c), GoogleNet (Figure 4.13d), Inception V3 (Figure 4.13e), Xception (Figure 4.13f), ResNet50 (Figure 4.13g), ResNet101 (Figure 4.13h), EfficientNet B0 (Figure 4.13i), and MobileNet V2 (Figure 4.13j). These ROC curves illustrate the trade-off between sensitivity (true positive rate) and specificity (1 - false positive rate) for different classification thresholds across the models.

Most models exhibit ROC curves close to the upper-left corner of the graph, signifying strong discriminative performance. Notably, ResNet101 and Xception achieve higher Area Under the Curve (AUC) values, reflecting superior classification accuracy compared to other models. The observed variability in performance among the models highlights the influence of architectural differences and their feature extraction capabilities. The consistent trends across multiple models emphasize the effectiveness of pre-trained CNNs, fine-tuned for Alzheimer's Disease (AD) classification tasks.

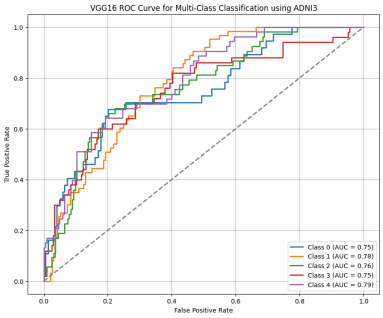
Among the evaluated models, the Xception model demonstrates particularly strong performance in classifying the stages of Alzheimer's Disease using the ADNI3 database. Its ROC curve is prominently positioned near the upper-left corner, indicating high sensitivity and specificity across various thresholds. This is corroborated by the AUC values for each class: Early Mild Cognitive Impairment (EMCI) with an AUC of 99%, Mild Cognitive Impairment (MCI) with an AUC of 96%, and Late Mild Cognitive Impairment (LMCI) with an AUC of 98%. These results confirm the model's robust ability to differentiate between cognitive impairment stages.

The Xception model's strong performance can be attributed to its advanced architecture. This architecture employs depthwise separable convolutions to enhance feature extraction while minimizing computational complexity. This enables the model to capture subtle and discriminative patterns in the data, which are crucial for accurately identifying Alzheimer's Disease stages. Furthermore, the layer-wise fine-tuning process optimized the pre-trained weights for the specific characteristics of the ADNI3 dataset, further enhancing its classification performance.

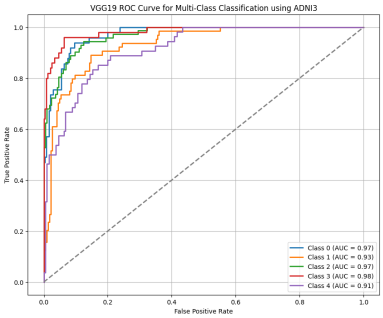
In the context of Alzheimer's Disease classification, the Xception model achieves an exceptional balance between sensitivity and specificity for each class, as evidenced by its consistently high AUC values.



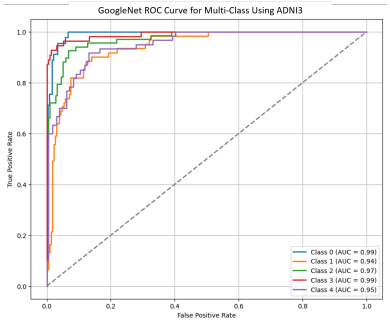
(a) AlexNet ROC curve results



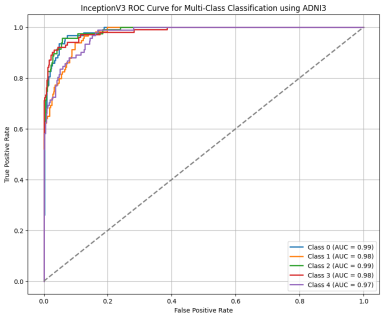
(b) VGG16 ROC curve results



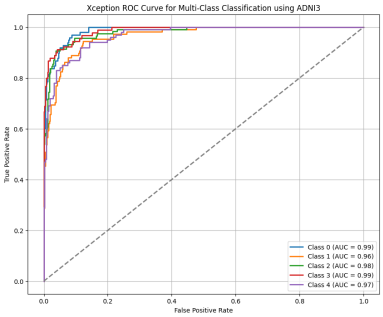
(c) VGG19 ROC curve results



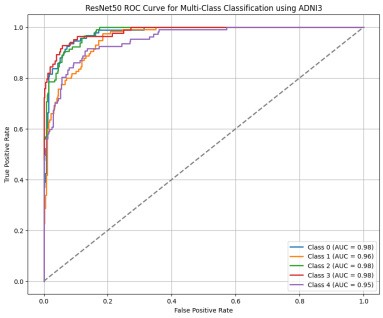
(d) GoogleNet ROC curve results



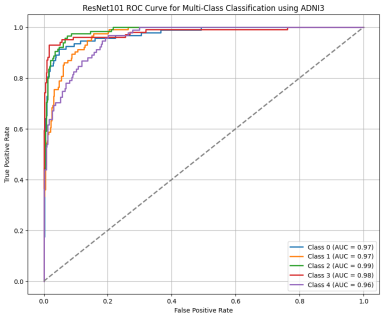
(e) Inception V3 ROC curve results



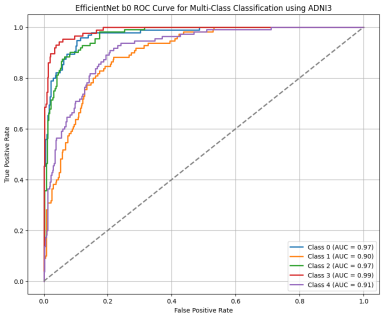
(f) Xception ROC curve results



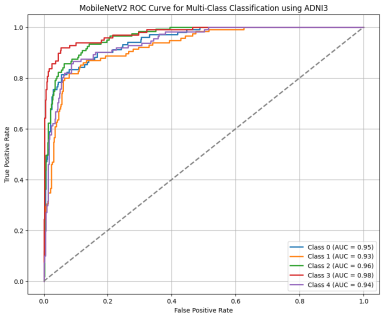
(g) ResNet 50 ROC curve results



(h) ResNet 101 ROC curve results



(i) EfficientNet B0 ROC curve results



(j) MobileNet V2 ROC curve results

Figure 4.13: layer-wise fine-tuning pre-trained CNN models ROC curves results using ADNI3 database

4.5 Discusion

The comparative evaluation of pre-trained CNN models across two distinct datasets, Kaggle and ADNI3, provides valuable insights into the performance and generalizability of these architectures for Alzheimer's disease classification. The results from both datasets highlight significant trends and emphasize the importance of model architecture, dataset characteristics, and fine-tuning strategies in achieving robust classification performance.

The data initially presented a significant class imbalance for the Kaggle dataset, which contains four classes. This imbalance posed challenges for training deep learning models, as underrepresented classes risk being overshadowed by dominant ones, leading to biased predictions. A rigorous data augmentation process was applied to balance all classes (See Figure 4.11) to address this. Despite the improved class distribution, only a slight improvement in model performance was observed. This indicates that while data augmentation mitigates imbalance-related issues, other factors may constrain its benefits. Furthermore, the rigorous augmentation process (see algorithm 2) introduces the risk of overfitting, as the models may overly adapt to the synthetic data rather than generalise to unseen data.

On the other hand, the ADNI dataset, comprising five classes, presented a clinically diverse and more balanced distribution. This allowed the models to leverage a broader and more representative set of features for classification, contributing to generally higher performance metrics. Models such as Xception and InceptionV3 demonstrated exceptional performance on this dataset, achieving accuracy above 85% and AUC values exceeding 97%. These results underscore the effectiveness of advanced architectures in extracting meaningful features even in complex multi-class settings.

Comparing the results across datasets reveals consistent trends in model performance. Advanced architectures such as Xception, InceptionV3, and ResNet101 consistently ranked among the top-performing models. This consistency highlights their ability to generalize effectively across datasets with varying characteristics. However, the slight variations in precision, recall, and AUC between Kaggle and ADNI datasets indicate that dataset-specific factors, including image quality, data distribution, and class complexity, significantly influence performance. For instance, the Kaggle dataset's initial imbalance and the risk of overfitting due to extensive data augmentation may have limited the models' ability to achieve optimal performance.

Layer-wise fine-tuning also played a crucial role in enhancing model performance across both datasets. By adapting pre-trained features to the specific task of Alzheimer's disease classification, fine-tuning enabled models to achieve higher accuracy and better generalization. Nonetheless, the results emphasize the limitations of particular architectures. Models like VGG16 and EfficientNet consistently underperformed, likely due to their relatively shallow or less specialized designs, which struggle to capture the intricate patterns associated with Alzheimer's disease classification tasks.

In summary, the findings in this chapter highlight the strengths and limitations of various pre-trained CNN models for Alzheimer's disease classification. While advanced architectures such as Xception, InceptionV3, and ResNet101 exhibit strong potential for achieving state-of-the-art performance, the observed variations between Kaggle and ADNI datasets underscore the critical impact of dataset characteristics on model outcomes. Furthermore, the challenges associated with class imbalance and the potential risks of overfitting emphasize the need for careful dataset preparation and the need for data augmentation adapted to medical data. Future work could explore strategies to further mitigate overfitting, enhance generalization, and validate the robustness of these models.

4.6 Summary

In this chapter, we applied ten pre-trained CNN models to perform a multiclassification task, addressing Alzheimer's disease (AD) as a multiway problem. Additionally, we fine-tuned these pre-trained CNN models for our specific task using a layer-wise strategy across all models. We also investigated how sensitive the pre-trained CNN models are to the unbalanced datasets we used. One dataset was oversampled to achieve balance, while the other had slightly unbalanced classes. The results indicate that oversampling the data significantly with synthetic data does not substantially enhance model performance, presenting another issue we will address in the next chapter. We also compared the performance of the pre-trained CNN models.

CHAPTER 5

ALZHEIMER'S DISEASE CLASSIFICATION BASED ON HIGH LEVEL FEATURES AND MACHINE LEARNING

5.1 Overview

Deep learning is a promising technology; however, it faces challenges in feature interpretability, data availability, and the need for large datasets, especially in critical fields such as medical research. Training deep learning models effectively requires substantial data to capture intricate patterns, particularly when diagnosing complex conditions like

Data scarcity is an inherent challenge in the medical field due to strict data privacy regulations, making it difficult to obtain and use large datasets for model training. This constraint limits the feasibility of training deep learning models from scratch. Although transfer learning techniques offer a potential solution by leveraging pre-trained models, they introduce a new concern: the interpretability of the pre-learned features. Precisely, what features does the model inherently capture from its pre-trained knowledge, and to what extent do these features accurately represent the discriminative characteristics needed for AD classification?

This chapter addresses these concerns, explicitly aiming to answer the research questions posed in Chapter 2 and to explore further the interpretability of CNN features used in Chapter 4. Here, we leverage the advantages of the approaches discussed in Chapters 3 and 4 to address key issues: overcoming data limitations, enhancing the feature extraction process by minimizing the need for extensive feature engineering, and, crucially, determining which CNN architecture provides the most discriminative features.

5.2 Proposed Approach

The proposed methodology outlined in this chapter focuses on overcoming the challenges mentioned in the previous chapters by presenting an automated detection and diagnosis of AD. The fourth proposed methodology in this thesis is structured into four primary steps: preprocessing, where the database is prepared for input into a pre-trained CNN architecture; Deep Features Extractions (DFE), which involves the extraction of high-level features encompassing a broad range of characteristics; Baysian Optimization (BO), a crucial phase that personalizes the model based on the variability of the deep features

and automatically adjusts its hyperparameters; and, finally, classification [1].

The procedure begins with extracting Fifteen unique characteristic vectors using several networks. Nine vectors (designated F1 to F9) are obtained utilizing the VGG16, VGG19, and AlexNet models. In comparison, six supplementary vectors (F10 to F15) are acquired from architectures such as GoogleNet, Inception V3, Xception, EfficientNet B0, ResNet50, and ResNet101, as shown in Figure 5.1. Every attribute vector serves to the establishment of a prediction (P1 to P15), and an efficient machine learning technique is then used to categorize the data (C1 through C15) and discriminate between the various phases of AD. To determine the usefulness of this approach, it is applied to two different datasets.

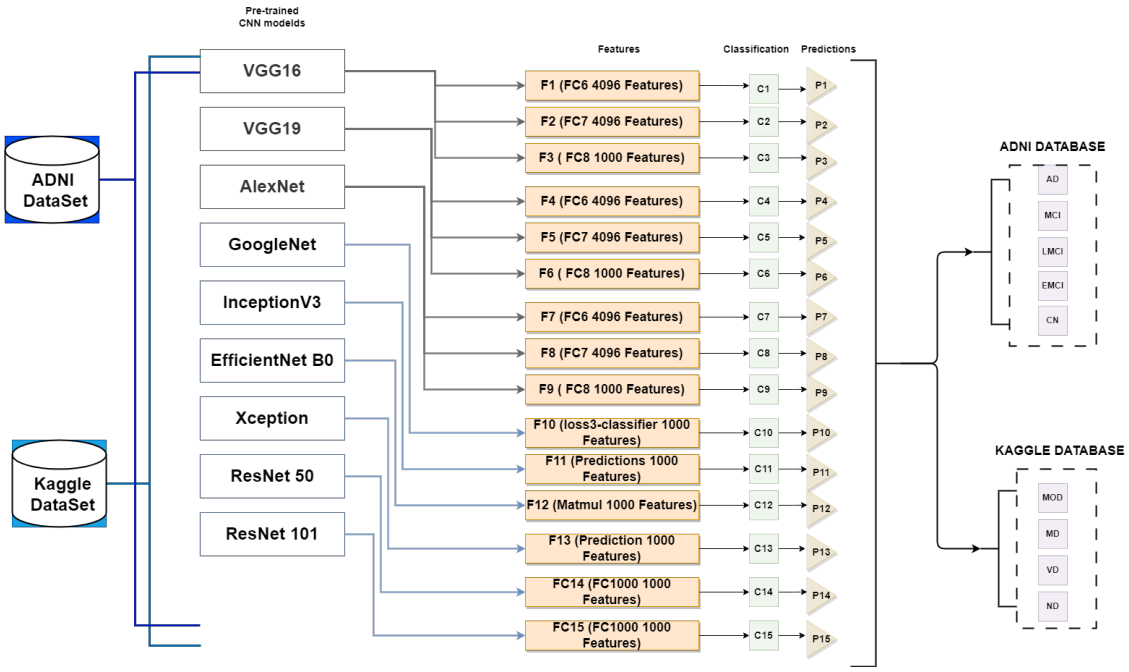


Figure 5.1: Proposed Approach diagram [19].

5.3 High-level features extraction: deep features

The primary advantage of deep learning techniques is their ability to extract features automatically. CNNs streamline the labour-intensive process of traditional classification methods, particularly feature extraction, which often demands extensive engineering and can be time-consuming. Consequently, CNNs are recognized as powerful feature extractors, as the features they learn are hierarchically structured, reflecting the dataset's complexity based on the architecture's depth.

In this hierarchical structure, the lower layers extract low-level features such as edges and textures. These initial layers act as edge detectors, with convolutional filters detecting horizontal and vertical lines, textures, and simple patterns. As training progresses to higher layers, the network combines these low-level features into mid-level ones, recognizing basic shapes and colors. Mid-level features are more abstract than edges or textures but

[1]Saim, M., & Feroui, A. (2024). Automated Alzheimer's disease detection and diagnosis method based on Bayesian optimization and CNN-based pre-trained features. Multimedia Tools and Applications, 1-41.

not as specific as object parts. They play an essential role in building the complexity of the network's understanding.

The middle layers take the edges, textures, and basic shapes detected by the lower layers and assemble them into more recognizable patterns. For instance, these layers might recognize shapes like circles and squares or more complex textures. The neurons in the middle layers have a larger receptive field than in the lower layers, allowing them to detect combinations of lower-level features over a larger area. This structure helps the network transition from essential visual elements to more structured patterns. Moreover, they contain features that are not necessarily specific to a single object but rather patterns or textures that could belong to multiple objects. Mid-level features are often transferable across related tasks or domains because they capture reusable information that applies to various images, such as shapes and textures.

High-level features in a CNN are found in the deeper layers, closer to the output, where the network captures abstract, task-specific information. These features result from lower and mid-level features' progressive combination and abstraction. High-level layers learn complex representations that correspond to specific parts of objects or even whole objects, representing more meaningful and recognizable aspects of the input. Therefore, in this chapter, we will utilize the CNN models as a deep feature extractor so the extracted deep features are the high-level representation of raw data.

Employing deep characteristics derived utilizing a pre-trained network confers a substantial benefit by supplying exclusive and distinguishing information. This can substantially enhance the performance of various Machine Learning (ML) algorithms in task-specific applications. This study explores a range of high-level deep features to comprehensively understand how different pre-trained models contribute to extracting robust features for improving ML task performance.

5.4 Machine learning optimization based on DF variation

Achieving a high accuracy rate in classification problems is a primary objective for researchers worldwide. However, this desirable outcome is contingent upon various factors, including data quality and the selection of discriminative features. Additionally, identifying the appropriate machine learning model for a specific problem presents its challenges. These challenges can be addressed through experimental investigation; however, the most complex aspect lies not in choosing the right model but in determining the optimal combination of hyperparameters to achieve the best results. Hyperparameters govern the learning process from the data and influence the complexity of the model.

Hyperparameter Tuning Optimization (HPO) begins with selecting hyperparameters that require tuning to maximize the ML model outcome by finding the global optimum, followed by defining a range of potential values for each hyperparameter. Subsequently, the search process is initiated to identify the best combination that maximizes accuracy. In this Chapter, the search process was conducted using Bayesian optimization search. This methodology was chosen because it can select the best hyperparameters based on the previous interaction. Therefore, we aim to investigate various features extracted from the different pre-define CNN models in chapter 4 and fine-tune the machine learning algorithms based on variations in these features to identify the optimal hyperparameters that maximize classification performance.

In this work, we categorized the extracted Deep Features (DF) into four distinct types

of features, resulting in four experiments. Each experiment was focused on two aspects: the classification results and the combinations of hyperparameters.

5.4.1 Bayesian optimization

Bayesian optimization (BO) is an efficient approach that employs Bayesian statistics to determine optimal hyperparameters. This approach is probabilistic and employs Bayes' theorem. In contrast to other hyperparameter optimization methods, Bayesian Optimisation generates probability distributions over the hyperparameter space and refines them progressively depending on previous examinations.

Bayesian optimization directs the search efficiently by focusing on areas with the highest probability of improvement, making it an invaluable tool for solving challenging optimization problems with limited evaluations of the function [115].

In this chapter, finding the optimal hyperparameters to maximize model performance is our optimization problem. To solve it, we start by describing our objective function (eq.5.1) as follows:

$$\max_{x \in \mathcal{X}}(Objectif\_f(x)) \tag{5.1}$$

Where:

- $Objectif\_f(x)$: is the model performance, in our case, the accuracy metric.
- x: The hyperparameter set of the classification model.
- $\mathcal{X}$: the space of various value.

The challenge is to find the best values within space $\mathcal{X}$ because each evaluation of $Objectif\_f(x)$ (training and testing the model) can be time-consuming. Therefore, applying Bayes' theorem (see eq.5.2) helps us to explore the hyperparameter space as follows: systematically

$$P(f \mid T) \propto P(T \mid f)P(f) \tag{5.2}$$

Here, Bayes' theorem helps to refine our understanding of f where $P(f)$ represents the prior belief about the possible behaviors of f, and T represents the observations or the results of previous model evaluations with different hyperparameters[115],[116].

Each evaluation adjusts our starting data point (belief about the shape of f guiding the solution toward the regions in $\mathcal{X}$ where the model's performance is most likely to improve.

Modeling f as accurately as possible is done using the Gaussian process. Where the prediction is made for the performance of unexplored hyperparameter values but also the uncertainty of those predictions. By balancing exploration (trying uncertain regions) and exploitation (focusing on promising regions)[115].

5.4.1.1 Gaussian process

A GP is a non-parametric statistical model frequently employed in machine learning and statistics for predicting unknown functions[115],[117]. It is especially effective in scenarios where the objective is to estimate a function based on a limited number of observations, such as in Bayesian optimization for hyperparameter tuning. Generally, a GP assumes that any collection of points or inputs within the function's domain has a joint Gaussian (normal) distribution for the corresponding outputs. This framework establishes a distribution over potential functions, enabling the GP to adapt flexibly to the data without adhering to a fixed structure [118].

In practice, we aim to find the optimal configuration x of hyperparameters to maximize ML model performance. Thereby we considered the function $f(\mathrm{x})$ as an unknown function that follows the Gaussian process prior. Our observations are collected as pairs $\{\mathbf{x}_n, y_n\}_{n=1}^{N}$ .where each observation y_n is distributed as:

$$y_n \sim \mathcal{N}\left(f\left(\mathbf{x}_n\right), \nu\right) \tag{5.3}$$

Here, ν represents the variance of noise in our observations, accounting for randomness in model performance evaluations.

The GP prior, combined with our observations, produces a posterior distribution over possible functions, enabling us to estimate $f(x)$ in regions we have not yet sampled. We use an *acquisition function* $a : \mathcal{X} \rightarrow \mathbb{R}^{+}$ to select the next hyperparameter setting to evaluate by solving:

$$\mathbf{x}_{\text{next}} = \operatorname{argmax}_{\mathbf{x}} a(\mathbf{x}) \tag{5.4}$$

The acquisition function balances exploration and exploitation based on previous observations and GP hyperparameters, denoted by :

$$a\left(\mathbf{x}; \left\{\mathbf{x}_n, y_n\right\}, \theta\right) \tag{5.5}$$

In this work, we utilize the function of the probability of improvement(PI), which is a function that searches to maximize the probability of improving upon the current best observation. For a Gaussian process, PI can be calculated as:

$$a_{\mathrm{PI}}\left(\mathbf{x}; \left\{\mathbf{x}_n, y_n\right\}, \theta\right) = \Phi(\gamma(\mathbf{x})) \tag{5.6}$$

Where:
$\gamma(\mathbf{x}) = \frac{f(\mathbf{x}_{\text{best}}) - \mu(\mathbf{x}; \{\mathbf{x}_n, y_n\}, \theta)}{\sigma(\mathbf{x}; \{\mathbf{x}_n, y_n\}, \theta)}$ and $\phi(\cdot)$ is the cumulative distribution function of the standard normal.

In this chapter, the MSVM, KNN, and DT are the ML models that we aim to tun using BO to maximize the test set's accuracy.

We aim to identify the best hyperparameter configuration x that maximizes the accuracy function $f(x)$. Therefore, we begin with baseline settings in the hyperparameter space, where each value of the hyperparameter is the starting point for the data.(see eq 5.1)

We establish a baseline performance score, or "original score," calculated by the target score function $f(x)$ before any optimization. This score corresponds to the initial configuration and provides our starting data point. At each BO iteration, a new hyperparameter combination is eventually selected by maximizing an acquisition function. (see eq.5.5, eq.5.6).

This function balances exploration with exploitation by steering the search to beneficial areas of the hyperparameter range. Upon picking every option, it is assessed using the target score function, and the observation set is updated to include the most recent configuration and its associated value.This updated data refines the GP_{PI} surrogate model, enhancing its ability to estimate performance across potential configurations.

The process continues until a termination criterion, in our case a maximum of 50 iterations, is reached. Once finished, the optimal hyperparameters are used to train new models. Their accuracy on the test set is then measured and recorded as the final evaluation.

The chosen configuration was assessed using the objective scoring algorithm, with the data points updated to reflect the new settings and their corresponding scores. The objective scoring algorithm surrogate model, which approximates the performance of hyperparameter configurations, was subsequently refined using the latest data. This iterative process continued until a predefined stopping criterion was satisfied, such as reaching a maximum of 50 iterations. Upon completion, the optimized hyperparameters were identified and used to construct new machine learning models, which were then evaluated on the test dataset. The accuracy of these models was computed and documented. The Bayesian optimization (BO) hyperparameter tuning algorithm is illustrated below.

Algorithm 3 Bayesian Optimization for Hyperparameter Tuning

1: **Input:** Search space $\mathcal{X}$, Gaussian Process (GP) prior, acquisition function $a : \mathcal{X} \to \mathbb{R}^+$
2: **Initialize:** Select an initial set of points $\{\mathbf{x}_n\}_{n=1}^N \subset \mathcal{X}$ and evaluate objective function $f(\mathbf{x}_n)$ to get observations $y_n \sim \mathcal{N}(f(\mathbf{x}_n), \nu)$, where ν is the noise variance.
3: **for** $t = N + 1$ to T **do**
4: **Compute Posterior:** Update the GP_{PI} posterior over $f(\mathbf{x})$ using observed data $\{\mathbf{x}_n, y_n\}_{n=1}^{t-1}$.
5: **Select Next Point:** Choose the next point $\mathbf{x}_{\text{next}}$ by maximizing the acquisition function:

$$\mathbf{x}_{\text{next}} = \text{argmax}_{\mathbf{x}\in\mathcal{X}}\, a\left(\mathbf{x}; \{\mathbf{x}_n, y_n\}_{n=1}^{t-1}, \theta\right)$$

6: **Evaluate Objective:** Compute $y_{\text{next}} = f(\mathbf{x}_{\text{next}}) + \epsilon$ where $\epsilon \sim \mathcal{N}(0, \nu)$.
7: **Update Data:** Add $(\mathbf{x}_{\text{next}}, y_{\text{next}})$ to the observed data set.
8: **end for**
9: **Output:** Return the best observed configuration $\mathbf{x}_{\text{best}} = \text{argmin}_{\mathbf{x}_n} f(\mathbf{x}_n)$.

5.4.2 Machine learning hyperparameters tuning via BO based on the fully connected deep features

As mentioned, we aim to finetune the ML models using BO based on the database features. Starting with the DFs obtained from the Fully Connected Layers (FC) layers of AlexNet, VGG16, and VGG19.

The tables 5.1, 5.2, and 5.3 provide a comprehensive analysis of hyperparameter tuning results for three distinct machine learning models—Multiclass Support Vector Machine (MSVM), K-Nearest Neighbors (KNN), and Decision Tree (DT)—applied to features ex-

tracted from fully connected layers (FC6, FC7, FC8) of convolutional neural networks (CNNs) such as VGG16 and VGG19, across two datasets (Dataset 1 and Dataset 2).

Table 5.1 details the hyperparameter optimization for the Bo MSVM model, focusing on the regularization parameter (C), kernel coefficient (Gamma), and kernel type (Kernel). For Dataset 1, the highest accuracy of 95.37% is achieved using FC6 features with a polynomial kernel, where the optimal hyperparameters are C=1.0 and Gamma=0.01. Similarly, for VGG19, the best accuracy of 91.69% is obtained with FC6 features using a polynomial kernel (C=0.44, Gamma=0.008). In contrast, Dataset 2 shows lower performance, with the best accuracy of 74.24% for FC6 features using a polynomial kernel (C=0.06, Gamma=0.01). The results indicate that polynomial kernels are generally optimal for higher-level features (FC6, FC7), while linear kernels are preferred for lower-level features (FC8). This suggests that the complexity of the kernel function significantly influences model performance, with more complex kernels (polynomial) better suited for capturing intricate patterns in higher-level features and simpler kernels (linear) sufficient for lower-level features. These findings underscore the importance of tailored hyperparameter selection based on dataset and feature characteristics.

Table 5.2 presents the results for the BO KNN model, with a focus on the number of neighbors (N neighbors) and the weighting method (Weights). For Dataset 1, the highest accuracy of 97.69% is achieved using FC6 features with 4 neighbors, indicating that this configuration optimally balances bias and variance for this dataset. Similarly, FC7 and FC8 features yield high accuracies of 95.42% and 97.68%, respectively, with the same number of neighbors. In contrast, Dataset 2 achieves its best accuracy of 65.45% for FC6 features using 3 neighbors, with lower but consistent performance across FC7 and FC8 features. The results suggest that Dataset 1 benefits from a broader local context (4 neighbors) to capture complex decision boundaries, while Dataset 2, with its simpler structure, performs optimally with fewer neighbors (3). This highlights the critical role of the number of neighbors in KNN performance and the variability in model effectiveness across datasets.

Table 5.3 outlines the hyperparameter tuning results for the BO Decision Tree (DT) model, focusing on the splitting criterion (Criterion), maximum depth (Max_depth), and granularity parameters (Min_samples_leaf, Min_samples_split). For Dataset 1, the highest accuracy of 69.25% is achieved using FC7 features, with entropy as the preferred splitting criterion for higher-performing configurations. The maximum depth ranges from 61 to 100, indicating that deeper trees are beneficial for capturing complex patterns in Dataset 1. In contrast, Dataset 2 achieves its best accuracy of 50.53% using FC6 features with entropy and a maximum depth of 42. Gini impurity also appears in several configurations. Still, the overall lower accuracies suggest that Dataset 2 has fewer distinct class boundaries or higher noise, making it more challenging for the DT model to achieve high performance. The variability in granularity parameters further emphasizes the importance of fine-tuning hyperparameters to optimize model performance.

Table 5.1: The hyperparameter setting for MSVM utilizing FC features via BO.

Database	DFs		Best score	Best parameters			Execution Time(s)
				C	Gamma	Kernel	
Dataset 1	AlexNet	FC6	95.37%	1.0.	0.01.	Poly.	7654.35.
		FC7	92.63%		0.006.		8299.04.
		FC8	71.97%	0.9.	0.007.	Linear.	3448.019.
	VGG16	FC6	81.26%	0.2.	0.007.	Poly.	11806.03.
		FC7	88.66%	0.4.	0.001.		8090.73.
		FC8	69.03%	0.9.		Linear.	3211.90.
	VGG19	FC6	91.69%	0.44.	0.008.	Poly.	6556.23.
		FC7	88.17%	1.0.	0.007.		7540.10.
		FC8	69.18%	1	0.002.	Linear.	3605.24.
Dataset 2	AlexNet	FC6	74.24%	0.06.	0.01.	Poly.	1762.02.
		FC7	71.80%	0.9.	0.05.		1671.26.
		FC8	62.03%	0.8.	0.01.	Linear.	355.15.
	VGG16	FC6	69.18%	0.1.	0.009.	Poly.	1832.87.
		FC7	67.49%	1.0.	0.005.		3552.46.
		FC8	53.64%	0.83.	0.001.	Linear.	708.28.
	VGG19	FC6	69.40%	0.07.	0.01.	Poly.	1833.53.
		FC7	67.22%	0.15.			1843.58.
		FC8	52.97%	1.0.		Linear.	832.07.

Table 5.2: The hyperparameter setting for KNN utilizing FC features via BO

Database	DFs		Best score	Best parameters		Execution Time(s)
				N_neighbors	Weights	
Dataset 1	AlexNet	FC6	97.69%	4	Distance.	178.86
		FC7	95.42%			246.83
		FC8	97.68%			243.96
	VGG16	FC6	81.59%			192.46
		FC7	82.95%			127.29
		FC8	79.10%			66.36
	VGG19	FC6	94.89%			100.4
		FC7	93.90%			110.12
		FC8	78.65%			71.86
Dataset 2	AlexNet	FC6	65.45%	3		106.92.
		FC7	65.94%			101.29.
		FC8	64.29%			80.35.
	VGG16	FC6	65.71%			72.26.
		FC7	65.72%			69.37.
		FC8	60.74%			49.36.
	VGG19	FC6	63.49%			71.72.
		FC7	63.05%			74.67.
		FC8	60.47%			51.19.

Table 5.3: The hyperparameter setting for DT utilizing FC features via BO.

Database	DFs		Best score	Best parameters					Execution Time(s)
				Criterion	Max_depth	Max_features	Min_samples_leaf	Min_sample_split	
Dataset 1	AlexNet	FC6	63.19%	Entropy.	100	None	1	2	1487.31
		FC7	62.25%						1676.86
		FC8	61.12%	Gini.	61		61		643.71
	VGG16	FC6	62.29%	Entropy.	3		42	5	1542.55
		FC7	55.83%		67		23	4	1568.54
		FC8	62.36%		5		1	84	296.38
	VGG19	FC6	57.61%	Entropy.	84		24	11	1270.05
		FC7	58.32%		24				1838.08
		FC8	57.55%	Gini.	72				540.3
Dataset 2	AlexNet	FC6	50.53%	Entropy.	42				971.35
		FC7	48.88%		88				821.76
		FC8	47.95%	Gini.	34				278.47
	VGG16	FC6	44.98%	Entropy.	49			2	957.52
		FC7	45.82%		71		1		483.33
		FC8	45.56%	Gini.	48				150.65
	VGG19	FC6	46.75%	Entropy.	100				831.37
		FC7	47.02%		71				855.96
		FC8	45.11%	Gini.	100				267.72

5.4.3 Machine learning hyperparameters tuning via BO based on the Inception-based features

We aim to optimize the machine learning model in the second experiment utilizing the inception-based CNN DFs. The tables 5.4, 5.5, 5.6, present the results of hyperparameter tuning using BO for three machine learning models applied to deep features extracted from architectures such as GoogleNet, Inception V3, and Xception across two datasets

Table 5.4 present BO MSVM Hyperparameter Tuning For the MSVM model. In Dataset 1, the highest accuracy of 69.84% is achieved using Inception V3 features with a polynomial kernel (C=1, Gamma=0.01). Similarly, GoogleNet and Xception features yield accuracies of 66.88% and 69.05%, respectively, with the same hyperparameters. This suggests the polynomial kernel is well-suited for capturing complex patterns in Kaggle. In contrast, ADNI3 shows lower achievement with the best accuracy of 57.36% for Xception features using a linear kernel (C=0.8, Gamma=0.003). The shift to a linear kernel for Dataset 2 indicates that simpler decision boundaries are more effective for this dataset.

Table 5.5 presents BO KNN Hyperparameter Tuning. For Dataset 1, the highest accuracy of 88.31% is achieved using Xception features with 4 neighbors and a distance-based weighting method. GoogleNet and Inception V3 features also perform well, with accuracies of 76.74% and 78.03%, respectively, using 4 neighbors. This indicates that Dataset 1 benefits from a broader local context to capture intricate patterns. For Dataset 2, the best accuracy of 64.78% is achieved using Xception features with 3 neighbors, suggesting that a smaller number of neighbors is sufficient for this dataset, likely due to its simpler decision boundaries.

Table 5.6 shows the BO DT Hyperparameter Tuning results. In Dataset 1, the highest accuracy of 61.70% is achieved using Inception V3 features with Gini impurity and a maximum depth of 100. GoogleNet and Xception feature yield accuracies of 60.05% and 55.00%, respectively, with varying hyperparameters. For Dataset 2, the best accuracy of 42.80% is achieved using Inception V3 features with entropy as the splitting criterion and a maximum depth of 100. The lower accuracies in Dataset 2 suggest that the dataset's less distinct class boundaries or higher noise levels make it more challenging for the DT model to achieve high performance.

Table 5.4: The hyperparameter setting for MSVM utilizing Inception features via BO.

Database	Deep Features	Best score	Best parameters			Execution Time(s)
			C	Gamma	Kernel	
Dataset1	GoogleNet	66.88%	1	0.01	poly	3059.84
	Inception V3	69.84%				2966.14
	Xception	69.05%				2960.54
Dataset2	GoogleNet	55.15%		0.01	linear	651.45
	Inception V3	57.19%		0.003		574.14
	Xception	57.36%	0.8	0.005		351.02

Table 5.5: The hyperparameter setting for KNN utilizing Inception features via BO.

	Deep Features	Best score	Best parameters		Execution Time(s)
			n_neighbors	Weights	
Dataset1	GoogleNet	76.74%	4	distance	68.77
	Inception V3	78.03%			77.15
	Xception	88.31%			69.7
Dataset2	GoogleNet	61.36%	3		50.44
	Inception V3	63.18%			51.78
	Xception	64.78%			47.27

Table 5.6: The hyperparameter setting for DT utilizing Inception features via BO.

	Deep Features	Best score	Best parameters					Execution Time(s)
			Criterion	Max_depth	Max_features	Min_samples_leaf	Min_sample_split	
Database 1	GoogleNet	60.05%	Entropy	60	None.	100	2	437.3
	Inception V3	61.70%	Gini	100		85		467.95
	Xception	55.55%	Entropy			48	31	1605.93
Database 2	GoogleNet	42.01		52		1	2	197.86
	Inception V3	42.80%		100				307.69
	Xception	44.18%		63				320.97

5.4.4 Machine learning hyperparameters tuning via BO based on Residual Connections-Based features

Our third trial examined to determine how the ML hyperparameters set changed when we used DFs with the residual connection-based CNN. The tables 5.7, 5.8, and 5.9 show the outcomes of optimizing the parameters of the three machine learning algorithms that were used to extract deep features from residual connection-based architectures, namely ResNet 50 and ResNet 101, on two different datasets.For the MSVM model (table 5.7), in first database, the maximum accuracy of 70.25% is achieved using ResNet 101 features with C=1 and Gamma=0.002, while ResNet 50 features yield 69.80% accuracy with C=1 and Gamma=0.001. For the second database, the best accuracy of 61.58% is achieved using ResNet 50 features with C=0.97 and Gamma=0.001, indicating that slightly lower regularization and kernel coefficients are optimal for this dataset.

The KNN model's performance (Table 5.8) is influenced by the number of neighbors (n) and the weighting method. In Dataset 1, the highest accuracy of 78.67% is achieved using

ResNet 50 features with 4 neighbors and a distance-based weighting method, while ResNet 101 features yield 77.86% accuracy with the same number of neighbors. For Dataset 2, the best accuracy of 64.29% is achieved using ResNet 101 features with 3 neighbors, suggesting that fewer neighbors are sufficient for this dataset.

Several hyperparameters determine the DT model's performance (table (Table 5.9). In Dataset 1, the highest accuracy of 62.57% is achieved using ResNet 101 features with entropy as the splitting criterion and a maximum depth of 5. In contrast, ResNet 50 features yield 61.98% accuracy with a maximum depth of 15. For Dataset 2, the best accuracy of 43.73% is achieved using ResNet 50 features with Gini impurity and a maximum depth of 56, indicating that deeper trees are more effective for this dataset despite its lower overall performance. The results highlight the importance of dataset-specific hyperparameter tuning and the impact of feature extraction from residual connection-based architectures on model performance.

Table 5.7: The hyperparameter setting for MSVM utilizing Residual Connections-Based features via BO.

	Deep Features	Best score	Best parameters			Execution Time(s)
			C	Gamma	Kernel	
Dataset1	ResNet 50	69.80%	1	0.001	Linear	3303.77
	ResNet 101	70.25%		0.002		2094.53
Dataset2	ResNet 50	61.58%	0.97	0.001		586.31
	ResNet 101	58.79%	1			639.67

Table 5.8: The hyperparameter setting for KNN utilizing Residual Connections-Based features via BO.

	Deep Features	Best score	Best parameters		Execution Time(s)
			n_neighbors:	weights:	
Dataset1	ResNet 50	78.67%	4	distance	127.27
	ResNet 101	77.86%			118.6
Dataset2	ResNet 50	62.25%	3		87.58
	ResNet 101	64.29%			89.11

Table 5.9: The hyperparameter setting for DT utilizing Residual Connections-Based features via BO.

	Deep Features	Best score	Best parameters					Execution Time(s)
			Criterion	max_depth	max_features	min_samples_leaf	min_samples_split	
Dataset1	ResNet 50	61.98%	Entropy.	15	None	75	94	557.81
	ResNet 101	62.57%		5		100	2	532.57
Dataset2	ResNet 50	43.73%	Gini.	56		1	2	201.76
	ResNet 101	37.07%		59		64	24	146.9

5.4.5 Machine learning hyperparameters tuning via BO based on efficientNet b0-Based features

The objective here is to find the hyperparamets set based on the EfficientNet B0 deep features variation.The tables 5.10, 5.11, 5.12 present the results of hyperparameter tuning

using BO for the three ML algorithms across two datasets (Dataset 1 and Dataset 2). For the MSVM model (see table 5.10), in the first Data, the 76.80% accuracy is achieved using EfficientNet b0 features with a polynomial kernel (C=1, Gamma=0.01), indicating that the polynomial kernel effectively captures the complex patterns in this dataset. In contrast, the second Data achieves its best accuracy of 68.79% using a linear kernel, suggesting that simpler decision boundaries are more suitable for this dataset. Utilizing the KNN model (see table 5.11) for Dataset 1, the highest accuracy of 93.20% is achieved using EfficientNet b0 features with 4 neighbors and a distance-based weighting method, highlighting the importance of a broader local context for capturing intricate patterns. In Dataset 2, the best accuracy of 65.62% is achieved with 3 neighbors, indicating that fewer neighbors are sufficient for this dataset, likely due to its simpler structure. When fine-tuning the DT hyperparameter (see table 5.12) for Dataset 1, the highest accuracy of 58.31% is achieved using entropy as the splitting criterion, a maximum depth=39, and a minimum samples=2 necessary to divide an internal node. This configuration suggests that Dataset 1 benefits from deeper trees and entropy-based splitting, effectively capturing complex patterns and nuanced decision boundaries in the data.

In contrast, for Dataset 2, the best accuracy of 44.84% is achieved using Gini impurity as the splitting criterion, a shallower tree with a maximum depth=11, and a minimum=19 samples needed to divide an internal node. Utilizing Gini impurity and a more conservative splitting strategy indicates that ADNI3 has a simpler structure, with less distinct class boundaries or higher noise levels, making it more challenging for the DT model to achieve high performance. Additionally, considering all features (max_features=None) for the ADNI3 Dataset suggests that the model benefits from leveraging the full feature space, while the Kaggle Dataset may not require explicit feature limitation.

Table 5.10: The hyperparameter setting for MSVM utilizing efficientNet b0 features via BO.

	Deep Features	Best score	Best parameters			Execution Time(s)
			C	Gamma	kernel	
Dataset1	EfficientNet b0	76.80%	1	0.01	poly	2033.45
Dataset2		68.79%			Linear	531.58

Table 5.11: The hyperparameter setting for KNN utilizing efficientNet b0 features via BO.

	Deep Features	Best score	Best parameters		Execution Time(s)
			n_neighbors	weights	
Dataset1	EfficientNet b0	93.20%	4	distance	121.2
Dataset2		65.62%	3		74.86

Table 5.12: The hyperparameter setting for DT utilizing efficientNet b0 features via BO.

	Deep Features	Best score	Best parameters					Execution Time(s)
			Criterion:	max_depth	max_features	min_samples_leaf:	min_samples_split	
Dataset1	Efficient Net b0	58.31%	entropy	39	None	1	2	703.48
Dataset2		44.84%	Gini	11			19	214.06

5.5 Machine learning classification results using deep features

This section presents the outcome of the tuned machine-learning algorithms.

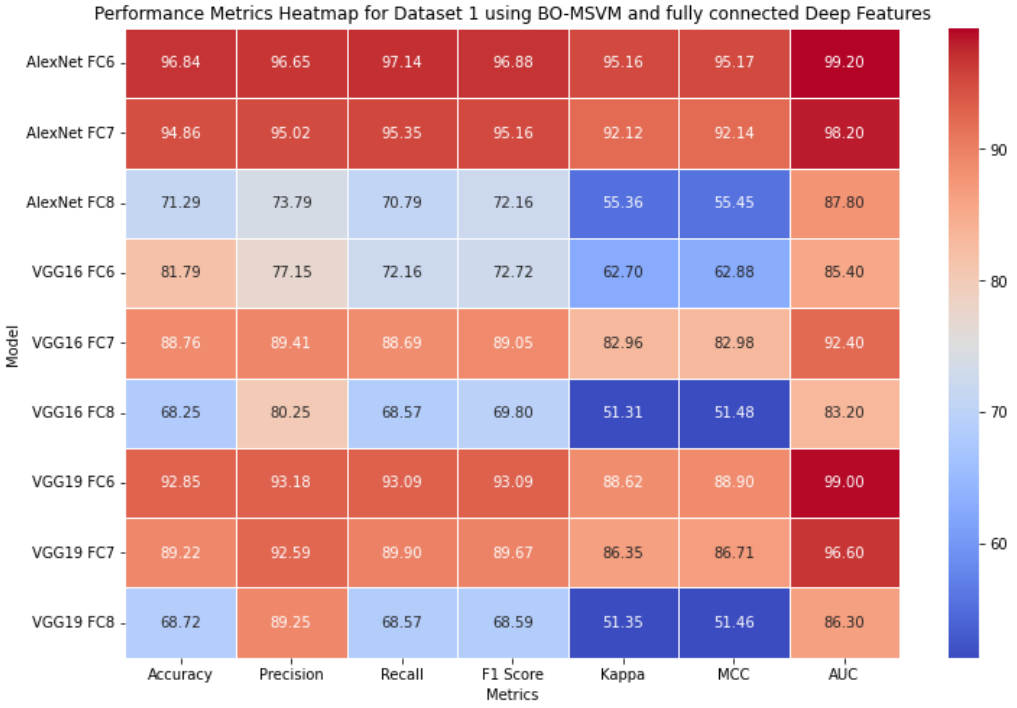
5.5.1 Classification results using fully connected DF

The heatmap figures 5.2, 5.4, and 5.6 visually show the categorization outcomes achieved by running the feature vectors through the fine-tuned ML algorithms. Darker reds indicate higher values and better performance, while lighter blues represent lower metrics; these figures illustrate each of the pre-defined evaluation measures associated with every ML algorithm and FC layer.

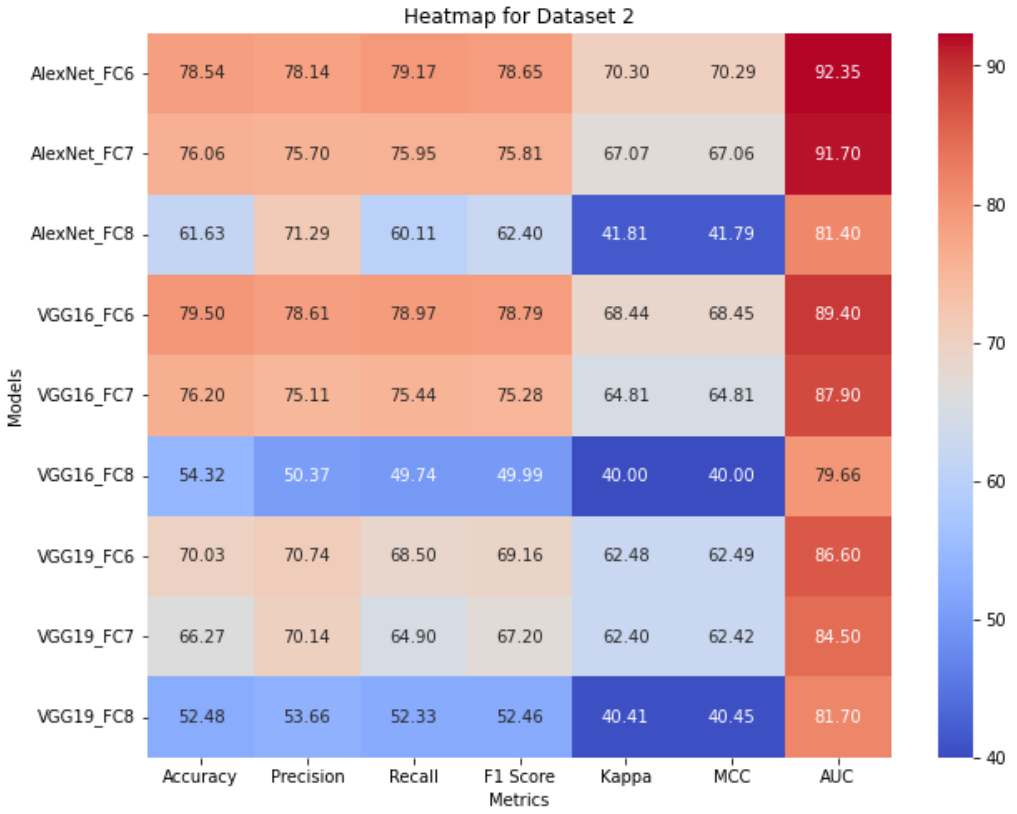
By applying the fine-tuned MSVM (Figures 5.2a and 5.2b) to the FC layers, we evaluated the DFs from the FC layers on two datasets.

In Dataset 1 subfigure 5.2a , AlexNet's FC6 achieved outstanding results, with an accuracy of 96.84% and an exceptionally high AUC of 99.20%, as reflected by the darkest red tones in the heatmap. VGG16's FC7 also demonstrated strong performance, with an accuracy of 88.76% and an AUC of 89.90%, represented by lighter red shades. In contrast, the FC8 layers for most architectures showed lower performance, with accuracy and AUC values ranging from 83.20% to 89.90%, highlighted by lighter shades.

For Dataset 2 subfigure 5.2b, the performance was generally lower. AlexNet's FC6 maintained relatively good results, with an accuracy of 78.54% and an AUC of 93.30%, evident in the red tones. VGG16's FC7 showed moderate performance, achieving an accuracy of 76.06% and an AUC of 92.30%. However, the FC8 layers displayed weaker results, with accuracy ranging from 54.32% to 69.50% and AUC values between 81.70% and 89.70%, represented by significantly lighter tones in the heatmap.



(a) BO_MSVM classification results utilizing the FC features obtained from dataset 1.



(b) BO_MSVM classification results utilizing the FC features obtained from dataset 2.

Figure 5.2: Heatmaps classification performance of BO_MSVM utilizing the FC features.

The subsequent illustration displays ROC curves as part of thoroughly examining these findings. The figures are categorized by architecture and deep feature layers (DF), enabling straightforward performance comparisons across various configurations and phases.

Figure 5.3 shows the ROC curves for the BO-MSVM classifier on the two sets with different FC_DFs. It shows that AlexNet's FC6 layer distinguished Alzheimer's disease (AD) classes best across all architectures and deep feature layers, with the greatest AUC values. All designs show strong performance, with AUC values over 85%, highlighting the BO-MSVM classifier's capacity to use FC_DFs for Database 1. The ROC curves confirm our method's AD detection and diagnosis efficacy.

In contrast to AlexNet_FC8 (Figure 5.3c), AlexNet_FC6 (Figure 5.3a) demonstrates near-perfect classification across all categories, whereas BO-MSVM (Figure 5.3c) shows reduced discrimination, especially for class 3 (VMD), as shown in red. The ROC curves for VGG16 and VGG19 show that using FC8_DF in these designs results in the lowest AUC values. Comparing VGG16_FC8 to VGG19_FC8, the ROC curve shows that VGG16_FC8 has the poorest class 0 (MD) discrimination. ROC curves for Database 2 show a similar tendency.

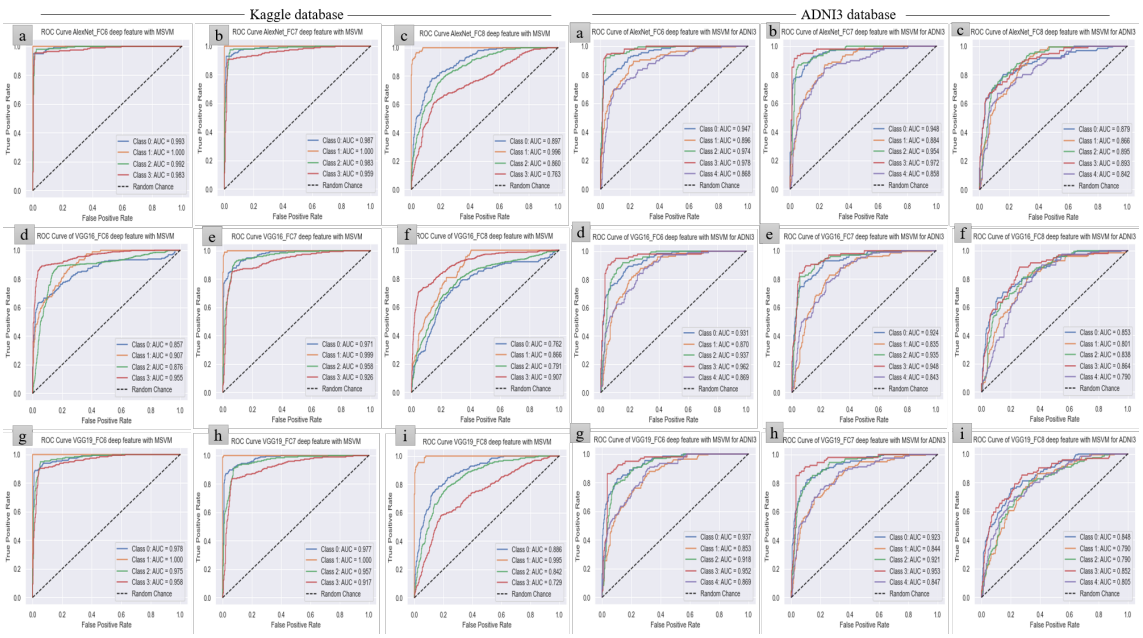
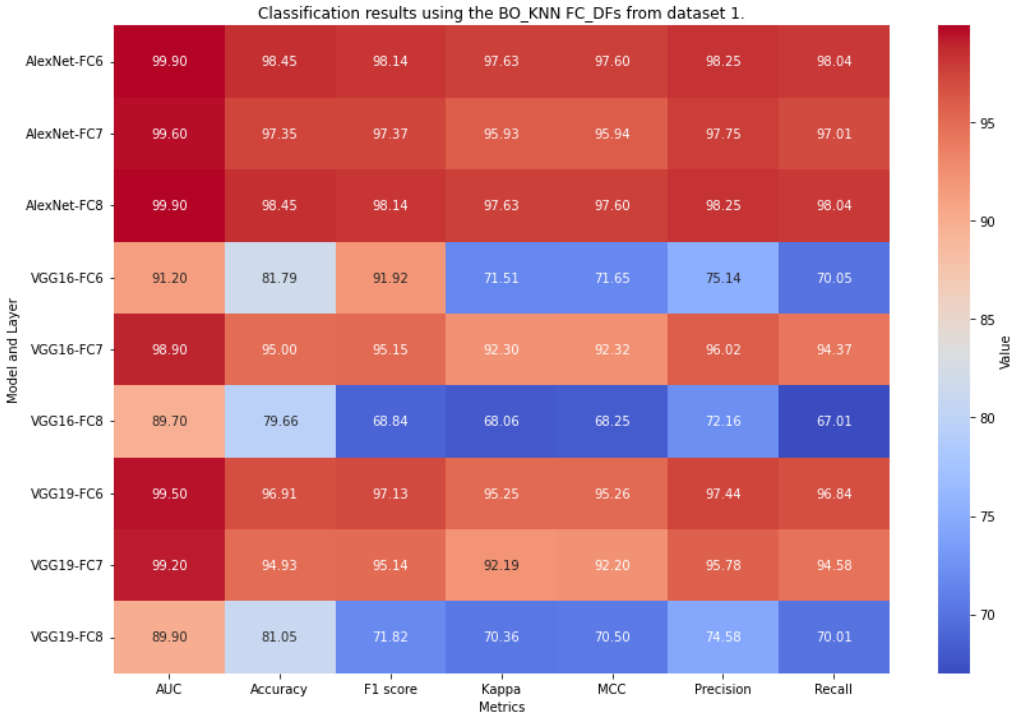


Figure 5.3: The ROC curve Application of BO-MSVM classifier to both databases using FC_DFs. Images from (a) to (c) show the roc curve of the AlexNet DFs for FC6 to FC8, (d) to (f) the VGG16 DFs, and (g) to (i) the VGG19 DFs for both databases.

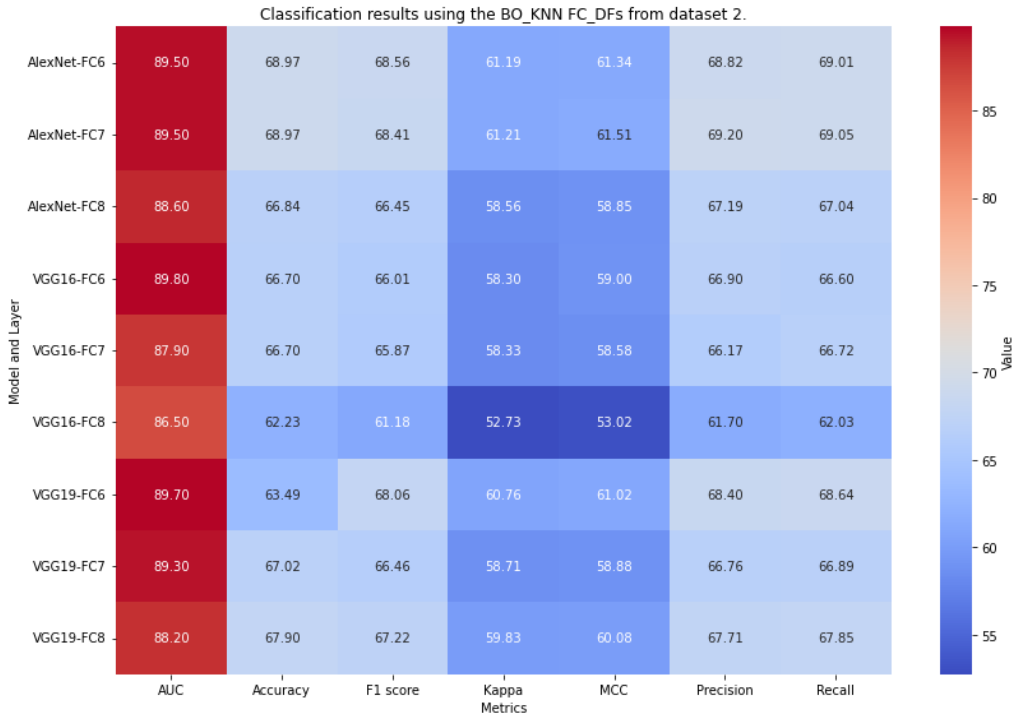
Figure 5.4 displays the classification performance utilizing FC_DFs and the BO-KNN classifier.

The first heatmap subfigure 5.4a , derived from dataset 1 (Kaggle), showcases the classification results of the BOKNN model applied to FC layers of pre-trained deep learning architectures, including AlexNet, VGG16, and VGG19. The performance metrics reveal exceptional performance across most layers. Notably, the AlexNetFC6 and AlexNetFC8 layers achieved near perfect accuracy and AUC scores (99.90%), along with high precision and recall values, indicating outstanding discriminative power. Similarly, the VGG19-FC6 and VGG19-FC7 layers demonstrated remarkable performance, with accuracy exceeding

94% and strong F1 scores. However, the VGG16-FC8 and VGG19-FC8 layers exhibited comparatively lower performance, particularly in precision and recall, suggesting potential challenges in generalizing these specific layers. The heatmap underscores the model's strong adaptability to dataset 1, achieving near-optimal results for most layers and highlighting consistent limitations with certain FC layers. In contrast, the second heatmap subfigure 5.4b, derived from dataset 2 (ADNI3), presents a more varied performance profile for the same BOKNN model. While the AlexNet-FC6 and AlexNet-FC7 layers maintained relatively high accuracy and AUC scores (89.50%), the overall performance metrics were lower than dataset 1. For instance, the VGG16-FC8 and VGG19-FC7 layers showed significantly reduced precision and recall, indicating challenges in generalizing to this dataset. The VGG19-FC6 layer, however, demonstrated relatively strong performance, with accuracy and AUC scores comparable to the top-performing layers in dataset 1. The heatmap reveals that while certain layers (e.g., AlexNet-FC6 and VGG19-FC6) perform well across most metrics, others (e.g., VGG16-FC8) struggle, potentially due to differences in dataset characteristics, such as sample size, feature distribution, or noise levels.



(a) BO_KNN classification results utilizing the FC features obtained from dataset 1.



(b) BO_KNN classification results utilizing the FC features obtained from dataset 2.

Figure 5.4: Heatmaps classification outcome of BO_KNN with the FC_DFs.

Refer to figure 5.5 for BO-KNN ROC curve results. Figures 3a–3g show the maximum sensitivity of BO-KNN ROC curves. These figures show the ideal ROC curve for all classes using AlexNet FC6, FC7, and FC8 and VGG19 FC6 for Database 1. Database 2's ROC curves show modest effectiveness across all decision functions, demonstrating how decision

functions and data properties affect the quality of classification.

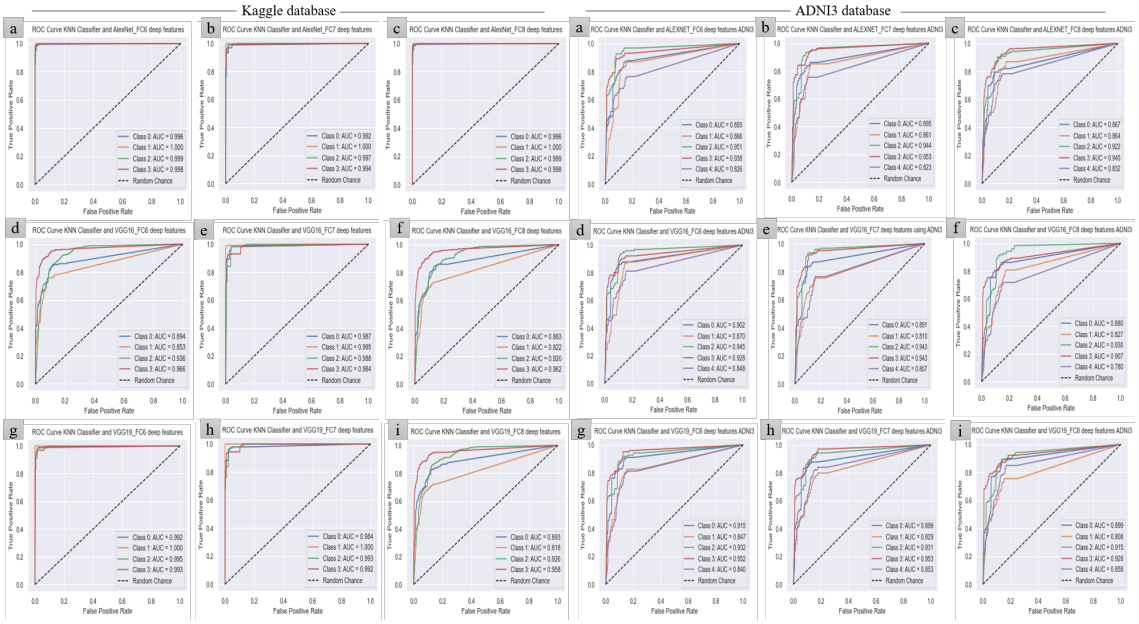
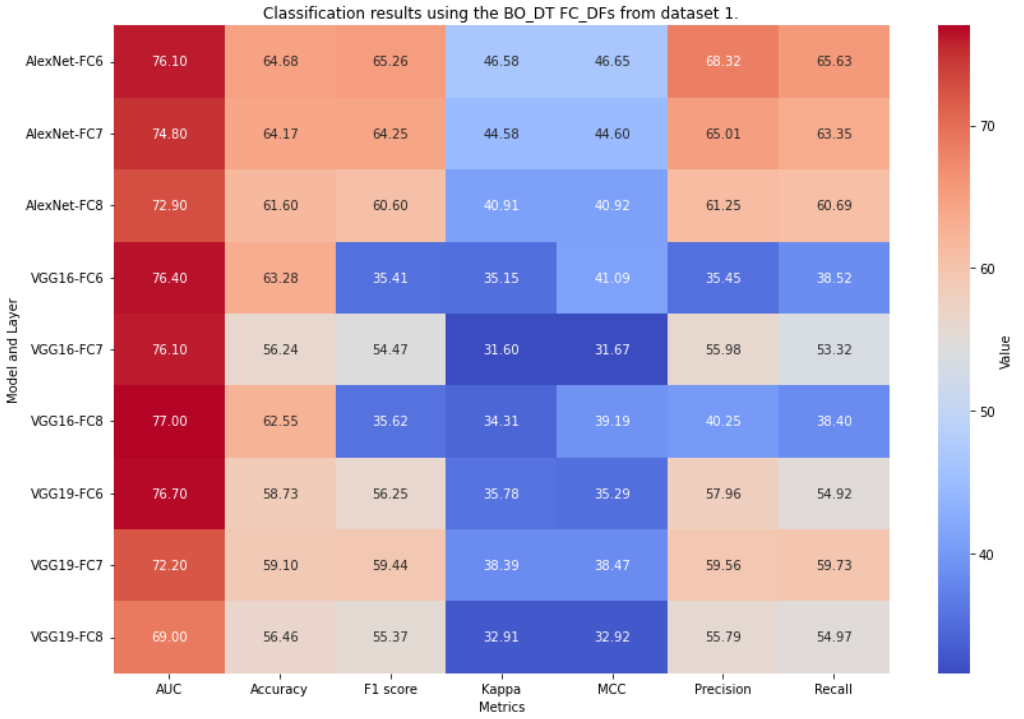
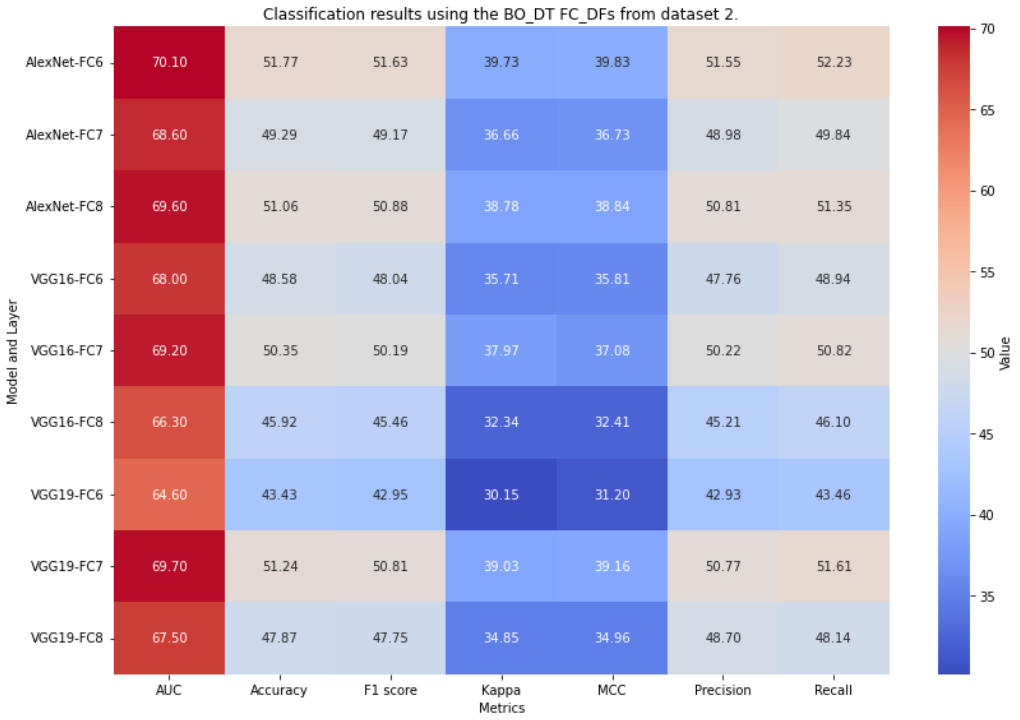


Figure 5.5: The ROC curve of the BO-KNN classifier for the two data sets using FC_DFs. Images from (a) to (c) show the roc curve of the AlexNet DFs for FC6 to FC8, from (d) to (f) the VGG16 DFs, and from (g) to (i) the VGG19 DFs.

Figure 5.6 shows heatmaps of BO-DT classification results for two datasets. In Dataset 1, AlexNet's FC6 earned the highest accuracy (64.68%) and demonstrated promising recall, F1 score, and precision (see Figure 5.6a). It had a remarkable 76.10% AUC. Despite a lesser accuracy of 63.28%, VGG16's FC6 has a superior precision of 35.45% and recall of 38.52% with an AUC of 76.40%. Figure 5.6b demonstrates AlexNet's FC6's high recall, precision, and accuracy of 51.77% in Dataset 2. It had a 70.10% AUC. Both datasets illustrate that DFs affect classification performance; AlexNet's FC6 consistently beats others. Moderate accuracy ratings must be noted.



(a) BO_DT classification results using the FC_DFs from dataset 1.



(b) BO_DT classification results using the FC_DFs from dataset 2.

Figure 5.6: Heatmaps classification results of BO_DT with the FC_DFs.

For both datasets, the ROC curves for the BO-DT classifier using different FC_DFs are shown in figure 5.7. Across all datasets, AlexNet's FC6 has the greatest AUC values, demonstrating its remarkable ability to differentiate between AD and normal classes. All DFs show reasonable performance when combined with BO-DT, with AUC values usually

falling between 60% and 75% for the two datasets. Although the BO-DT classifier and the FC_DFs are effective in detecting and diagnosing AD, their performance is not up to par with other classifiers like BO-MSVM and BO-KNN, as seen by these ROC curves.

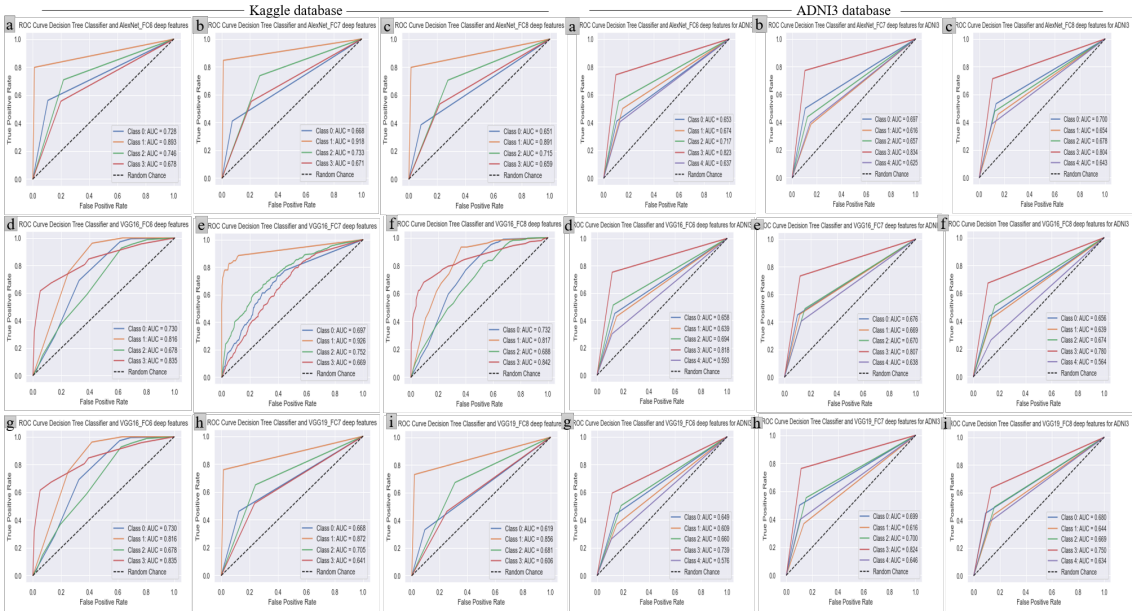


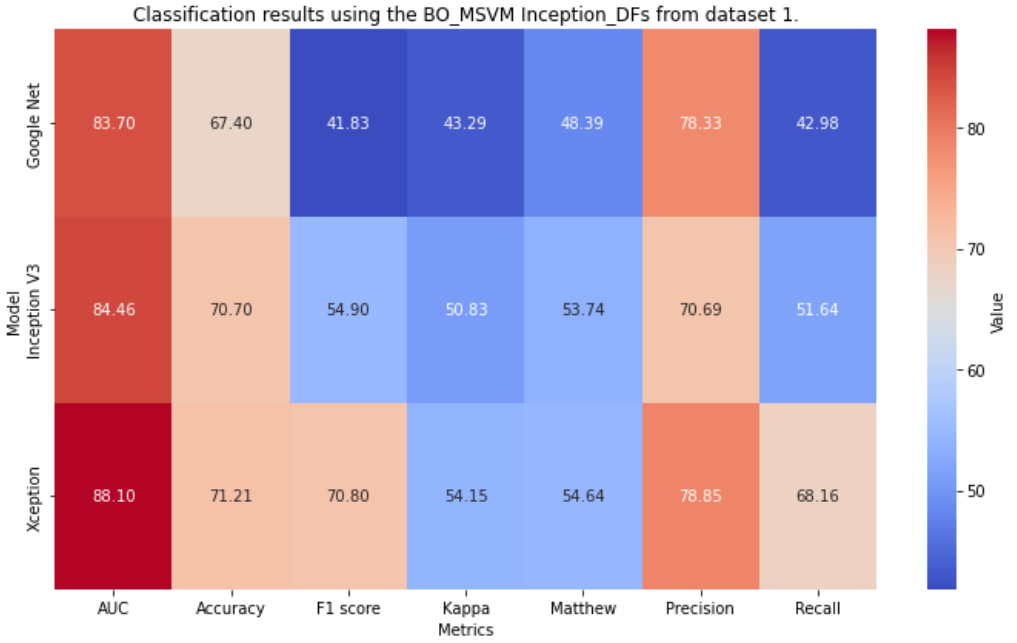
Figure 5.7: The ROC curve of the BO-DT classifier for both databases with the FC_DFs Images from (a) to (c) present the roc curve of the AlexNet DFs for FC6 to FC8, from (d) to (f) present the roc curve of the VGG16 DFs for FC6 to FC8, and from (g) to (i) present the roc curve of the VGG19 DFs for FC6 to FC8 for both databases.

5.5.2 Classification results using inception DFs.

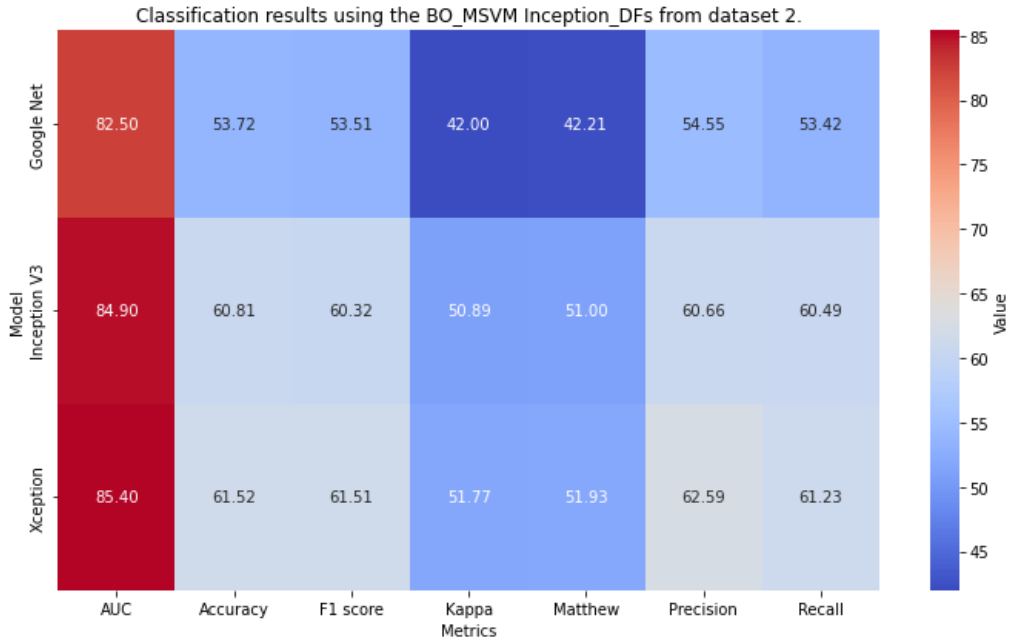
The heatmap figures 5.8, 5.9, and 5.10 present an in-depth examination of the classification outcomes obtained through inception-based features for datasets 1 and 2.

The Xception features produced the greatest findings for both databases (Figure 5.8a and 5.8b), according to the classification findings provided by BO-MSVM in figure 5.8. The second database had an accuracy of 61.52%, whereas Dataset 1 had a maximum of 71.21%. The figures below display the remaining metrics. Similarly, the Xception DFs also yielded the best results using the BO_KNN model for the database(Figure 5.9); the accuracy achieved is 92.21%

The figure 5.10 illustrates the classification results of the BO_DT classifier for datasets 1 and 2, as represented by colors (see Figures 5.10a and 5.10b). Interestingly, dataset 1 outperforms dataset 2 in terms of total performance. In Dataset 1, GoogleNet obtained 58.73% accuracy, Inception V3 earned 59.17%, while Xception achieved 56.24% accuracy. In Dataset 2, GoogleNet, Inception V3, and Xception achieved accuracy rates of 44.68%, 43.97%, and 48.58%, respectively. It is worth noting that Dataset 1 for Inception V3 had the greatest accuracy of all DF tested. Similarly, dataset 1 consistently produced higher AUC values than dataset 2. Xception scored a remarkable AUC of 76.90% in Dataset 1 and 68.10% in Dataset 2. These results demonstrate Dataset 1's higher performance, as well as the efficiency of Inception V3 and Xception deep features in classification tasks.

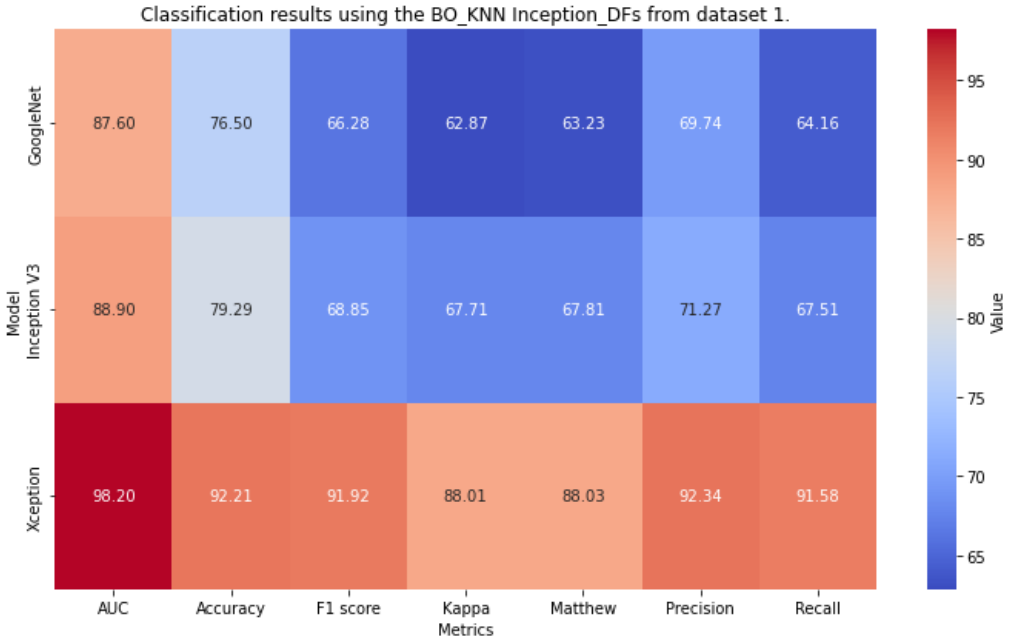


(a) BOMSVM classification results utilizing the Inception features obtained from dataset 1.

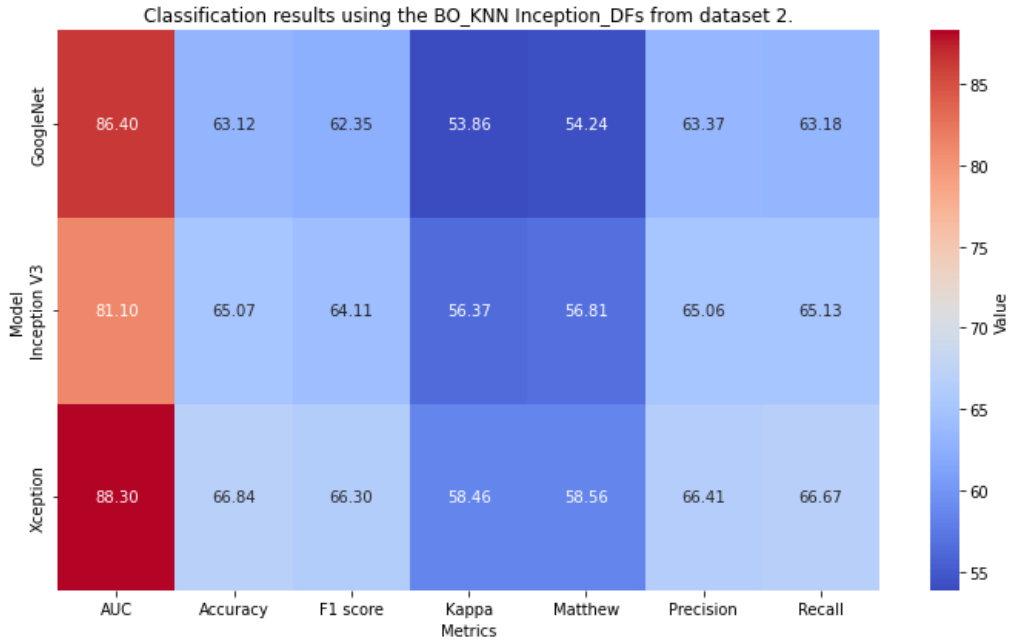


(b) BOMSVM classification results utilizing the Inception features obtained from dataset 2.

Figure 5.8: Heatmaps classification results of BO_MSVM with the Inception features.

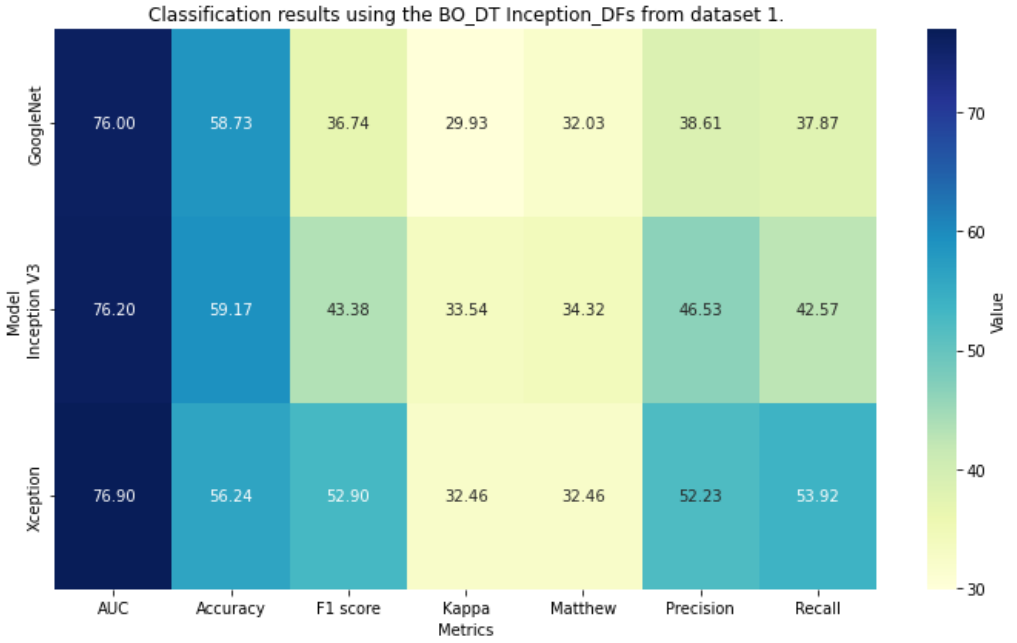


(a) BOKNN classification results utilizing the Inception features obtained from dataset 1.

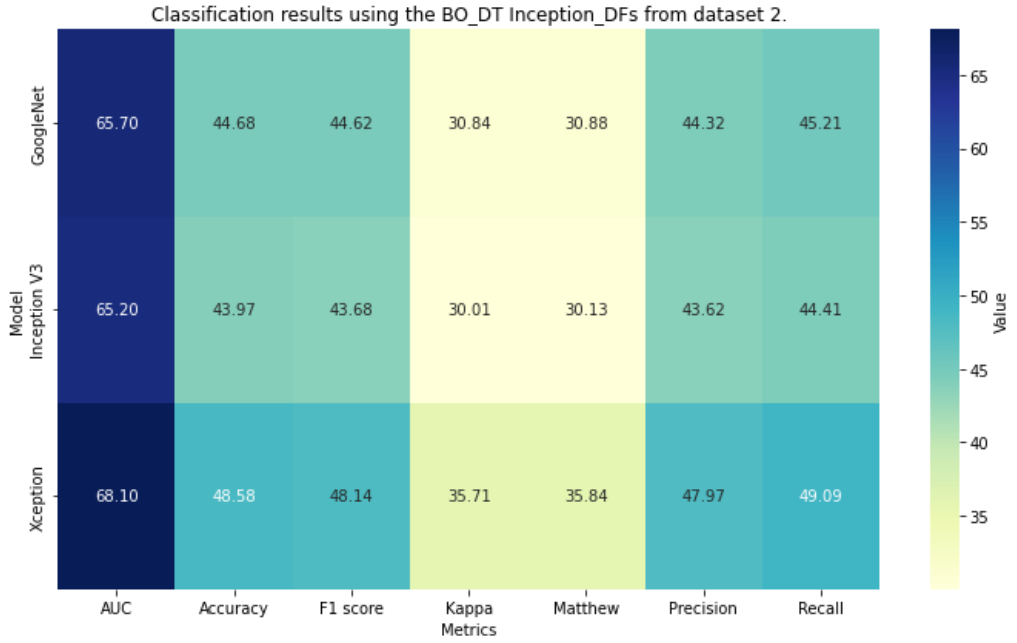


(b) BOKNN classification results utilizing the Inception features obtained from dataset 2.

Figure 5.9: Heatmaps classification results of BO_KNN with the Inception features.



(a) BODT classification results utilizing the Inception features obtained from dataset 1.



(b) BODT classification results utilizing the Inception features obtained from dataset 2.

Figure 5.10: Heatmaps classification results of BO_DT with the Inception features.

The ROC curves of BO_MSVM, BO_KNN, and BO_DT are displayed in Figures 5.11, 5.12, and 5.13, respectively are obtained when utilizing the Inception features.

In Figure ,5.11,The ROC curves for GoogleNet and Xception show a similar pattern across all classes and databases. The curves fall between the ideal curve and the random chance line, indicating noticeable fluctuations. Remarkably, the Class 1 (MOD) curve in Database 1 is nearly perfect, achieving an impressive AUC of 99.9%.

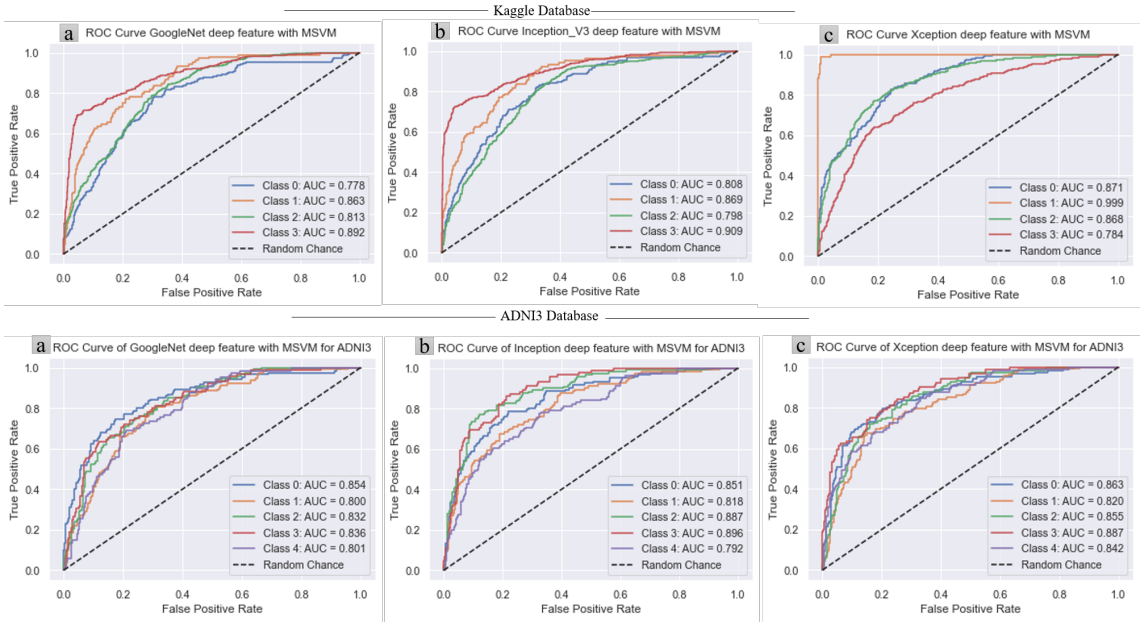


Figure 5.11: The ROC curve Of the BO-MSVM using both databases with the Inception features

Figure 5.12 shows that BOKNN ROC curve approaches the ideal curve, especially when using Xception deep features, with a strong True Positive Rate (TPR). Figure 5.12c shows that the BOKNN classifier is quite sensitive when applied to the Kaggle database.

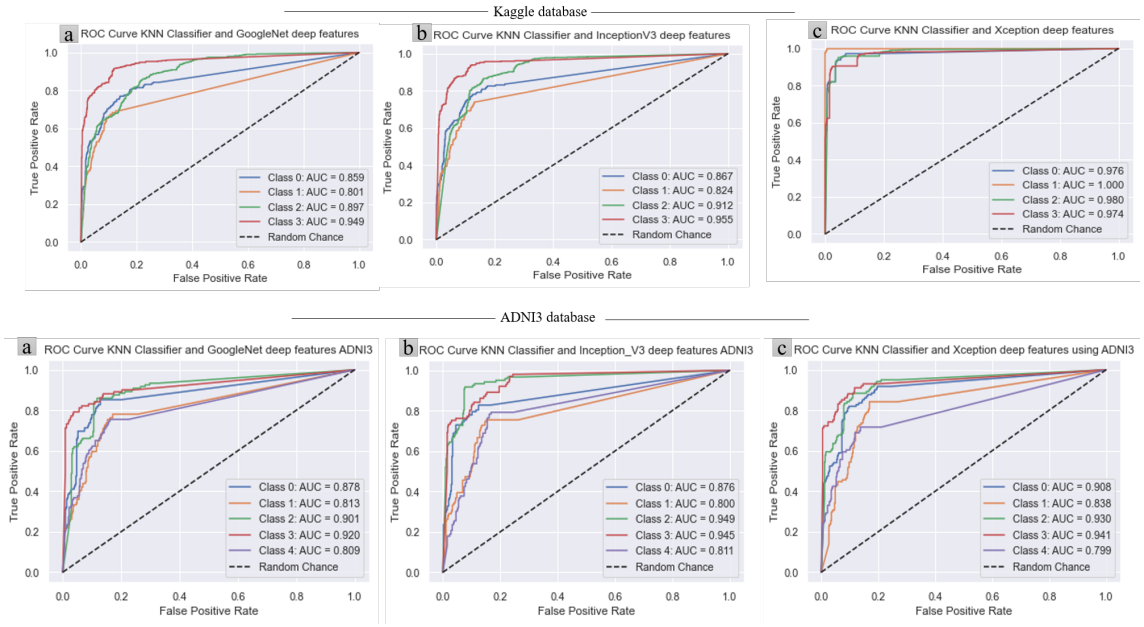


Figure 5.12: The ROC curve Of the KNN using both databases with the Inception features.

Meanwhile, the ROC curve for the BO-DT classifier, presented in Figure 5.13, reveals a high False Positive Rate (FPR) as it nears the random chance line.

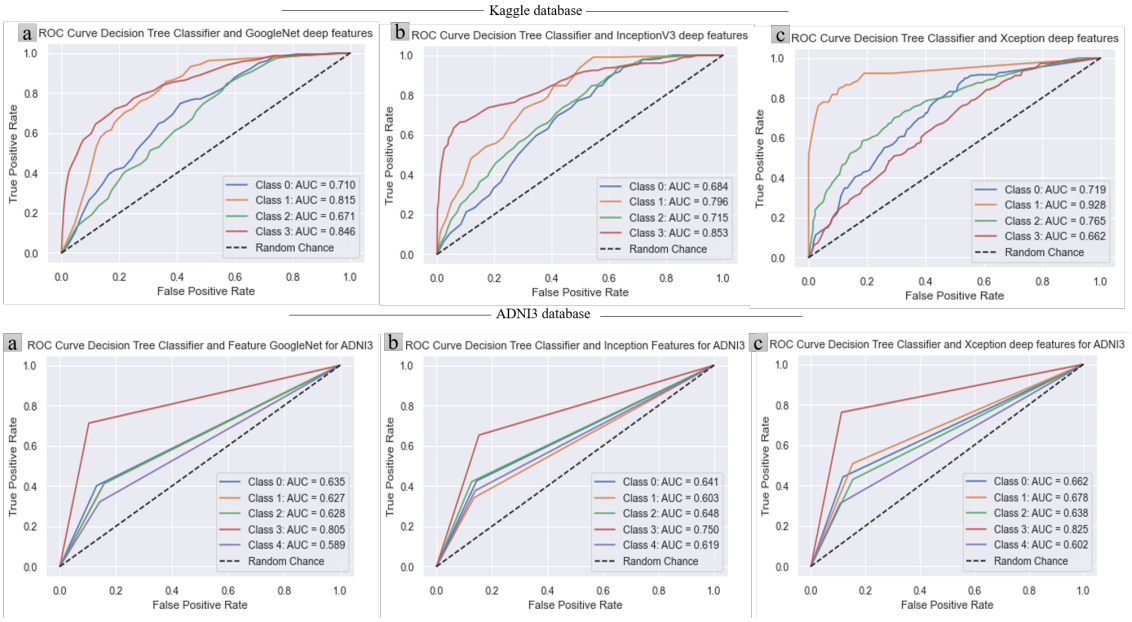
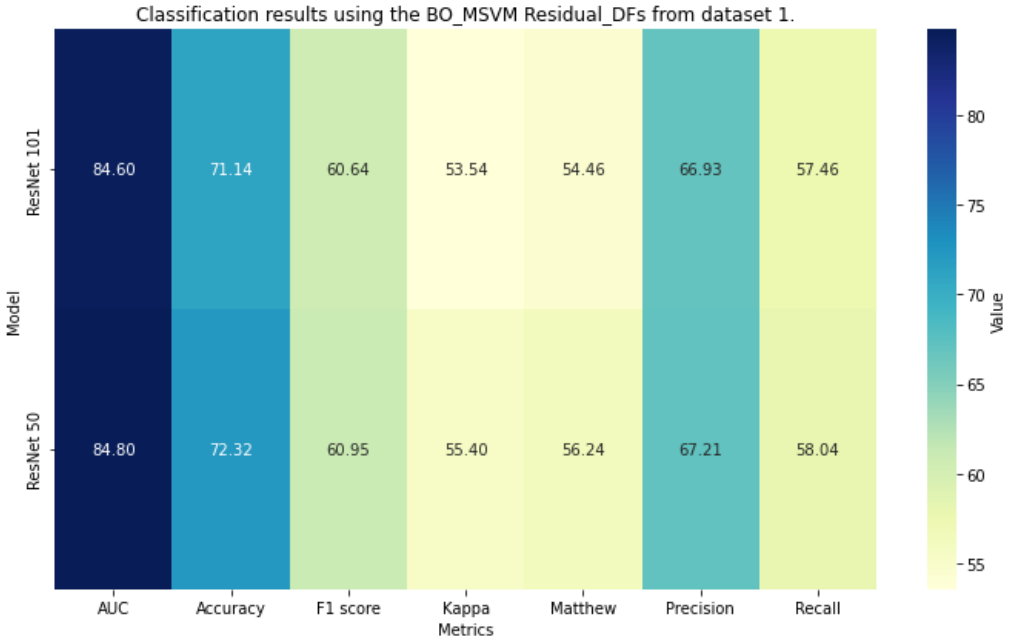


Figure 5.13: The ROC curves Of the DT using both databases with the inception features.

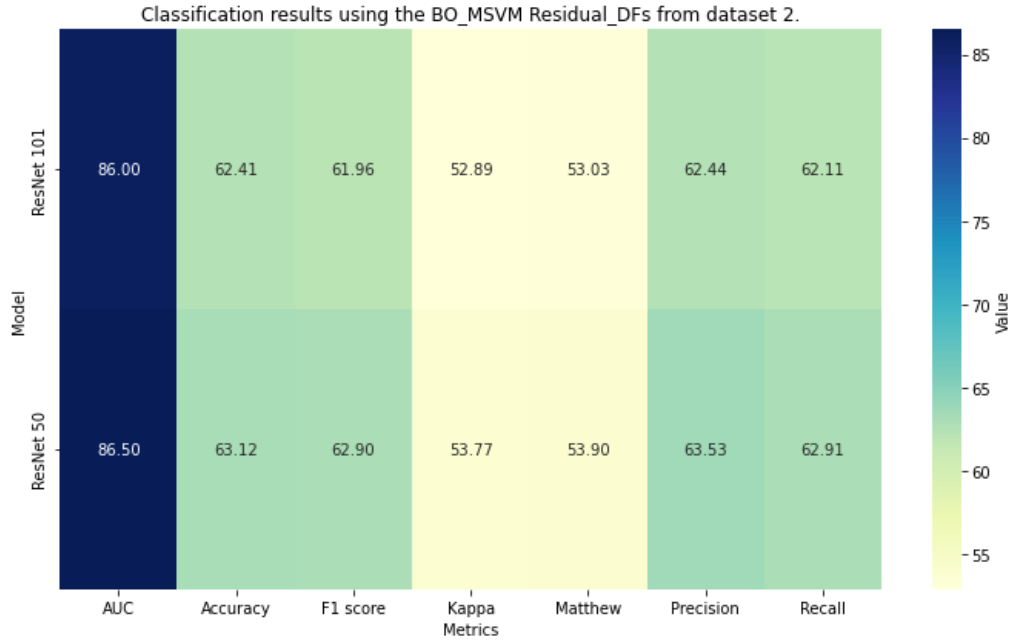
5.5.3 Classification results using Residual Connections DFs.

In the following part, we will present the outcomes of the classifier utilizing the residual deep features.The heatmap in Figure 5.14 illustrates the BO_MSVM classifier performance using Residual-derived features, with two subfigures highlighting key metrics.

In the first subfigure5.14a represents results for dataset 1, achieving an accuracy of 71.16% and an AUC of 86.01%. The second subfigure 5.14b, with a gradient towards yellow and orange, shows moderate performance for dataset 2, with an accuracy of 69.18% and an AUC of 86.60%. Across both datasets, the BO_MSVM model demonstrates consistent performance, with slightly higher AUC values, indicating robust discriminatory power. However, dataset 1 exhibits marginally better accuracy than dataset 2, suggesting varying effectiveness of the residual DFs across the datasets.



(a) BOMSVM classification results using the Residual features from dataset 1.

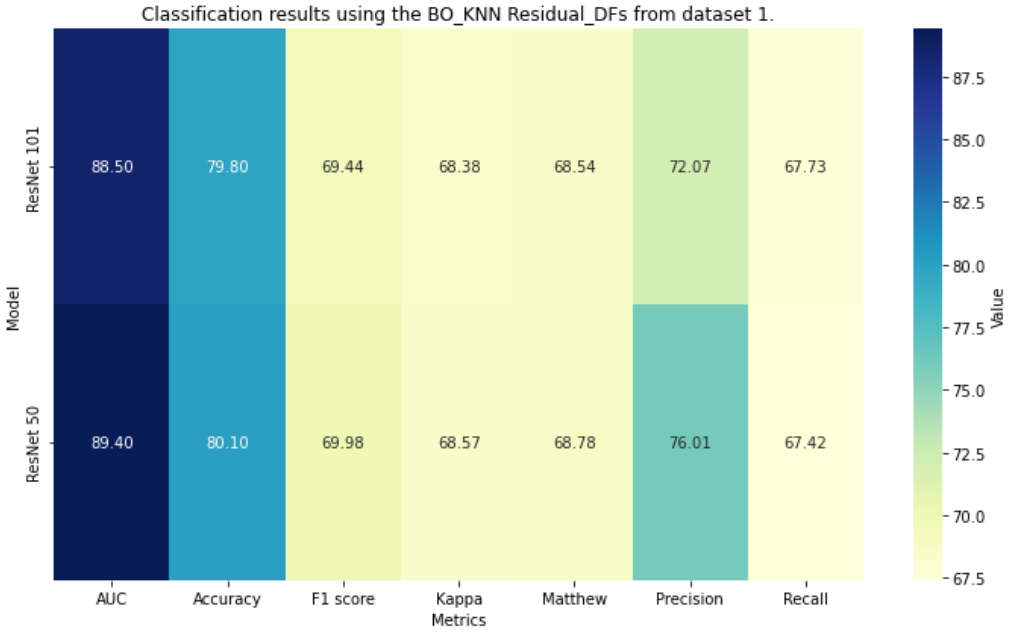


(b) BOMSVM classification results using the Residual features from dataset 2.

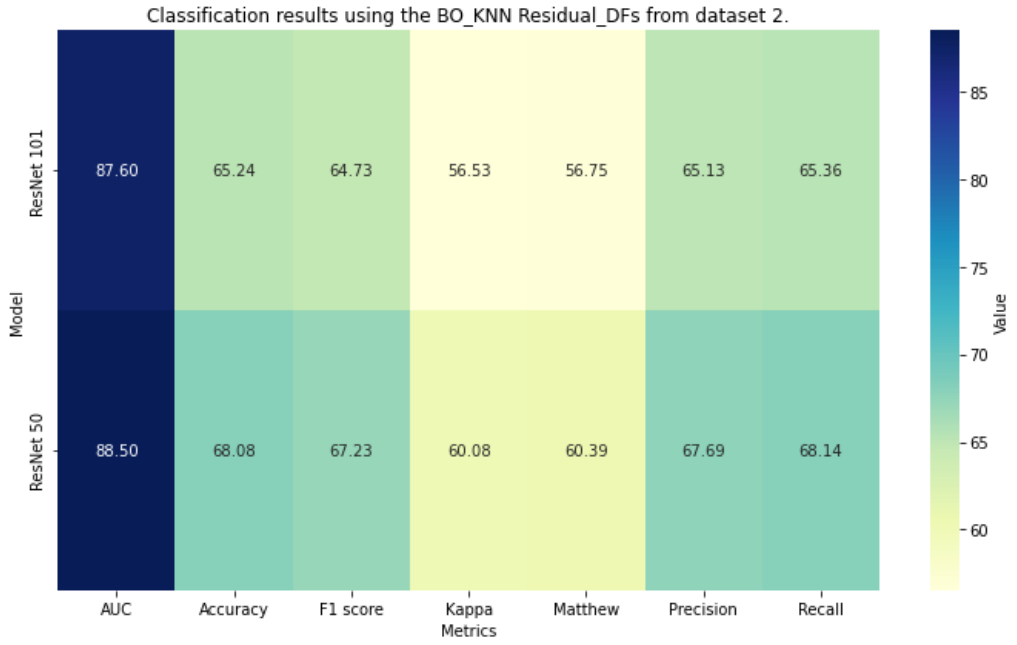
Figure 5.14: Heatmaps classification results of BO_MSVM with the Residual connections features.

The heamap figure 5.15 illustrates the classification results of the BO_KNN model using Residual Deep Features from two datasets, visualized as heatmaps with a blue-to-yellow color gradient indicating performance values. Subfigure 5.22a, corresponding to dataset 1, achieves an accuracy of 79.00% and an AUC of 89.50%, with the darker blue shades highlighting high AUC and accuracy. Subfigure 5.15b, for dataset 2, shows an accuracy of 65.24% and an AUC of 87.60%, with a shift toward lighter blue-green shades for both metrics. The heatmaps reveal stronger model performance on dataset 1 compared

to dataset 2, as evidenced by the higher accuracy and AUC values in the darker regions.



(a) BOKNN classification results using the Residual features using dataset 1.



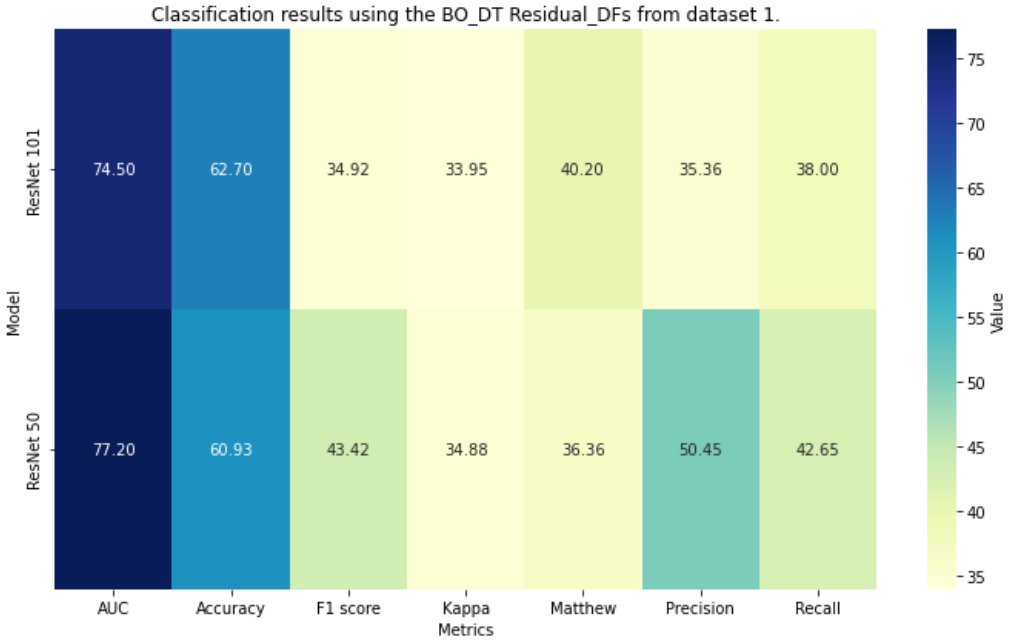
(b) BOKNN classification results using the Residual features from dataset 2.

Figure 5.15: Heatmaps classification results of BO_KNN with the Residual connections features.

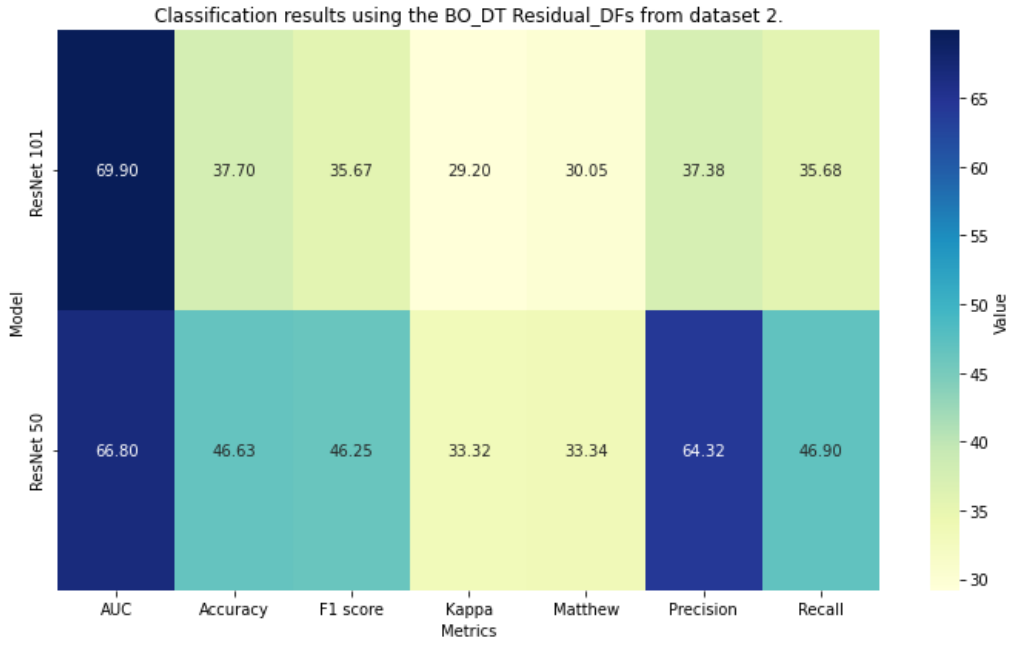
Figure 5.16 showcases the classification performance of the BO_DT model using Residual Deep Features for two datasets. The heatmaps employ a blue-to-yellow gradient, where darker blue represents higher values and yellow signifies lower values.

Subfigure 5.16a corresponds to dataset 1, achieving an accuracy of 62.70% and an AUC of 74.50%, with relatively darker shades highlighting these metrics. Subfigure 5.16a

represents dataset 2, with a higher accuracy of 46.63% and an AUC of 66.80%, demonstrating weaker performance compared to dataset 1, as reflected in the lighter shades. These results indicate the model performs better on dataset 1, particularly in terms of AUC and accuracy.



(a) BODT classification results using the Residual features from dataset 1.



(b) BODT classification results using the Residual features from dataset 2.

Figure 5.16: Heatmaps classification results of BO_DT with the Residual connections features.

The subsequent figures illustrate (Figures 5.17, 5.18, and 5.19) the ROC curve outcomes for BO_MSVM, BO_KNN, and BO_DT, respectively, employing the residual DFs.

The figures display a similar pattern, characterized by a quick initial increase that indicates effective identification of positive instances. However, a closer look at the curves in Figures 8 and 9 reveals that BO_KNN slightly outperforms BO_MSVM. This is evident from the ROC curve of BO_KNN, which closely aligns with the optimal curve, indicating improved sensitivity.

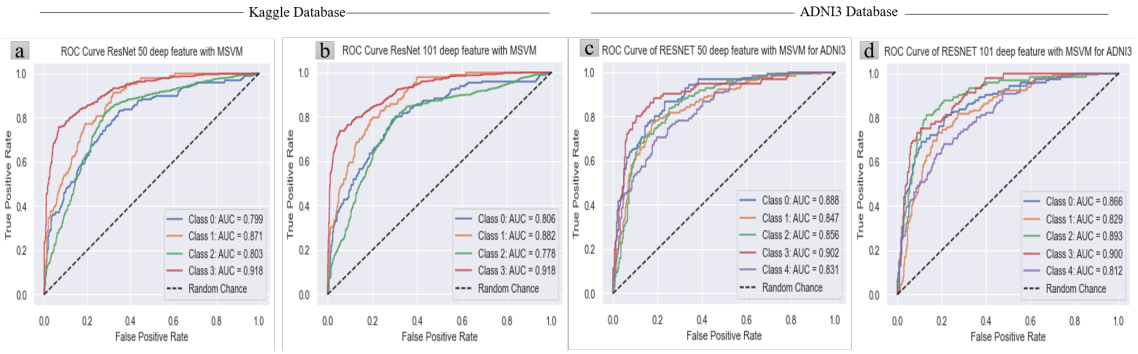


Figure 5.17: The ROC curve Of the MSVM classifier for each datasets with the residual features.

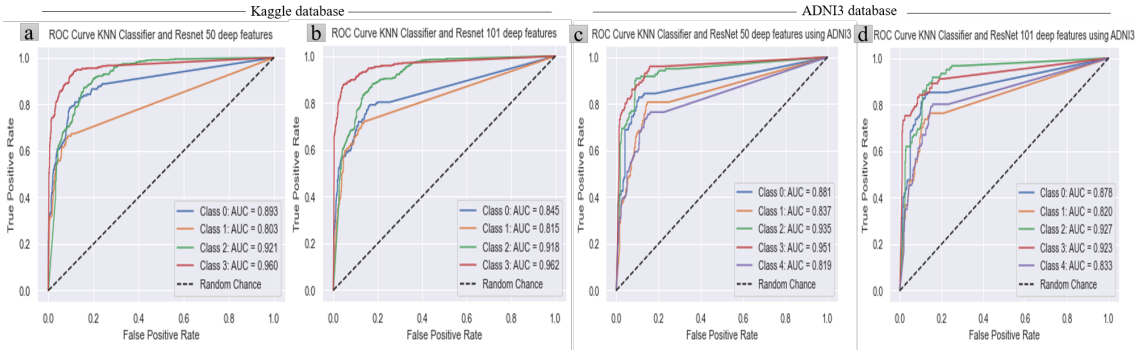


Figure 5.18: The ROC curve Of the KNN classifier for each datasets with the residual features.

Conversely, Figure 5.19 demonstrates that the ROC curve for BO-DT closely resembles that of a random classifier, exhibiting equal true positive rate (TPR) and false positive rate (FPR), indicating that this classifier struggles to detect positive instances.

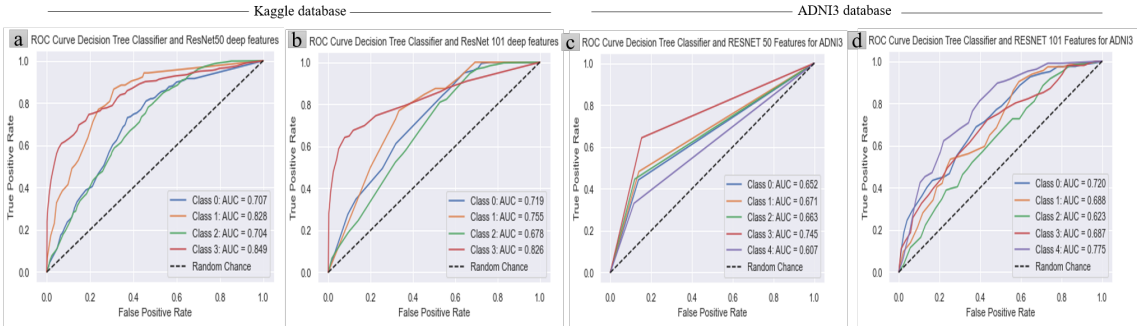


Figure 5.19: The ROC curve of the decision tree classifier for each datasets with the residual features.

5.5.4 Classification results using EfficientNet b0 DFs.

In this section, the fine-tuned ML classifier results with efficientNet b0 DFs were presented in heatmaps colors 5.20, 5.21, and 5.22 for BO_MSVM, BO_KNN, and BO_DT, respectively.

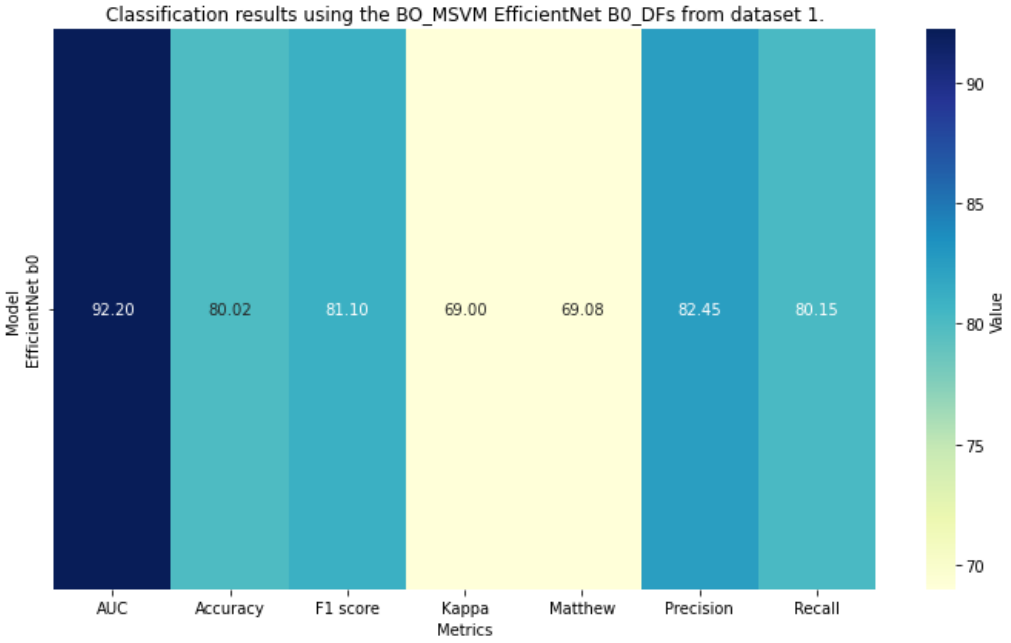
The results presented in figure 5.20 demonstrate the performance of the EfficientNet B0 deep features in datasets 1 and 2.

The heatmap in subfigure 5.20a illustrates the classification results of the BO_MSVM classifier using EfficientNet B0-derived features for dataset 1. The AUC value of 92.20, displayed in blue, indicates excellent discriminative ability, reflecting the model's effectiveness in distinguishing between classes. The accuracy of 80.02, shown in warm yellow-green, demonstrates strong classification performance. The F1 score of 81.10, also in yellow-green, highlights a balanced trade-off between precision and recall.

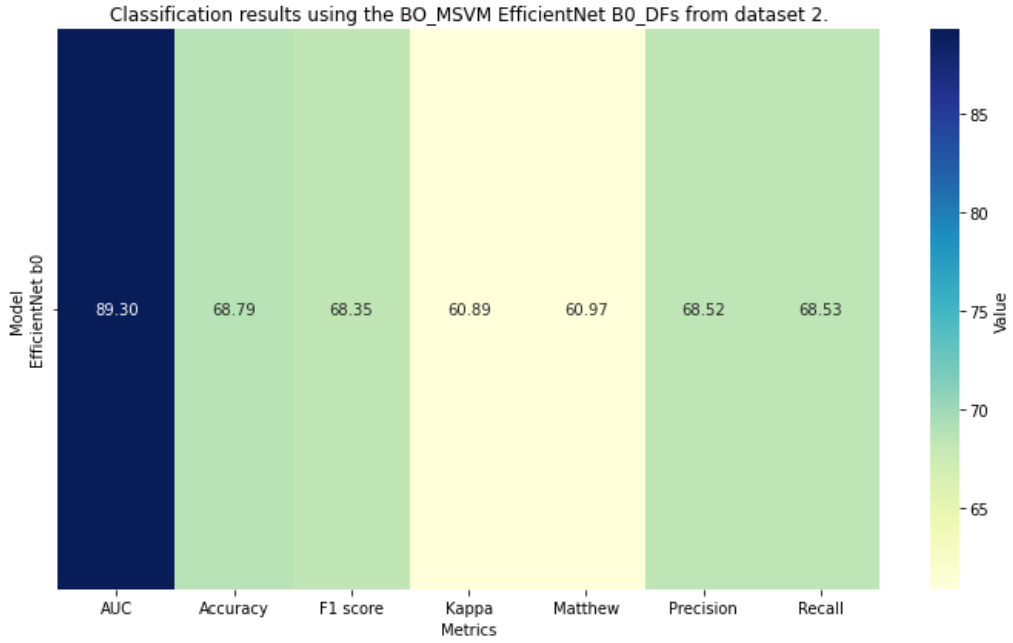
The Kappa metric (69.00) and Matthew's correlation coefficient (69.08), both in lighter yellow, reveal substantial agreement and correlation between predicted and actual labels. Precision, at 82.45 in brighter yellow, shows a high rate of true positive predictions, while the recall of 80.15, also in yellow, indicates a strong ability to correctly identify positive instances. Overall, the heatmap's gradient from green to yellow-green visually captures the model's high performance across all metrics, with the green AUC value standing out as a particular strength. This visualization underscores the BO_MSVM classifier's effectiveness on dataset 1, with consistent results across accuracy, precision, recall, and other key metrics.

The heatmap in subfigure 5.20b presents the classification results of the BO_MSVM classifier using EfficientNet B0-derived features for dataset 2. The AUC value of 89.30, depicted in green, indicates strong discriminative ability, reflecting the model's effectiveness in distinguishing between classes. The accuracy of 68.79, shown in yellow, suggests moderate performance with room for improvement. The F1 score of 68.35, also in yellow, highlights a balanced trade-off between precision and recall but indicates potential for further optimization.

The Kappa metric (60.89) and Matthew's correlation coefficient (60.97), both in lighter yellow, reveal moderate agreement and correlation between predicted and actual labels, pointing to areas where the model performs adequately but could be enhanced. Precision, at 68.52 in yellow, shows a moderate rate of true positive predictions, while the recall of 68.53, also in yellow, indicates a decent ability to correctly identify positive instances. Overall, the heatmap's gradient from green to yellow visually encapsulates the model's performance, with the green AUC value standing out as a strength, while the yellow tones in accuracy, F1 score, and recall highlight areas for potential refinement. This visualization underscores the BO_MSVM classifier's solid performance on dataset 2, with consistent results across key metrics but opportunities for further improvement.



(a) BO_MSVM classification results using the EfficientNet B0 DFs from dataset 1.



(b) BO_MSVM classification results using the EfficientNet B0 DFs from dataset 2.

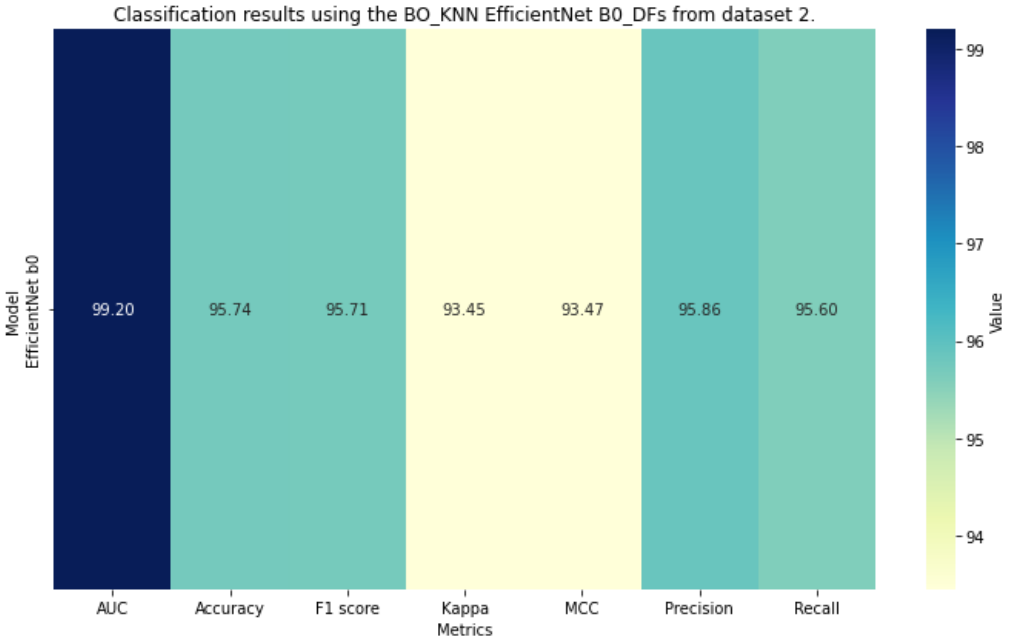
Figure 5.20: Heatmaps classification output of BO_MSVM with the EfficientNet B0 DFs.

Figure 5.21 displays the classification outcomes of the EfficientNet B0 DFs with BO_KNN by the color variations where the blue indicate the highest performance and the light yellow indicated the lowest performance.

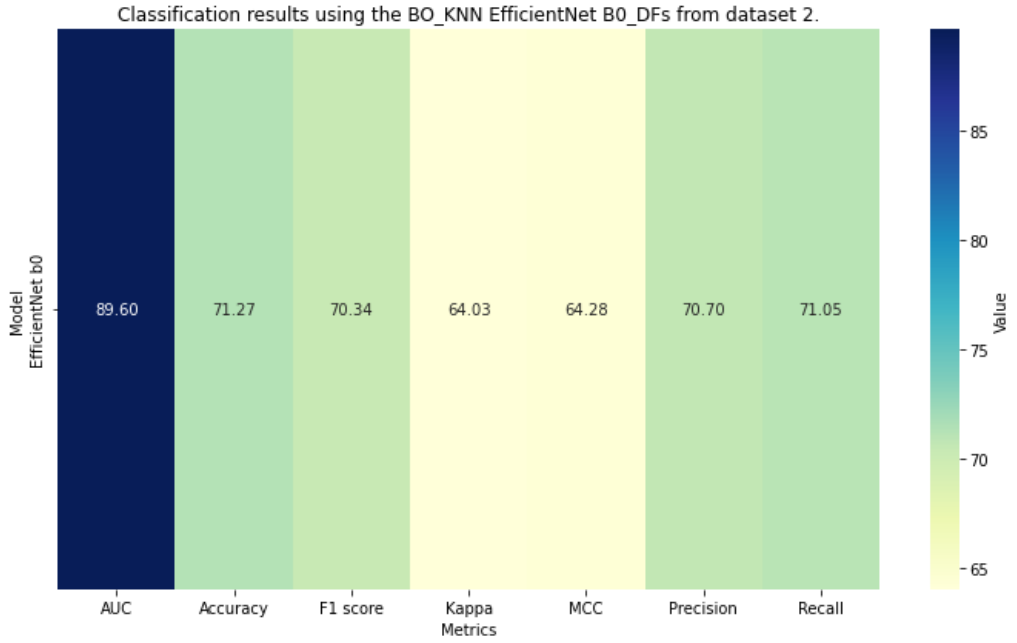
The heatmap in sunfigure 5.21a shows the BO_KNN classifier's performance using EfficientNet B0 features for dataset 1. The AUC of 99.20, in vibrant green, indicates near-perfect class discrimination. Accuracy (95.74) and F1 score (95.71), in green-yellow, reflect outstanding classification and a strong balance between precision and recall. The Kappa

(93.45) and MCC (93.47), in lighter green-yellow, show high agreement and correlation. Precision (95.86) and recall (95.60) further confirm the model's excellent ability to identify true positives. Overall, the heatmap's green to green-yellow gradient highlights the model's exceptional performance across all metrics.

The heatmap in subfigure 5.21b displays the BO_KNN classifier's performance using EfficientNet B0 features for dataset 2. The AUC of 89.60, indicates class discrimination. Accuracy (71.27) and F1 score (70.34), in yellow-green, show solid but improvable performance. The Kappa (64.03) and MCC (64.28), in lighter yellow, reveal moderate agreement and correlation. Precision (70.70) and recall (71.05), also in yellow, suggest decent but not optimal identification of true positives.



(a) BO_KNN classification results using the EfficientNet B0 DFs from dataset 1.



(b) BO_KNN classification results using the EfficientNet B0 DFs from dataset 2.

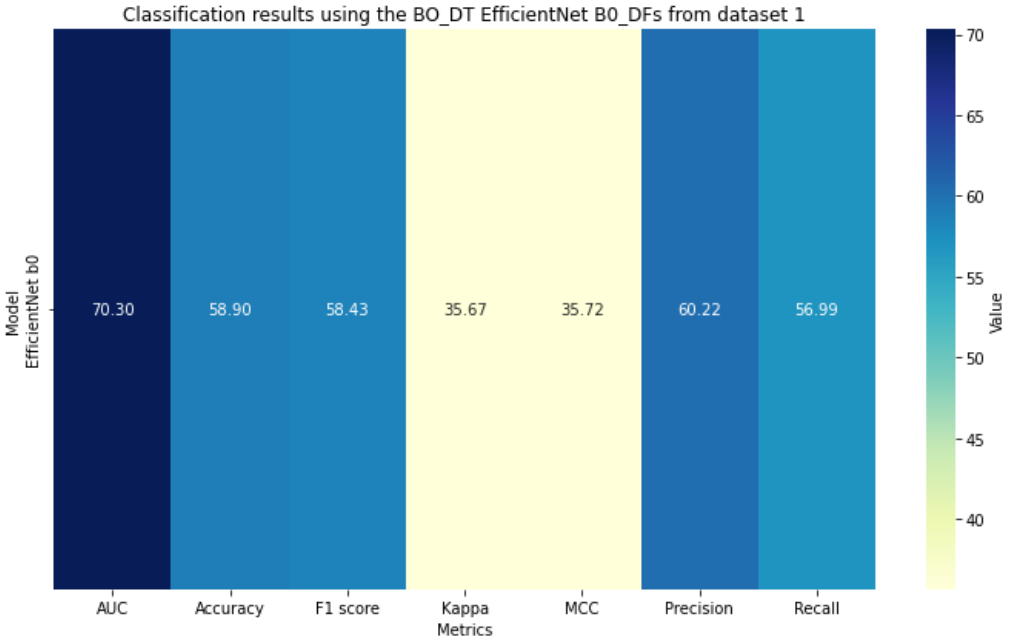
Figure 5.21: Heatmaps classification results of BO_KNN with the EfficientNet B0 DFs.

The heatmaps in Figure 5.22 deliver an extensive graphical representation of the classification outcome using the BO_DT classifier with EfficientNet B0-derived features for the two distinct datasets.

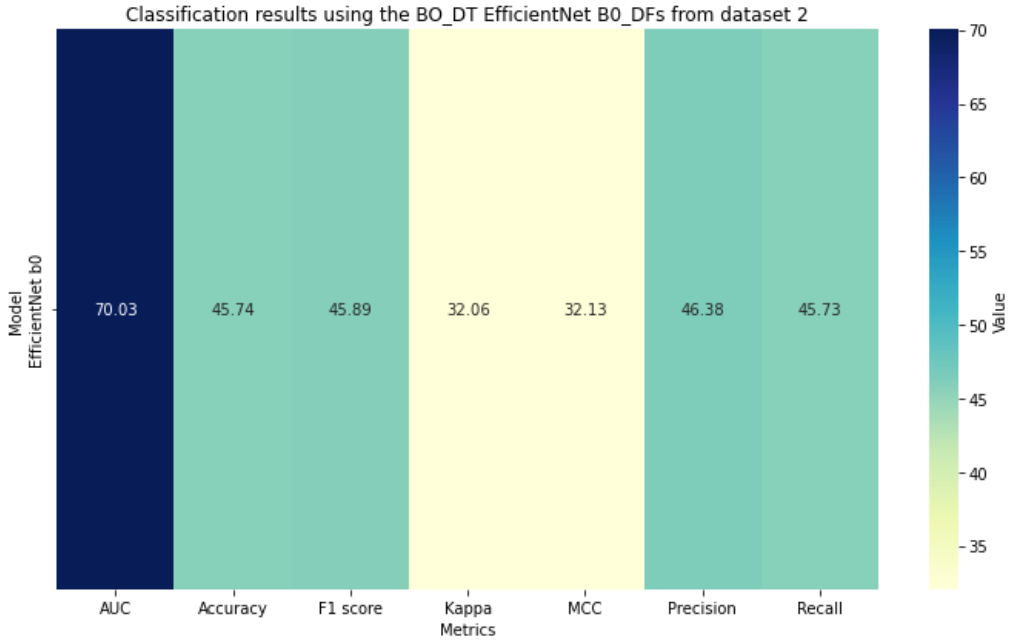
The heatmap in subfigure 5.22a shows the BO_DT classifier's performance using EfficientNet B0 features for dataset 1. The AUC of 70.30, in vibrant green, indicates strong class discrimination. Accuracy, at 58.90 in yellow-green, reflects moderate performance with room for improvement. The F1 score of 58.43, in lighter yellow, shows a balance be-

tween precision and recall but suggests further optimization is needed. The Kappa (35.67) and MCC (35.72), in cooler yellow, indicate moderate agreement and correlation between predictions and actual labels. Precision, at 60.22 in brighter yellow, shows a high rate of true positives, while recall (56.99, yellow) suggests decent but improvable identification of positive instances.

The heatmap in subfigure 5.22b shows the BO_DT classifier's performance using EfficientNet B0 features for dataset 2. The AUC of 70.03, in vibrant green, indicates strong class discrimination, while the accuracy of 45.74, in yellow, suggests moderate performance with room for improvement. The F1 score of 45.89, also in yellow, reflects a balance between precision and recall but signals potential for optimization. The Kappa (32.06) and MCC (32.13), in lighter yellow and orange, reveal moderate agreement and correlation, highlighting areas of difficulty. Precision (46.38) and recall (45.73), both in orange, indicate challenges in true positive predictions and identifying positive instances. The heatmap's gradient from green to orange emphasizes the model's strong AUC but underscores limitations in accuracy, precision, and recall, pointing to areas needing refinement.



(a) BO_DT classification results using the EfficientNet B0 DFs from dataset 1.



(b) BO_DT classification results using the EfficientNet B0 DFs from dataset 2.

Figure 5.22: Heatmaps classification results of BO_DT with the EfficientNet B0 DFs.

The ROC curves presented below offer a comprehensive analysis of the classification performance achieved using EfficientNetB0-derived features combined with optimized machine learning algorithms. Figures 5.23, 5.24, and 5.25 illustrate the ROC curves for the BO_MSVM, BO_KNN, and BO_DT classifiers, respectively.

In Figure 5.23a, Class 1 stands out with an exceptional True Positive Rate (TPR) and a perfect Area Under the Curve (AUC) of 100%, outperforming the other classes, which also exhibit strong TPRs. Figure 5.23b shows a ROC curve that closely follows

the optimal curve, highlighting a robust balance between TPR and False Positive Rate (FPR). This alignment underscores the classifier's effectiveness in distinguishing between classes with high precision and reliability.

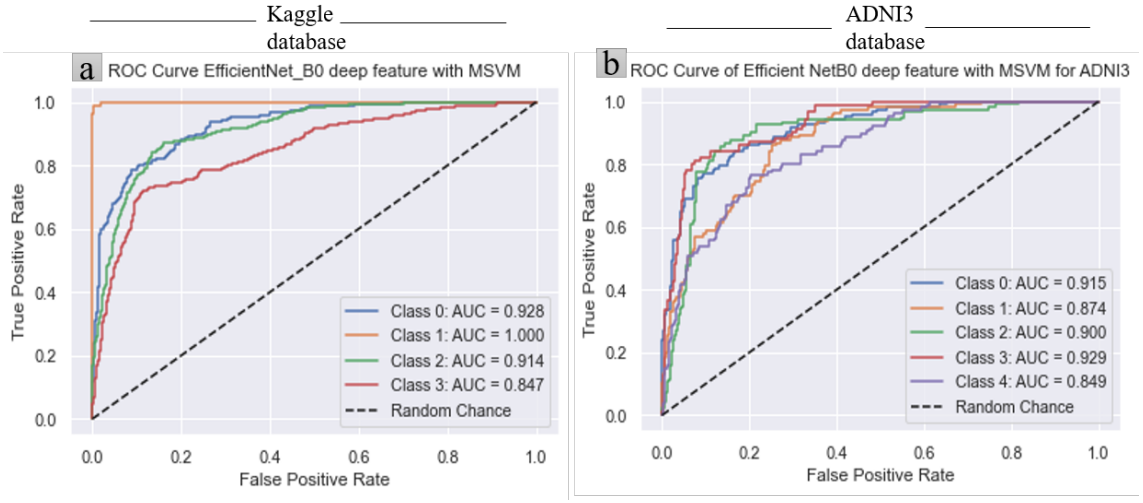


Figure 5.23: ROC Curve Analysis of the BO_MSVM Classifier Using EfficientNet B0 Features Across Both Databases.

Figure 5.24b demonstrates a comparable pattern to Figure 5.23b, yet showcases a more consistent convergence across different thresholds on the ROC curve. In contrast, Figure 5.24a exhibits an nearly flawless ROC curve for the Kaggle classes, signifying a minimal false positive rate (FPR).

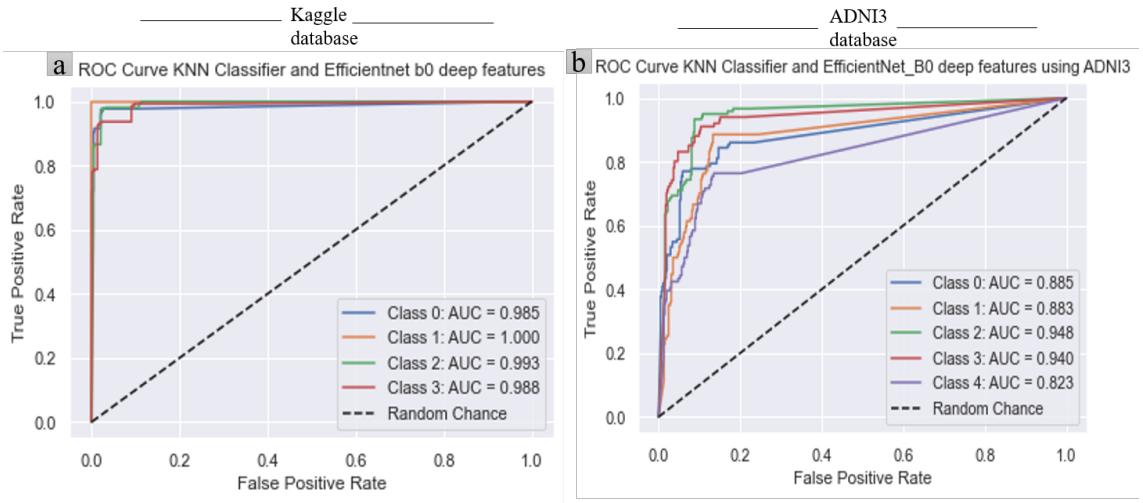


Figure 5.24: ROC Curve Analysis of the BO_KNN Classifier Using EfficientNet B0 Features Across Both Databases.

In Figure 5.25, both Figures 5.25a and 5.25b imply that the ROC curves corresponding to the BO-DT classifier exhibit a performance that is only slightly superior to random chance, suggesting that the disparity between the True Positive Rate (TPR) and the False Positive Rate (FPR) is negligible.

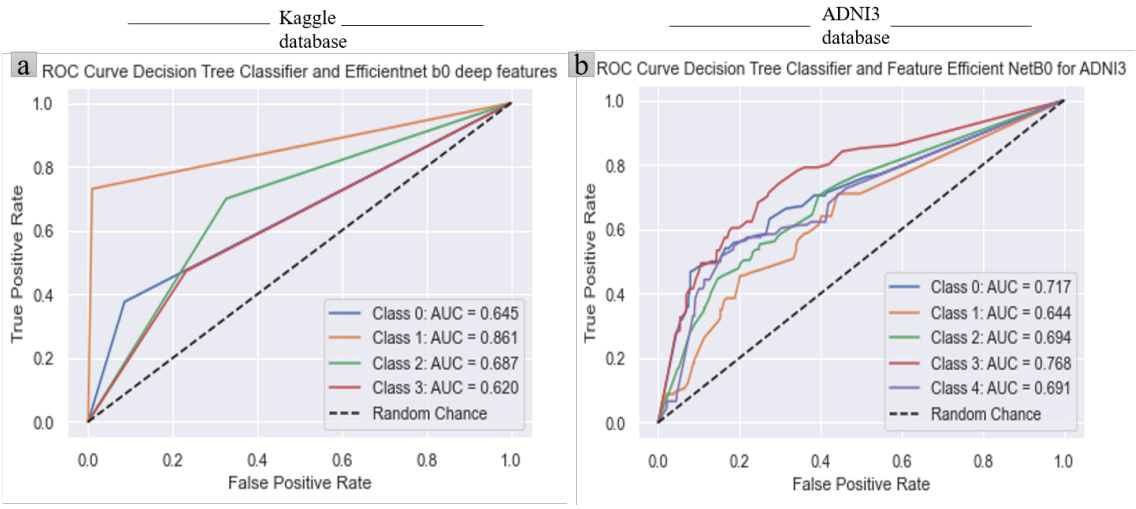


Figure 5.25: ROC Curve Analysis of the BO_DT Classifier Using EfficientNet B0 Features Across Both Databases.

5.6 Discussion

This chapter introduces a machine learning approach designed to classify the stages of Alzheimer's Disease (AD) with enhanced accuracy. Our objective is to streamline the feature extraction process and optimize machine learning algorithms by analyzing variations in feature vectors. To achieve this, we conduct an in-depth exploration of transfer learning techniques, focusing on deep feature extraction. Our goal is to identify the most effective transfer learning architecture for extracting meaningful features that can accurately differentiate between closely related stages of AD.

As this methodology is a subsequent development of the previously proposed methods in this thesis, it is important to highlight that our proposed methodology aims to address the challenges encountered earlier. Therefore, we utilized the previous CNN models for deep feature extraction and grouped these features based on the nature of the utilized architecture. This resulted in four sets of features obtained from four different experiments. Based on these experiments, we found that the fully connected deep features hold the most discriminative characteristics, leading to the highest performance of the optimized ML algorithms, particularly the optimized MSVM and KNN.

The results show that both the BOMSVM and BOKNN classifiers perform optimally when utilizing the fully connected layers of the AlexNet architecture. Additionally, the BOKNN achieved the highest accuracy for the Kaggle database, whereas the BOMSVM excelled in performance for the ADNI database.

All other metrics further support these findings, indicating strong predictive capabilities and reliable identification of positive instances.

Utilizing the Kaggle database, the BOKNN classifier attained 98.45%, reflecting its robust predictive abilities.

When employing the ADNI3 database, the BOMSVM algorithm outperformed the other algorithms, yielding an accuracy of 78.54%. This indicates an acceptable level of predictive capacity, further validated by an AUC value of 93.30%, proving the BO-MSVM classifier's excellent discriminative capacity.

Understanding the differences between the two databases that may affect performance

is essential. Kaggle has more samples, resulting in higher accuracy. In contrast, ADNI3 includes a wider range of disease categories, notably cognitive decline, which complicates the classification task. Furthermore, the cranial images in ADNI3 present extraneous information that could impede performance, especially compared to the samples in Kaggle database, which consist solely of brain tissue. The difficulty in differentiating between closely related categories in ADNI3, including MCI, LMCI, and EMCI, further diminishes its total score.

We evaluate the optimal classifier for each dataset by examining factors including dataset size, DF attributes, and the complexity of decision boundaries. These criteria are crucial for the precise classification of instances into their appropriate categories. Therefore, we aim to determine the decision boundaries which successfully distinct between stages. Since model hyperparameters significantly influence this process, we employ Bayesian Optimization to find the optimal hyperparameters. Across all experiments and databases, the KNN hyperparameter identified was the distance for weighting, which ensures that closer neighbors have a greater impact on predictions. The k values are 4 for the first dataset and 3 for the second, improving the model's responsiveness to localized patterns. The MSVM hyperparameters were also set to C = 1, gamma=0.01, and a polynomial kernel, those hyperparametrs are beneficial for capturing intricate relationships due to multiple classes and potentially overlapping stages.Here, a polynomial kernel was utilized for FC6 and FC7, whereas a linear kernel was implemented for FC8.

This study examines transfer learning dense features for classifying Alzheimer's disease. The models ResNet50, Xception, and EfficientNet B0 demonstrated strong outcome utilizing KNN and MSVM classifiers, achieving AUCs of 89.40% and 88.50% for KNN and 84.80% and 86.50% for MSVM with ResNet50; 98.20% and 88.50% for KNN and 88.10% and 85.40% for MSVM with Xception; and 99.20% and 89.60% for KNN and 92.20% and 89.30% for MSVM with EfficientNet B0, respectively.

Transfer learning CNNs are designed to automatically learn specific features relevant to the problem. Initially, these models utilize filters to extract low-level features, In contrast, the deeper layers capture more intricate characteristics. The feature learning process generates numerical representations of data patterns that assist in classifying the stages of AD. Each architecture possesses attributes such as depth, capacity, number of layers, kernel size, and training strategy that influence performance outcomes.

Numerous imaging studies have focused on AD classification and detection, tackling binary and multi-classification challenges with varying results. However, these studies varied in effectiveness and reliability depending on sample size, data quality, extracted features, and method implementation. These variances emphasize the significance of thoroughly investigating these elements when evaluating and comparing findings with ours. Consequently, we have compared our results with studies employing TL in conjunction with ML algorithms as presented in the following table 5.13, which state and highlight the outperformance of our proposed methodology.

Table 5.13: Comparative study of the performance outcomes of the suggested approach with the latest research.

References	Aproches	database	Accuracy
[119]	VGG-C	Kaggle	80.13%
[120]	neural network with VGG16	Kaggle	90.40%
			71.10%
[121]	sparse feature	ADNI	59.16%
[122]	Modified AlexNet	Kaggle	91.70%
[75]	TL	Kaggle	VGG-19= 92.86%, VGG-16,= 92.83%, Inception-V3= 91.04%, Xception,= 90.57% ResNet-50=85.99%, DenseNet169=88.64%
[74]	ResNet 50	Kaggle	93.00%
	VGG19		94.00%
	Xception		89.00%
	EfficientNetB7		82.00%
	DensNet201		96.00%
[123]	USVM-RFE1	ADNI	CN vs. AD: 100%
			CN vs.MCI:90%
			MCI vs.AD:73.68%
Our Approach [19]	AlexNetFC6+BO-KNN	Kaggle	98.45%
	AlexNetFC6_BO_MSVM		96.84%
	AlexNetFC6_BO_MSVM	ADNI3	78.54%

5.7 Thesis approaches comparison

In this dissertation, our work focuses on classifying various stages of Alzheimer's Disease (AD), aiming to enhance overall disease stage classification using 2D MRI images. As is well known, several factors significantly influence the performance of any machine learning system in classification tasks. These include, first, the quality of the input images; second, the usefulness of the extracted features; and third, the choice of the machine learning algorithm employed. While these factors directly impact the results, a deeper analysis reveals the presence of additional indirect influences that also shape classification outcomes.

To address these challenges, this dissertation systematically considers all direct impact factors and examines them experimentally. Among these factors, the quality of the extracted features emerges as the most critical determinant of classification success. Poor features fail to yield satisfactory results, even when paired with advanced classifiers. Consequently, a significant focus of this research is on extracting and utilizing meaningful features. To this end, we categorize the features into two groups: low-level features and high-level features.

The existing literature offers extensive work regarding low-level features, particularly since the advent of medical imaging and computer vision. However, our literature review revealed a significant gap in using the Histogram of Oriented Gradients (HOG) technique for feature extraction in AD. Although HOG is widely applied in fields such as face recognition, its potential in AD research remains largely unexplored. Considering the pathophysiology of AD, where the disease progression leads to irregular cortical contours, asymmetry, and loss of structural symmetry, HOG is well-suited for capturing these irregularities. Specifically, HOG identifies changes in gradient direction along cortical boundaries, detects shrinking gyri and widening sulci, and highlights asymmetry by comparing gradient features across hemispheres. However, applying HOG alone is insufficient, as it struggles to differentiate abnormal variations from normal ones, particularly in the early stages of AD.

To enhance HOG's performance, we propose integrating it with the BCFCM technique to emphasize subtle contour changes. This combination allows our first proposed approach (M1) to address a notable gap in the literature while achieving our objectives. Despite its potential, this methodology has limitations: HOG focuses primarily on local edge patterns, missing non-contour-related biomarkers such as textural changes and intensity differences. These shortcomings motivated the exploration of our second approach, which shifts the focus to textural features.

MRI images are inherently rich in texture, and detecting only contours for AD stage classification proves insufficient. Consequently, our second approach (M2) targets textural features, moving from the spatial domain to the frequency domain due to its promising outcomes in the literature. While wavelet transforms have been extensively utilized for AD classification, we identified a gap concerning the Fast Finite Shearlet Transform (FFST). Unlike wavelets, FFST remains underexplored in AD research. In this approach, we extracted textural features in the FFST domain, investigated directional significance, and identified that the fourth direction provided the most meaningful features. Furthermore, we performed two-way and four-way classifications, demonstrating competitive performance compared to existing studies. However, this methodology also requires further investigation, as addressing textural features alone does not fully capture the complexities of AD progression.

The first and second approaches rely on low-level features and face similar challenges. These include the need for extensive feature engineering and limited applicability to large-scale datasets. We recognized these limitations and transitioned to high-level features in our third proposed methodology (M3).

High-level feature extraction involves deep learning, which can automatically capture both low-level and high-level features. However, insufficient data for training deep learning models from scratch posed a significant challenge. To overcome this, we employed inductive transfer learning and conducted an extensive literature review. We identified the need to investigate layer-wise fine-tuning strategies in AD applications further. In Chapter 4, we applied this strategy to ten pre-trained CNN models, achieving the best results with the Xception and VGG19 architectures. Nevertheless, the limited and unbalanced nature of the available datasets remained a challenge. To address this, we applied data augmentation techniques to the Kaggle dataset while avoiding augmentation for the ADNI dataset to assess its impact. The results highlighted that specific augmentation techniques, such as rotation and scaling, fail to eliminate overfitting in medical imaging applications.

Despite achieving our objectives, the third approach is highly prone to overfitting and revealed that manually fine-tuning layers is suboptimal and requires automation to account for dataset variations. This limitation led to our fourth proposed methodology

(M4). Here, we focused on using feature representations from pre-trained CNNs as inputs to machine learning classifiers. Additionally, we introduced hyperparameter tuning to optimize classifiers based on feature variations, yielding promising results. Notably, the fully connected layer 6 of the AlexNet architecture provided the best outcomes for both databases. The proposed methodologies are summarized in Table 5.14.

Table 5.14: Summary of the thesis results.

Chapter	Method	Database	Features	Model/Classifier	Accuracy(%)
Chapter 03	M1	Kaggle	18900	MSVM	96.8
		ADNI2		MSVM	96
	M2	Kaggle	30	MSVM	72.8
Chapter 04	M3	Kaggle	/	VGG19	77.23
		ADNI3		Xception	85.19
Chapter 05	M4	Kaggle	4096	KNN	98.45
		ADNI3		MSVM	78.54

5.8 Summary

This chapter addresses various types of features, which we refer to as deep features. We deeply investigate the impact of these features on classification results. Additionally, we optimize the classifier used in this research based on different variations of the feature set. The optimization process was conducted using the BO method, which helps achieve maximum accuracy by automatically selecting the best hyperparameters for each ML model.

We also found that the best results were obtained when using the FC deep feature, which indicates that this feature holds discriminative information that aids in better classifying the AD stages. Although the other features exhibited reliable results compared to the FC features, the FC features outperformed the others when using all the different models. Our experiments show that the SVM and KNN models showcased good performance, especially during optimization.

Answering the questions dedicated to this chapter demonstrates that this approach overcomes the major drawback of traditional machine learning: feature extraction and engineering. We also identified which features are more discriminative when dealing with AD.

At the end of this chapter, we present a review and comparison of the approaches proposed and investigated in this thesis, highlighting the flow and progress of this work.

CONCLUSION

This thesis aims to address Alzheimer's Disease (AD) as a multiclass classification problem and accurately identify each class. Each step in this work presented unique challenges. We proposed several methodologies to tackle these, starting with handcrafted features and machine learning algorithms and progressing to deep learning techniques. While each method has strengths and limitations, they were designed to address specific research questions and improve upon previously proposed approaches. The major focus of this thesis was to classify various stages of AD simultaneously, as this is a key challenge identified in the state-of-the-art literature.

Our first proposed methodology investigated two key aspects: the impact of Regions of Interest (ROIs) and the effectiveness of computer vision feature extraction techniques as discriminators for AD stage classification. Notably, we observed that Histogram of Oriented Gradients (HOG) features, though well-established in face detection systems, were rarely explored in the context of AD feature discrimination. We incorporated an additional step to boost classification accuracy and enhance their performance. This methodology also included integrating the bias correction fuzzy c-means (BCFCM) method, which segments different brain matters. By coupling these techniques, we assessed the utility of HOG features and highlighted cortical region contour variations, achieving satisfactory results.

However, recognizing that HOG features are not textural and that brain tissue exhibits strong textural properties, we explored textural features for AD classification. MRI effectively captures soft tissue information, making textural features more informative. Despite promising findings in the literature, spatial domain-based textural features often yielded suboptimal results.

To overcome this, we shifted to the frequency domain. We proposed using the Shearlet transform—specifically, the Fast Finite Shearlet Transform (FFST)—which had not been utilized for AD feature representation.

The Shearlet's ability to capture substantial directional information led us to calculate the Gray-Level Co-occurrence Matrix (GLCM) in the frequency domain, synchronized with the FFST's directional coefficients. We extracted 30 textural features from each FFST direction in this second methodology, improving classification by leveraging directional and textural synergies.

While these methodologies advanced the feature extraction phase, a key limitation remained: reliance on handcrafted, labour-intensive features requiring domain expertise.

To address this, we transitioned to deep learning, which automates feature extraction. Despite its advantages, deep learning posed challenges, particularly the need for large

datasets—a constraint in medical imaging research. We employed transfer learning to mitigate this and investigated various pre-trained Convolutional Neural Networks (CNNs).

Our study uniquely explored the impact of layer-wise fine-tuning strategies for AD classification, comparing the outcomes of ten pre-trained CNN models. Although this approach improved feature extraction, it introduced another challenge: overfitting due to the imbalanced nature of the dataset. Despite efforts such as data augmentation, overfitting persisted.

We proposed a fourth methodology that combines their strengths to bridge the gap between traditional and deep learning methods. Machine learning algorithms excel with smaller datasets but require extensive feature engineering, while deep learning automates feature extraction but relies on large datasets. We created a hybrid approach by leveraging deep learning for automatic feature extraction and machine learning for classification. Significantly, we extended beyond feature extraction to optimize the machine learning algorithms. Using Bayesian optimization, we fine-tuned the hyperparameters of the machine learning models, adapting them to the variations in extracted features. This innovative methodology effectively combined the benefits of both approaches, aiming to enhance classification performance and address the limitations of each individually.

In summary, this dissertation addresses several challenges and introduces various methodologies for classifying the stages of Alzheimer's disease (AD). The last proposed methodology successfully overcomes all the challenges simultaneously. However, studying AD requires multiple sources of information. Although MRI is considered a strong source of diagnostic information, it is essential to explore additional avenues of research when focusing on the early stages of the disease.

Therefore, there are several interesting lines of research to consider for future work:

First, incorporating other modalities and fusing multiple sources of information will undoubtedly enhance accuracy and minimize detection errors.

Second, an intriguing line of research involves disease prognosis, explicitly predicting disease progression over time. This represents a promising area of investigation.

While both research avenues mentioned for future work are valuable, the primary focus should be on employing alternative data augmentation solutions. Traditional data augmentation techniques like rotation and shear are unsuitable for medical data, as many systems require exact images. Training a model on such altered data may result in erroneous decisions. Thus, a primary future objective is to address this issue, with one potential solution being the use of generative adversarial networks (GANs) to generate MRI images.

Additionally, although using the Fast Fourier Transform (FFT) in this work is promising, it has not been extensively explored. Therefore, in our future work, we intend to complete that system by concatenating all directions and integrating deep learning to create a comprehensive system.

BIBLIOGRAPHY

[1] Lucinda hampton. Brain part, 2020. URL https://www.physio-pedia.com/index.php? Last accessed 03 February 2025.

[2] Cleveland clinic. Anatomy of the cerebral cortex, 2022. URL https://www.physio-pedia.com/index.php? Last accessed 03 February 2025.

[3] Sara Adars. Definition and mechanics of structural brain plasticity, 2022. URL https://www.qualialife.com/what-is-neuroplasticity-mechanisms-of-functional-and-structural-brain-plasticity. Last accessed 03 February 2025.

[4] Katy McLaughlin. The three types of neurons are motor neurons, interneurons, and sensory neurons, 2020. URL https://biologydictionary.net/sensory-neuron/. Last accessed 03 February 2025.

[5] Teva Bracha. Images of an amyloid-beta plaque and a neurofibrillary tangle. the first hampers communication between neurons and the second disrupts the integral structure of the neuron leading to its decomposition., 2018. URL https://medium.com/@teampenguinucsd/from-curse-to-cure-a-look-at-the-neuroprotective-effects-of-amyloid-beta-62aaf21e5a Last accessed 03 February 2025.

[6] Hanzhang Lu, Lidia M Nagae-Poetscher, Xavier Golay, Doris Lin, Martin Pomper, and Peter CM Van Zijl. Routine clinical brain mri sequences for use at 3.0 tesla. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 22(1):13–22, 2005.

[7] S Mahalakshmi, T Velmurugan, et al. Detection of brain tumor by particle swarm optimization using image segmentation. *Indian Journal of Science and Technology*, 8(22):1, 2015.

[8] Meriem Saim and Amel Feroui. A new hybrid method based on bias-correction fuzzy c means and histogram of oriented gradient for alzheimer disease detection. In *2022 First International Conference on Computer Communications and Intelligent Systems (I3CIS)*, pages 31–36. IEEE, 2022.

[9] Meriem Saim and Amel Feroui. An efficient computer system for alzheimer diseases classification using fast finite shearlet transform domain and support vector machine classifier. In *2022 IEEE 9th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)*, pages 337–342. IEEE, 2022.

[10] Matthew Stewart. The convolution operation, 2019. URL `https://medium.com/towards-data-science/simple-introduction-to-convolutional-neural-networks-cdf8d3077bac`. Last accessed 03 February 2025.

[11] Hasan Maher Ahmed and Manar Younis Kashmola. Performance improvement of convolutional neural network architectures for skin disease detection. *International Journal of Computing and Digital Systems*, pages 189–201, 2023.

[12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[13] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[14] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[16] Kazi Tanvir, Akinul Islam Jony, Mohamed Kaisarul Haq, Farzana Nazera, Mahaaganapathy Dass, and Valliappan Raju. Clinical insights through xception: A multiclass classification of ocular pathologies. *Tuijin Jishu/Journal of Propulsion Technology*, 44(04):2023, 2023.

[17] Haikel Alhichri, Asma S Alswayed, Yakoub Bazi, Nassim Ammour, and Naif A Alajlan. Classification of remote sensing images using efficientnet-b3 cnn model with attention. *IEEE access*, 9:14078–14094, 2021.

[18] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.

[19] Meriem Saim and Amel Feroui. Automated alzheimer's disease detection and diagnosis method based on bayesian optimization and cnn-based pre-trained features. *Multimedia Tools and Applications*, pages 1–41, 2024.

[20] World Health Organization et al. Global status report on the public health response to dementia. 2021.

[21] Wieslaw L Nowinski. Introduction to brain anatomy. *Biomechanics of the Brain*, pages 5–40, 2011.

[22] Birte U Forstmann, Max C Keuken, and Anneke Alkemade. An introduction to human brain anatomy. *An introduction to model-based cognitive neuroscience*, pages 71–89, 2015.

[23] Peter J Whitehouse, Konrad Maurer, and Jesse F Ballenger. *Concepts of Alzheimer disease: Biological, clinical, and cultural perspectives*. JHU Press, 2000.

[24] Barbara J Grabher. Effects of alzheimer disease on patients and their family. *Journal of nuclear medicine technology*, 46(4):335–340, 2018.

[25] Wangchen Tsering and Stefan Prokop. Neuritic plaques—gateways to understanding alzheimer's disease. *Molecular Neurobiology*, 61(5):2808–2821, 2024.

[26] Michel Goedert, David S Eisenberg, and R Anthony Crowther. Propagation of tau aggregates and neurodegeneration. *Annual review of neuroscience*, 40(1):189–210, 2017.

[27] David S Eisenberg and Michael R Sawaya. Taming tangled tau. *Nature*, 547(7662): 170–171, 2017.

[28] Grace J Lee and Julie A Suhr. Principles and practices of neuropsychological assessment. In *Comprehensive Clinical Psychology, Second Edition*, pages 167–178. Elsevier, 2022.

[29] Marshal F Folstein, Susan E Folstein, and Paul R McHugh. "mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*, 12(3):189–198, 1975.

[30] Jessica L Wood, Sandra Weintraub, Christina Coventry, Jiahui Xu, Hui Zhang, Emily Rogalski, Marek-Marsel Mesulam, and Tamar Gefen. Montreal cognitive assessment (moca) performance and domain-specific index scores in amnestic versus aphasic dementia. *Journal of the International Neuropsychological Society*, 26(9): 927–931, 2020.

[31] Wilma G Rosen, Richard C Mohs, and Kenneth L Davis. A new rating scale for alzheimer's disease. *The American journal of psychiatry*, 141(11):1356–1364, 1984.

[32] Jana Podhorna, Tillmann Krahnke, Michael Shear, John E Harrison, and Alzheimer's Disease Neuroimaging Initiative. Alzheimer's disease assessment scale–cognitive subscale variants in mild cognitive impairment and mild alzheimer's disease: change over time and the effect of enrichment strategies. *Alzheimer's research & therapy*, 8:1–13, 2016.

[33] Molly E. Zimmerman. *Battery Approach*, pages 354–355. Springer New York, New York, NY, 2011. ISBN 978-0-387-79948-3. doi: 10.1007/978-0-387-79948-3_1175. URL `https://doi.org/10.1007/978-0-387-79948-3_1175`.

[34] BP Kolitz. 11-the usefulness of a validated neuropsychological test battery in scientifically supporting and communicating forensic evaluation results. editor (s): Elbert w. russell, the scientific foundation of neuropsychological assessment. *The Scientific Foundation of Neuropsychological Assessment. Elsevier*, pages 341–357, 2012.

[35] Berit Agrell and Ove Dehlin. The clock-drawing test. *Age and ageing*, 27(3):399–404, 1998.

[36] Kenneth I Shulman. Clock-drawing: is it the ideal cognitive screening test? *International journal of geriatric psychiatry*, 15(6):548–561, 2000.

[37] Zeshu Shao, Esther Janse, Karina Visser, and Antje S Meyer. What do verbal fluency tasks measure? predictors of verbal fluency performance in older adults. *Frontiers in psychology*, 5:772, 2014.

[38] Juliana V Baldo and Arthur P Shimamura. Letter and category fluency in patients with frontal lobe lesions. *Neuropsychology*, 12(2):259, 1998.

[39] Tom N Tombaugh. Trail making test a and b: normative data stratified by age and education. *Archives of clinical neuropsychology*, 19(2):203–214, 2004.

[40] ClJ Bench, CD Frith, PM Grasby, KJ Friston, E Paulesu, RSJ Frackowiak, and Raymond J Dolan. Investigations of the functional anatomy of attention using the stroop test. *Neuropsychologia*, 31(9):907–922, 1993.

[41] Yu-Te Lin and Yi-Hsiu Lai. Stroop color-word test performance of chinese-speaking persons with alzheimer's dementia. *International Journal of Gerontology*, 18(2), 2024.

[42] Tooba Altaf, Syed Muhammad Anwar, Nadia Gul, Muhammad Nadeem Majeed, and Muhammad Majid. Multi-class alzheimer's disease classification using image and clinical features. *Biomedical Signal Processing and Control*, 43:64–74, May 2018. ISSN 1746-8094. doi: 10.1016/j.bspc.2018.02.019. URL `http://dx.doi.org/10.1016/j.bspc.2018.02.019`.

[43] R. Sampath and J. Indumathi. Earlier detection of alzheimer disease using n-fold cross validation approach. *Journal of Medical Systems*, 42(11), October 2018. ISSN 1573-689X. doi: 10.1007/s10916-018-1068-5. URL `http://dx.doi.org/10.1007/s10916-018-1068-5`.

[44] Baskar Duraisamy, Jayanthi Venkatraman Shanmugam, and Jayanthi Annamalai. Alzheimer disease detection from structural mr images using fcm based weighted probabilistic neural network. *Brain imaging and behavior*, 13:87–110, 2019.

[45] K Shankar, SK Lakshmanaprabu, Ashish Khanna, Sudeep Tanwar, Joel JPC Rodrigues, and Nihar Ranjan Roy. Alzheimer detection using group grey wolf optimization based features with convolutional classifier. *Computers & Electrical Engineering*, 77:230–243, 2019.

[46] Krishnakumar Vaithinathan, Latha Parthiban, Alzheimer's Disease Neuroimaging Initiative, et al. A novel texture extraction technique with t1 weighted mri for the classification of alzheimer's disease. *Journal of neuroscience methods*, 318:84–99, 2019.

[47] Gokce Uysal and Mahmut Ozturk. Hippocampal atrophy based alzheimer's disease diagnosis via machine learning methods. *Journal of Neuroscience Methods*, 337: 108669, 2020.

[48] Heba Soliman Zaina, Samir Brahim Belhaouari, Tanya Stanko, and Vladimir Gorovoy. An exemplar pyramid feature extraction based alzheimer disease classification method. *IEEE Access*, 10:66511–66521, 2022. ISSN 2169-3536. doi: 10.1109/access.2022.3183185. URL http://dx.doi.org/10.1109/access.2022.3183185.

[49] Sumit Salunkhe, Mrinal Bachute, Shilpa Gite, Nishad Vyas, Saanil Khanna, Keta Modi, Chinmay Katpatal, and Ketan Kotecha. Classification of alzheimer's disease patients using texture analysis and machine learning. *Applied System Innovation*, 4 (3):49, 2021.

[50] Ben Nicholas, Akhil Jayakumar, Basil Titus, and T. Remya Nair. *Comparative Study of Multiple Feature Descriptors for Detecting the Presence of Alzheimer's Disease*, page 331–339. Springer Singapore, October 2021. ISBN 9789811636752. doi: 10.1007/978-981-16-3675-2_25. URL http://dx.doi.org/10.1007/978-981-16-3675-2_25.

[51] Amrutha Veluppal et al. Differentiation of alzheimer conditions in brain mr images using bidimensional multiscale entropy-based texture analysis of lateral ventricles. *Biomedical Signal Processing and Control*, 78:103974, 2022.

[52] Aimei Dong, Guodong Zhang, Jian Liu, and Zhonghe Wei. Latent feature representation learning for alzheimer's disease classification. *Computers in Biology and Medicine*, 150:106116, 2022.

[53] Karissa Chan, Corinne Fischer, Pejman Jabehdar Maralani, Sandra E Black, Alan R Moody, and April Khademi. Alzheimer's and vascular disease classification using regional texture biomarkers in flair mri. *NeuroImage: Clinical*, 38:103385, 2023.

[54] Lucas José Cruz de Mendonça and Ricardo José Ferrari. Alzheimer's disease classification based on graph kernel svms constructed with 3d texture features extracted from mr images. *Expert Systems with Applications*, 211:118633, January 2023. ISSN 0957-4174. doi: 10.1016/j.eswa.2022.118633. URL http://dx.doi.org/10.1016/j.eswa.2022.118633.

[55] Debesh Jha, Saruar Alam, Jae-Young Pyun, Kun Ho Lee, and Goo-Rak Kwon. Alzheimer's disease detection using extreme learning machine, complex dual tree wavelet principal coefficients and linear discriminant analysis. *Journal of Medical Imaging and Health Informatics*, 8(5):881–890, June 2018. ISSN 2156-7018. doi: 10.1166/jmihi.2018.2381. URL http://dx.doi.org/10.1166/jmihi.2018.2381.

[56] Ni Gao, Li-Xin Tao, Jian Huang, Feng Zhang, Xia Li, Finbarr O'Sullivan, Si-Peng Chen, Si-Jia Tian, Gehendra Mahara, Yan-Xia Luo, Qi Gao, Xiang-Tong Liu, Wei Wang, Zhi-Gang Liang, and Xiu-Hua Guo. Contourlet-based hippocampal magnetic resonance imaging texture features for multivariant classification and prediction of alzheimer's disease. *Metabolic Brain Disease*, 33(6):1899–1909, September 2018. ISSN 1573-7365. doi: 10.1007/s11011-018-0296-1. URL http://dx.doi.org/10.1007/s11011-018-0296-1.

[57] Putri Wulandari, Dian Candra Rini Novitasari, and Ahmad Hanif Asyhar. Identification of alzheimer's disease in mri data using discrete wavelet transform and support vector machine. In *Proceedings of the International Conference on Mathematics and Islam*. SCITEPRESS - Science and Technology Publications, 2018. doi: 10.5220/0008519301980204. URL http://dx.doi.org/10.5220/0008519301980204.

[58] U. Rajendra Acharya, Steven Lawrence Fernandes, Joel En WeiKoh, Edward J. Ciaccio, Mohd Kamil Mohd Fabell, U. John Tanik, V. Rajinikanth, and Chai Hong Yeong. Automated detection of alzheimer's disease using brain mri images– a study with various feature extraction techniques. *Journal of Medical Systems*, 43(9), August 2019. ISSN 1573-689X. doi: 10.1007/s10916-019-1428-9. URL http://dx.doi.org/10.1007/s10916-019-1428-9.

[59] Jinwang Feng, Shao-Wu Zhang, and Luonan Chen. Identification of alzheimer's disease based on wavelet transformation energy feature of the structural mri image and nn classifier. *Artificial Intelligence in Medicine*, 108:101940, August 2020. ISSN 0933-3657. doi: 10.1016/j.artmed.2020.101940. URL http://dx.doi.org/10.1016/j.artmed.2020.101940.

[60] Jinwang Feng, Shao-Wu Zhang, Luonan Chen, and Jie Xia. Alzheimer's disease classification using features extracted from nonsubsampled contourlet subband-based individual networks. *Neurocomputing*, 421:260–272, January 2021. ISSN 0925-2312. doi: 10.1016/j.neucom.2020.09.012. URL http://dx.doi.org/10.1016/j.neucom.2020.09.012.

[61] Jinwang Feng, Shao-Wu Zhang, Luonan Chen, and Chunman Zuo. Detection of alzheimer's disease using features of brain region-of-interest-based individual network constructed with the smri image. *Computerized Medical Imaging and Graphics*, 98:102057, June 2022. ISSN 0895-6111. doi: 10.1016/j.compmedimag.2022.102057. URL http://dx.doi.org/10.1016/j.compmedimag.2022.102057.

[62] Sadiq Alinsaif and Jochen Lang. 3d shearlet-based descriptors combined with deep features for the classification of alzheimer's disease based on mri data. *Computers in Biology and Medicine*, 138:104879, November 2021. ISSN 0010-4825. doi: 10.1016/j.compbiomed.2021.104879. URL http://dx.doi.org/10.1016/j.compbiomed.2021.104879.

[63] Shabana Urooj, Satya P. Singh, Areej Malibari, Fadwa Alrowais, and Shaeen Kalathil. Early detection of alzheimer's disease using polar harmonic transforms and optimized wavelet neural network. *Applied Sciences*, 11(4):1574, February 2021. ISSN 2076-3417. doi: 10.3390/app11041574. URL http://dx.doi.org/10.3390/app11041574.

[64] Manu Raju, Varun P. Gopi, V S Anitha, and Abishek Sherawat. Early diagnosis of alzhiemer's disease using wavelet-pooling based deep convolutional neural network. *Sādhanā*, 48(3), August 2023. ISSN 0973-7677. doi: 10.1007/s12046-023-02219-8. URL http://dx.doi.org/10.1007/s12046-023-02219-8.

[65] Joana Silva, Bruno C. Bispo, and Pedro M. Rodrigues. Structural mri texture analysis for detecting alzheimer's disease. *Journal of Medical and Biological Engineering*, 43(3):227–238, April 2023. ISSN 2199-4757. doi: 10.1007/s40846-023-00787-y. URL http://dx.doi.org/10.1007/s40846-023-00787-y.

[66] Xiaowang Bi, Wei Liu, Huaiqin Liu, and Qun Shang. Artificial intelligence-based mri images for brain in prediction of alzheimer's disease. *Journal of Healthcare Engineering*, 2021:1–7, October 2021. ISSN 2040-2295. doi: 10.1155/2021/8198552. URL http://dx.doi.org/10.1155/2021/8198552.

[67] Ahsan Bin Tufail, Yong-Kui Ma, and Qiu-Na Zhang. Binary classification of alzheimer's disease using smri imaging modality and deep learning. *Journal of Digital Imaging*, 33(5):1073–1090, July 2020. ISSN 1618-727X. doi: 10.1007/s10278-019-00265-5. URL http://dx.doi.org/10.1007/s10278-019-00265-5.

[68] Meie Fang, Zhuxin Jin, Feiwei Qin, Yong Peng, Chao Jiang, and Zhigeng Pan. Re-transfer learning and multi-modal learning assisted early diagnosis of alzheimer's disease. *Multimedia Tools and Applications*, 81(20):29159–29175, 2022.

[69] Mosleh Hmoud Al-Adhaileh. Diagnosis and classification of alzheimer's disease by using a convolution neural network algorithm. *Soft Computing*, 26(16):7751–7762, 2022.

[70] Mian Muhammad Sadiq Fareed, Shahid Zikria, Gulnaz Ahmed, Saqib Mahmood, Muhammad Aslam, Syeda Fizzah Jillani, Ahmad Moustafa, Muhammad Asad, et al. Add-net: an effective deep learning model for early detection of alzheimer disease in mri scans. *IEEE Access*, 10:96930–96951, 2022.

[71] Taliah Tajammal, Syed Khaldoon Khurshid, Abdul Jaleel, Samyan Qayyum Wahla, and Riaz Ahmad Ziar. Deep learning-based ensembling technique to classify alzheimer's disease stages using functional mri. *Journal of Healthcare Engineering*, 2023(1):6961346, 2023.

[72] Hina Nawaz, Muazzam Maqsood, Sitara Afzal, Farhan Aadil, Irfan Mehmood, and Seungmin Rho. A deep feature-based real-time system for alzheimer disease stage detection. *Multimedia Tools and Applications*, 80(28–29):35789–35807, June 2020. ISSN 1573-7721. doi: 10.1007/s11042-020-09087-y. URL http://dx.doi.org/10.1007/s11042-020-09087-y.

[73] Faizal Hajamohideen, Noushath Shaffi, Mufti Mahmud, Karthikeyan Subramanian, Arwa Al Sariri, Viswan Vimbi, Abdelhamid Abdesselam, and Alzheimer's Disease Neuroimaging Initiative. Four-way classification of alzheimer's disease using deep siamese convolutional neural network with triplet-loss function. *Brain Informatics*, 10(1):5, 2023.

[74] Pushpendra Singh Sisodia, Gaurav Kumar Ameta, Yogesh Kumar, and Neelam Chaplot. A review of deep transfer learning approaches for class-wise prediction of alzheimer's disease using mri images. *Archives of Computational Methods in Engineering*, 30(4):2409–2429, 2023.

[75] Ayoub Assmi, Khaoula Elhabyb, Achraf Benba, and Abdelilah Jilbab. Alzheimer's disease classification: a comprehensive study. *Multimedia Tools and Applications*, pages 1–24, 2024.

[76] Akbar Asgharzadeh-Bonab, Hashem Kalbkhani, and Sina Azarfardian. An alzheimer's disease classification method using fusion of features from brain magnetic resonance image transforms and deep convolutional networks. *Healthcare Analytics*, 4:100223, 2023.

[77] Amar Shukla, Rajeev Tiwari, and Shamik Tiwari. Alz-convnets for classification of alzheimer disease using transfer learning approach. *SN Computer Science*, 4(4):404, 2023.

[78] Daniel Georges Olle Olle, Julien Zoobo Bisse, and Ghislain Abessolo Alo'o. Application and comparison of k-means and pca based segmentation models for alzheimer disease detection using mri. *Discover Artificial Intelligence*, 4(1):11, 2024.

[79] AM El-Assy, Hanan M Amer, HM Ibrahim, and MA Mohamed. A novel cnn architecture for accurate early detection and classification of alzheimer's disease using mri data. *Scientific Reports*, 14(1):3463, 2024.

[80] Xin Zhang, Le Gao, Zhimin Wang, Yong Yu, Yudong Zhang, and Jin Hong. Improved neural network with multi-task learning for alzheimer's disease classification. *Heliyon*, 10(4), 2024.

[81] Mohamed N Ahmed, Sameh M Yamany, Nevin Mohamed, Aly A Farag, and Thomas Moriarty. A modified fuzzy c-means algorithm for bias field estimation and segmentation of mri data. *IEEE transactions on medical imaging*, 21(3):193–199, 2002.

[82] James C Bezdek, Robert Ehrlich, and William Full. Fcm: The fuzzy c-means clustering algorithm. *Computers & geosciences*, 10(2-3):191–203, 1984.

[83] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. Ieee, 2005.

[84] Oscar Déniz, Gloria Bueno, Jesús Salido, and Fernando De la Torre. Face recognition using histograms of oriented gradients. *Pattern recognition letters*, 32(12):1598–1603, 2011.

[85] Laboni Akter et al. Early identification of parkinson's disease from hand-drawn images using histogram of oriented gradients and machine learning techniques. In *2020 Emerging Technology in Computing, Communication and Electronics (ETCCE)*, pages 1–6. IEEE, 2020.

[86] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20:273–297, 1995.

[87] Asmita Dixit and Aparajita Nanda. Mr brain image tumor classification via kernel svm with different preprocessing techniques. In *Advances in Computing and Data Sciences: Third International Conference, ICACDS 2019, Ghaziabad, India, April 12–13, 2019, Revised Selected Papers, Part I 3*, pages 221–230. Springer, 2019.

[88] Mayank Singh, Pawan K Gupta, Vipin Tyagi, Jan Flusser, Tuncer Ören, and Rekha Kashyap. *Advances in Computing and Data Sciences: Third International Conference, ICACDS 2019, Ghaziabad, India, April 12–13, 2019, Revised Selected Papers, Part II*, volume 1046. Springer, 2019.

[89] Zhe Wang, Changming Zhu, Zengxin Niu, Daqi Gao, and Xiang Feng. Multi-kernel classification machine with reduced complexity. *Knowledge-Based Systems*, 65:83–95, 2014.

[90] Abdulhamit Subasi and Saqib Ahmed Qureshi. Artificial intelligence-based skin cancer diagnosis. In *Applications of Artificial Intelligence in Medical Imaging*, pages 183–205. Elsevier, 2023.

[91] Gongde Guo, Hui Wang, David Bell, Yaxin Bi, and Kieran Greer. Knn model-based approach in classification. In *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings*, pages 986–996. Springer, 2003.

[92] Rohini Srivastava, Shailesh Kumar, and Basant Kumar. Classification model of machine learning for medical data analysis. In *Statistical Modeling in Machine Learning*, pages 111–132. Elsevier, 2023.

[93] RON BRINKMANN. *Basic Image Manipulation*, page 93–148. Elsevier, 2008. ISBN 9780123706386. doi: 10.1016/b978-0-12-370638-6.00004-3. URL `http://dx.doi.org/10.1016/b978-0-12-370638-6.00004-3`.

[94] Suhel Dhanani and Michael Parker. *Introduction to Digital Filtering*, page 19–28. Elsevier, 2013. ISBN 9780124157606. doi: 10.1016/b978-0-12-415760-6.00004-0. URL `http://dx.doi.org/10.1016/b978-0-12-415760-6.00004-0`.

[95] Rafid Mostafiz, Mohammad Shorif Uddin, Nur-A Alam, Md Mahmodul Hasan, and Mohammad Motiur Rahman. Mri-based brain tumor detection using the fusion of histogram oriented gradients and neural features. *Evolutionary Intelligence*, 14: 1075–1087, 2021.

[96] A.A.C. Cruz, S.V. Carneiro, S.M.A. Pontes, J.J.P. Oliveira, J.P.O. Lima, V.M. Costa, L.M.U.D. Fechine, C.S. Clemente, R.M. Freire, and P.B.A. Fechine. *Sensing Materials: Optical Sensing Based on Carbon Quantum Dots*, page 542–559. Elsevier, 2023. ISBN 9780128225493. doi: 10.1016/b978-0-12-822548-6.00025-x. URL `http://dx.doi.org/10.1016/b978-0-12-822548-6.00025-x`.

[97] James S. McKenzie, James A. Donarski, Julie C. Wilson, and Adrian J. Charlton. Analysis of complex mixtures using high-resolution nuclear magnetic resonance spectroscopy and chemometrics. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 59(4):336–359, November 2011. ISSN 0079-6565. doi: 10.1016/j.pnmrs.2011.04.003. URL `http://dx.doi.org/10.1016/j.pnmrs.2011.04.003`.

[98] Devvi Sarwinda and Alhadi Bustamam. 3d-hog features–based classification using mri images to early diagnosis of alzheimer's disease. In *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*, pages 457–462. IEEE, 2018.

[99] Eero P Simoncelli, William T Freeman, Edward H Adelson, and David J Heeger. Shiftable multiscale transforms. *IEEE transactions on Information Theory*, 38(2): 587–607, 1992.

[100] Roberto H Bamberger and Mark JT Smith. A filter bank for the directional decomposition of images: Theory and design. *IEEE transactions on signal processing*, 40 (4):882–893, 1992.

[101] J-P Antoine, Pierre Carrette, Romain Murenzi, and Bernard Piette. Image analysis with two-dimensional continuous wavelet transform. *Signal processing*, 31(3): 241–272, 1993.

[102] Emmanuel J Candès and David L Donoho. New tight frames of curvelets and optimal representations of objects with piecewise c2 singularities. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 57(2):219–266, 2004.

[103] Minh N Do and Martin Vetterli. The contourlet transform: an efficient directional multiresolution image representation. *IEEE Transactions on image processing*, 14 (12):2091–2106, 2005.

[104] Kanghui Guo, Gitta Kutyniok, and Demetrio Labate. Sparse multidimensional representations using anisotropic dilation and shear operators. *Wavelets and splines*, 14:189–201, 2006.

[105] Demetrio Labate, Wang-Q Lim, Gitta Kutyniok, and Guido Weiss. Sparse multidimensional representation using shearlets. In *Wavelets XI*, volume 5914, pages 254–262. SPIE, 2005.

[106] Sören Häuser and Gabriele Steidl. Fast finite shearlet transform. *arXiv preprint arXiv:1202.1773*, 2012.

[107] Lauge Sørensen, Mads Nielsen, Alzheimer's Disease Neuroimaging Initiative, et al. Ensemble support vector machine classification of dementia using structural mri and mini-mental state examination. *Journal of neuroscience methods*, 302:66–74, 2018.

[108] Sadiq Alinsaif, Jochen Lang, Alzheimer's Disease Neuroimaging Initiative, et al. 3d shearlet-based descriptors combined with deep features for the classification of alzheimer's disease based on mri data. *Computers in Biology and Medicine*, 138: 104879, 2021.

[109] Sakshi Indolia, Anil Kumar Goswami, S.P. Mishra, and Pooja Asopa. Conceptual understanding of convolutional neural network- a deep learning approach. *Procedia Computer Science*, 132:679–688, 2018. ISSN 1877-0509. doi: 10.1016/j.procs.2018.05.069. URL `http://dx.doi.org/10.1016/j.procs.2018.05.069`.

[110] Vladimír Kunc and Jiří Kléma. Three decades of activations: A comprehensive survey of 400 activation functions for neural networks. *arXiv preprint arXiv:2402.09092*, 2024.

[111] Lisa Torrey and Jude Shavlik. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI global, 2010.

[112] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, October 2010. ISSN 1041-4347. doi: 10.1109/tkde.2009.191. URL `http://dx.doi.org/10.1109/tkde.2009.191`.

[113] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[114] Karim Aderghal, Karim Afdel, Jenny Benois-Pineau, and Gwénaëlle Catheline. Improving alzheimer's stage categorization with convolutional neural network using transfer learning and different magnetic resonance imaging modalities. *Heliyon*, 6 (12), 2020.

[115] Wenlong Chen, Tudor Paraschivescu, and Xu Can. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 4:2951–59, 2012.

[116] Xilu Wang, Yaochu Jin, Sebastian Schmitt, and Markus Olhofer. Recent advances in bayesian optimization. *ACM Computing Surveys*, 55(13s):1–36, 2023.

[117] Xilu Wang, Yaochu Jin, Sebastian Schmitt, and Markus Olhofer. Recent advances in bayesian optimization. *ACM Computing Surveys*, 55(13s):1–36, July 2023. ISSN 1557-7341. doi: 10.1145/3582078. URL `http://dx.doi.org/10.1145/3582078`.

[118] Vu Nguyen. Bayesian optimization for accelerating hyper-parameter tuning. In *2019 IEEE second international conference on artificial intelligence and knowledge engineering (AIKE)*, pages 302–305. IEEE, 2019.

[119] Batzaya Tuvshinjargal and Heejoung Hwang. Vgg-c transform model with batch normalization to predict alzheimer's disease through mri dataset. *Electronics*, 11 (16):2601, 2022.

[120] Shagun Sharma, Kalpna Guleria, Sunita Tiwari, and Sushil Kumar. A deep learning based convolutional neural network model with vgg16 feature extractor for the detection of alzheimer disease using mri scans. *Measurement: Sensors*, 24:100506, 2022.

[121] Lina Xu, Zhijun Yao, Jing Li, Chen Lv, Huaxiang Zhang, and Bin Hu. Sparse feature learning with label information for alzheimer's disease classification based on magnetic resonance imaging. *IEEE Access*, 7:26157–26167, 2019.

[122] Taher M Ghazal and G Issa. Alzheimer disease detection empowered with transfer learning. *Computers, Materials & Continua*, 70(3):5005–5019, 2022.

[123] Bharat Richhariya, Muhammad Tanveer, Ashraf Haroon Rashid, Alzheimer's Disease Neuroimaging Initiative, et al. Diagnosis of alzheimer's disease using universum support vector machine based recursive feature elimination (usvm-rfe). *Biomedical Signal Processing and Control*, 59:101903, 2020.