Ministry of higher education and scientific research of Algeria

The high national school of statistics and applied economics

E.N.S.S.E.A

Applied statistics and econometrics Department

A thesis submitted in partial fulfillment of the requirements of

Master Degree in statistics and data science

Early Detection of Alzheimer's Disease Using Deep Learning: A Comparative Study

Presented by :

Wafa Nourelhouda Ghlissat

Supervised by:

Dr. Assri

Board Members:

Name	Scientific Degree	Role

Class 2023/2024

Ministry of higher education and scientific research of Algeria

The high national school of statistics and applied economics

E.N.S.S.E.A

Applied statistics and econometrics Department

A thesis submitted in partial fulfillment of the requirements of

Master Degree in statistics and data science

Early Detection of Alzheimer's Disease Using Deep Learning: A Comparative Study

Presented by :

Wafa Nourelhouda Ghlissat

Supervised by:

Dr. Assri

Board Members:

Name	Scientific Degree	Role

Class 2023/2024

Dedication

This thesis is dedicated to:

My father, Guettaf, who has always supported and guided me. He has shown me how to work hard, and never quit. He has always trusted me and encouraged me to reach my goals and been the best father ever existed.

My mother, Zohra, who have been my rock, my source of strength, and my comfort throughout my life. Her unconditional love, care, and guidance have shaped every part of who I am today, standing by me through both joyous and challenging times, she has shown me the true meaning of resilience and love.

To my sisters ,Fatna, who has been my steadfast companion in my academic journey since I was just three years old. her belief in me have been my guiding light.Mebarka, Oumelkheir, Khaoula, Djihade, Sana, and Khadidja, who have been not only my sisters but also my best friends. Your love, friendship, and wisdom have shaped me into the person I am today.

My brothers,Mohammed, Abdelkader, and Abdelnasser, for their constant support and protection,who have loved and cared for me with such devotion.

My best friends, Tourkia and Kheira , who have always listened and understood me, who have accepted me for who I am, giving me a safe place to talk about my dreams and problems.

To my roommates and best friends ,Meriem and Wiaam, for all the good memories we share. This journey has been amazing with you, filled with laughter, support, and unforgettable moments. Your friendship has made this experience truly special.

To My life mentor that I can not imagine this journey without you by my side.

Those special people, who have loved me, inspired me, and supported me in this journey, I dedicate this thesis with a lot of thanks and love.

Acknowledgment

To begin with, my heartfelt gratitude to Almighty Allah who gave me the chance to work on this thesis and pull it out successfully. It has been an amazing learning experience for me. Moreover, I wholeheartedly show my gratitude towards My supervisor Mr Ayoub Asri for his guidance and contribution.Without his proper counseling, I would have not been able to complete my thesis work.A special gratitude to my co-supervisor Mr Abdelouahab Hocine for sharing his knowledge and experience with me and giving me the necessary assistance and motivation to achieve my desired goal.

I would like to extend my deepest gratitude to my mentor Souissi Mohamed Amine for his unwavering support, guidance, and motivation in every step of my research.

In addition, I am also very thankful to Hamza Mekid for his patience and encouragement throughout the internship of this thesis. Last but not the least, thanks to my parents for making sure this work goes as well as it could, as well as friends and loved ones who supported me in my research.

Abstract

In this thesis, a comprehensive comparative study on CNNs for Alzheimer's Disease (AD) classification was conducted. We evaluated various CNN architectures, including InceptionV3, EfficientNet, and four custom CNNs, using the ADNI dataset for training/testing and the OASIS dataset for external testing. Detailed data preprocessing techniques were implemented to improve classification accuracy.

The performance of the models was assessed using metrics such as validation accuracy, testing accuracy, AUC, loss, F1 score, recall, and precision. EfficientNet showed the highest overall accuracy and AUC, but lower F1 score and recall, indicating potential class imbalance issues. CNN 3 stood out with high performance across most metrics with 97% testing accuracy and exceptional OASIS test accuracy as a new data with 88%.While EfficientNetB3 excelled in many metrics, CNN 3's robustness on external data makes it more suitable for clinical applications. CNN 3’s consistent accuracy and reliability on external datasets highlight its potential for clinical use, warranting further enhancement and validation.

Keywords: Alzheimer's Disease , Convolutional Neural Networks , EfficientNet, ADNI dataset, Classification Accuracy.

الخلاصة

في هذه الأطروحة، تم إجراء دراسة مقارنة شاملة حول استخدام الشبكات العصبية التلافيفية **(CNNs)** لتصنيف مرض الزهايمر .**(AD)** قمنا بتقييم مختلف هياكل الشبكات العصبية التلافيفية، بما في ذلك **InceptionV3** و **EfficientNet**وأربع شبكات عصبية تلافيفية مخصصة، باستخدام مجموعة بيانات **ADNI** للتدريب والاختبار ومجموعة بيانات **OASIS** للاختبار الخارجي. تم تنفيذ تقنيات معالجة البيانات بالتفصيل لتحسين دقة التصنيف.

تم تقييم أداء النماذج باستخدام مقاييس مثل دقة التحقق، ودقة الاختبار، ومساحة تحت المنحنى**(AUC)** ، والخسارة، ودقة **F1**، والاسترجاع، والدقة. أظهرت **EfficientNet** أعلى دقة شاملة و**AUC**، ولكن درجة **F1** والاسترجاع المنخفضة أشارت إلى وجود مشكلات محتملة في توازن الفئات. تميزت **CNN 3** بأداء عالي عبر معظم المقاييس ودقة استثنائية في اختبار .**OASIS**بينما تفوقت **EfficientNetB3** في العديد من المقاييس، يجعل أداء **CNN 3** القوي على البيانات الخارجية أكثر ملاءمة للتطبيقات السريرية. تسلط دقة وموثوقية **CNN 3** المتسقة على البيانات الخارجية الضوء على إمكاناته للاستخدام السريري، مما يستدعي المزيد من التحسين والتحقق.

الكلمات المفتاحية : مرض الزهايمر ,**(AD)** لشبكات العصبية التلافيفية**(CNN)** ,مجموعة بيانات , **ADNI** دقة التصنيف , **EfficientNet**.

Table of content:

List of Abbreviation:

AD	**Alzheimer's Disease**
ADAM	**Adaptive Moment Estimation**
ADNI	**Alzheimer's Disease Neuroimaging Initiative**
ADRDA	**The Alzheimer's Disease and Related Disorders Association**
AI	**Artificial Intelligence**
API	**Application Programming Interface**
AUC	**Area under the ROC Curve**
CAD	**computer-aided diagnostic**
CNN	**Convolutional Neural Network**
CT	**Computed Tomography**
DL	**Deep learning**
DNN	**Deep neural net**
FC	**Fully connected**
FP	**False Positive**
FPR	**False Positive rate**
FN	**False Negative**
fMRI	**Functional magnetic Resonance Imaging**
ILSVRC	**ImageNet Large Scale Visual Recognition Challenge**
KNN	**k-nearest neighbor**
MCI	**Mild Cognitive Impairment**
MCIc	**Mild Cognitive Impairment converter**
MCInc	**Mild Cognitive Impairment non converter**
MD	**Mild dementia**
ML	**Machine learning**
MOD	**Moderate Dementia**
MRI	**Magnetic Resonance Imaging**

NC	**Normal Control**
ND	**No Dementia**
NINCDS	**National Institute of Neurological Disorders and Stroke**
NN	**Neural net**
OASIS	**Open Access Series Of Imaging Studies**
PET	**Positron emission tomography**
PPV	**Positive Predictive Value**
ReLU	**Rectified Linear Unit**
RF	**Random forest**
ROC	**Receiver Operating Characteristic**
SGD	**Stochastic Gradient Descent**
SVM	**support vector machine**
TCN	**temporal convolutional network**
TN	**True negative**
TP	**True Positive**
TPR	**True Positive rate**
VMD	**Very Mild Dementia**

List of figures :

List of tables

General Introduction

Background of the study

As the population ages, Alzheimer's disease (AD), for which there is presently no cure, is becoming an epidemic-scale healthcare burden. The development of primary preventive strategies is being aided by unambiguous epidemiological evidence regarding AD risk factors and disappointing outcomes from clinical studies conducted in mild-to-moderate AD dementia.

One important strategy is that the long asymptomatic stage of AD is characterized, which makes it possible to start intervention studies and secondary prevention programs on asymptomatic at-risk individuals before there is considerable irreversible neuronal loss and dysfunction.

These days, physicians are expected to identify Alzheimer's disease even before dementia manifests. Physicians must be able to assess, diagnose, and treat patients with Alzheimer's disease with speed and accuracy. This includes understanding the illness's biology and symptoms. Additionally, it provides an opportunity for patients and their caregivers to positively alter their daily routines, perhaps enhancing their quality of life for years to come.

Early and accurate diagnosis of Alzheimer’s disease (AD) is essential for disease management and therapeutic choices that can delay disease progression.

Problem Statement:

Despite advancements in medical research, the early and accurate diagnosis of Alzheimer's disease (AD) remains a significant challenge. The long asymptomatic stage of AD provides a critical window for intervention

and secondary prevention, yet current diagnostic methods are often insufficient. Physicians need to diagnose AD with speed and precision before dementia manifests, but the complexity of AD, combined with the limitations of traditional diagnostic tools, makes this difficult.

The potential of machine learning (ML) models to improve AD diagnosis is promising, but their effectiveness varies widely depending on the model and the metrics used for evaluation. To date, **there has been no comprehensive comparison of different ML models in the context of AD diagnosis, leaving a gap in our understanding of which approaches are most effective.**

This study aims to address this gap by systematically testing various ML models across different evaluation metrics. The goal is to identify the most accurate and reliable models for early AD detection, thereby enhancing diagnostic capabilities and improving outcomes for patients at risk of Alzheimer's disease.

Significance of the study

To further highlight the development and efficacy of Alzheimer's disease detection systems, performance comparison against baseline techniques and current state-of-the-art algorithms is crucial. These comparisons help researchers identify the strengths and limitations of their proposed approaches and highlight the progress made in the field

The early detection and diagnosis of Alzheimer’s disease have significance in clinical and industrial use. For prompt interventions, early detection is an essential component of therapeutic treatment. These techniques also enhance patient outcomes and management. Studies demonstrate that

patients' condition advances when the illness is discovered early. This detection also makes it possible for the healthcare professional to plan and make early decisions for recovery or hold the spread of the disease. For prompt interventions, early detection is an essential component of therapeutic treatment. These techniques also enhance patient outcomes and management.

This research is beneficial for students studying neurology and psychology who might also make use of this research methodology by observing the patient's behavior during early detection.

Aim of the study

This systematic review summarizes the main findings and offers a thorough overview of the literature with an emphasis on CNNs in AD detection. By looking at several methods, approaches, methodologies, evaluation measures, and other information, we hope to demonstrate CNNs' potential as an AD diagnostic tool.

Limitations :

Detecting Alzheimer's Disease (AD) using artificial intelligence (AI) faces several significant limitations that impede the development of accurate and reliable detection models. These limitations include issues related to limited data, reliance on expert input, domain knowledge, coding skills, and the time required for implementing hybrid approaches.

1. Limited Data

- **Dataset Availability**: There is a scarcity of comprehensive datasets that cover a wide range of subjects and biomarkers across different stages of AD. This limitation restricts the ability to train and evaluate AI models effectively.

2. Still Relies on an Expert

- **Expert Knowledge**: The accurate identification of specific brain regions relevant to AD detection requires expertise. Limited availability of such expert knowledge hampers the development of precise AI algorithms.

3. Domain Knowledge

- **Specialized Knowledge Requirement**: Developing effective AI models for AD detection requires in-depth domain knowledge, particularly in understanding the intricate structures and variations present in medical images like MRI and PET scans.

4. Coding Skills

- **Technical Expertise**: Implementing sophisticated AI models necessitates advanced coding skills. The complexity of creating and optimizing these models poses a barrier for researchers without substantial programming expertise.

5. Time for Hybrid Approach

- **Development Time**: Employing hybrid approaches, which combine multiple techniques and methodologies, requires considerable time and effort. This extended development period can slow down the progress of creating robust AI-based AD detection systems.

6. Hardware challenges

Computational Resources: Analyzing complex medical images and running sophisticated AI algorithms demand significant computational power. Limited access to high-performance hardware can impede the processing and analysis of large datasets, affecting the efficiency and scalability of AD detection models.

Research questions

In the quest for more effective diagnostic tools for Alzheimer's disease (AD), the performance of machine learning models stands out as a crucial area of investigation. This study aims to determine which machine learning models exhibit the highest performance in the early detection of AD. By evaluating various methodologies, approaches, and metrics, this research seeks to identify the models that offer the greatest accuracy and reliability in diagnosing AD at its initial stages. This inquiry is essential for advancing our understanding of how machine learning can contribute to timely and precise AD diagnosis, thereby potentially improving patient outcomes and healthcare management strategies. In direct words:

- Which machine learning models demonstrate the highest performance in the early diagnosis of Alzheimer's disease?

Hypothesis

- This study hypothesizes that the EfficientNet-B3 model will be the most effective among all tested models in the early detection of Alzheimer's disease using MRI images.

Thesis Outline

To test the previous hypothesis we suggested the following outline to fulfill the research's requirements, and it was as following:

Chapter 1: introduction

This chapter provides an overview of Alzheimer's disease (AD), discussing its background, symptoms, risk factors, stages, and the significance of early detection. It explores traditional detection limitations and introduces the role of deep learning, specifically convolutional neural networks (CNNs) and transfer learning, in enhancing AD detection from MRI data.

Chapter 2: AI in Medical Imaging

This chapter delves into the foundational concepts of machine learning and deep learning, emphasizing their applications in computer vision and medical imaging. It provides an overview of CNNs, their architecture, and transfer learning principles, highlighting popular pretrained models used in AD detection.

Chapter 3: Methodology

This chapter details the methodology for conducting a comparative study between custom CNN architectures and pretrained models for AD detection. It outlines the experimental setup, data collection and preprocessing methods, and the implementation specifics of the models

Chapter 4: Results and Discussion

Here, the chapter presents the findings from the experimental phase, analyzing the performance of each model architecture. It includes detailed results of the CNN models and pretrained models supported by metrics such as accuracy, precision, recall, F1 score, and ROC-AUC. The discussion interprets these results in the context of AD detection efficacy and potential clinical applications.

Chapter 1

Alzheimer's disease

1.1 Alzheimer's Disease

1.1.1 Overview

Dementia is an umbrella term for brain diseases that cause a decline in a person's ability to think, remember, and general behavior that is significant enough to interfere with daily tasks (Joseph et al, 2017)

The prevalence of dementia worldwide is estimated to be over 45 million people (Martin et al, 2015) and is predicted to triple by 2050 as a consequence of increased life expectancy, establishing dementia as one of the biggest global public health challenges (Crous-Bou et al, 2017) . There are several types of dementia, with Alzheimer's disease (AD) being the most common. AD accounts for about 60-80% of all cases of dementia (Martin et al, 2015). This gives an idea about the huge number of affected individuals around the world and the case of uncertainty of many of those who have not been diagnosed.

Alzheimer's disease is a neurodegenerative brain disorder in which neurons in specific parts of the brain are irreparably lost as the disease progresses. Memory and behavioral problems result from this neural loss (Backès and Lucas, 2021). Many of the patients are unable to remember their most precious moments with their beloved ones, and most of the time the disease affects their relationship with people around them as a result of its symptoms. The process of brain degeneration may be slow or fast (Burns andS.Iliffe, 2009). Changes in the brain may occur before noticeable symptoms, such as the formation of amyloid plaques and neurofibrillary tangles (Perl, 2010). The focus of research is creating remedies that can slow down early symptom incidence; however, there are no medications that can prevent its development in the human body. (Passeri et al,2022) .As of now, there is no treatment for Alzheimer's disease.

An estimated 5.3 million Americans have AD; 5.1 million are aged ≥ 65 years, and approximately 200,000 are aged < 65 and have younger onset AD. By mid-century, the number of people living with AD in the United States is projected to grow by nearly 10 million, fueled in large part by the aging baby boom generation. Today, someone in the country develops AD every 67 seconds. By 2050, one new case of AD is expected to develop every 33 seconds, resulting in nearly 1 million new cases per year, and the

estimated prevalence is expected to range from 11 million to 16 million. (Zaven et al, 2017) According to the possible risk factors of AD, the new generations are more likely to be more affected.

By 2013,in the United States alone, official death certificates recorded 84,767 deaths from AD, making AD the sixth leading cause of death in the United States and the fifth leading cause of death in Americans aged ≥ 65 years. Between 2000 and 2013, deaths resulting from heart disease, stroke, and prostate cancer decreased by 14%, 23%, and 11%, respectively, whereas deaths from AD increased by 71%. The actual number of deaths to which AD contributes (or deaths with AD) is likely much larger than the number of deaths from AD recorded on death certificates (Zaven et al, 2017).

The initial silent and asymptomatic stage, referred to as preclinical AD, is characterized by a sequence of pathophysiological hallmarks that start to appear about 20 years before the onset of symptoms (Dubois et al, 2014). Unfortunately, none of the drugs tested to date in clinical trials to change the course of the disease have shown effective results in AD dementia (Schneider et al,2014). Research is still ongoing, yet nothing serious is looming on the horizon.

1.1.2 Symptoms

The symptoms of Alzheimer's can vary from one person to another. Memory problems are typically one of the first signs of the disease. A decline in non-memory aspects of cognition, such as finding the right word, trouble understanding visual images and spatial relationships, and impaired reasoning or judgment, may also signal the early stages of Alzheimer's. As the disease progresses, symptoms become more severe and include increased confusion and behavior changes. (NIA, 2024). AD disease can affect both memory and cognition. cognitive abilities.

Some people with memory problems have a condition called mild cognitive impairment (MCI). With MCI, people have more memory problems than normal for their age, but their symptoms do not interfere with their everyday lives. Movement difficulties and problems with the sense of smell have also been linked to MCI. Older people with MCI are at greater risk of developing Alzheimer's, but not all of them do so. Some may even revert to

normal recognition. (NIA, 2024). One of the most dangerous results of the given symptoms is stepping away from hobbies, social gatherings, or other commitments, which can turn into a reason for severe depression.

1.1.3 Risk factors

Since AD develops over a long preclinical stage that can last for several decades, the extent to which risk factors assessed in late life or shortly before the onset of clinical symptoms are a result of pathological changes rather than having a causal relationship has been discussed intensively (Crous-Bou et al, 2017). In other words, some people could be having hidden symptoms of AD, but they can not tell about it until advanced stages are developed.

In recent years, scientists have made tremendous progress in better understanding Alzheimer’s and the momentum continues to grow. Still, scientists don’t yet fully understand what causes Alzheimer’s disease in most people. The causes probably include a combination of age-related changes in the brain, along with genetic, environmental, and lifestyle factors. The importance of any one of these factors in increasing or decreasing the risk of developing Alzheimer’s may differ from person to person. (NIA, 2024) Despite current advances, the exact reason of Alzheimer's remains unclear. It is likely a mix of age-related brain changes, genetics, environment, and lifestyle, with each factor's impact varying per individual. This brings up the need for more personalized research and treatment strategies.

After attaining the age of 65, the risk of developing the disease increases twice every five years. While the exact cause of AD remains unknown, genetic factors are believed to play a significant role (Blaikie et al, 2019), this research indicates that if a parent or a sibling with AD , one is more likely to develop the disease.

Vascular disease creates a link between increased risk of dementia and risk factors during midlife. Traumatic brain injury can cause a 2.3 times greater risk of developing Alzheimer's. Environmental issues such as exposure to pollution, smoking, and infections could cause AD (Breijyeh & Karaman, 2020). Understanding the issues, brain injuries, and environmental factors elevate Alzheimer's risk emphasizes the need for proactive plans.

The fact that a third of AD cases are potentially attributable to modifiable risk factors highlights the potential of risk factor reduction for disease prevention. (Crous-Bou et al, 2017). The relationship between body weight and AD seems to be a consequence of midlife obesity, which could increase the risk of AD by 60%. (Crous-Bou et al, 2017) Reducing modifiable risk factors, like midlife obesity, could significantly lower Alzheimer's disease risk.

The US National Institutes of Health highlighted diabetes mellitus, smoking, depression, mental inactivity, physical inactivity and poor diet as being associated with increased risk of cognitive decline, AD, or both (Daviglus et al, 2010). Later on, this list was further extended to include hypertension, obesity and low educational attainment. (Barnes and Yaffe, 2011) All these more can increase the possibility of cognitive decline and Alzheimer's risk

1.1.4 Stages of AD

There are three distinct phases in the evolution of Alzheimer's disease: cognitively normal, mild cognitive impairment (MCI), and dementia. According to Petersen et al. (1999), the MCI phase is a transitional stage that occurs between normal aging and the preclinical phase of dementia. It results in cognitive impairments that have little effect on instrumental daily activities.

AD can occur in multiple stages, including mild cognitive impairment (MCI), mild AD, moderate AD, and severe AD (Caselli et al, 2013).

1.1.4.1 Normal Control (NC)

In Normal Control intelligence and memory power levels are considered normal. A healthy brain is one that can carry out mental functions like learning and judgment without any trouble, according to the U.S. Center for Disease Control and Prevention.

1.1.4.2 Mild Cognitive Impairment (MCI):

MCI is a stage of AD that is almost complete (Aljunid et al, 2014) characterized by a person's difficulty performing daily tasks, including

memory loss, difficulty reasoning, difficulty making decisions, and language difficulties. Common problems include having difficulty pronouncing words correctly, forgetting new people's details, losing and misplacing objects, and forgetting what was just read. (Karlsson et al, 1989).These cognitive changes are noticeable to the affected individual and their close ones while still allowing for the performance of daily tasks.

Not every person with MCI will advance to AD. About 15-20% of people 65 years of age or older have MCI, and 30–40% of those with MCI will develop AD within five years. (Aljunid et al, 2014). It takes about 18 months to convert from MCI to AD. After that, MCI patients are categorized as either non-converters (MCInc) or converters (MCIc).

1.1.4.3 Mild Alzheimer’s disease :

People with Alzheimer's disease have more memory loss and other cognitive problems as the disease progresses. Issues may include straying and getting lost, having problems managing finances and making payments, asking the same questions repeatedly, taking longer than usual to finish everyday tasks, and experiencing behavioral and personality changes. At this point, people are frequently diagnosed.

1.1.4.4 Moderate Alzheimer’s disease :

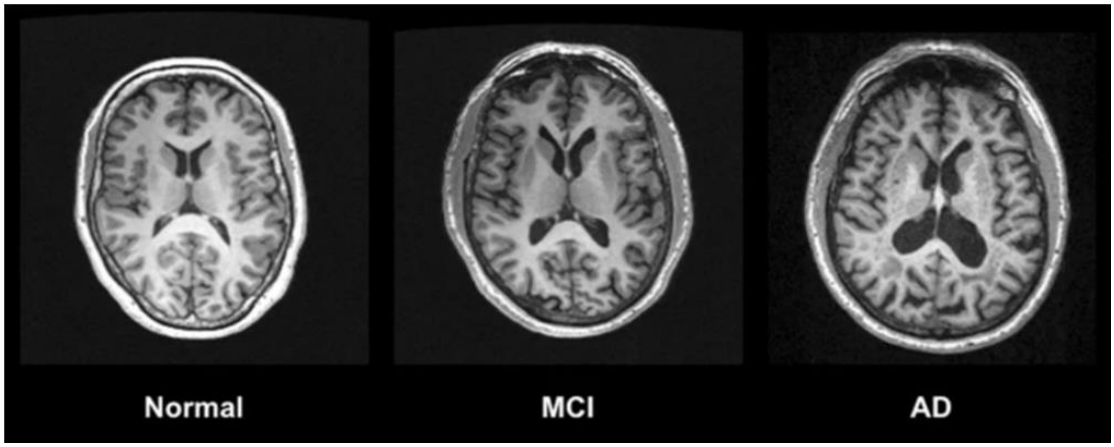
This stage is characterized by damage to brain regions responsible for language, reasoning, conscious thought, and sensory processing, including accurate sound and smell detection. As confusion and memory loss worsen, people start to have trouble identifying friends and family. They might not be able to pick up new skills, complete complicated tasks like putting on clothes, or deal with unfamiliar circumstances. At this point, a person may also experience delusions, paranoia, hallucinations, and impulsive behavior. (NIA, 2024)

1.1.4.5 Severe Alzheimer’s disease:

Severe AD represents the stage's ultimate level. In the end, the brain becomes completely covered in plaques and tangles, and the brain tissue gets much smaller. Individuals suffering from severe Alzheimer's disease are

utterly reliant on others for care and are unable to communicate. (NIA, 2024).

Figure N° 01: MRI imaging sequence:



(Avinash et al, 2019)

The figure shows decreased gray matter volume in an AD patient (right) compared to a healthy control (left) and intermediate gray matter decline in a patient with MCI (middle).

1.2 AD Detection:

1.2.1 AD Detection Techniques

The diagnosis of AD cannot be made with a single test; rather, a number of methods must be used in conjunction with one another due to the significant individual variation in the disease's pathological mechanisms and clinical manifestations. Unfortunately, its clinical applicability is limited because an autopsy is the only method that can confirm a definitive detection.
(Alsubaie et al, 2024)

Some of these methods include looking at the patient's medical history, family history, and family members' perspectives on the patient's behavioral and cognitive changes (Zaven et al, 2017).

The Alzheimer's Disease and Related Disorders Association (ADRDA) and the National Institute of Neurological Disorders and Stroke (NINCDS) developed clinical diagnostic criteria for AD, which include memory

impairment in addition to extra features like abnormal cerebrospinal fluid biomarkers or abnormal neuroimaging (MRI and PET) Dubois, (Duboiset al, 2007)

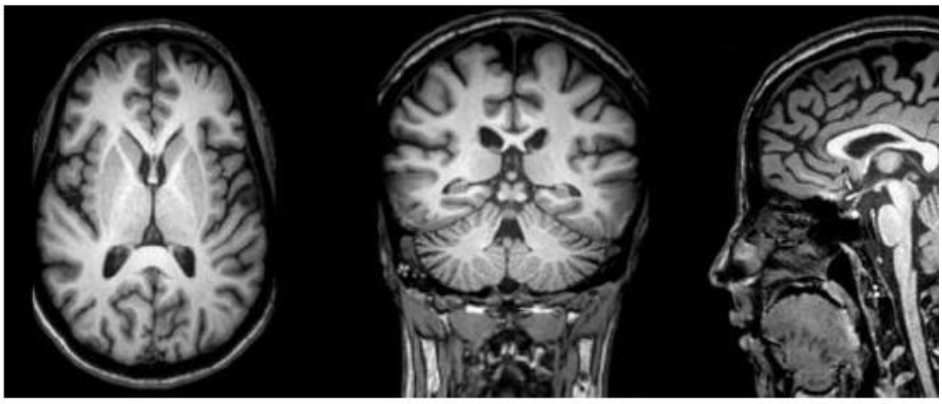
In order to gather vital information about the structural and functional alterations linked to AD, brain scans are a vital component of the AD detection system. Such as positron emission tomography (PET) and magnetic resonance imaging (MRI) (Shenton et al, 2012). Additionally, cortical width shrinkage in specific brain regions, such as the orbitofrontal, parietal, and activist areas is identified using medical imaging technology (Jason et al, 2005). These scans are also capable of identifying neurofibrillary tangles and amyloid plaques. It is crucial to remember that deciphering brain scans calls for experience and familiarity with neuroimaging. In order to guarantee accurate and trustworthy scan interpretation, radiologists and neurologists frequently work together as mentioned in the previous reference.

1.2.2 MRI as a medical data

Magnetic resonance imaging (MRI), a diagnostic method that uses non-invasive imaging technology to identify a variety of diseases and structural abnormalities, is one of the imaging techniques used to acquire detailed images of the brain. MRI uses magnetic fields and radio waves to generate resolution brain scan images, revealing anatomical features of the brain (Matthews et el, 2004). MRI is an essential part of the clinical assessment of patients with suspected Alzheimer dementia (Karikari et al, 2022), as they provide excellent anatomical detail and high diagnostic accuracy, and crucial information about structural and functional changes associated with AD (Waldemar et al, 2007). Many brain-related diseases, including Alzheimer's disease, stroke, multiple sclerosis, and brain tumors, can be detected and evaluated early with its help (AbdulAzeem et al, 2021).

Without the harmful ionizing radiation present in other imaging modalities like computed tomography and x-rays, magnetic resonance imaging (MRI) creates finely detailed three-dimensional anatomical images. The brain, muscles, fat, and bodily fluids all exhibit excellent soft tissue contrast in MRI images (Radue et al, 2016). One advantage is that it can be used to visualize anatomy in three dimensions (see Figure 2), such as along the coronal (from front to back), sagittal (from side to side), and axial (from top to bottom) planes.

Figure N° 02: MRI head scan Axial (left) Coronal (middle), and Sagittal (right)



(Sriramakrishnan et al, 2020)

1.2.3 Importance of Early Detection

Despite the fact that there are currently no disease-modifying medications on the market that can reverse the illness' initial clinical symptoms (Sreelakshmi et al, 2022), early detection of AD is crucial for a number of reasons. It can be concluded that early diagnosis and treatment can help to prevent or slow the progression of a disease and its onset of symptoms.

1.2.4 Limitations of Traditional detection

The most typical symptoms of AD and other forms of dementia have been previously discussed; additional symptoms include depression, disorientation, behavioral issues, mood and personality changes, and behavioral abnormalities. Since these symptoms typically manifest at advanced stages of AD development, traditional diagnosis presents significant challenges for early detection. Moreover, some of these symptoms are also closely associated with normal aging.

Healthcare workers often have limited time, and clinicians need more time because they have a large number of patients to see. Physicians may not always have the time they would like to spend with each patient due to the

demanding nature of medicine, which could have an impact on the accuracy of diagnoses.

The symptoms of other disorders may coexist with those of Alzheimer's disease, a neurological disorder that is quite complex. A thorough assessment of a patient's cognitive abilities is necessary for an accurate diagnosis of AD, and this process can be complex and time-consuming. The difficulty in diagnosing is increased by the variation in symptom presentation and progression.

This made the early diagnosis and treatment of Alzheimer's difficult from a medical perspective. As the prevalence of the disease increases, the current approach to diagnosing and treating patients with Alzheimer's disease will need to be modified in order to integrate care across clinical specialties and the disease cycle.

Beginning with medical care for patients. For those who suffer from Alzheimer's dementia, their loved ones, caregivers, the social and medical systems, and society at large, it is an incredibly challenging condition. Because people are living longer globally, the severity of Alzheimer's disease is increasing, and there is an urgent need to find ways to stop or postpone the onset of the illness and the ensuing dementia.

After delving into the intricacies of diagnosing Alzheimer's disease, we now focus on a crucial query: what if we could use contemporary technology to identify this illness in its early stages? More specifically, can state-of-the-art AI methods adequately analyze the fine details captured by MRI scans? We create exciting opportunities for early diagnosis by merging the rich imaging data from MRIs with deep learning's potent pattern recognition capabilities. This strategy invites us to explore the possibilities of deep learning as a game-changing remedy and has the potential to completely change how we identify and treat Alzheimer's.

1.3 Deep Learning in Medical and brain Imaging

In healthcare, diagnostic predictions include binary classification, multi-class classification, and regression-based predictions. (Alsubaie et al, 2024) Image classification can shed light on brain anatomy and function, indicating underlying pathology and illness progression. (Hcini et al,2024) Through the analysis of large amounts of medical images, it assists in early detection and

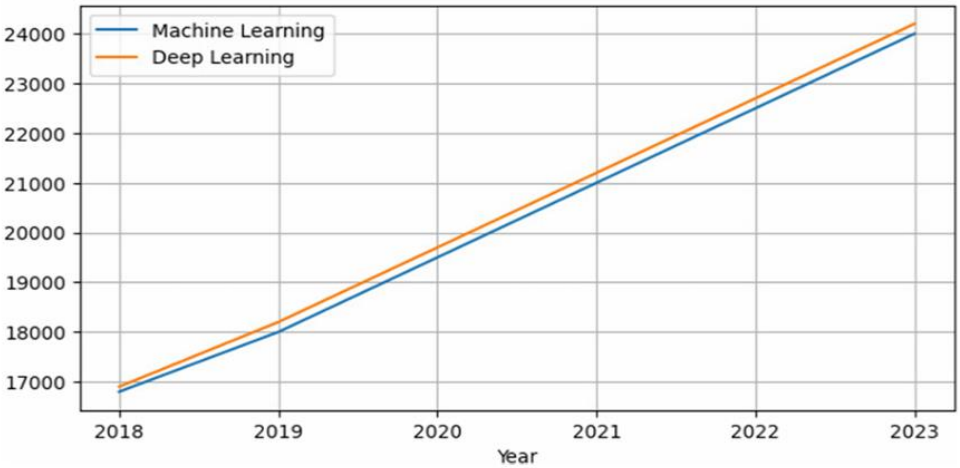
diagnosis, improves patient outcomes, and allows for improved therapies and disease management.

Research was done to evaluate the discriminative accuracy of deep learning algorithms to pathologists' diagnosis in identifying lymph node metastases in tissue sections of patients with breast cancer. In a challenge competition, certain deep learning algorithms outperformed a panel of 11 pathologists, and their performance was equivalent to that of an experienced pathologist evaluating slides with no time limits. (Bejnordi et al,2017) Whether this method has clinical relevance will be evaluated in a clinical environment.

In the realm of medical imaging, convolutional neural networks (CNNs) have achieved amazing results in tasks such as organ segmentation (Mohiuddin et al, 2023) and illness identification (Ganesh et al, 2023). These models may discover hidden representations, build links between diverse components of an image, and find disease-related patterns using neuroimaging data (Nirthika et al, 2022). They have been effectively used to a variety of medical imaging modalities, including structural MRI, FMRI, PET, and DTI (Estudillo-Romero, et al, 2022).

Distinguishing particular aspects within brain picture patterns is critical, but difficult. Significant advances in deep learning algorithms have been made in recent years,(Wang et al, 2020) which offers considerable potential for medical image analysis.

Figure N° 03: The number of publications on AD diagnosis using ML and DL from 2018 to 2023.



(Hcini et al, 2024).

The figure shows the increasing interest and research activity in both fields, based on the results of a web search using the terms "Alzheimer's disease diagnosis", "machine learning", and "deep learning". The figure also reveals that DL papers are rising faster than ML articles, implying that DL is a more new and promising strategy for AD diagnosis.

1.3.1 Deep learning application for Alzheimer’s early detection

Deep learning models have emerged as powerful tools in Alzheimer’s disease detection, improving diagnosis, prognosis, and treatment, leveraging their ability to learn complex patterns and representations from large-scale imaging datasets. To detect presence of AD for a given patient and in which stage, As a result, scientists may detect early stages of Alzheimer's disease and respond quickly to avert the effects of later stages. (Alsubaie, 2024) These approaches have the potential to alter clinical practice and help develop personalized and targeted therapies for people at risk of or afflicted by AD.

1.3.2 Convolutional Neural Networks (CNN) for AD detection:

Convolutional neural networks (CNNs) have been widely adopted as powerful tools in Alzheimer’s disease early detection due to their effectiveness in analyzing spatial relationships within brain images. (Alsubaie, 2024) including brain MRI and PET scans (Bin Bae, 2020) CNNs adeptly extract features from these images, enabling precise, reliable, data-driven AD classification and prediction.

However, they face particular challenges in Alzheimer's disease research. One such problem is the requirement for vast and diverse datasets in order to train CNN models efficiently. AD data are frequently restricted in quantity and might have class imbalances, necessitating rigorous data augmentation tactics and techniques to solve these difficulties. (Alsubaie, 2024).

1.3.3 Transfer learning for AD detection:

Transfer learning has shown promise in Alzheimer’s disease detection by leveraging pre-trained deep learning models on large-scale natural image datasets, such as ImageNet. By fine-tuning the pre-trained models on brain image data, transfer learning allows for effective knowledge transfer and improved performance, even with limited labeled training samples (Pan, 2010).

Furthermore, by employing transfer learning with pretrained CNNs Pretrained Models On Large-scale datasets, such as ImageNet, can be fine-tuned using AD- tasks compensating for limited AD-specific data .Transfer learning enables the utilization of knowledge from diverse domains and can enhance the generalization ability of AD detection models (Hcini et al, 2024)

Chapter 2

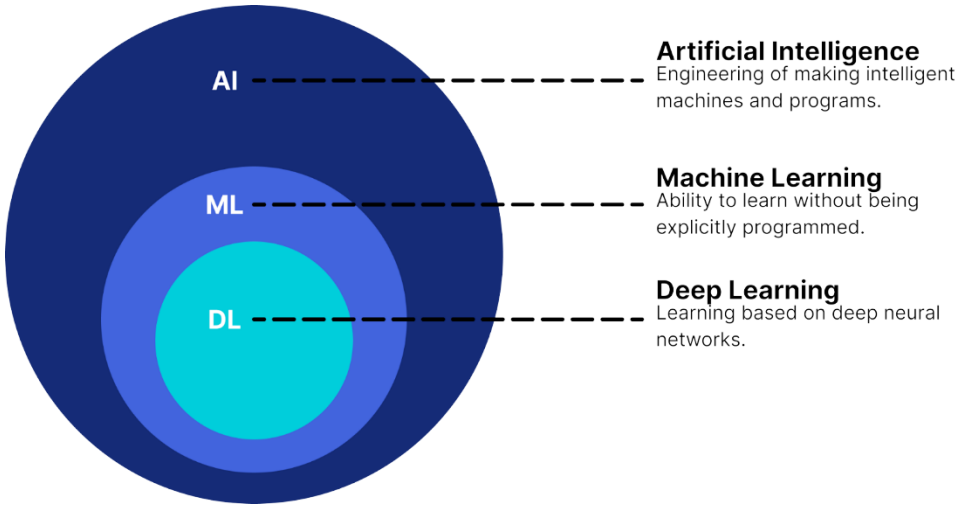
Artificial Intelligence

2.1 Artificial intelligence

2.1.1 Background

Artificial intelligence (AI) is a general term used to describe training machines to model intelligent behavior and perform human tasks with minimal human intervention (Yingge al, 2020).AI algorithms attempt to generate expert systems by making predictions or classifications based on incoming data. The combination of computer science and robust datasets allows for problem-solving. Intelligent technologies that can mimic human behavior have emerged as an effective and promising technology. The ability to reason independently and make decisions for oneself drove the exponential expansion of this novel concept.

Figure n° 04: A simple explanation of AI



(Amani, 2024)

AI is the widest category, including any technologies that simulate human intelligence. Artificial intelligence encompasses the subfields of machine learning and deep learning, which are commonly associated with AI. ML includes techniques that allow systems to learn from data. Deep learning, a further subset within ML, leverages neural networks to address more complex and deeper problems.

2.1.2 Machine learning :

Bringing up the concept of Machine Learning (ML). ML is a specific subset of AI that is a set of algorithms that can be used to complete tasks without being specifically programmed. These algorithms learn and improve from input data to make future predictions (Taeho, 2021) . Machine learning aims to improve problem-solving efficiency and effectiveness. Machine learning is today used in a variety of computer-related applications, including image classification, computer vision, object identification, language processing, speech recognition, and medicine. This ground-breaking strategy consists of looking for patterns and forming conclusions without being explicitly programmed, relying solely on past examples. Learning algorithms and complicated models have been constructed by observing past data linkages and patterns, resulting in dependable decisions and high accuracy results.

ML is often divided into supervised, unsupervised, and reinforcement learning (Taeho, 2021).In supervised learning, the algorithm learns from the training data's labels or targets. In unsupervised learning, the input is not classified or labeled. The algorithm needs to find patterns in the data. In reinforcement learning, an agent learns to make decisions through trial and error to maximize rewards.

2.1.2.1 Supervised learning

On the one hand, supervised learning requires a priori knowledge of what the result should be (Shalev & Ben-David, 2014). During the learning phase, the machine learning system must be presented with pairs of data inputs and data outputs (labeled datasets). The learning process will focus on attempting to predict the output for a specific input, which will then be compared to the actual output. As a result, the ML model will learn by computing this process for each corresponding input, then measuring its accuracy and learning over time. Supervised problems can be classified as regression or classification problems, depending on whether the output is continuous or discrete.

2.1.2.2 Unsupervised learning

On the other hand, unsupervised learning algorithms are not provided pre-assigned labels or scores for the training data.(Shalev & Ben-David, 2014).They must then discover any naturally occurring patterns in the training data set. This new family of machine learning methods includes among others:

- Clustering, in which the algorithm attempts to divide its training examples into categories with similar features.
- Principal component analysis, where the algorithm finds ways to compress the training data set by identifying which features are most useful for discriminating between different samples

There is also Semi supervised learning that combines supervised and unsupervised learning by using both labeled and unlabeled data to train AI models for classification and regression tasks (Ouali et al, 2020)

2.1.2.3 Reinforcement learning:

Reinforcement learning is a goal-oriented algorithm (Taeho, 2021). The method involves an agent with a goal, and when action is taken, the agent receives a reward or punishment based on how the action compares to the goal. The algorithm will develop a behavior that seeks to maximize the reward.

Deep learning is also a subset of machine learning. Deep learning refers to algorithms that learn from examples by mimicking the human brain using neural networks.

2.2 Deep learning

2.2.1 An overview

Deep learning is a subset of machine learning that uses multilayered neural networks to recognize complex patterns in the data. The neural networks are inspired by the structure of the human brain's biological nervous system (Zaven et al, 2022) specifically how the layers are connected with nodes, allowing information to flow through the network and be transformed and processed at each step . According to Kohonen (Zaven et al, 2022), NN's

hierarchical organization allows them to understand and interact with real-world objects.

DL has established a revolutionary approach in comparison to other ML methods. Because it can extract information directly from images via neural networks (Yinjun & Lingzhi, 2024). To address this limitation, DL techniques have been developed (Torres-Velazquez et al, 2020) by eliminating the feature extraction stage from the training process, saving time and effort for each machine learning and domain experts.

2.2.2 Deep learning potential applications:

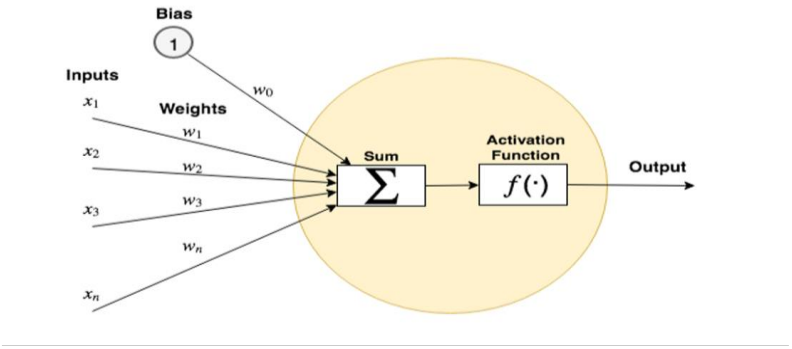
Deep learning enables computer models with numerous processing layers to learn various degrees of abstraction for data representation. These approaches have significantly advanced the state-of-the-art in object identification, voice recognition, visual object recognition, and a range of other domains, including drug development and genomics.

Computer vision is an artificial intelligence domain that enables computers to comprehend and interpret visual data, similar to human vision. It involves capturing images or videos, processing them, analyzing the data, and interpreting the results to make decisions or take actions (Klette, 2014).The potential of computer vision is vast, with applications across various industries such as autonomous vehicles, facial recognition, medical imaging, and industrial automation.

2.2.3 Neural network

Neurons, or nodes, are the basic units of a neural network. They are inspired by the neurons in the brain (Iqbal, 2021), which receive input, do computation, and generate an output.

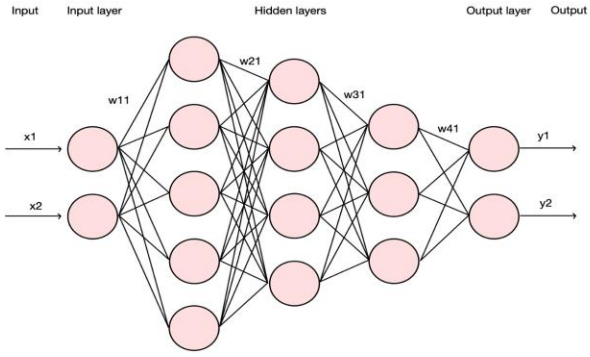
Figure n° 05: An artificial neuron.



(Nguyen & Julia , 2023)

In a neural network, the input is fed into the network's input layer. This layer contains the same number of nodes as the dataset's features. After the input layer, the data is routed to one or more hidden layers for additional processing. These are the layers that find the pattern in the data. The more hidden layers the network has, the more complex patterns can be found (Iqbal, 2021). The final hidden layer produces the neural network output. The number of layers equals the network's depth. A network with only one layer is known as a shallow network, whereas a network with multiple layers is known as a deep neural network. The neural network's prediction is based on the patterns it has learned from training data.

Figure n° 06: An illustration of a simple neural network with an input layer, hidden layers and an output layer An artificial neuron.



(Nguyen & Julia Cat-Vy, 2023)

2.2.3.1 Training of Neural Networks :

The NN training process is divided into two stages: forward and back propagation. Forward propagation is the process of passing an input through a network in the forward in image classification applications, the most common cost function is cross entropy loss. a threshold before applying the activation function to determine whether the neuron is active or inactive.
The backpropagation algorithm backward-passes to adjust the model's parameters, improve prediction, and thus minimize the loss function. It updates the weights of the NN based on its performance relative to the predicted values, which is evaluated by comparing the difference between the NN's output and actual labels using an appropriate loss function and optimizer (Ekuma, 2023).

Gradient descent (SGD) and its variants, such as Adam , are commonly used optimization algorithms (Konur et al, 2015) The selection of this loss function depends on the performed task. In the case of image classification application, cross entropy loss is the most common cost function.

2.2.3.1.1 Activation Functions :

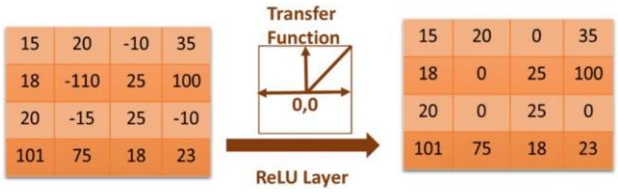
The activation function is significant because it provides non-linearity into the network and determines a neuron's output based on its inputs and weights (Kim, 2021) There are various activation functions available, such as tanh, sigmoid, etc., but the rectified linear unit (ReLU) and softmax are likely the most prominent activation functions in the hidden layers.

2.2.3.1.2 Rectified Linear Unit (ReLU) :

ReLU is utilized as an activation function in the model to increase nonlinearity and improve performance. Each convolutional block gets the ReLU activation layer. This layer significantly reduces the linearity of the extracted features. It also makes a wider range of input data predictions possible for the model. (Kim et al, 2021) . As a result, all negative feature values are set to zero.

$$f(net_j) = max(o, net_j)$$

Figure n° 07: Process of ReLU as a transfer function .



(Brownlee, 2022)

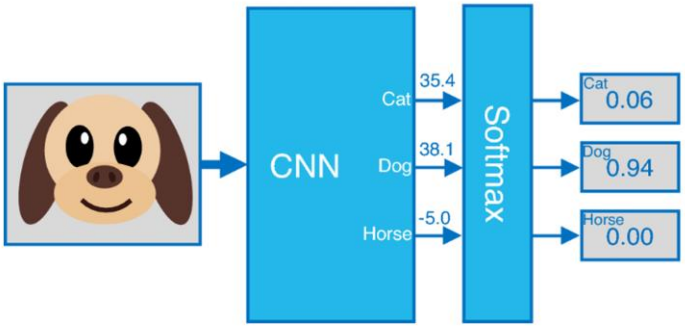
2.2.3.1.3 Softmax

For classification problems, the softmax activation function is often used in the final output layer to normalize the outputs, It gives each element in the output vector a value between 0 and 1, weighted sum values into probabilities that total is equal to 1 (Brownlee, 2022). The element values can therefore be read as probabilistic values, indicating the probability that any of the classes are true. The equation for the softmax is:

$$f(netj) = \frac{e^{net\ j}}{\sum_{k=1}^{k} e^{net\ j}}$$

Where k is the number of outputs is, $net\ j$ is the pre-activation value from neuron, and k is all the outputs from that layer.

Figure n° 08: The Softmax activation function



(Brownlee, 2022).

2.2.3.1.4 Regularization

Neural networks are prone to overfitting. Overfitting occurs when a model agrees very well with available data but fails to perform well when presented with unseen or new data. The ideal method to improve the generalization of a model is to increase the quantity of training data, but this is not always achievable, thus alternative strategies must be used(Srivastava et al , 2014) . A variety of regularization approaches, such as dropout or batch normalization, are used to prevent overfitting, increase generalization performance, and lower the risk of overfitting.

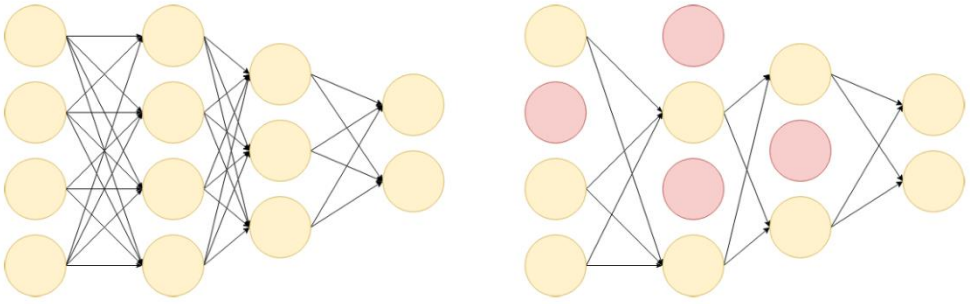
These are the techniques that we are going to use mostly in our implementation of data, The next coming parts in the next chapter will implicitly discuss the usage of these techniques.

2.2.3.1.5 Dropout

Dropout is a powerful regularization technique to reduce overfitting and improve generalization error. (Cardarilli et al 2021), .by ensuring that the model does not heavily rely on a single or small number of features (Srivastavaet al , 2014). This technique is applied by randomly removing units and their connections during training.

For every iteration (mini-batch, for example), a layer's neurons are kept with a probability of p. For the neurons that are not kept, all the connections related to the neurons are dropped.

Figure n° 09: A simple example of dropout



(Brownlee, 2022)

* The connection to the red-colored neurons is dropped.

2.2.3.1.6 Batch normalization

Batch normalization was originally developed to address the problem of Internal Covariate Shift that occurs in deep neural networks. Where ReLu is not zero-centered due to which the input distribution shifts over time. As a result, the deep neural network learns very slowly. Batch normalization addresses this issue and thus makes the model learn faster(Alberola et al, 2019)this technique normalizes the output of a previous activation layer by subtracting the batch mean and dividing by the batch standard deviation.

2.2.3.1.7 Early stopping

Early stopping is a regularization technique that requires very little modification to the cost function, the underlying training procedure, or the range of permissible parameter values. (Goodfellow et al, 2016). In an attempt to improve test results, training should be stopped when the validation error is at its lowest. This can be accomplished in practice by saving the model each time the validation error falls below the lowest validation error that has been recorded thus far. The model with the lowest validation error is tested using the test result and restored when the training is finished.

2.2.3.2 Evaluation

The performance of deep learning models is usually evaluated using metrics such as accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC). The models' capacity for generalization is evaluated using cross-validation or independent test sets (Jiang et al,2020). While sensitivity and specificity evaluate the system's capacity to correctly identify positive and negative cases, respectively, accuracy measures the system's overall correctness in making predictions (Powers, 2020). Capturing the trade-off between true positive rate and false positive rate, AUC-ROC offers a thorough assessment of the system's discriminating power (Fawcett, 2006).

2.3 CNN :

2.3.1 An overview :

Convolutional Neural Networks (CNNs) is a type of DL algorithm that are commonly used for image classification, object detection, and image segmentation (Chimakurthi, 2020) Cnn manages grid format data like 2D images and displays significant spatial dependencies within local regions of the grid (Aggarwal, 2018)

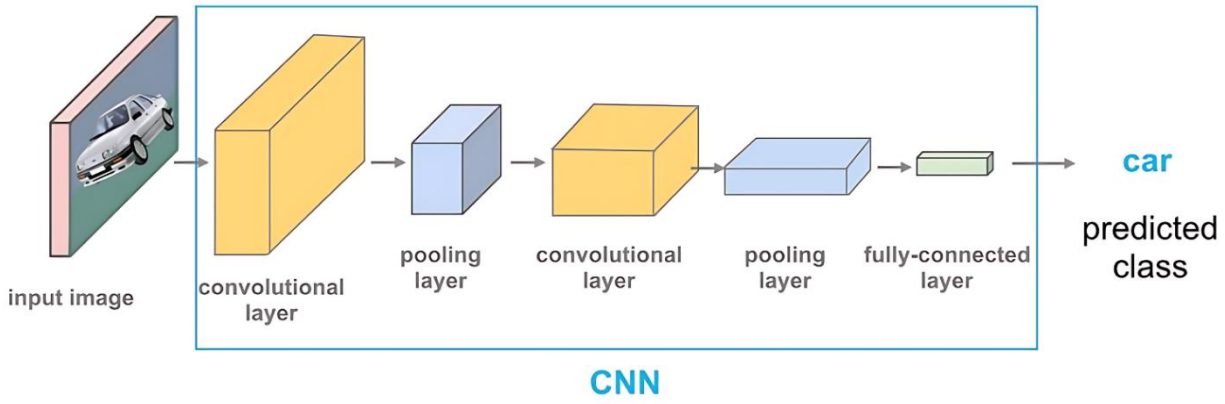
CNNs have experienced exponential growth in terms of data computation and applicability over the last decade. These days, their applications span a wide range of domains, with object detection, classification, and segmentation having the biggest effects in image analysis, speech and facial recognition, computer vision, and medical diagnosis.

2.3.2 Architecture

A Convolutional Neural Network (CNN) is a kind of neural network made up of millions of neurons arranged in multiple layers, each with learnable weights and biases. A CNN typically consists of an initial convolution layer, multiple pooling layers, fully connected layers, and a final fully-connected layer known as the classifier (Yamashita et al, 2018).

Sliding a tiny window, known as a kernel or filter (weight metrics), over an input image is the primary goal of the first convolution layer (HciniGetal,2021). This is done while the convolutional layers train the filters to identify specific elements of the input image, like corners or edges. Subsequently, the pooling layers downsample the convolutional layers' feature maps, producing smaller spatial datasets and a more computationally efficient network.(Hcini et al,2022) The convolutional layers identify high-level features that the fully connected layers use to make predictions, which are then used for classification (Figure 10)(Hcini et al, 2023). CNN can have multiple convolutional, activation, and pooling layers to learn more complex features and improve classification accuracy.

Figure n°10: Architecture of a CNN showing convolution, pooling, and FC (dense) layers.



(Hasan, 2021)

The model can be trained to extract from low-level to high-level abstractions in order to interpret the image by capturing spatial and temporal dependencies (Bernal et al, 2019). A detailed description of each layer that makes up a CNN's structure is provided below:

2.3.2.1 Input Layer

This layer only works with image data. An image has three dimensions (width, height, and depth), which is a matrix of pixel values (Aggarwal et al, 2018). Depth refers to RGB channels. Furthermore, the input layer needs to be divisible by two. For example, 32, 64, 224, 384, and 512. To deal with the input layer, the three-dimensional matrix must be transformed into a single column. For example, if a picture has a size of 24x24=576, it must first be transformed to 576x1 before being passed to the input layer. If there are n training instances, the dimension of input will be (576, n).

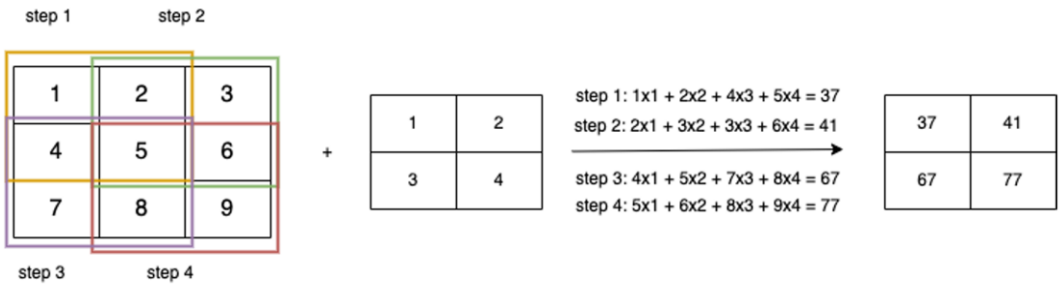
2.3.2.2 Convolutional Layer :

Convolutional layers are the fundamental building element of CNNs. This is the layer where the convolution process occurs and the Convolutional Neural Network learns (Aggarwal et al, 2018).The convolution layer handles the

majority of the computational hard work. It has certain parameters and hyperparameters.

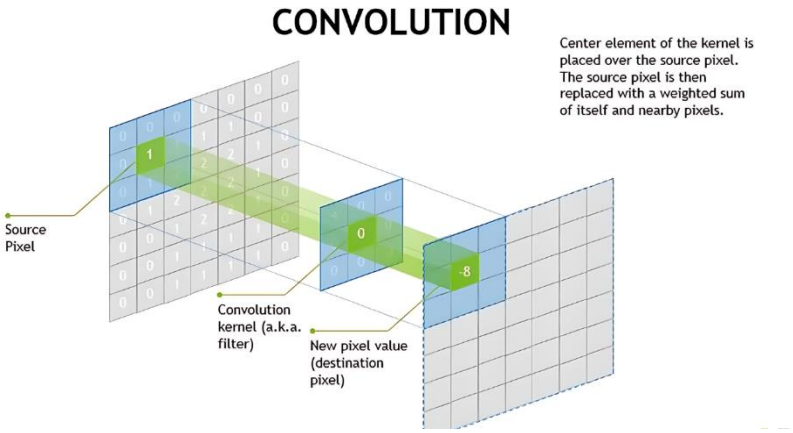
During the convolution process, the filter is placed in every feasible spot inside the picture or hidden layer to ensure that it completely covers the image, and a dot product is computed between the filter's parameters and the matching grid of data in the input volume. To do this dot product, the relevant entries inside the filter's three-dimensional region and input volume are seen as equal-sized vectors (Aggarwal et al, 2018).As shown in Figure 4, in the forward pass, each filter will move from left to right first, then up to down. The distance between each move is called stride.

Figure n°11 : Graphical depiction of convolution operation with a 2x2 filter.



(Hasan, 2021)

Figure n°12: The convolutional operation.



(Huang et al ,2024)

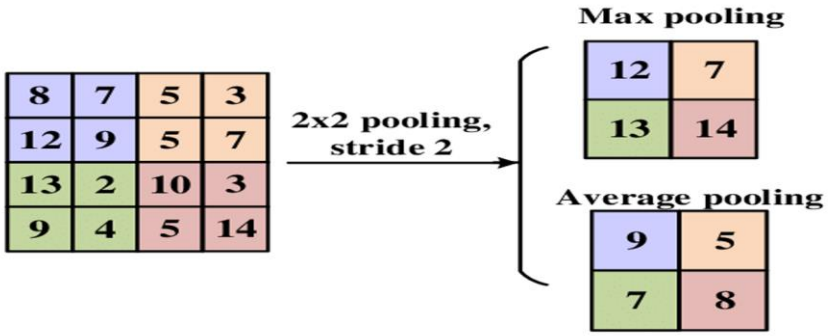
2.3.2.3 Activation Layer.

The activation layer applies the activation function to the feature maps. It is used at the end of a hidden unit or convolution to add nonlinear complexity and generate thresholded values (Aggarwal et al, 2018). The resulting values are transmitted to the next layer.

2.3.2.4 Pooling Layer:

Two convolutional layers are frequently sandwiched with a pooling layer in between. The pooling layer reduces the amount of processing and parameters when it works on small areas of a layer by down sampling the representation, which usually takes the shape of a square grid (Aggarwal et al, 2018). One of two possible pooling functions is max or average. Because max pooling performs better, it is frequently used.

Figure n°13: Max pooling & average pooling.



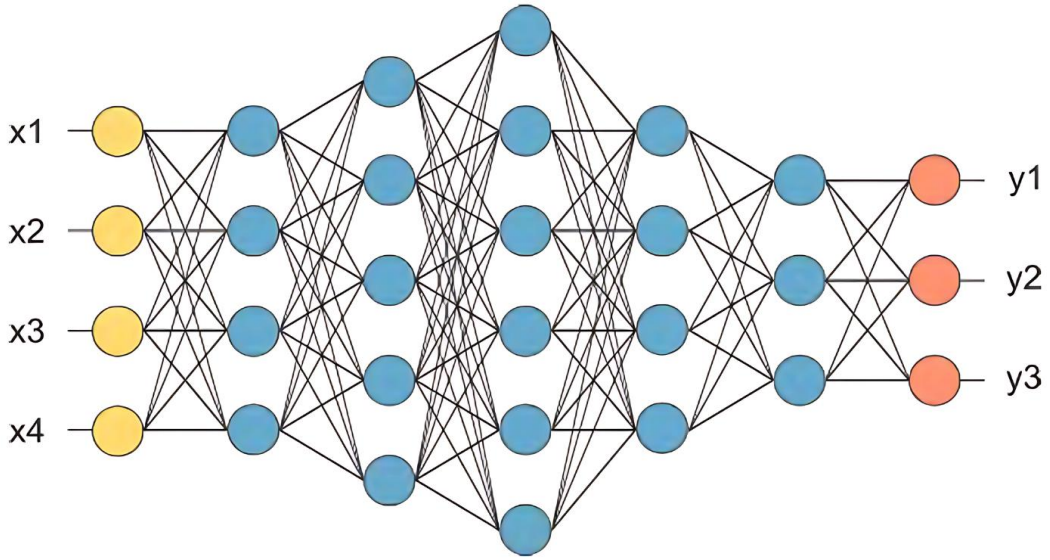
(Yingge et al, 2020).

2.3.2.5 Fully Connected Layer :

Also called dense layers, which are CNN's final layers. Each neuron in the first FC layer is densely connected to every feature map in the last spatial layer. FC layers are made to fit the specific issue at hand (Aggarwal et al, 2018). In this layer of a classification task, input from previous layers is transformed from matrix to vector. Similar to a neural network, the fully connected layer receives the volume of the preceding layer after it has reduced.

The output layer (last FC layer) is connected to every neuron in the second to last layer and has a weight assigned to it. By looking at the output of the previous layer, this layer decides which features mostly match a particular class. Using an output activation function such as Softmax, the fully connected layer provides the correct probabilities for the different classes as it computes the products between the weights and the previous layer. By using the activation function outputs are classified.

Figure n°14: Fully connected layer illustration.



(Karlsson et al, 1989).

2.4 Transfer Learning

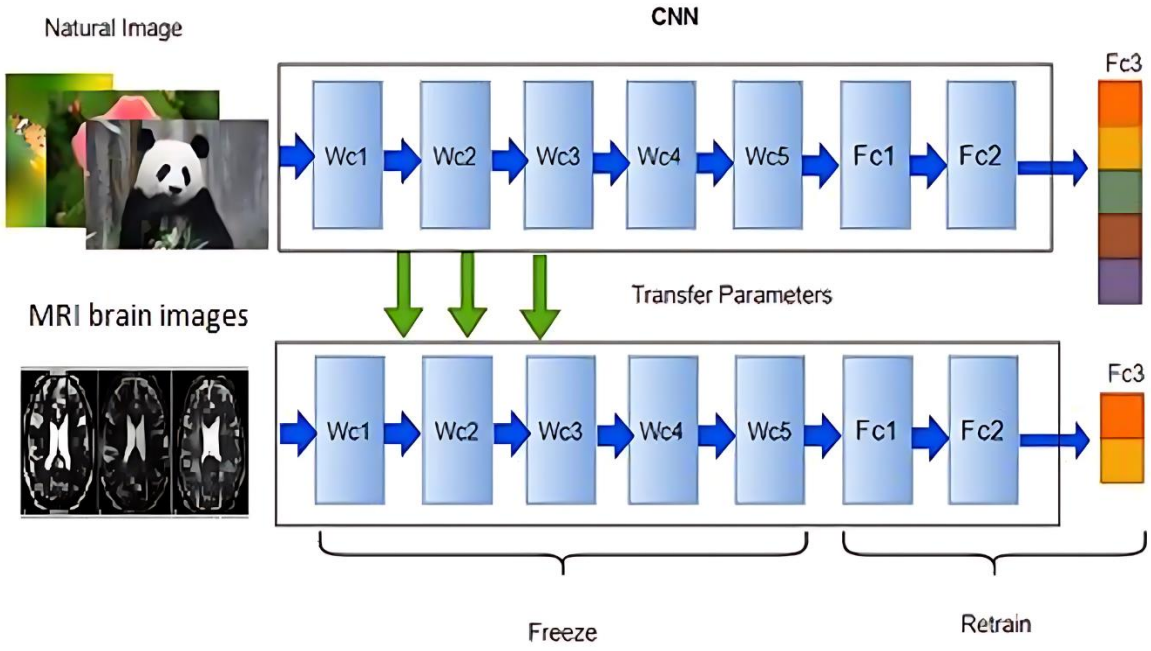
2.4.1 An overview

Transfer learning, a technique that applies information learned from one task to improve performance on another related task has shown promise in research (Pan, 2010), To enhance performance on smaller datasets with less training data, we can pre-train CNNs on massive datasets like ImageNet.

Fine-tuning is a popular transfer learning technique for deep neural networks. where a pre-trained model's parameters are adjusted through a few training cycles to make them suitable for a new task. (Guo et al, 2020) Typically, when fine-tuning, we either optimize all pre-trained parameters or freeze some layers of the pre-trained model by keeping their weights fixed

so they don't update during training. This is typically done for initial layers that capture basic features like edges or textures then we optimize the remaining layers to achieve the desired result.

Figure n° 15: The process of transfer learning technique



(Hcini et al,2024)

2.4.2 Popular Pre-Trained Models :

Several CNN models have been developed over the years that have won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al,2015), which tests techniques for object detection and picture classification on a large scale. Examples of these models are in table bellow:

Table n°: 01 Pre-trained CNN models applied to Image Net dataset.

Model	Size (MB)	Parameters	Depth	Time (ms) per step (CPU)	Time (ms) per step (GPU)
Xception	88	22.9M	81	109.4	8.1
VGG16	528	138.4M	16	69.5	4.2
VGG19	549	143.7M	19	84.8	4.4
ResNet50	98	25.6M	107	58.2	4.6
ResNet50V2	98	25.6M	103	45.6	4.4
ResNet101	171	44.7M	209	89.6	5.2
ResNet101V2	171	44.7M	205	72.7	5.4
ResNet152	232	60.4M	311	127.4	6.5
ResNet152V2	232	60.4M	307	107.5	6.6
InceptionV3	92	23.9M	189	42.2	6.9
InceptionResNetV2	215	55.9M	449	130.2	10.0
MobileNet	16	4.3M	55	22.6	3.4
MobileNetV2	14	3.5M	105	25.9	3.8
DenseNet121	33	8.1M	242	77.1	5.4
DenseNet169	57	14.3M	338	96.4	6.3
DenseNet201	80	20.2M	402	127.2	6.7
EfficientNetB0	29	5.3M	132	46.0	4.9
EfficientNetB1	31	7.9M	186	60.2	5.6
EfficientNetB2	36	9.2M	186	80.8	6.5
EfficientNetB3	48	12.3M	210	140.0	8.8
EfficientNetB4	75	19.5M	258	308.3	15.1
EfficientNetB5	118	30.6M	312	579.2	25.3
EfficientNetB6	166	43.3M	360	958.1	40.4
EfficientNetB7	256	66.7M	438	1578.9	61.6

(Hcini et al,2024).

2.5 Related Work

Deep learning has received a lot of interest for its application in identifying Alzheimer's disease. Several deep-learning algorithms have recently been presented as diagnostic tools for Alzheimer's disease, which will help clinicians make educated treatment decisions. These solutions address the limitations of many medical computer-aided diagnostic (CAD) systems.In this section, ,We present some of theses studies:

2.5.1 Alzheimer's Detection with Deep Learning Previous Studies

(Odusami et al.2021) developed a deep learning-based approach that can predict AD, late MCI, and early MCI.The 138-subject fMRI dataset from the Alzheimer's Disease Neuroimaging Initiative (ADNI) was utilized for testing.

(Ebrahimi et al. 2021) utilized an ImageNet dataset to train a CNN known as ResNet-18. The sequence-based models that were employed were the temporal convolutional network (TCN) and several forms of recurrent neural

networks. Many deep sequence based models and combinations were employed for AD detection.

(Shanmuga, 2022) employed three pre-trained networks—GoogLeNet, AlexNet, and ResNet-18—for the classification job. These networks were trained and evaluated using 6,000 photos taken from the ADNI database.

(Nancy & Priyadarshini, 2020) introduced an artificial neural network that combines feed forward and reverse propagation. The input layer computes the output values before sending them to the hidden layer. Each hidden layer collects signals from each input layer using its own set of weights and bias functions. These values are sent to the output layer, which uses them to generate the output. An activation function is used to get the output after it has been mixed with the inputs. To ascertain if an each input symbol is above or below the threshold value, a threshold value is employed.

2.5.2 State of Art

Showing existing state-of-the-art algorithms is important to demonstrate the advancement and effectiveness of Alzheimer's disease detection systems. This section presents recent proposed approaches and highlights the progress made in the field:

(AlShehri, 2022) presented a DL-based method for the identification and categorization of AD utilizing brain MRI pictures in a comparative investigation. The images were classified into four phases of AD by the study using two CNN designs, DenseNet-169 and ResNet 50: non-dementia, very mild dementia, mild dementia, and moderate dementia. According to the study, DenseNet-169 performed better than ResNet-50 in terms of testing and training accuracy, attaining 97.7% and 83.82%.Respectively, the suggested approach may be applied to the analysis and categorization of AD in real time. The work advances the area of computer-aided diagnostics by using deep learning techniques to tackle a difficult but significant issue.

Using the OASIS dataset, (Rajendiran et al. 2022) investigated the use of many deep CNNs, such as AlexNet, VGG-16 Net, ResNe-50, and Google Net models, for the classification of the early stages of AD. They contrasted their findings with those of other currently used techniques, including random

forest (RF), k-nearest neighbor (KNN), and support vector machine (SVM). They found that, with an accuracy of 97.54%, precision of 97.67%, recall of 97.54%, and F1-score of 97.55%, the Google Net model performed the best out of all the techniques.

(Sisodia et al. 2023) have developed deep transfer learning models with MRI (Magnetic Resonance Imaging) pictures to diagnose the several phases of AD, such as "Very-Mild Demented," "Mild-Demented," "Moderate-Demented," and "No-Demented." The model determines the correct type of AD by utilizing strategies for data preprocessing and augmentation. Then, additional deep transfer learning models (Resnet50, VGG19, Xception, DenseNet201, and EfficientNetB7) are used to classify and forecast the early stages of AD.

AlzheimerNet, a sophisticated CNN classifier that can differentiate between the Normal Control (NC) category and all five phases of AD, was presented by (Shamrat et al. 2023). For the purpose of training and testing the suggested model, the MRIscandataset from the ADNI database is obtained. To prepare the raw data for analysis, they employed the CLAHE image enhancement technique. Data augmentation was used to rectify the dataset's imbalance, and the end product included 60,000 picture data from all 6 classes.

(Raza et al, 2023), they particularly employed pictures segregated by gray matter (GM) in the brain to show how MRI of AD may be segmented and classified using convolutional neural network (CNN) customisation and transfer learning. They used a pre-trained DL model as their foundation model and then used transfer learning.

(Balaji et al, 2023) offered a hybrid approach to deep learning. A method for early AD detection that combines multimodal imaging and convolutional neural network with the Long Short-Term Memory algorithm combines findings from magnetic resonance imaging (MRI), positron emission tomography (PET), and traditional neuropsychological tests.

The Dual-3DM-AD model's primary contribution is its novel method of combining MRI and PET scans to diagnose AD accurately and early 2024 (Ahmad Khan et al 2024).Through preprocessing techniques like noise

reduction and skull stripping, followed by 3D image conversion, the model improves the quality of images. To reduce complexity, it uses a Mixed-transformer with Furthered U-Net architecture for semantic segmentation. In order to use information from all modalities, a multi-scale feature extraction module also extracts pertinent features from segmented pictures. These characteristics are then aggregated using a densely linked feature aggregator module. Through a softmax layer, a multi-head attention mechanism helps reduce feature dimensionality, enabling multi-class Alzheimer's diagnosis. Together, these architectural elements improve the model's efficacy and accuracy in diagnosing AD.

The development of a unique hybrid deep learning model is the main contribution of (Mahim et al, 2024) study. With the help of a Gated Recurrent Unit (GRU) and a Vision Transformer (ViT), this model improves the ability to identify and categorize Alzheimer's disease (AD) in brain MRI images. Class imbalances and computational constraints are two issues that the ViT-GRU design successfully handles in clinical contexts. The model exhibits enhanced generalizability across various populations and imaging techniques by including spatial and temporal information from multiple datasets, hence broadening its use in the diagnosis of AD.

This work undertakes a focused comparative analysis of four custom convolutional neural network (CNN) architectures alongside pretrained models EfficientNetB3 and InceptionV3 for Alzheimer's disease detection using MRI data. The study contrasts the efficacy of transfer learning with custom model building approaches in this context. Unlike previous studies (Odusami et al., 2021; Ebrahimi et al., 2021; Shanmugam et al., 2022; Nancy & Priyadarshini, 2020; AlShehri, 2022; Rajendiran et al., 2022; Sisodia et al., 2023; Shamrat et al., 2023; Raza et al., 2023; Balaji et al., 2023; Ahmad Khan Arafat et al., 2024; Mahim et al., 2024), that have explored a broader range of deep learning methodologies and pretrained networks such as ResNet, DenseNet, GoogleNet, and AlexNet, this research confines its investigation to these specific models. By concentrating on this select group, the thesis aims to provide a detailed and comparative assessment of their performance characteristics, highlighting the advantages and limitations of transfer learning versus custom model construction in optimizing diagnostic accuracy and clinical utility for Alzheimer's disease diagnosis from MRI scans.

Chapter 3

Methodology

Introduction

In the previous chapter, we have discussed the use of the DL in AD early detection by using the suggested method to identify neutrally degenerated brain MRI regions, the severity and course of the illness can be approximated. In this chapter we are describing our methodology of searching and researching to get into our aim of this study

3.1 Methodology and design:

In this research, we used a posttest-only control group design. Models are assigned to two samples, exposed to the same data, and post-tested. Posttest scores are then compared to determine the effectiveness of the outputs. The posttest-only control group design can be expanded to include more than two models. The combination of random assignment (data) and the use of the same preprocessing steps and environments serve to control for all threats to internal validity.

3.1.1 The comparative study approach :

The comparative study approach is chosen to provide a thorough evaluation of various deep learning models for early Alzheimer's detection. It makes it possible to compare model performance objectively, pointing out advantages and disadvantages and guaranteeing reliable, broadly applicable findings. The study establishes thorough performance baselines and identifies efficient architectures by incorporating both custom CNNs and pretrained models.

This methodology advances the field of medical imaging research by offering insightful information, supporting the creation of trustworthy diagnostic instruments, and guiding the distribution of healthcare resources. In the end, it guarantees the choice of the best models for clinical use, enhancing patient outcomes and early diagnosis.

3.1.2 The selection of DL models :

The project addresses the problem of Alzheimer's disease detection using supervised learning, aiming to determine if a patient has Alzheimer's based on MRI scans. Two primary approaches are chosen: transfer learning and custom convolutional neural networks (CNNs).

1. **Transfer Learning**: Utilizes pre-trained models (Inception V3 and EfficientNet) to enhance performance with limited data.

2. **Custom CNNs**: Four unique CNN architectures are specifically designed for multiclass classification of Alzheimer's disease stages.

This dual approach addresses data limitations, improves model robustness, and provides comprehensive insights into effective architectures by combining custom CNNs with pre-trained models.

3.1.3 Workflow

The experiments that are conducted in this work perform multiclass classification for AD stages. The diagram of an AD detection system is shown in **Figure 12.** In this part, every block of the system is presented in detail.

Figure n°16: Block diagram of an AD detection system

The first step of this workflow includes the input data. Starting from data collection to pre-processing to enhance the quality of the images where the data is modified to be clean and ready for the classification task.

The second step contains different classification models where A supervised deep learning approach was used, where we have a set of training image data and its corresponding label. Four custom CNN models are designed, along with two pre-trained models, InceptionV3 and EfficientNet adapted through transfer learning.

After that, the classification consists of training and testing steps. In the training step, the input data is labeled to train the algorithm. In the testing phase, the classifier is validated by giving output prediction and comparing it to the true data labels.

Finally ,The best-performing model is chosen based on the evaluation metrics, including accuracy, precision, recall, F1-score, and AUC-ROC. The model with the highest overall performance is selected as the best architecture that aids in Alzheimer's early stage detection.

3.2 Experimental Setup :

In this work, Python programming language has been chosen as a working environment,A virtual working environment is a way of creating an isolated space for projects, where we can install specific packages and dependencies without affecting the global system. Therefore, we use Colab with Python 3. Colab is a cloud-based platform that allows you to write and execute Python code in your browser, using Google's computational resources.

A few models have been developed using Python 3.6 and tested on Google Colab, the cloud-hosted environment designed to run Python programmes on a 12 GB Tesla T4 GPU. Keras 4 is the deep learning framework utilized, and TensorFlow 5 1.15 serves as the backend.

A couple more models implemented in Python 2.10 were tested on a local Jupyterlab machine with 8 cores and16 GB 3060 RTX GPU. TensorFlow 2.16 is the backend for the deep learning framework Keras 3.3.3.

3.3 Data

3.3.1 Data Collection

One of the key aspects of creating value by means of an automatic learning system is collecting the right input data to work with. In the case of this project, the information should be focused on all the data that is commonly used by the professional practitioners in the diagnosis of Alzheimer's disease.it should be noted that the requirements of the data are : the chosen dataset must be labeled, with a field representing the diagnostic result; and it must be large enough, containing at least 6.000 entries and with good quality .

Several data sets are available on Kaggle for AD classification. Dedicated organizations like ADNI and OASIS provide access to their datasets for research and educational purposes, but it is obtained from Kaggle due to the organization and the ease of technical downloading.

In this study, The MRI dataset from (ADNI) was dedicated for training , validation and initial testing. The OASIS dataset was dedicated to testing external data that the models had not seen like before.

3.3.1.1 ADNI

The Alzheimer's Disease Neuroimaging Initiative (ADNI,2024) database. Launched in 2003 by Dr. Michael W. Weiner (Weiner et al,2010), this public-private partnership seeks to track the progression of MCI and early-stage AD using multimodal data, including serial MRI, PET scans, and clinical assessments. It collects, validates and utilizes this data, The primary goals of ADNI are to embrace research to detect Alzheimer's disease (AD) early using various biomarkers and assessments, support advances in AD prevention and treatment, and provide open access to all data. It is one of the most common datasets used in machine learning projects in the field of Alzheimer's disease, being present in a lot of projects. In the same manner,

3.3.1.2 OASIS

OASIS stands for Open Access Series of Imaging Studies (OASIS, 2024). Oasis aims to offer brain neuroimaging datasets that are freely accessible for data analysis and sharing. We selected samples of this data to test out models. This is the second dataset used in this project, which includes longitudinal MRI data in older adults who are demented and those who are not.

3.3.2 Data description and samples:

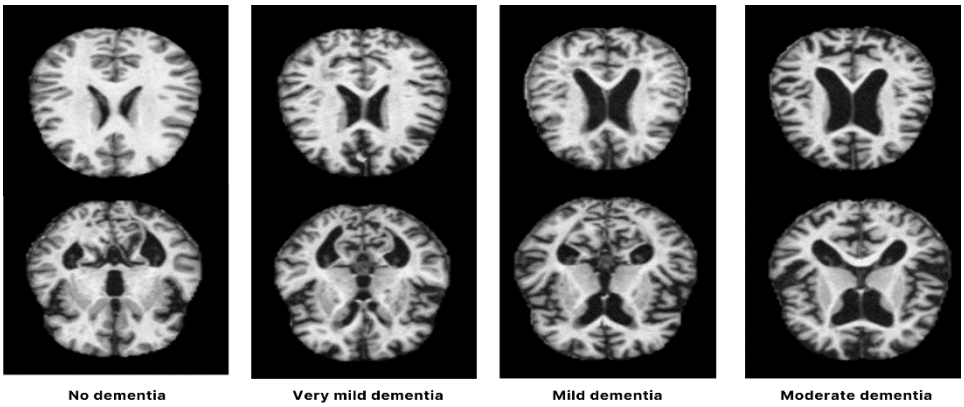
These are two independent imaging data with different sources. The following provides a brief explanation of these datasets and the diagnostic labels that were supplied:
Our ADNI dataset (4 classes of Images), which is dedicated to brain scans for dementia diagnosis, contains 6400 images. The dataset is already divided into training and testing sets, with 5121 axial images for training and 1279 for testing. The samples themselves are individual three-channel (RGB) images of 127 x 208 pixels. Table 11 provides dataset distribution with several images in the obtained dataset

Table n°2: Dataset distribution classes

Class	Train and validation	test
No Dementia (ND)	2560	640
Very Mild Dementia (VMD)	1792	448
Mild Dementia (MID)	717	179
Moderate Dementia (MOD)	52	12
%	80%	20%

Each image is labeled with the corresponding dementia level: No Dementia, Very Mild Dementia, Mild Dementia, and Moderate Dementia. While patient age and other information are not provided. However, both of these data sets have enormously large sample sizes and are formatted as three-dimensional images. The OASIS data set is 128 x 128 pixels with 18 gigabytes, while the ADNI dataset is 450 gigabytes

Figure n° 17: Classes of data



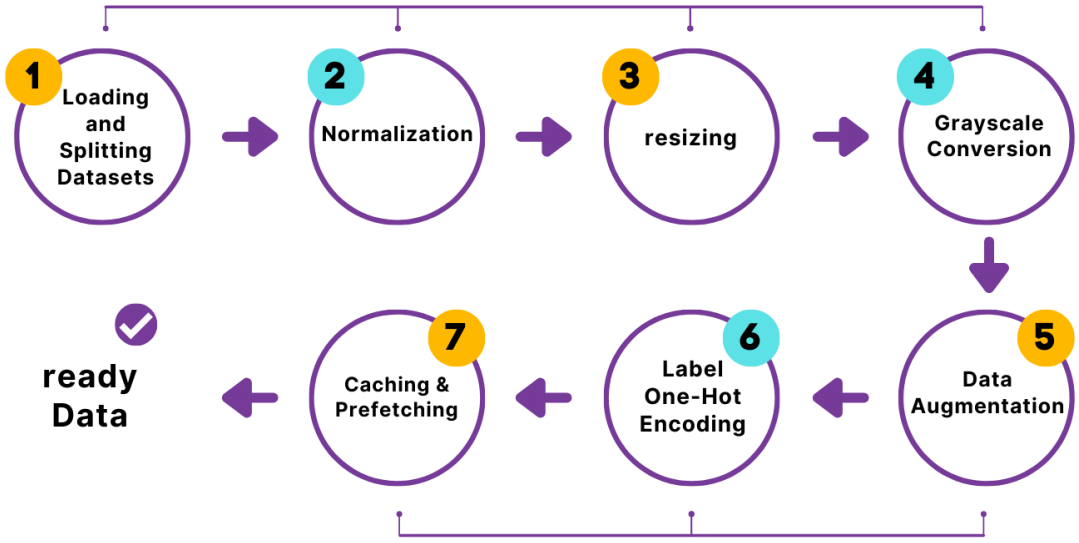
Source: From our data

Describing classes in the order of dementia level, No Dementia refers to individuals with no signs of cognitive impairment, functioning normally in daily activities. Very Mild Dementia involves subtle cognitive decline, but individuals can still manage most daily tasks independently. Mild Dementia presents noticeable cognitive impairments affecting daily activities and memory, necessitating some assistance. Moderate Dementia shows significant cognitive decline, impairing daily activities and personal care, requiring considerable help and supervision.

3.3.3 Data Pre-Processing :

Image preprocessing is a key stage of an AD detection system and a crucial step when training an artificial neural network. Depending on the data type, algorithm, and application, there are various preprocessing techniques. The steps involved in preprocessing the data are shown in the graph, which enables the model to produce more accurate and meaningful results.

Figure n°18: Data pre-processing steps



3.3.3.1 Data splitting

The dataset will be separated as one of the first tasks in this module. These sets consist of a test set, a validation set, and a training set. The dataset has already been divided into two subsets: one for training and one for testing. Now, our task is to further split the training subset into two parts: one for actual training and the other for validation. This will allow us to train our model on one portion of the data and evaluate its performance on another before applying it to the test subset. For the ADNI dataset ,within the training set, we further employed a nested split, allocating 80% for training and 20% for validation.

3.3.3.2 Normalization

Normalization is the primary step in the preprocessing procedure. By scaling the data between a maximum and a minimum, this widely used technique aims to produce zero mean and unit variance. By utilizing a higher learning rate and fewer epochs, standardizing the data enables faster training. A neural network's learning rate is a programmable hyperparameter that regulates how quickly the model adapts to the task at hand during the training phase. An epoch is a run through the complete dataset. Stated differently, the neural network converges more quickly when the input image data is distributed similarly. The normalization formula is defined as:

$$z = \frac{x_i - \mu}{\sigma}$$

where μ is the mean and σ is the standard deviation from the mean.

3.3.3.3 Resizing

Resizing the images to a consistent width and height eases computation and lowers data variability. The process of resizing involves employing reduction or interpolation algorithms to alter the image's dimensions.

The term "image resizing" in OpenCV (Python) refers to image scaling. Numerous applications in machine learning and image processing benefit from scaling. It aids in pixel reduction, which has a number of benefits for images. For example, It can reduce the time of training of a neural network as more is the number of pixels in an image more is the number of input nodes that in turn increases the complexity of the model. It is also useful for enlarging images. It is frequently necessary to resize the image, either to shirk it or to scale it up to fit the required size. For resizing an image, OpenCV offers us a number of interpolation techniques. Choosing the most pertinent brain regions for each view, reducing the amount and complexity of the data, and simplifying the process of aligning and comparing various images are just a few benefits of this step.

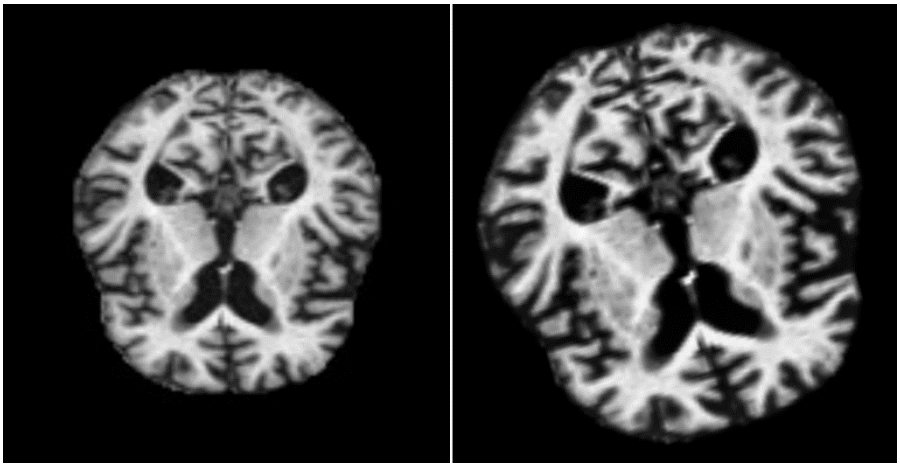
3.3.3.4 Grayscale conversion :

Grayscale conversion is a common preprocessing step in image processing and computer vision tasks. It involves converting a color image into a grayscale image, which contains only shades of gray ranging from black to white. This process is often necessary for certain algorithms that work with single-channel images or when color information is not relevant for the specific task at hand.As there are no colors for the MRI data, we will use grayscale images.this way it will simplifies the data, reduces memory usage, and preserves essential structural information for accurate analysis.

3.3.3.5 Data augmentation

Simple image manipulations like flipping, cropping, rotating, resizing, altering brightness, contrast, or color, adding noise, or erasing specific areas of the image are examples of data augmentation. By creating more, richer, and sufficient datasets for model training, this technique improves the performance of the recognition system and increases the accuracy and loss of my model on the test set.

Figure n°19: Same data sample with augmentation from our data



Source: From our data

3.3.3.6 Encoding the categorical features :

Categorical features are the ones that have labels instead of numeric values, e.g.The classes in our data have four possible values : No Dementia, Very Mild Dementia, Mild Dementia, and Moderate Dementia. Such features do not have an order i.e., one value cannot be ranked higher or lower than the other; these are called nominal categories. If a categorical feature can be ordered e.g., a feature size can be small, medium, large etc.; it is called an ordinal category.

In this thesis, we only have to pre-process nominal categories, so we restrict our discussion to such features. We use one-hot encoding to encode nominal categorical features. For example, No Dementia has the value 0, Very Mild Dementia has the value 1, Mild Dementia has the value 2, and Moderate Dementia is classed 3, Notice that one-hot encoded features have numerical values and they can now be standardized

3.3.3.7 Optimizing Data Pipeline: Caching and Prefetching

Caching and prefetching are designed to improve the performance and efficiency of the data pipeline by reducing data loading and preprocessing time, and ensuring that the model always has data ready for processing without waiting. It performs the following actions:

Cache (): This step caches the dataset in memory after it has been loaded and preprocessed for the first time. This means that subsequent iterations over the dataset will be much faster, as the data doesn't need to be reloaded and reprocessed each time.

Pre-fetch (buffer size=AUTOTUNE): This step overlaps the preprocessing and the model's execution by fetching the next batch of data while the current batch is being processed by the model. The AUTOTUNE argument automatically tunes the buffer size to balance between CPU and GPU performance.

3.4 Implementation of CNN Models for Alzheimer's Detection

Once the input data has been processed and split, it’s ready to be fed into the neural network that will be in charge of learning from this data to be able to make predictions on new input data in the future.

Using the sequential model, different CNN architectures are implemented from scratch in this project. When the structure is created, the model will be compiled**.**

A set of parameters must be specified during compilation, including the optimization algorithm to be used for training the network and the loss function to assess the network that the optimization minimizes. Selecting the right algorithm and hyperparameters for the model is a crucial step in the project. After the model has been put together**,** the network will be trained. To do so, training data must be provided as input. Validation data is also provided to analyze the evolution of the process of the training. The number of epochs (the number of times the entire set of training data will be traversed) and the batch size (the number of records to use in each iteration) must be specified as parameters.

The architectures and training parameters of the models are elucidated in detail below, moreover, to facilitate comprehension, the models are visualized utilizing the Visualkeras library.

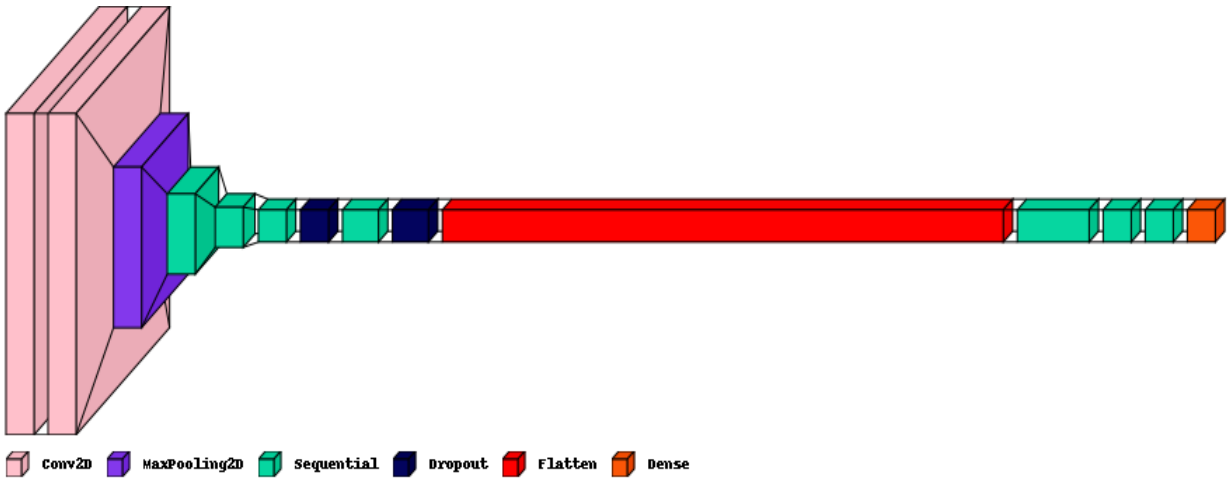
The specified neural networks architectures are designed for Alzheimer's disease classification from MRI images, employing a sequential layer arrangement as follows :

3.4.1 CNN 01 model :

- The model starts with two convolutional layers, each with 32 filters, followed by batch normalization and max-pooling operations to enhance feature extraction and reduce spatial dimensions.
- Two subsequent convolutional blocks are employed, each characterized by two separable convolutional layers with filter sizes increasing to 64 and 128, respectively, alongside max-pooling and dropout regularization to mitigate overfitting.
- Another convolutional block follows, integrating a larger filter size of 256 alongside dropout regularization.
- The resultant feature maps are flattened and forwarded into a densely connected network featuring three dense layers with diminishing units of 512, 128, and 64, respectively, supplemented by dropout regularization for enhanced generalization.
- The final dense layer, employing softmax activation, yields classification outputs across four classes.

This is the block diagram illustrating the model architecture, showcasing its components and connections:

Figure n°20: The block Diagram illustrating CNN 01 architecture.



- The model's architecture accommodates distributed training by adjusting the batch size to 16 multiplied by the number of replicas in sync, ensuring optimal hardware utilization and efficient convergence.
- The Adam optimizer with a learning rate decay strategy and categorical cross-entropy loss function is leveraged for optimization.
- Callbacks such as model checkpointing and early stopping are used to deliver robust performance.
- With approximately 4.94 million trainable parameters encompassing convolutional, dense, batch normalization, and dropout layers. This is parameters summary of this model :

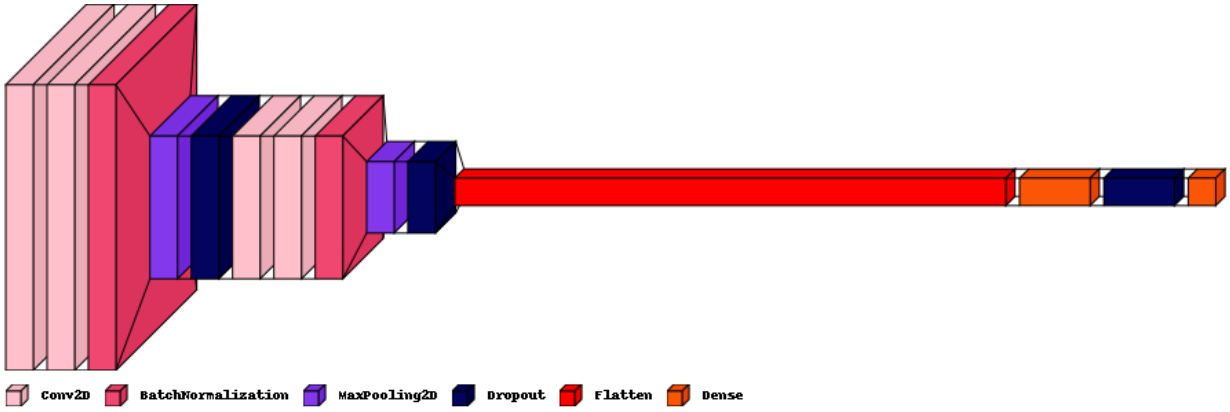
```
Model: "cnn_01"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
 conv2d_2 (Conv2D)           (None, 200, 200, 16)      160

 conv2d_3 (Conv2D)           (None, 200, 200, 16)      2320

 Max_pooling2d_5 (MaxPoolin  (None, 100, 100, 16)      0
 g2D)

 sequential_5 (Sequential)   (None, 50, 50, 32)        2160

 sequential_6 (Sequential)   (None, 25, 25, 64)        7392

 sequential_7 (Sequential)   (None, 12, 12, 128)       27072

 dropout_2 (Dropout)         (None, 12, 12, 128)       0

 sequential_8 (Sequential)   (None, 6, 6, 256)         103296

 dropout_3 (Dropout)         (None, 6, 6, 256)         0

 flatten_1 (Flatten)         (None, 9216)              0

 sequential_9 (Sequential)   (None, 512)               4721152

 sequential_10 (Sequential)  (None, 128)               66176

 sequential_11 (Sequential)  (None, 64)                8512

 dense_3 (Dense)             (None, 4)                 260

=================================================================
Total params: 4938500 (18.84 MB)
Trainable params: 4936132 (18.83 MB)
Non-trainable params: 2368 (9.25 KB)
_________________________________________________________________
```

3.4.2 CNN 02 model :

- The model begins with two convolutional layers, each comprising 32 filters and utilizing a kernel size of 2x2, followed by rectified linear unit (ReLU) activation and padding to maintain spatial dimensions.
- Batch normalization and max-pooling operations are subsequently applied to enhance feature extraction and reduce spatial dimensions.
- Dropout regularization is applied to mitigate overfitting.
- Two additional convolutional layers, each with 64 filters, are employed, maintaining the same activation, padding, and regularization schemes.
- Following flattening, the feature maps are fed into a dense layer with 512 units, activated by ReLU, and further regularized by dropout.
- The final dense layer with softmax activation produces classification outputs across four classes.

This is the block diagram illustrating the model architecture, showcasing its components and connections:

Figure n°21: The block Diagram illustrating CNN 02 architecture.



- The model is compiled with categorical cross-entropy loss and Adam optimizer, optimizing for accuracy metrics.
- The model is trained for 100 epochs with a batch size of 10.
- Callbacks for model checkpointing and early stopping are leveraged to ensure optimal performance and prevent overfitting during training.
- The architecture comprises approximately 81.95 million trainable parameters, distributed across convolutional, dense, batch

normalization, and dropout layers.This is parameters summary of this model :

```
Model: "cnn_02"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d_4 (Conv2D)           (None, 200, 200, 32)      160

 conv2d_5 (Conv2D)           (None, 200, 200, 32)      4128

 batch_normalization_11                               (None,200,200,32)
128

 max_pooling2d_10 (MaxPooli  (None, 100, 100, 32)      0
 ng2D)

 dropout_7 (Dropout)         (None, 100, 100, 32)      0

 conv2d_6 (Conv2D)           (None, 100, 100, 64)      8256

 conv2d_7 (Conv2D)           (None, 100, 100, 64)      16448

 Batch_normalization                                  (None,100,100,64)
256

 Max_pooling2d_11 (MaxPooli  (None, 50, 50, 64)        0
 ng2D)

 dropout_8 (Dropout)         (None, 50, 50, 64)        0

 flatten_2 (Flatten)         (None, 160000)            0

 dense_4 (Dense)             (None, 512)               81920512

 dropout_9 (Dropout)         (None, 512)               0

 dense_5 (Dense)             (None, 4)                 2052

=================================================================
Total params: 81951940 (312.62 MB)
Trainable params: 81951748 (312.62 MB)
Non-trainable params: 192 (768.00 Byte)
_________________________________________________________________
None
```

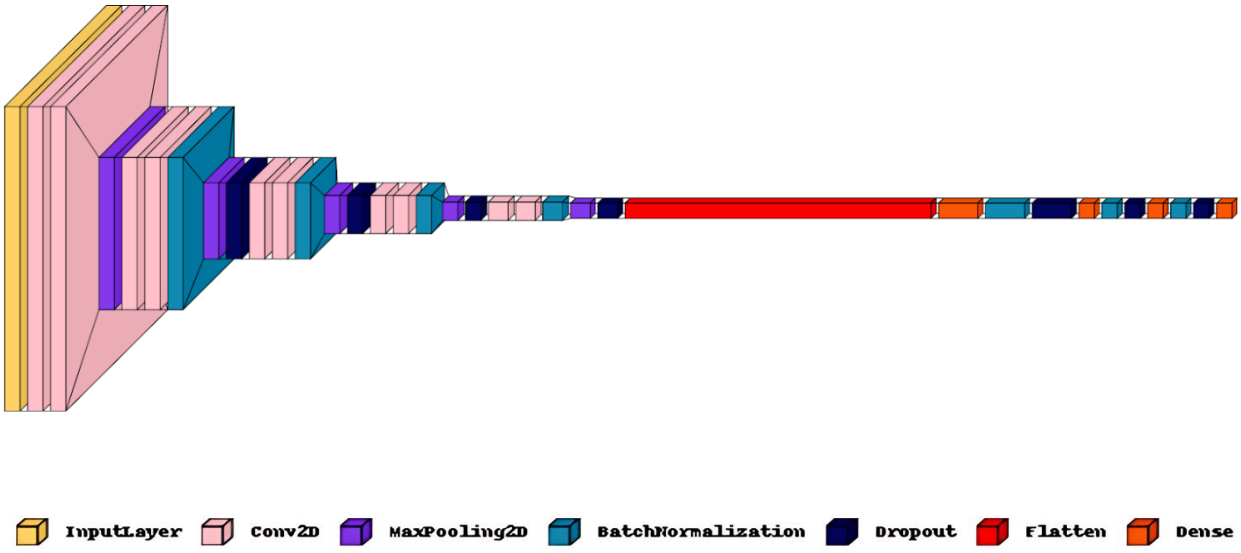
3.4.3 CNN 03 model :

- The model initiates with two convolutional layers, each with 16 filters and utilizing a kernel size of 3x3, activated by rectified linear units (ReLU) and padded to maintain spatial dimensions.
- Subsequent max-pooling operations are applied to downsample the feature maps.

- The model incorporates convolutional blocks, each comprising two convolutional layers with increasing filter sizes of 32, 64, and 128, respectively, supplemented by batch normalization and max-pooling.
- Dropout regularization is employed after the 128-filter convolutional block to prevent overfitting.
- The subsequent convolutional block with 256 filters is also followed by dropout regularization.
- Following flattening, the feature maps are passed through dense blocks, consisting of dense layers with decreasing units of 512, 128, and 64, respectively, activated by ReLU and regularized by dropout.
- The final dense layer with softmax activation produces classification outputs across four classes.

This is the block diagram illustrating the model architecture, showcasing its components and connections:

Figure n°22: The block Diagram illustrating CNN 03 architecture.



- The model is compiled with categorical cross-entropy loss and Adam optimizer, optimizing for accuracy metrics.
- Training is performed for 100 epochs, with an early stopping mechanism and a custom callback to terminate training upon achieving a validation accuracy threshold above 99%.

- The model's architecture encompasses approximately 5.97 million trainable parameters, distributed across convolutional, dense, batch normalization, and dropout layers.This is parameters summary of this model :

```
Model: "cnn_03"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 200, 200, 16)      448

 conv2d_1 (Conv2D)           (None, 200, 200, 16)      2320

 max_pooling2d(MaxPooling2D) (None, 100, 100, 16)
0

 sequential (Sequential)     (None, 50, 50, 32)        14016

 sequential_1 (Sequential)   (None, 25, 25, 64)        55680

 sequential_2 (Sequential)   (None, 12, 12, 128)       221952

 dropout (Dropout)           (None, 12, 12, 128)       0

 sequential_3 (Sequential)   (None, 6, 6, 256)         886272

 dropout_1 (Dropout)         (None, 6, 6, 256)         0

 flatten (Flatten)           (None, 9216)              0

 sequential_4 (Sequential)   (None, 512)               4721152

 sequential_5 (Sequential)   (None, 128)               66176

 sequential_6 (Sequential)   (None, 64)                8512

 dense_3 (Dense)             (None, 4)                 260

=================================================================
Total params: 5,976,788
Trainable params: 5,974,420
Non-trainable params: 2,368
_________________________________________________________________
```

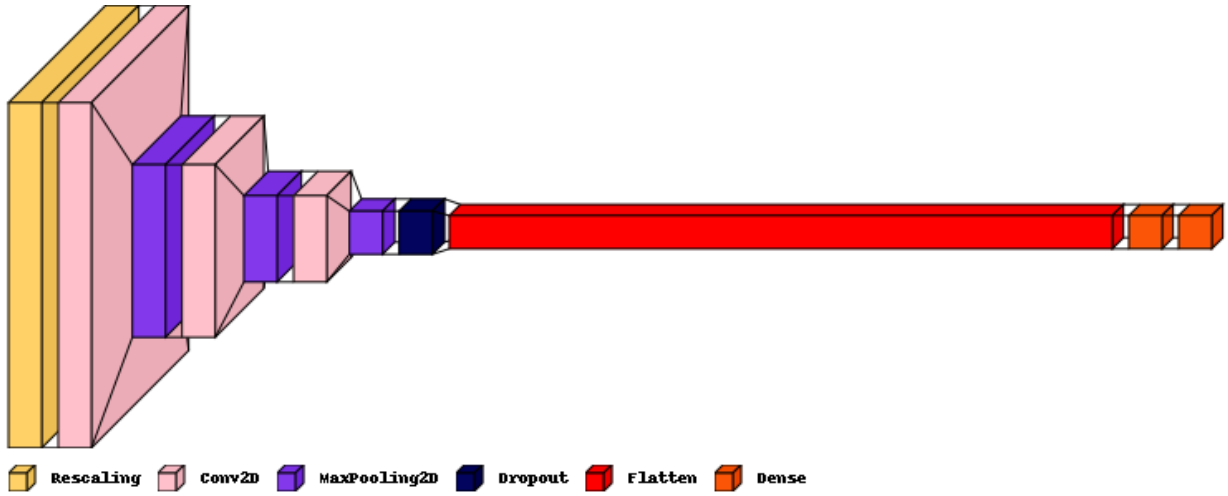
3.4.4 CNN 04 model :

- It begins with a rescaling layer to normalize pixel values between 0 and 1.
- The model then proceeds with three convolutional layers, each followed by max-pooling operations to downsample the feature maps and increase translation invariance.

- The convolutional layers consist of 16, 32, and 64 filters, respectively, with a kernel size of 3x3 and ReLU activation, aiding in feature extraction.
- A dropout layer with a dropout rate of 20% is incorporated after the third convolutional layer to mitigate overfitting.
- The resultant feature maps are flattened and passed through a dense layer with 128 units activated by ReLU.
- Finally, a dense output layer with softmax activation produces classification outputs across four classes.

This is the block diagram illustrating the model architecture, showcasing its components and connections:

Figure n°23: The block Diagram illustrating CNN 04 architecture.



- The model is compiled with sparse categorical cross-entropy loss and Adam optimizer, aiming to optimize accuracy metrics.
- Training is performed for 50 epochs, utilizing the provided training and validation datasets to facilitate model learning and evaluation.
- Through this architecture, the model endeavors to accurately classify Alzheimer's disease stages based on MRI images, contributing to early diagnosis and treatment strategies.
- The model architecture comprises approximately 5.1 million trainable parameters, primarily distributed across convolutional and dense layers.This is parameters summary of this model :

```
Model: "cnn_04"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 rescaling (Rescaling)       (None, 200, 200, 3)       0

 conv2d (Conv2D)             (None, 200, 200, 16)      448

 max_pooling2d (MaxPooling2D  (None, 100, 100, 16)     0
 )

 conv2d_1 (Conv2D)           (None, 100, 100, 32)      4640

 max_pooling2d_1 (MaxPooling  (None, 50, 50, 32)       0
 2D)

 conv2d_2 (Conv2D)           (None, 50, 50, 64)        18496

 max_pooling2d_2 (MaxPooling  (None, 25, 25, 64)       0
 2D)

 dropout (Dropout)           (None, 25, 25, 64)        0

 flatten (Flatten)           (None, 40000)             0

 dense (Dense)               (None, 128)               5120128

 dense_1 (Dense)             (None, 4)                 516

=================================================================
Total params: 5,144,228
Trainable params: 5,144,228
Non-trainable params: 0
_________________________________________________________________
```

3.5 Implementation of Pretrained Models for Transfer Learning

3.5.1 An overview of the two pretrained models used and the architecture

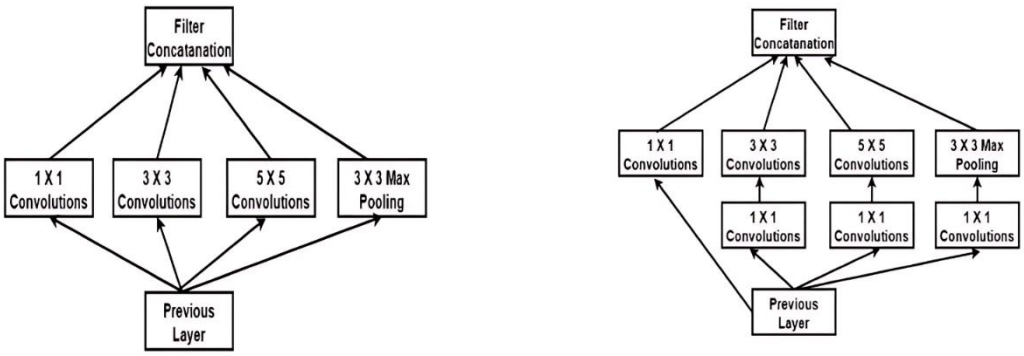
This part delves into the practical implementation of pretrained models for transfer learning in the realm of training a brain image classifier. Here, the focus is on two prominent CNN architectures: InceptionV3 and EfficientNet. Transfer learning. By harnessing pretrained models, we aim to exploit representations learned from extensive data in unrelated domains and adapt them to the specific task of classifying brain images. This part provides a concise overview of these pretrained models, detailing their architectures and underscoring their suitability for the classification task at hand, laying the groundwork for subsequent experimental evaluations.

3.5.1.1 Inception

3.5.1.1.1 Inception concept :

CNN classifiers have been significantly improved by the inception network. The architecture of the convolutional neural network is extremely complex. Following its invention, the convolutional layer of the most widely used convolutional neural networks was stacked deeper to achieve better performance. Its intricate architecture comes from the numerous methods it employs to increase precision and speed. The network has a 1x1 convolution. Additionally, global average pooling is used in place of fully connected layers (Tsang,2018). For the same input, it also includes a variety of convolution sizes and types. The visible portion of an image's size might vary greatly. Because of this large range, choosing the right kernel size is difficult. Information that is distributed locally is better served by a smaller kernel, while information that is distributed globally is better served by a larger kernel. In addition, overloading is a major problem for very deep networks, gradient updates find it difficult to spread throughout the network. In these circumstances, inception is helpful. Initially, different sized filters function at the same level (Hasan et al,2019)

Figure n° 24: Dimension reductions of Inception



(Raj, 2018).

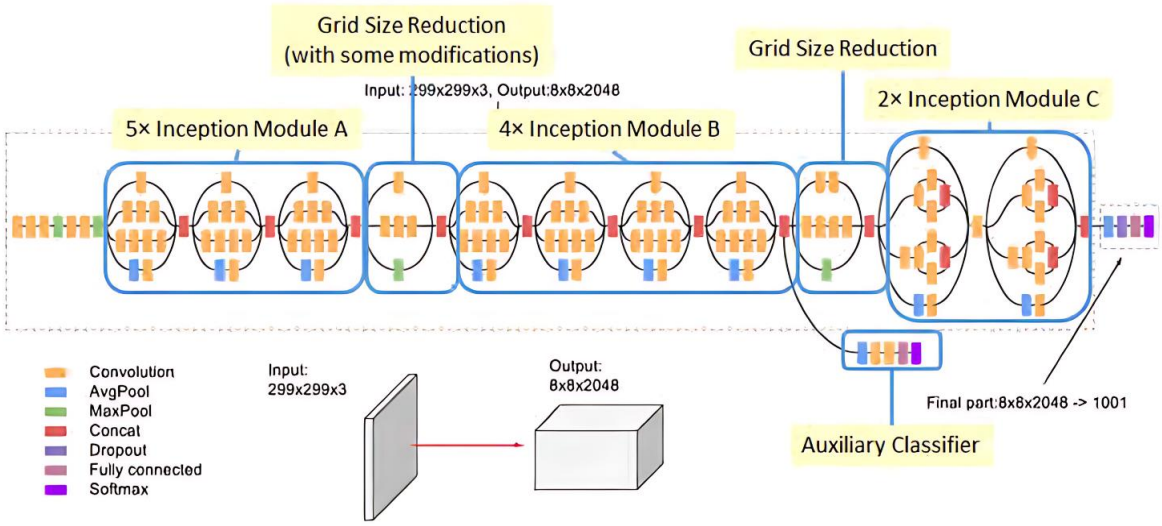
Three distinct filter sizes—1x1, 3x3, and 5x5)—are utilized to accomplish convolution on an input (Raj, 2018). It restricts the number of input channels by including an additional 1x1 convolution. By reducing the size of

the model, it also lessens the issue of over-fitting (Tsang, 2018). The 1x1 convolution is inserted before the max pooling. Concatenated outputs are sent into the subsequent module. Owing to its quick development, Inception has been implemented in multiple versions: Inception-v1, Inception-v2, Inception-v3, and Inception-v4 (Hasan et al,2019).

3.5.1.1.2 InceptionV3 :

As one of the pretrained models, we employed Inception-v3, the third iteration of the Inception architecture. It is the improved Inception-v2 version. In 2015, the architecture g: 4.8 was first unveiled . This well-liked CNN module is used to identify patterns in photos. Compared to VGG and ResNet, it is lighter. With 42 layers, this architecture achieves a lower error rate (Rosebrock et al,2017). InceptionV3 included modifications to the auxiliary classifiers that were proposed in InceptionV1. Auxiliary classifiers were employed in Inception V1 to create a deeper network, but they were used as a regularizer in Inception V3. When the accuracy gets close to a saturation condition at the end of the training procedure, that's when the auxiliary classifiers really start to contribute. In particular, they worked as regularizers in the presence of BatchNorm or Dropout operations. Thus, Inception-v3 was updated while maintaining all of Inception-v2's modifications. RMSProp Optimizer, Factorised 7x7 convolutions, BatchNorm in the Auxiliary Classifiers, Label Smoothing, and other features were included (Hasan et al,2019).

Figure n°25: InceptionV3 Architecture



(Hasan et al,2019)

3.5.1.2 EfficientNet

3.5.1.2.1 EfficientNet Family

EfficientNet is a convolutional neural network (CNN) family architectures ,which was proposed by Mingxing Tan and Quoc V. Le in 2019 (Tan et al ,2019) , The EfficientNet architectures are designed to achieve high accuracy while being computationally efficient. Here are key details about these models:

The EfficientNet CNN architecture scales all dimensions of depth, width, and resolution uniformly using a compound scaling method. This approach allows for better accuracy and efficiency compared to traditional methods of scaling individual dimensions.

The base EfficientNet model is built using inverted bottleneck residual blocks from MobileNetV2 and squeeze-and-excitation blocks. EfficientNet models are trained using RMSProp optimizer with a decay of 0.9 and momentum of 0.9, batch norm momentum of 0.99, and weight decay of 1e-5. The initial learning rate is 0.256, which decays by 0.97 every 2.4 epochs. The models use SiLU (Swish-1) activation, AutoAugment, and stochastic depth with a survival probability of 0.8. (Tan & Le, 2020)

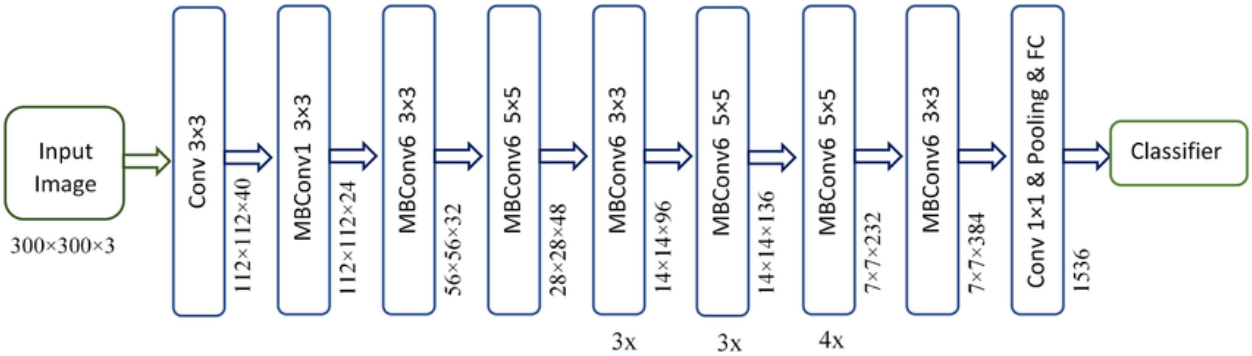
EfficientNet models have been used in various applications, including image classification, object detection, and semantic segmentation, and have achieved state-of-the-art accuracy on several datasets.

3.5.1.2.2 EfficientNet B3

EfficientNet B3 is a specific model in the EfficientNet family, which is designed to be computationally efficient while maintaining high accuracy.

EfficientNet B3 is based on the EfficientNet architecture and uses the same compound scaling method to uniformly scale depth, width, and resolution.Users can opt for pretrained weights trained on the ImageNet-1K dataset, with the flexibility to include them through the wights parameter. During training, the model achieves impressive accuracy, with top-1 and top-5 accuracy metrics reaching approximately 82.01% and 96.05%, respectively, on the ImageNet-1K dataset. Its extensive categorization capability spans numerous classes, with 997 categories omitted from the provided list and has 12 million parameters and 1.83 billion GFLOPS, with a manageable pretrained model file size of 47.2 MB. (PyTorch, 2024)
These characteristics make EfficientNet B3 suitable for various image classification tasks, ensuring high performance while conserving computational resources.the model is available as a pre-trained model in various frameworks such as PyTorch and Hugging Face.

Figure n° 26: EfficientNet B3 architecture.



(Inan, 2024)

3.5.2 Models Fine-tuning process

The process of fine-tuning transfer learning to make the models more appropriate for Alzheimer's detection. This technique involves the following steps (Zhang et al,2021):

1. Take a pre-trained source model and duplicate its parameters and weights to create a CNN model.

2. In the previously trained source model, unfreeze a few of the upper layers. Higher up in deep CNNs, the layers specialize more and more. The higher layers become increasingly specialized to the particular dataset that the model was trained on, while the lower levels acquire fundamental and universal features that can be applied to different kinds of images. Rather than substituting the general learning, the goal of fine-tuning is to modify these specialized features so they work with the new dataset.

3. Include an output layer with as many outputs as there are categories in the target dataset, along with additional trainable layers.

4.Use the target dataset to train the target model. All other layers' parameters are adjusted using the source model's parameters, but only the new trainable and output layers are trained from the beginning.(Ekuma, 2023)

3.5.2.1 Fine tuning InceptionV3 for Alzheimer's data

- The model is initialized with weights pre-trained on ImageNet, and all layers of the InceptionV3 base model are frozen to retain their learned features (functional).
- Additional layers are added for feature transformation, including Dropout layers for regularization to prevent overfitting.
- A GlobalAveragePooling2D layer is employed for dimension reduction, followed by Dense layers with ReLU activation to further transform features.
- Batch Normalization layers are included for normalization, enhancing model stability and performance.

This is the block diagram illustrating the model architecture, showcasing its components and connections:

Figure n°27: The block diagram illustrating InceptionV3 architecture.

- Custom callbacks are defined to monitor and control the training process: early stopping is implemented when accuracy exceeds a specified threshold, the learning rate is adjusted dynamically during training to improve convergence based on validation metrics.
- The model utilizes categorical cross-entropy loss function to optimize training, evaluating performance metrics such as Categorical Accuracy.
- Training proceeds for 100 epochs on the training data, with validation data used to monitor and evaluate model performance iteratively. This is parameters summary of this model :

```
Model: "inception_cnn_model"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 inception_v3 (Functional)   (None, 4, 4, 2048)        21802784

 dropout (Dropout)           (None, 4, 4, 2048)        0

 global_average_pooling2d (G  (None, 2048)             0
 lobalAveragePooling2D)

 flatten (Flatten)           (None, 2048)              0

 batch_normalization_94 (Bat  (None, 2048)             8192
 chNormalization)

 dense (Dense)               (None, 512)               1049088

 batch_normalization_95 (Bat  (None, 512)              2048
 chNormalization)

 dropout_1 (Dropout)         (None, 512)               0

 dense_1 (Dense)             (None, 256)               131328

 batch_normalization_96 (Bat  (None, 256)              1024
 chNormalization)

 dropout_2 (Dropout)         (None, 256)               0

 dense_2 (Dense)             (None, 128)               32896

 batch_normalization_97 (Bat  (None, 128)              512
 chNormalization)

 dropout_3 (Dropout)         (None, 128)               0

 dense_3 (Dense)             (None, 64)                8256

 dropout_4 (Dropout)         (None, 64)                0

 batch_normalization_98 (Bat  (None, 64)               256
 chNormalization)
```

```
 dense_4 (Dense)             (None, 4)                 260

=================================================================
Total params: 23,036,644
Trainable params: 1,227,844
Non-trainable params: 21,808,800
_________________________________________________________________
```

3.5.2.2 Fine tuning EfficientNetB3 for Alzheimer's data.

- The model is loaded and initialized with pre-trained weights from ImageNet, and all layers are initially frozen to retain their learned features.

Additional layers are appended to the model architecture:

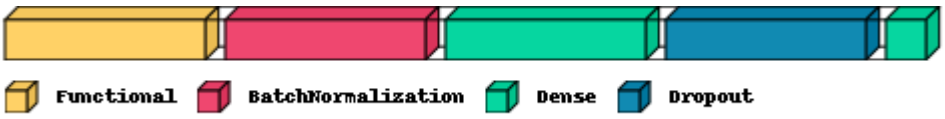
- BatchNormalization layer is added for normalization.
- A Dense layer with 256 units and ReLU activation function follows, enhancing feature extraction capabilities.
- Dropout regularization with a dropout rate of 0.45 is applied after the Dense layer to mitigate overfitting.
- Another Dense layer with softmax activation is added for final classification outputs across multiple classes.

And regularization techniques are implemented:

- L2 regularization with a coefficient of 0.016 is applied to the first Dense layer to reduce overfitting by penalizing large weights.
- L1 regularization with a coefficient of 0.006 is also applied on the first Dense layer, further promoting sparsity in the learned weights.

This is the block diagram illustrating the model architecture, showcasing its components and connections:

Figure n°28: The block diagram illustrating EfficientNetB3 architecture.



- The model is compiled using the Adamax optimizer with a learning rate set to 0.001, which adjusts the learning rate dynamically during training to optimize convergence.
- Categorical cross-entropy is employed as the loss function, suitable for multi-class classification tasks, to measure the model's performance.

- Early stopping mechanism is implemented during training to prevent overfitting and to halt training when the model's performance on the validation set no longer improves.
- The model undergoes training for 50 epochs on the training data, with validation data used to monitor and assess the model's learning progress and generalization ability. This is parameters summary of this model :

```
Model: "EfficientNetB3"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 efficientnetb3 (Functional)  (None, 1536)             10783535

 batch_normalization (BatchN  (None, 1536)             6144
 ormalization)

 dense (Dense)               (None, 256)               393472

 dropout (Dropout)           (None, 256)               0

 dense_1 (Dense)             (None, 4)                 1028

=================================================================
Total params: 11,184,179
Trainable params: 11,093,804
Non-trainable params: 90,375
_________________________________________________________________
```

3.6 Evaluation metrics selection :

In the realm of classification models, evaluation techniques and metrics are used to accurately measure performance. Through this part , we aim to provide a structured approach for comparing and contrasting different models, facilitating informed decision-making in the pursuit of optimal model selection and deployment.

A deep learning algorithm's performance is determined by its ability to reduce the error in the training phases and the gap between the training and test errors.

When evaluating the performance of our classification models, different methods and metrics can be used. We included classification accuracy, recall, Precision, F1 score, confusion matrix , and Receiver operating characteristic (ROC) . Every metric presents unique insights on the model's performance, for a comprehensive understanding of its abilities and limitations. For this the test dataset is used to evaluate the model. It is crucial to stress that the model has never seen this set of data before, which was split off at the start of the code.

3.6.1 Accuracy

Accuracy is one of the immediate metrics to evaluate. Accuracy measures how effectively the model can predict true positives and negatives within the classes, and it can be calculated using Equation

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

True positive (TP) refers to the number of instances that are correctly predicted as positive by the model ,true negative (TN) refers to the number of instances that are correctly predicted as negative by the model ,false positive (FP) refers to the number of instances that are incorrectly predicted as positive by the model and false negative (FN) refers to the number of instances that are incorrectly predicted as negative by the model .

For example, if the model is a binary classifier that predicts whether an MRI image is of a brain disease or not, a false positive would be an MRI that is incorrectly predicted as a disease but is actually not a disease.

FN For example, if the model is a binary classifier that predicts whether an MRI image is of a diseased or not, a false negative would be an MRI of a heart that is incorrectly predicted as not a disease but is actually an MRI of a disease

3.6.2 Precision

Precision score, also named as positive predictive value (PPV), measures the proportion of predicted positives that are really positive. It is calculated using Equation 2.10:

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

precision answers the question, "What proportion of the predicted positives are positive?" For AD early detection situations, this would indicate what proportion of the very mild dementia cases we identified as very mild dementia. Naturally, we would want this to be as high as possible; we would not want the model to identify other stages as very mild dementia. A low precision would mean that many cases we predicted as very mild dementia were other classes,potentially leading to inefficiencies or inaccuracies in AD early detection.

3.6.3 Recall

Recall score answers the question, "What proportion of the positives did we detect?" This is particularly relevant when dealing with very mild demented cases, where the recall score tells us the proportion of very mild demented cases the model detects. In event detection situations (for example, disease detection tests), we would like the recall to be as high as possible. A high recall score would mean we could detect most very mild dementia for the AD early prediction case. It is calculated using **Equation 2.10:**

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

Ideally, when both precision and recall are high, the detected object is correct, and the model can detect all the occurrences of a class.

3.6.4 F1 Score

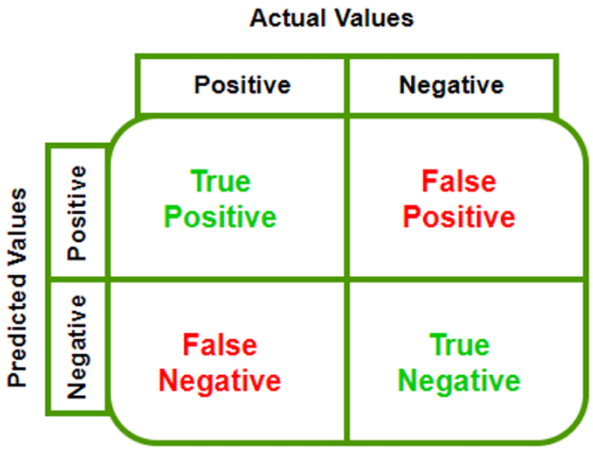
F1 Score seeks to balance precision and recall scores. But why is this equilibrium needed? Because precision and recall are trade-offs . Increasing the recall of the model frequently results in decreased precision; that is, when we try to increase the percentage of positives we classed, we may also increase the proportion of false positives, i.e., cases that are not positives but are predicted as such. Similarly, increasing precision leads to a decrease in recall. The F1 score, then, is a needed metric that measures the balance between precision and recall. Equation 2.11 is used to determine the F1 score.

$$F1 = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

3.6.5 Confusion matrix

The confusion matrix, shown in figure 18 is a table that summarizes the performance of a classification model by displaying the number of TP, TN, FP and FN for each class. It provides a visual overview of how well the model is distinguishing between different classes.

Figure n° 29 : Confusion Matrix



(Balamurugan et al,2022)

The confusion matrix shows insights of the errors being made by a classifier, but more importantly, it shows the types of errors that have been made (Balamurugan et al,2022). For this, sklearn.metrics offers a simple and fast way to present the confusion matrix of a model based on its predictions and the correct outputs of these predictions.like this, the confusion matrix would evaluate how well the predictions have been made.

3.6.6 Receiver operating characteristic (ROC)

The Receiver operating characteristic (ROC) can measure the classifier's ability to avoid a false classification , by plotting the True Positive Rate (recall) against the False Positive Rate (FPR) at a threshold.

The True Positive Rate (TPR) is shown on the Y-axis, and the False Positive Rate (FPR) forms the X-axis. The percentage of negatives the model got wrong is measured by the false positive rate, which is calculated as False Positives / Total Actual Negatives. The percentage of accurately predicted positives is measured by the true positive rate, which is calculated as True Positives / Total Actual Positives.

Consequently, FPR concentrates on the problems and TPR emphasizes the positives. Using several thresholds ranging from 0 to 1, the curve is created, and the acquired TPR and FPR are plotted on a graph. As a result, the ROC curve displays the model's ability to discriminate between the positive and negative classifications.A perfect model would have a true positive rate of 100% and no false positives (FPR of 0). The ideal classifier, which would

have 0 FPR and 100% TPR, would be represented by the dot in the upper left corner of Figure 4.9(Sikka, 2021).

The no skill, random classifier will have a 50% area under the curve (AUC) as it represents the diagonal line.Since a perfect classifier has 100% TPR, 0% FPR, and reaches the upper left corner right away, it will have 100% area under the curve. The area under the ROC curve for the majority of models will lie between 0.5 and 1, which corresponds to the range between the perfect classifier and the no skill model. Figure 4.9 provides a graphic representation of this. The model's skill and, consequently, the area under the ROC curve increase as the curve approaches the upper left corner (Sikka,2021).

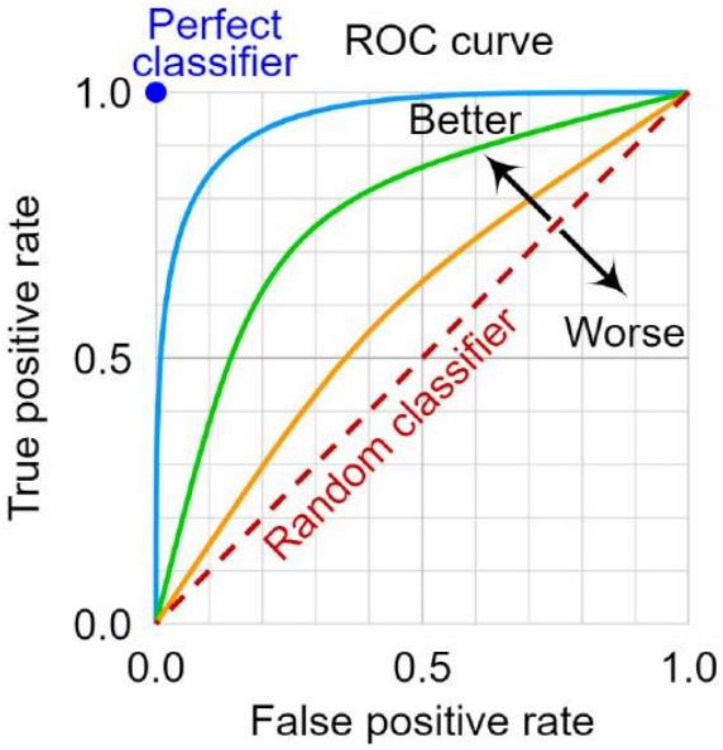


Figure n° 30: ROC curve

(Sikka, 2021)

3.7 Conclusion

In this chapter, we present the methodology adopted for the subsequent experimental work.that covers key aspects of the comparative study approach and detailed exploration of the data along with the CNNs architectures used, in the purpose of AD classification.and its key components, including the Transfer learning.In the next chapter, we validate the performance of the proposed model based on the implementation of the experiments ADNI and OASIS datasets.

Chapter 4

Results and Analysis

4.1 Results and Analysis

In order to evaluate the performance of the models, several metrics are considered, training validation and testing accuracies,AUC, precision, recall and F1 score.
In the ROC curves that we plotted, there are four curves, each representing the performance of a different class out of the four included:

- Mild Dementia (class 0)
- Moderate Dementia (class 1)
- Non-Dementia (class 2)
- Very Mild Dementia (class 3)

It's important to note that val accuracy is through the validation set of the ADNI data (the first column),while the other metrics are a result of the test data of the ADNI except the last test which is with the OASIS test set as an external completely dependent data set.

4.1.1 CNN Models' analysis of results:

4.1.1.1 CNN 1 model performance:

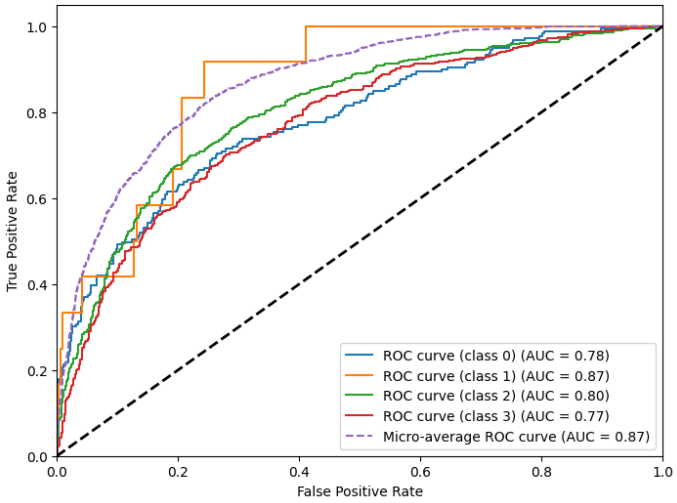
Table n° 3: CNN 1 model results:

val acc	acc	auc	loss	F1	recall	precision	OASIS test acc
0,93	0.849	0,87	1.1093	0.64	0.64	0.65	0,34

In this cnn model with 100 epochs and batch size 16 we got:
Validation accuracy (0.93): Indicates high accuracy on the validation set, with the model correctly classifying 93% of the data.
Testing accuracy (0.849): The test accuracy of 84.9% shows that the model maintains good performance on unseen data, though it's slightly lower than the validation accuracy.
AUC (0.87): An AUC of 0.87 signifies a good discriminative ability of the model across the four classes.
Loss (1.1093): Reflects the model's prediction error, with lower values being better. This is a moderately high loss.

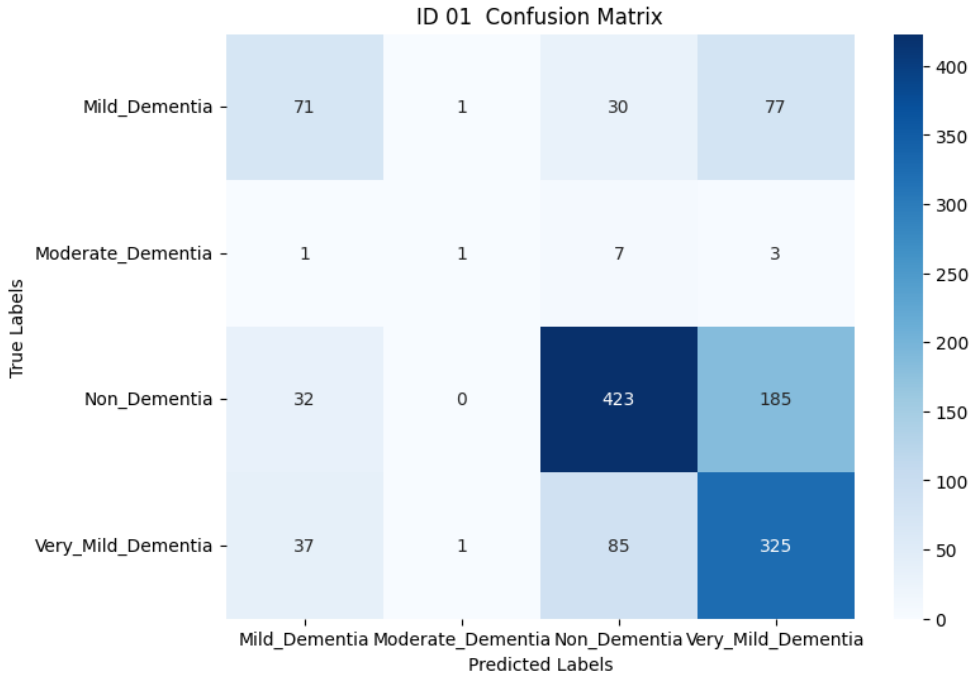
F1 Score (0.64): Indicates a moderate balance between precision and recall.
Recall (0.64): 64% of actual positives were correctly identified.
Precision (0.65): 65% of the model's positive predictions were correct.
OASIS test accuracy (0.34): Significantly lower at 34%, suggesting poor generalization to this specific test dataset.

Figure n° 31: ROC curve for CNN 01:



For Moderate Dementia (class 1, orange line), with a curve that closely approaches the top-left corner, indicating a high true positive rate and low false positive rate with the 0.87 AUC . The curves for Mild Dementia (class 0 , blue line) and Very Mild Dementia (class 3 , red line) are similar and show moderate performance with 0.78 AUC and 0.77 AUC respectively, staying closer to the diagonal. The Non-Dementia (class 2, green line) curve performs slightly higher than Mild and Very Mild Dementia with 0.80 AUC. The micro-average ROC curve reflects the overall good performance of the model, heavily influenced by its effectiveness in identifying Moderate Dementia.

Figure n°32: Confusion matrix for CNN 01 :



In this confusion matrix, we can see that the model performed well at classifying non-dementia and very mild dementia. There were 423 out of 640 non-dementia cases classified correctly,185 were misclassified as very mild dementia and 32 as mild dementia. Out of 448 very mild dementia cases, 325 were classified correctly, and 85 were misclassified as non dementia while 37 as mild dementia. Out of 179 mild dementia cases, only 77 were classified correctly, with 30 being misclassified as non-dementia and 77 as very mild dementia.

The initial CNN model demonstrated promising performance metrics on the validation dataset, achieving an accuracy of 84% and an area under the curve (AUC) of 0.87. Despite the relatively high accuracy and AUC, the model's F1 score, recall, and precision were moderate, each recorded at approximately 0.64. When evaluated on the external OASIS dataset, which the model had not previously encountered, the performance notably decreased, with an accuracy of only 34%. This significant drop in performance on unseen data indicates potential issues with the model's generalizability ,The significant loss value of 1.1093 further supports this, indicating potential issues with the model's learning process.

The ROC curves in the image suggest that the CNN model N1 is performing moderately well to well at classifying between positive and negative cases. Class 1 (Moderate Dementia) appears to have the best performance (AUC=0.87), while class 0 (Mild Dementia) has the worst performance (AUC=0.78). It indicates that the model can distinguish (Moderate Dementia)cases relatively well.
In the confusion matrix ,the model as well struggled more with very mild dementia and mild dementia.Overall, the model seems to be promising for detecting moderate dementia, but it may not be accurate enough to diagnose very mild or mild dementia on its own.

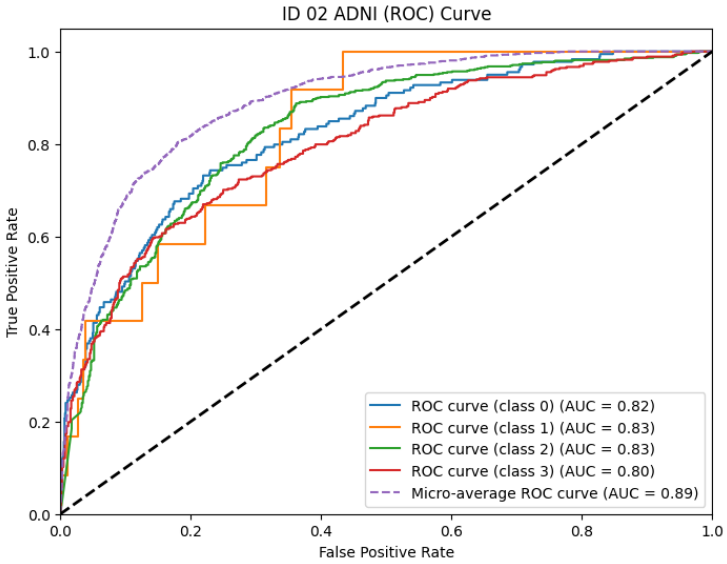
4.1.1.2 CNN 2 model performance :

Table n° 4 : CNN 2 model results:

val acc	acc	auc	loss	F1	recall	precision	OASIS test acc
0.95	0,69	0,89	2.2112	0.674	0.685	0.685	0,67

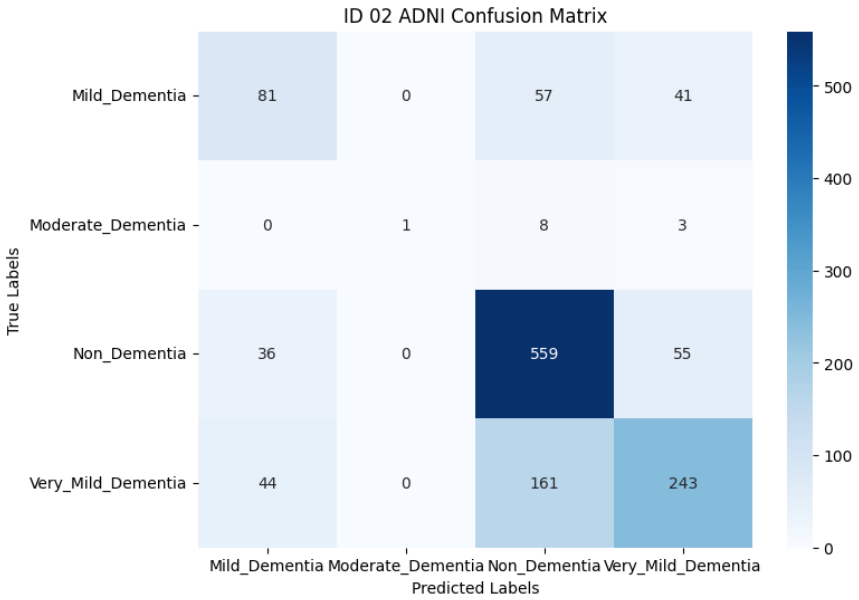
In this cnn model with 100 epochs and batch size 16 we got:
Validation Accuracy (0.95): Indicates strong performance on the validation dataset.with the model correctly classifying 95% of the data.
Test Accuracy (0.69): significantly lower than the validation accuracy, shows that the model maintains moderate performance on unseen data.
Test AUC (0.89): this results shows a strong ability to distinguish between classes on the test set.
Loss (2.2112): This is a significantly high loss, Implies that predictions are not as accurate as desired.
F1 Score (0.674): Reflects a balance between precision and recall .
Recall and Precision (0.685 each): Shows that 68.5% of actual positives are correctly identified and 68.5% of positive predictions are correct.
OASIS Test Accuracy (0.67): Significantly moderate at 67%, suggesting moderate generalization to this specific test dataset.

Figure n°33: ROC curve for CNN2 :



The curve for class 1 (Moderate Dementia, orange line) and class 2 (Non-Dementia, green line) are very similar, with 0.83 AUC for both. The curve for class 0 (Mild Dementia, blue line) is slightly less effective but still maintains a reasonably good performance with 0.82 AUC . The curve for class 3 (Very Mild Dementia, red line) shows the least performance with 0.80 AUC, staying closer to the diagonal line, which indicates more moderate discrimination. The micro-average ROC curve (purple dashed line) indicates overall good performance with 0.89 AUC .

Figure n°34: Confusion matrix for CNN2 :



In this confusion matrix, we can see that the model performed well at classifying non-dementia and very mild dementia. There were 559 out of 650 non-dementia cases classified correctly while 55 were misclassified as very mild dementia and 36 as mild dementia . and 243 out of 448 very mild dementia cases were classified correctly, 161 were misclassified as non dementia and 44 as mild dementia. Out of 179 mild dementia cases, only 81 were classified correctly, with 57 being misclassified as non-dementia and 41 as very mild dementia.

The second model exhibited a high validation accuracy of 95%, indicating strong performance on the training data. However, the test accuracy dropped to 69% Suggests the model may not generalize well to new data. The test AUC of 0.89 demonstrates the model's solid capability to differentiate between classes. Despite these positive metrics, the model's loss was notably high at 2.2112, indicating that the predictions may have significant errors and possible optimization issues .The F1 score, recall, and precision were all approximately 0.674 to 0.685, showing a balanced performance in terms of precision and recall but not outstanding performance. The OASIS test accuracy improvement to 67% is a positive sign, but not overall high generalizability to unseen data.

The AUC for moderate dementia (class 1) and non-dementia (class 2) are 0.83, which are the highest AUC among the four classes. This suggests that the model performs best at classifying non-dementia and moderate dementia cases.while The AUC for very mild dementia (class 3) is 0.80, which is the lowest AUC among the four classes. This suggests that the model performs less well at classifying very mild dementia cases.

In the Confusion matrix ,The model struggles to differentiate between Mild Dementia and Very Mild Dementia.There seems to be a significant number of misclassifications between Mild Dementia and Non-Dementia.Overall, this CNN model seems to have some promise in differentiating between people with Non-Dementia and those with some form of dementia. However, it does struggle to accurately classify the different stages of dementia.

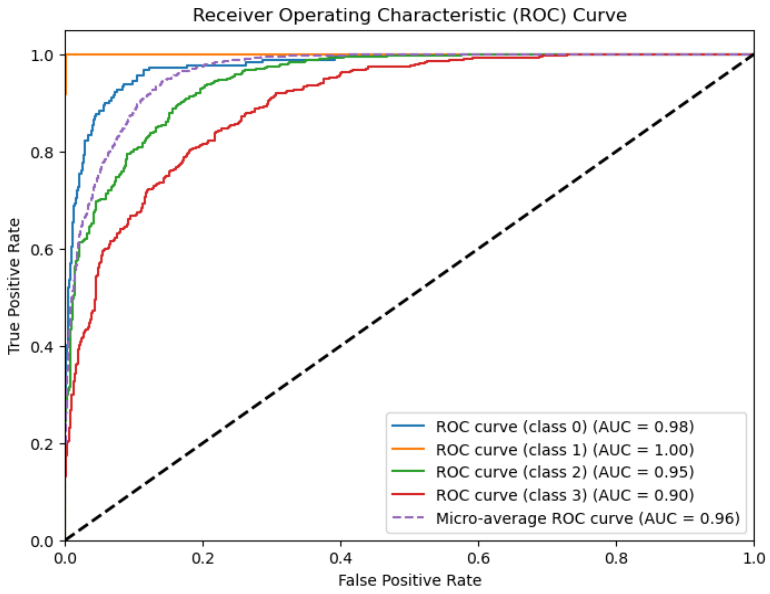
4.1.1.3 CNN 3 model performance :

Table n° 5 : CNN 3 model results:

val acc	acc	auc	loss	F1	recall	precision	OASIS test acc
0.9709	0.9050	0.9725	0.0998	0.87	0.905	0.91	0,88

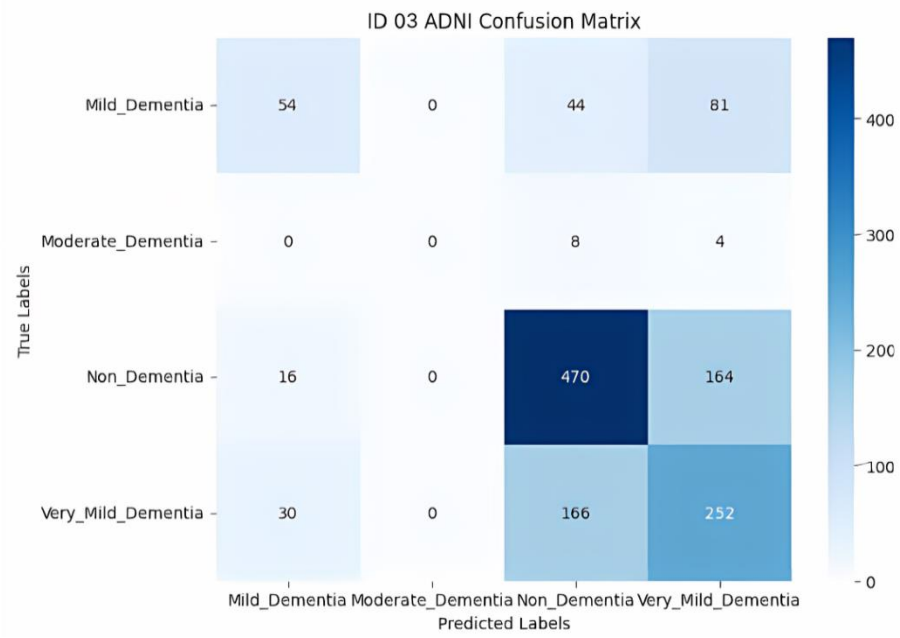
Validation Accuracy: The model achieved a high validation accuracy of 97%, indicating strong performance on the validation dataset.
Test Accuracy: The test accuracy of 90% suggests that the model generalizes well but slightly underperforms compared to the validation set.
Test AUC: A high test AUC of 97% implies excellent discriminatory power between classes.
Loss: The model's loss value of 0.0998 reflects low error, signifying effective learning during training.
Recall: The recall of 90% shows the model's effectiveness in correctly identifying relevant instances from the dataset.
Precision: The precision score of 91% reveals the model's accuracy in classifying positive samples correctly.
F1 Score: An F1 score of 87% indicates a balanced trade-off between precision and recall, demonstrating the model's robustness in handling both false positives and false negatives.
OASIS Test Accuracy: The test accuracy on the OASIS dataset is 88% suggesting the model maintains high performance even when applied to different data not seen during training.

Figure n°35: CNN 03 model ROC curve :



for mild Dementia (class 0, blue line), with the highest curve that closely approaches the top-left corner, indicating a high true positive rate and low false positive rate with the 0.98 AUC . The curves for Very Mild Dementia (class 3 , red line) show moderate performance with 0.78 AUC and 0.77 AUC respectively, staying closer to the diagonal. The Non-Dementia (class 2, green line) curve performs slightly higher than Mild and Very Mild Dementia with 0.80 AUC. The micro-average ROC curve (purple dashed line) reflects the overall good performance of the model with 0.96 AUC, heavily influenced by its effectiveness in identifying Moderate Dementia.

Figure n°36: CNN 03 model confusion matrix :



In this confusion matrix, we can see that the model performed well at classifying non-dementia and very mild dementia. There were 470 out of 650 non-dementia cases classified correctly with 164 were misclassified as very mild dementia and 16 were misclassified as mild dementia .252 out of 448 very mild dementia cases were classified correctly,and 166 were misclassified as non dementia while 30 were misclassified as mild dementia and Out of 179 mild dementia cases, only 54 were classified correctly, with 44 being misclassified as non-dementia and 81 as very mild dementia.

The very high validation accuracy (97.09%), test accuracy (90.50%), and test AUC (0.9725) reflect robust classification capabilities and an excellent ability to distinguish between classes.The low loss value (0.0998) indicates excellent prediction accuracy. The high F1 score (0.87), recall (0.905), and precision (0.91) suggest a well-balanced performance with few false positives and negatives. The improved OASIS test accuracy (88%) further confirms the model's strong generalization capabilities. Overall, this model is highly effective for the given classification task.

The ROC AUC analysis indicates the model's high effectiveness in detecting Mild Dementia (AUC 0.98) and overall (micro-average AUC 0.96), likely due to distinct features that strongly differentiate Mild Dementia from other classes. The moderate performance in identifying Very Mild Dementia (AUC 0.78 and 0.77) suggests overlapping characteristics with other classes, making it harder to distinguish. The slightly better performance for Non-Dementia (AUC 0.80) indicates clearer distinguishing features from dementia-related classes but still poses some challenges due to potential similarities with early-stage dementia.

The model performed well at classifying non-dementia and very mild dementia cases. However, it struggled significantly with classifying mild dementia, with many instances being misclassified as non-dementia or very mild dementia.

4.1.1.4 CNN 4 model performance :

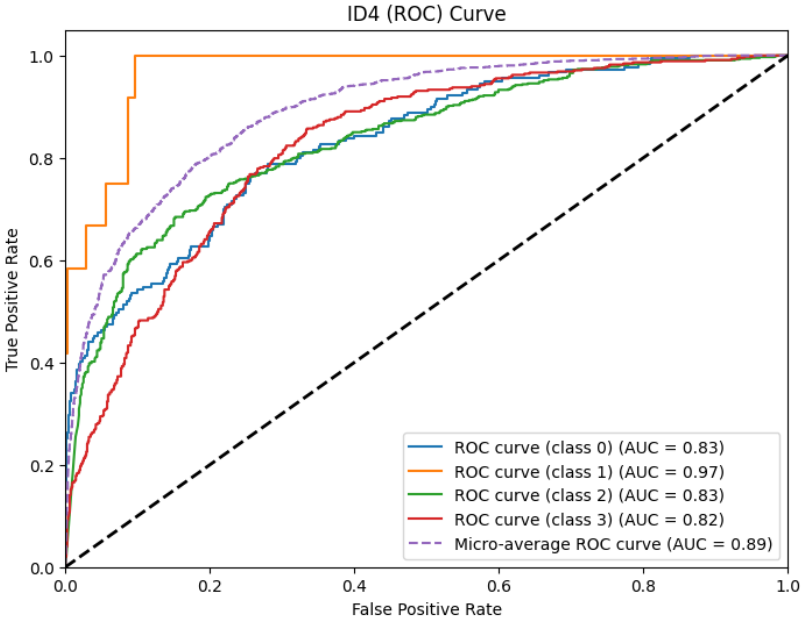
Table n°6 : CNN 4 model results:

val acc	acc	auc	loss	F1	recall	precision	OASIS test acc
0,99	0.6787	0.86	0.0269	0.68	0.67	0.68	0.25

Validation Accuracy: The validation accuracy of 99 % indicates excellent performance on the validation dataset.
Test Accuracy: The test accuracy is 67.8%, which is significantly lower than the validation accuracy.

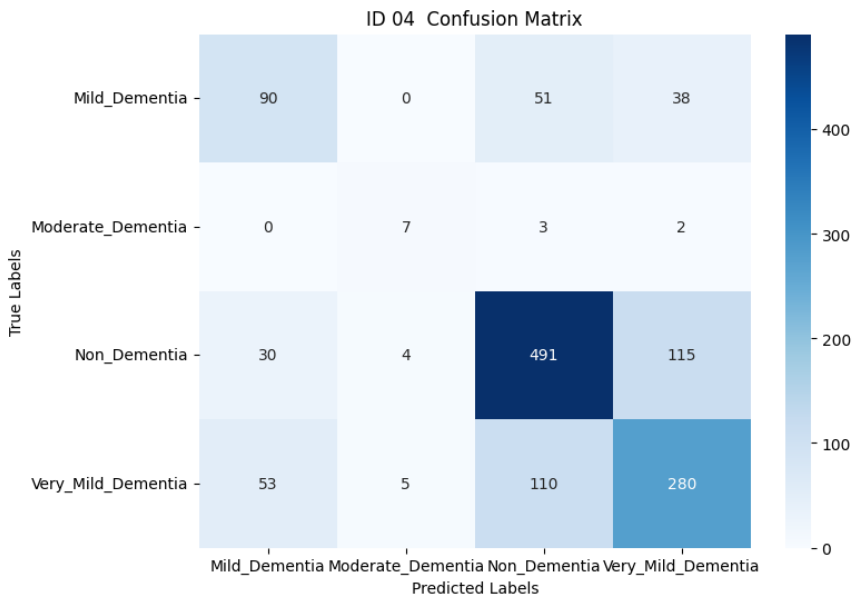
Test AUC: The AUC score of 86% suggests reasonable discriminatory ability between classes.
Loss: The model's loss value of 0.0269 is very low, indicating minimal error during training and good convergence.
Recall: The recall score of 67% indicates the model's effectiveness in identifying relevant instances from the dataset.
Precision: The precision score of 68% shows the model's accuracy in classifying positive samples correctly.
F1 Score: The F1 score of 68% indicates a moderate balance between precision and recall.
OASIS Test Accuracy: The test accuracy on the OASIS dataset is 25% which is very low.

Figure n°37: CNN 04 model ROC curve :



For Moderate Dementia (class 1, orange line), with the highest curve that closely approaches the top-left corner, indicating a high true positive rate and low false positive rate with the 0.97 AUC The curves for Very Mild Dementia (class 3 , red line) shows also a high performance with 0.82 AUC , staying closer to the diagonal. The Non-Dementia (class 2, green line) and the mild Dementia (class 0, blue line) curves perform slightly higher with 0.83 AUC for both. The micro-average ROC curve with 0.89 AUC reflects the overall good performance of the model, heavily influenced by its effectiveness in identifying Moderate Dementia.

Figure n°38: CNN 04 model confusion matrix:



In this confusion matrix, we can see that the model performed well at classifying non-dementia and very mild dementia. There were 491 out of 640 non-dementia cases classified correctly with 115 were misclassified as very mild dementia and 30 were misclassified as mild dementia .
280 out of 448 very mild dementia cases were classified correctly,and 110 were misclassified as non dementia while 53 were misclassified as mild dementia and Out of 179 mild dementia cases, only 90 were classified correctly, with 51 being misclassified as non-dementia and 38 as very mild dementia.

This model demonstrates exceptionally high validation accuracy (99%) and a low loss value (0.0269), indicating precise predictions on the validation dataset. However, the lower testing accuracy (67.87%) suggests potential overfitting or a need for better generalization to unseen data. The AUC of 0.86 indicates good performance in distinguishing between classes.
The F1 score, recall, and precision were all around 0.68, indicating a balanced performance in terms of precision and recall. When tested on the external OASIS dataset, the model's accuracy was 25%, significantly lower compared to its performance on the validation dataset. This substantial drop

in accuracy highlights a notable challenge in generalizing the model's performance to unseen datasets

The model excels in identifying moderate dementia (class 1) with an AUC of 0.97, indicating high accuracy. It also performs well for very mild dementia (class 3) with an AUC of 0.82, though there is moderate misclassification, particularly with non-dementia cases. Non-dementia (class 2) and mild dementia (class 0) both have an AUC of 0.83, showing good performance. The micro-average AUC of 0.89 reflects overall strong model effectiveness. The confusion matrix highlights the model's strength in correctly classifying non-dementia and very mild dementia cases but shows some misclassification, especially within closely related categories.

4.1.2 Transfer Learning's analysis Results

4.1.2.1 Efficient Performance:

Table n° 7: EfficientNetB3 model results

val acc	acc	auc	loss	F1	recall	Precision	OASIS test acc
0,9828	0,99	0,99	0.13	0.50	0.48	0.53	0.33

Validation Accuracy: The validation accuracy of 98% indicates strong performance on the validation dataset.

Test Accuracy: The test accuracy is 99%, demonstrating high performance on the test dataset, which aligns closely with the validation accuracy.

Test AUC: The AUC score of 99% indicates excellent discriminatory power between classes, suggesting the model's ability to distinguish between different classes effectively.

Loss: The model's loss value of 0.13 is relatively low, indicating effective learning during training with minimal error.

Recall: The recall score of 48% indicates the model's ability to correctly identify relevant instances from the dataset.

Precision: The precision score of 53% shows the model's accuracy in classifying positive samples correctly.

F1 Score: The F1 score of 50% suggests a moderate balance between precision and recall.

OASIS Test Accuracy: The test accuracy on the OASIS dataset is 33%, which is lower compared to the test accuracy on the model's original test

set, suggesting some drop in performance when applied to unseen data.

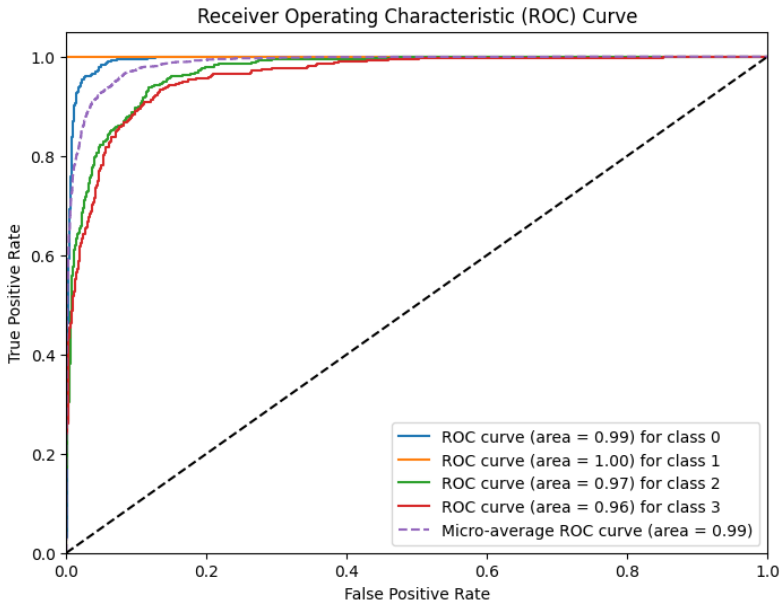
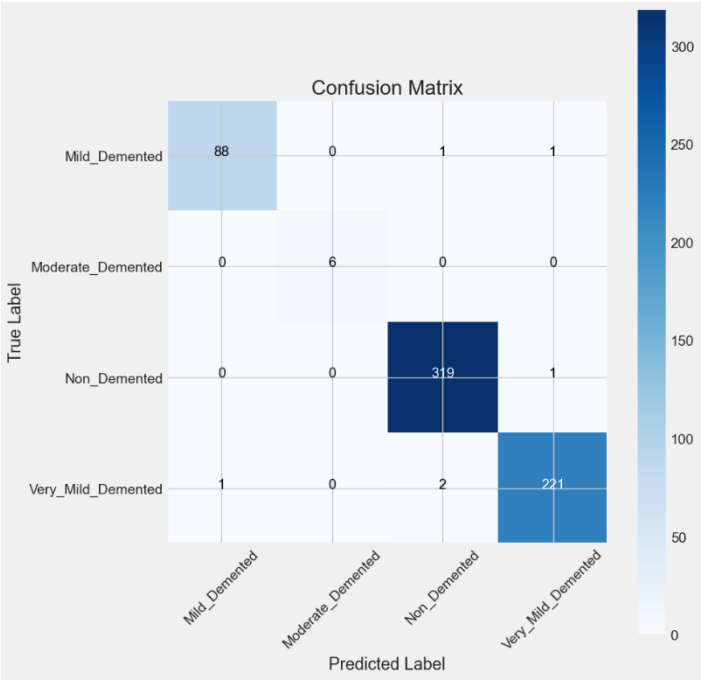


Figure n°39: Efficient net ROC curve :

For Moderate Dementia (class 1, orange line), with the highest curve that closely approaches the top-left corner, indicating a high true positive rate and low false positive rate.The curves for Very Mild Dementia (class 0, blue line) shows also a high performance with 0.99 AUC , staying closer to the diagonal. The Non-Dementia (class 2, green line) and the mild Dementia (class 0, blue line) curves perform slightly less with 0.97 and 0.96 AUC for both respectively. The micro-average ROC curve with 0.99 AUC reflects the overall high performance of the model, heavily influenced by its effectiveness in identifying Very Mild Dementia .

Figure n°40: Efficient net confusion matrix :



In this confusion matrix, we can see that the model performed well at classifying non-dementia , very mild dementia and mild dementia. There were 319 out of 318 non-dementia cases classified correctly with only one misclassified as very mild dementia .221 out of 224 very mild dementia cases were classified correctly,and only 2 were misclassified as non dementia ,and Out of 90 mild dementia cases,88 were classified correctly, with 2 being misclassified .

This model shows very high validation accuracy (98.28%) and testing accuracy (99%), along with an excellent AUC (0.99) and low loss (0.13), suggesting strong performance and well-fitted predictions on the training and validation data. However, the F1 score (0.50), recall (0.48), and precision (0.53) indicate moderate performance in balancing false positives and false negatives.The low OASIS test accuracy (33%) reveals poor generalization to the OASIS test set, suggesting potential issues with the model's ability to handle different data distributions , generalizability and overall performance on unseen data.
The model shows excellent performance across all classes, particularly for very mild dementia (class 0), with an AUC of 0.99 and minimal misclassification, heavily influencing the overall micro-average AUC of 0.99

that underscores the model's overall exceptional performance, particularly influenced by its ability to accurately identify Very Mild Dementia.Moderate dementia (class 1) also has a high true positive rate with a closely approaching top-left ROC curve. Non-dementia (class 2) and mild dementia (class 0) have AUCs of 0.97 and 0.96, respectively, indicating slightly lower but still strong performance.
The confusion matrix confirms this effectiveness, The model excelled at classifying non-dementia, very mild dementia, and mild dementia cases with high accuracy. There were minimal misclassifications across all categories, indicating strong performance.

4.1.2.2 InceptionV3 Performance:

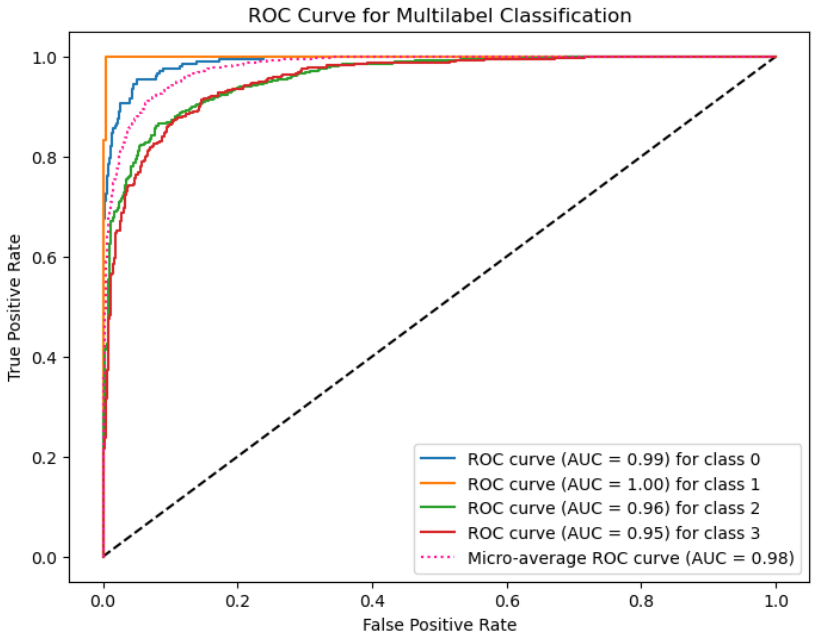
Table n° 8: InceptionV3 model results

val acc	acc	auc	loss	F1	recall	precision	OASIS test acc
0.98	0,90	0.98	0.2596	0.90	0.86	0.873	31.50

Validation Accuracy: The validation accuracy of 98% indicates strong performance on the validation dataset, showing the model's capability to learn from the training data effectively.
Test Accuracy: The test accuracy of 90% demonstrates good generalization to the test dataset,
Test AUC: The AUC score of 98% indicates excellent discriminatory power between classes, suggesting the model's ability to distinguish between different classes effectively.
Loss: The model's loss value of 0.2596 is moderate in reducing error during training.
Recall: The recall score of 86% indicates the model's effectiveness in correctly identifying relevant instances from the dataset.
Precision: The precision score of 87.3% shows the model's accuracy in classifying positive samples correctly.
F1 Score: The F1 score of 90% suggests a strong balance between precision and recall, demonstrating the model's robustness in handling both false positives and false negatives.

OASIS Test Accuracy: The test accuracy on the OASIS dataset is 31.5%, which is Significantly low, suggesting poor generalization to this new test dataset.

Figure n°41: InceptionV3 ROC curve :



For Moderate Dementia (class 1, orange line), with the highest curve that closely approaches the top-left corner, indicating a high true positive rate and low false positive rate.The curves for Very Mild Dementia (class 0, blue line) shows also a high performance with 0.99 AUC , staying closer to the diagonal. The Non-Dementia (class 2, green line) and the mild Dementia (class 0, blue line) curves perform slightly less with 0.96 and 0.95 AUC for both respectively. The micro-average ROC curve with 0.98 AUC reflects the overall high performance of the model, heavily influenced by its effectiveness in identifying Very Mild Dementia .

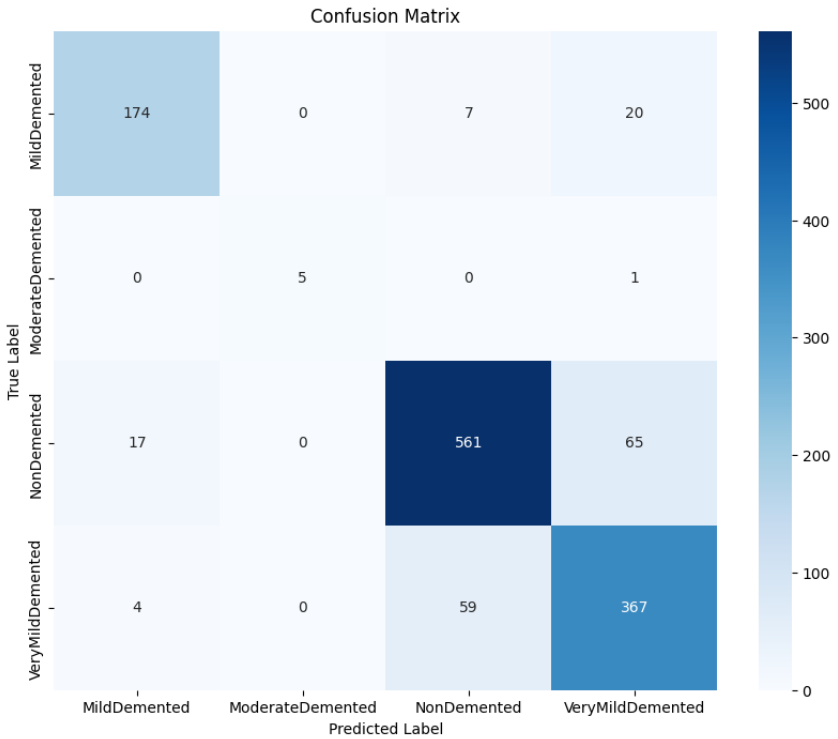


Figure n°42: InceptionV3 confusion matrix:

In this confusion matrix, we can see that the inceptionV3 model performed well at classifying non-dementia , very mild dementia and mild dementia. There were 561 out of 643 non-dementia cases classified correctly with 65 misclassified as very mild dementia and 17 as mild dementia .367 out of 430 very mild dementia cases were classified correctly,and 59 were misclassified as non dementia ,and Out of 201 mild dementia cases,174 were classified correctly, with 20 being misclassified as very mild dementia and 7 as non dementia.

This model shows strong validation accuracy (98%) and test accuracy (90%)although there is a slight drop in the testing accuracy, along with an excellent AUC (0.98), indicating strong overall performance and good discriminative ability. The relatively low loss value (0.2596) supports this.The high F1 score (0.90), recall (0.86), and precision (0.873) suggest a well-balanced performance with few false positives and negatives. However, the extremely low OASIS test accuracy (31.50%) raises concerns about the model's ability to generalize to different datasets, indicating poor generalization.

The ROC AUC analysis indicates strong performance of the model in identifying Moderate Dementia (AUC 0.99) and Very Mild Dementia (AUC

0.99), characterized by high true positive rates and low false positive rates for these classes. Non-Dementia (AUC 0.96) and Mild Dementia (AUC 0.95) also show good performance but slightly lower than the dementia-related classes. The micro-average ROC curve with 0.98 AUC reflects the model's overall effectiveness, primarily driven by its exceptional capability in detecting Very Mild Dementia.
In the confusion matrix, The InceptionV3 model performed well at classifying non-dementia and mild dementia cases but struggled more with differentiating very mild dementia from the other classes. Misclassification was primarily seen between non-dementia and very mild dementia cases.

4.2 Discussion: final comparison and interpretation:

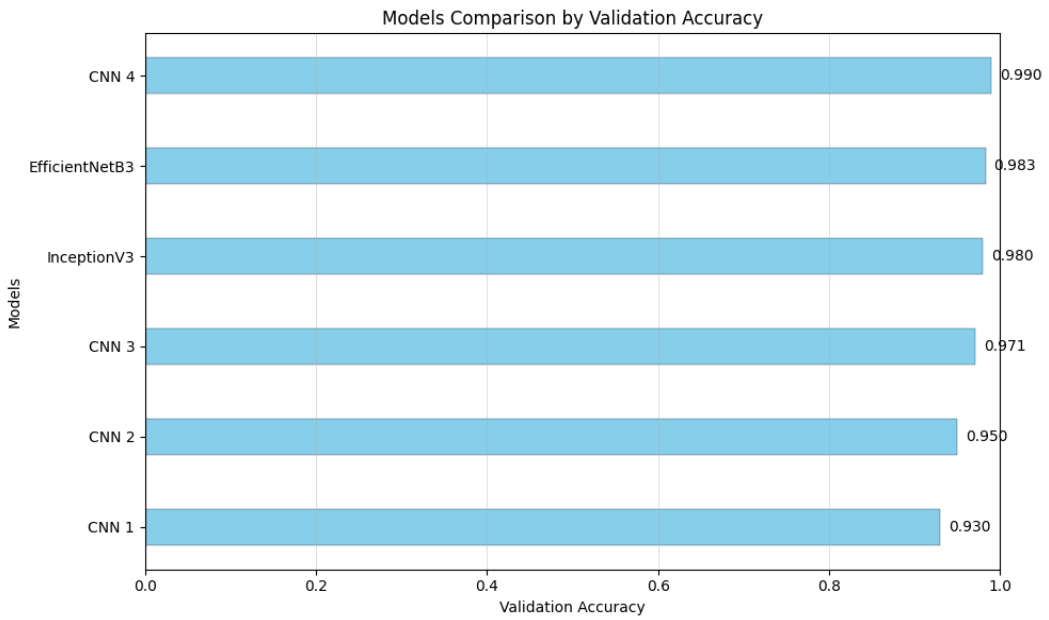
To perform a comparative analysis of the different CNN models, we considered this table.

Table n°9: All models results summary

Model	Val acc	Test acc	AUC	Loss	F1	Recall	Precision	OASIS test acc
Cnn 1	0,93	0.849	0,87	1.109	0.64	0.64	0.65	0,34
Cnn 2	0.95	0,69	0,89	2.211	0.674	0.685	0.685	0,67
Cnn 3	0.971	0.905	0.972	0.099	0.87	0.905	0.91	0,88
Cnn 4	0,99	0.678	0.86	0.027	0.68	0.67	0.68	0.25
EffiecientNetb3	0,983	0,99	0,99	0.13	0.50	0.48	0.53	0.33
InceptionV3	0.98	0,90	0.98	0.259	0.90	0.86	0.873	0.31

The plot shows the models from highest to lowest based on validation accuracy

Figure n° 43: models sorted by validation accuracy (Val Acc)

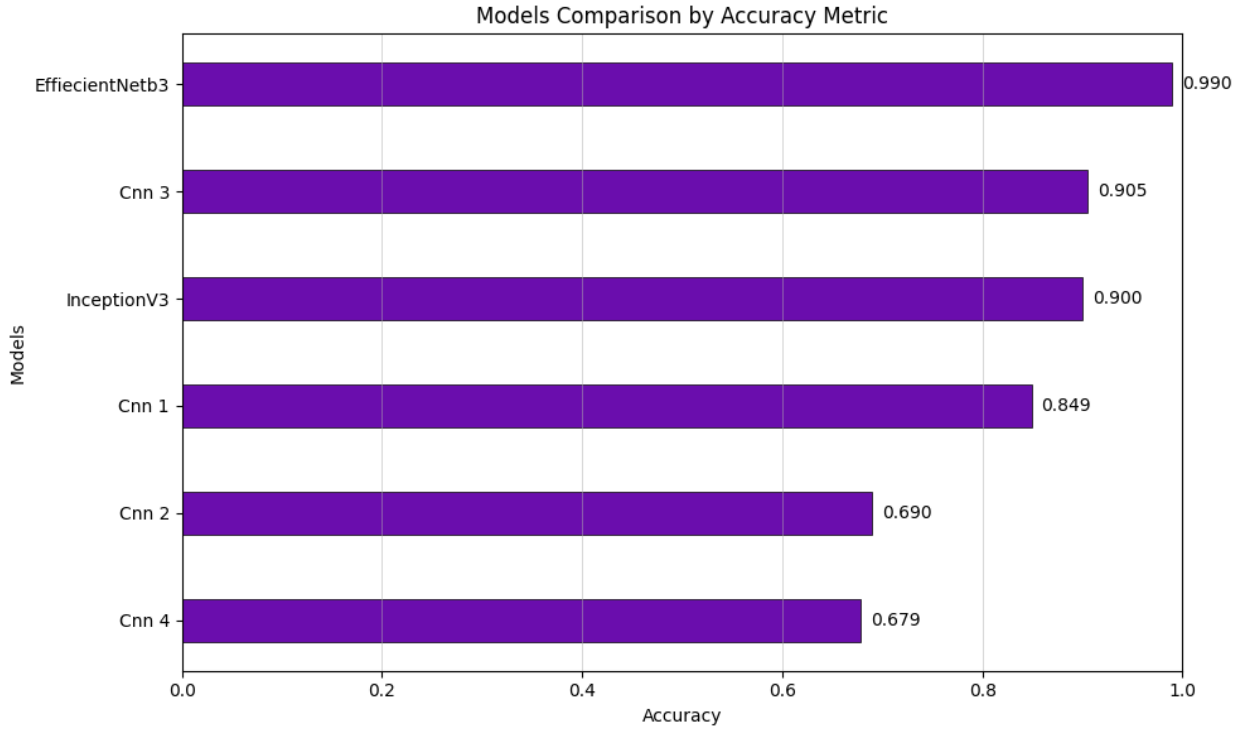


- **CNN 4** has the highest validation accuracy at 0.99, suggesting it performs exceptionally well during the training phase.
- **EfficientNet** and **InceptionV3** follow closely with validation accuracies of 0.9828 and 0.98, respectively.
- **CNN 3** also performs well with a validation accuracy of 0.9709.
- **CNN 1** and **CNN 2** have lower validation accuracies at 0.93 and 0.95, respectively.

4.2 .1 Accuracy (Acc)

The plot shows the models from highest to lowest based on testing accuracy.

Figure n° 44: models sorted by Accuracy (Acc)



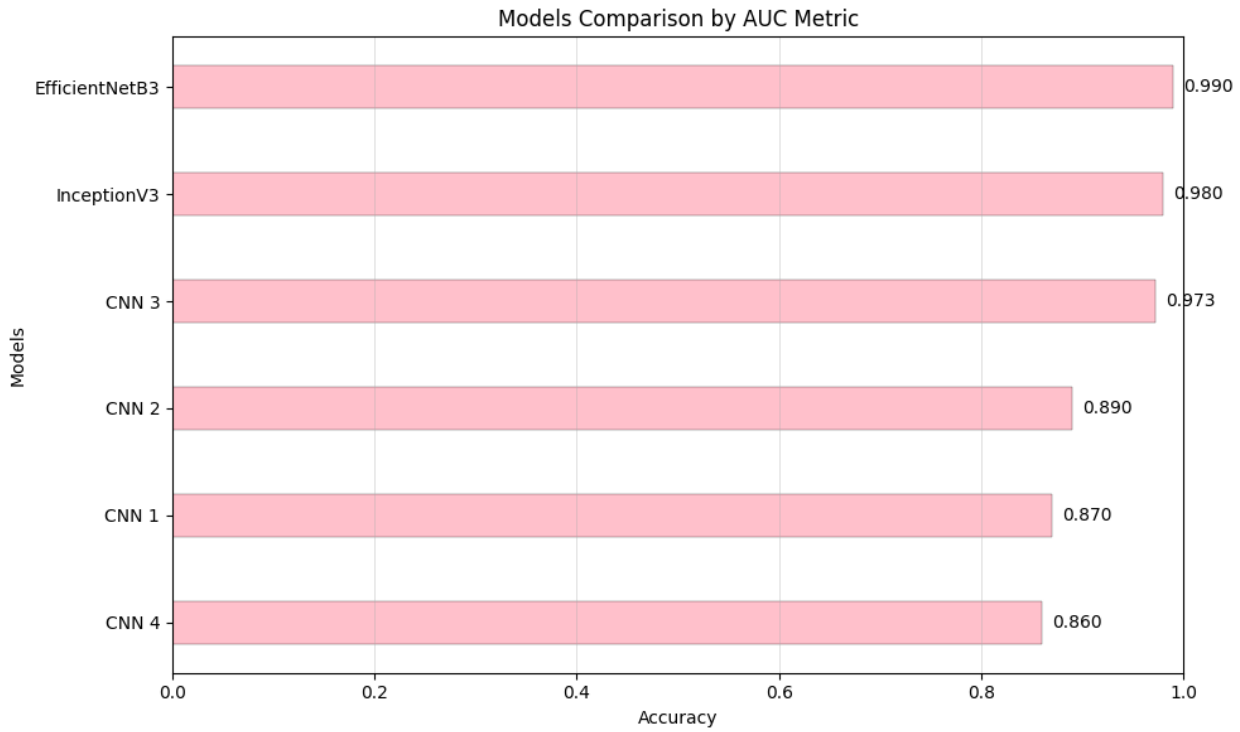
Overall accuracy measures the model's performance on the test set:

- **EfficientNet** achieves the highest accuracy at 0.99, indicating excellent generalization to the test data.
- **CNN 3** follows with an accuracy of 0.9050.
- **InceptionV3** also performs well with an accuracy of 0.90.
- **CNN 1**, **CNN 2**, and **CNN 4** show lower accuracies, with CNN 4 having the lowest at 0.6787

4.2.2 Area Under the Curve (AUC)

The plot shows the models from highest to lowest based on AUC.

Figure n°45: models sorted by AUC

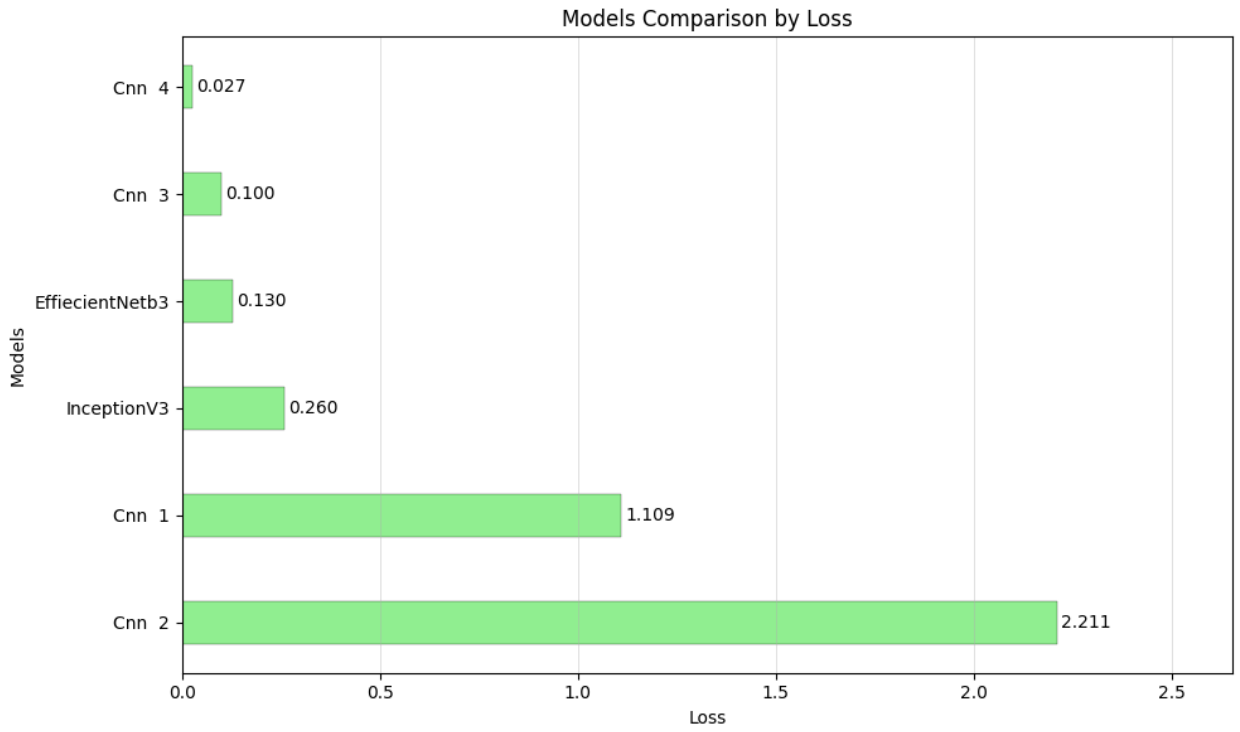


- **EfficientNetB3** achieves the highest AUC at 0.99, indicating superior discriminatory ability.
- **InceptionV3** follows closely with an AUC of 0.98.
- **CNN 3** also performs well with an AUC of 0.9725.
- **CNN 1**, **CNN 2**, and **CNN 4** have lower AUC values, with CNN 4 being the lowest at 0.86.

4.2.3 Loss

The plot shows the models from lowest to highest based on training loss

Figure n° 46: models sorted by loss

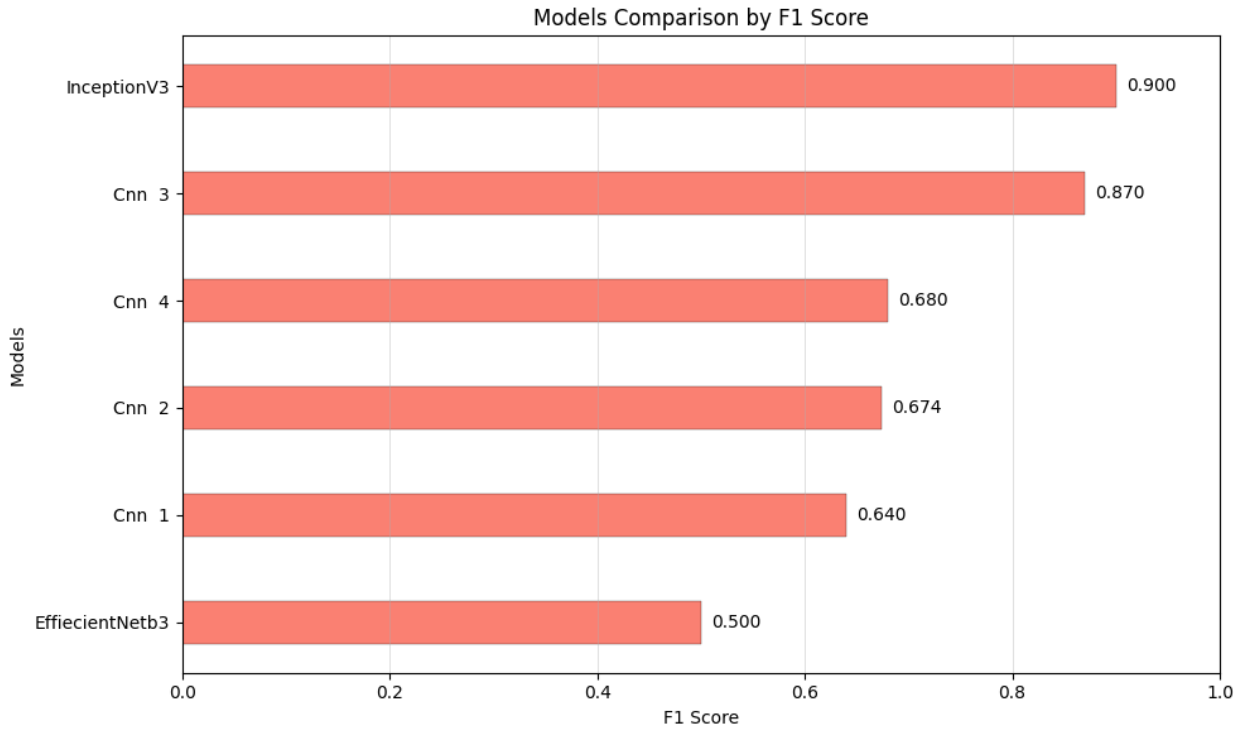


- **CNN 4** has the lowest loss at 0.0269, indicating high precision in predictions.
- **CNN 3** and **EfficientNet** also performed well with losses of 0.0998 and 0.13, respectively.
- **InceptionV3** has a moderate loss of 0.2596.
- **CNN 1** and **CNN 2** have higher losses, with CNN 2 having the highest at 2.2112.

4.2.4 F1 Score

The plot shows the models from highest to lowest based on F1 score

Figure n° 47: models sorted by F1 score

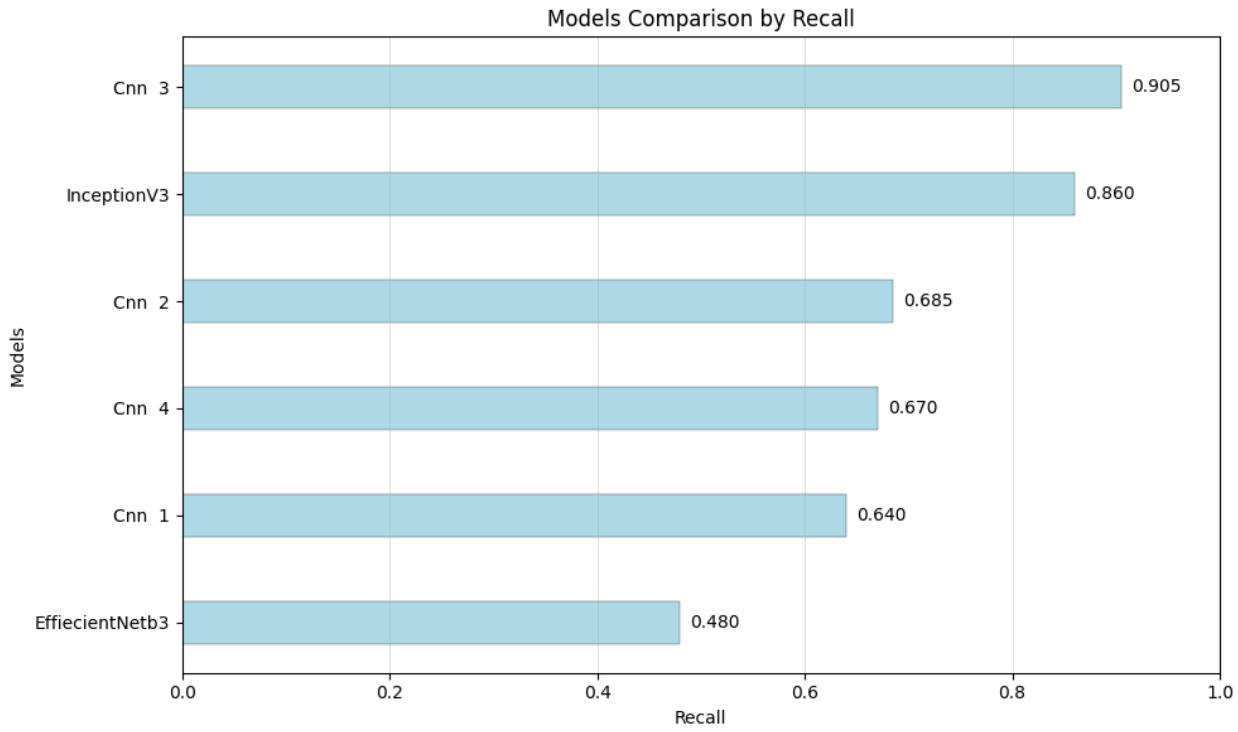


- **InceptionV3** achieves the highest F1 score at 0.90, suggesting a good balance between precision and recall.
- **CNN 3** also performs well with an F1 score of 0.87.
- **CNN 2**, **CNN 4**, and **EfficientNet** have moderate F1 scores, with EfficientNet being the lowest at 0.50.
- **CNN 1** has the lowest F1 score at 0.64.

4.2.5 Recall

The plot shows the models from highest to lowest based on testing accuracy

Figure n° 48: models sorted by Recall

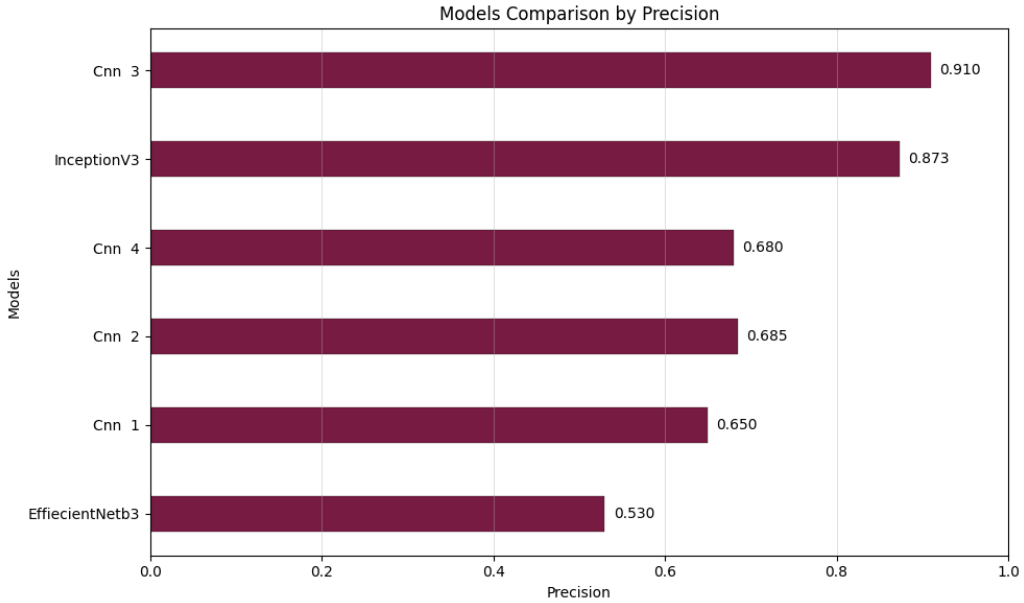


- **CNN 3** achieves the highest recall at 0.905.
- **InceptionV3** follows with recalls of 0.86 .
- **CNN 2 and CNN 4** moderate recall of 0.68 and 0.67 respectively .
- **CNN 1** and **EfficientNet** have lower recall values, with EfficientNet being the lowest at 0.48.

4.2.6 Precision

The plot shows the models from highest to lowest based on precision

Figure n° 49: models sorted by precision

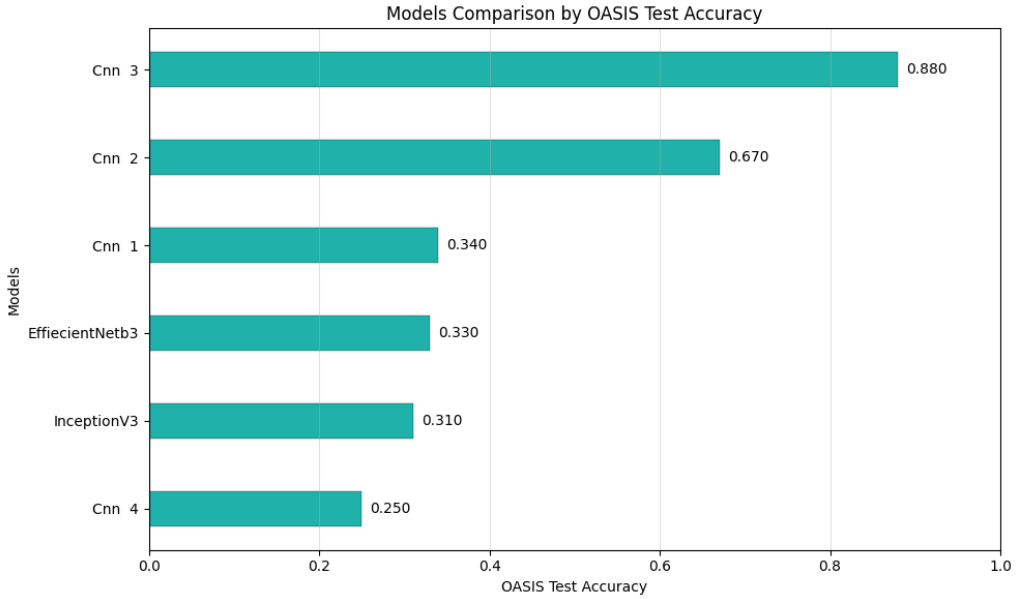


- **CNN 3** has the highest precision at 0.91
- **InceptionV3** also have a high precision 0.873.
- **CNN 2**, **CNN 4**, and **CNN 1** have moderate precision with CNN 1 being the low at 0.65.
- **EfficientNet** has the lowest precision at 0.53.

4.2.7 OASIS Test Accuracy

The plot shows the models from highest to lowest based on OASIS Test accuracy

Figure n° 50: models sorted by OASIS test accuracy



- **CNN 3** achieves the highest OASIS test accuracy at 0.88, suggesting it generalizes very well to external datasets.
- **CNN 2** follows with an OASIS test accuracy of 0.67.
- **CNN 1**, **CNN 3**, **CNN 4**, and **EfficientNetb3** have lower OASIS test accuracies, with CNN 4 being the lowest at 0.25 .

4.3 Summary of findings:

- **EfficientNet** demonstrates exceptional overall accuracy and AUC, indicating robust performance and generalizability. However, its F1 score , recall and its OASIS test accuracy are relatively low.
- **CNN 3** shows strong performance across most metrics, including a high validation accuracy, AUC, F1 score, recall, and precision,and excellent OASIS test accuracy.
- **InceptionV3** also performs well, with high F1 score, making it a well-rounded model though its OASIS test accuracy is relatively low.
- **CNN 4** has the highest validation accuracy and lowest loss but struggles with accuracy, AUC, and OASIS test accuracy.
- **CNN 1** and **CNN 2** show moderate performance, with CNN 2 having a notably high OASIS test accuracy.

Each model has its strengths and weaknesses, making them suitable for different aspects of Alzheimer's disease detection.

4.4 Conclusion

In this thesis, a comprehensive comparative study on the use of CNNs for Alzheimer's Disease (AD) classification using various architectures was conducted. The study involved the design, implementation, experimentation,evaluating and comparing the performance of different CNN models for four stages of the disease classification (4 classes), including InceptionV3, EfficientNet, and four custom CNNs, using the ADNI dataset for training and testing as well as OASIS dataset for testing only. The methodology included detailed data preprocessing techniques, to improve classification accuracy and aid medical specialists.

In this study, we evaluated the performance of the models using various metrics such as validation accuracy, testing accuracy , area under the curve (AUC), loss, F1 score, recall, and precision.

Our comparative analysis demonstrated that each model exhibited unique strengths and weaknesses. EfficientNet showed the highest overall accuracy and AUC, indicating robust performance and generalizability. However, its lower F1 score and recall suggested potential issues with class imbalance or misclassification. CNN 3 emerged as a well-rounded model with high performance across most metrics, including an outstanding OASIS test accuracy. Efficient NetB3 also achieved high validation accuracy and AUC but lower OASIS test accuracy.

In conclusion,selecting a single model from the competitive set proved challenging. While EfficientNetB3 excels in numerous performance metrics, CNN 3 demonstrates superior performance on external, unseen data, underscoring its robustness and reliability. This distinction makes CNN 3 particularly advantageous for clinical applications, where generalization to novel data is critical. Additionally, CNN 3 exhibits as well high performance in the other metrics closely approaching those of EfficientNetB3, further reinforcing its competitive standing and potential suitability for clinical deployment.

The empirical evidence suggests that CNN 3’s ability to maintain high accuracy and reliability on external datasets positions it as a more suitable candidate for clinical use, warranting further enhancement and validation in this domain. paving the way for more accurate and reliable diagnostic tools.

4.5 Future Work

4.5.1 Data and Training Process Improvements

A promising avenue for future work involves expanding the dataset used for model training. While the model demonstrates robust performance on validation and test datasets, its suboptimal performance on the OASIS test set warrants further investigation. Analyzing the training process and addressing potential discrepancies within the dataset may enhance the model's robustness and its ability to generalize across diverse datasets.

4.5.2 Preprocessing Enhancements

Exploring novel and more effective preprocessing techniques can potentially improve model performance. Enhanced preprocessing could help in extracting more relevant features and reducing noise, thereby improving the overall accuracy of the model.

4.5.3 Model Architecture and Hyper-parameter Optimization

In terms of Convolutional Neural Network (CNN) models, future work could focus on refining the network architecture. This includes tuning hyperparameters and employing cross-validation techniques to identify optimal configurations. Additionally, investigating hybrid approaches that combine CNNs with other machine learning techniques may yield significant improvements.

4.5.4 Multi-Class Classification and Alternative Imaging Modalities

Future research could also explore the implementation of multi-class classification to differentiate between three stages: Normal Cognition (NC), Alzheimer’s Disease (AD), and Mild Cognitive Impairment (MCI). Furthermore, comparing the performance of different architectures using various types of brain imaging, such as Positron Emission Tomography (PET) scans, could provide deeper insights and improve diagnostic accuracy.

4.5.5 Clinical Validation

Despite achieving high classification accuracy, it is imperative to seek confirmation from medical specialists due to the inherent possibility of error. Ensuring clinical validation is essential to confirm the reliability and safety of the developed system for practical applications.

In conclusion, Future work should focus on expanding the dataset to improve model robustness and generalizability, exploring advanced preprocessing techniques to enhance feature extraction, and refining CNN architectures through hyperparameter optimization and cross-validation. Additionally, incorporating multi-class classification to distinguish between

Normal Cognition (NC), Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD), as well as exploring alternative imaging modalities such as PET scans, could provide deeper insights and improve diagnostic accuracy. Clinical validation remains crucial to ensure the reliability and safety of these models for practical applications.addressing these areas in future work can lead to significant advancements in the detection, diagnosis, and classification of Alzheimer’s Disease, ultimately contributing to more reliable and effective clinical tools.

References

Alberola, M., Molina Gallego, A., Garay Maestre, G., & Unai. (2019). *Artificial Vision and Language Processing for Robotics*. Packt Publishing, BPB Publications.

Aljunid, S. M., et al. (2014). Development of clinical pathway for mild cognitive impairment and dementia to quantify cost of age-related cognitive disorders in Malaysia. *Malays. J. Public Health Med.*, 14, 88–96.

Alsubaie, M. G., & Luo, S., Shaukat, K. (2024). Alzheimer's Disease Detection Using Deep Learning on Neuroimaging: A Systematic Review. *Mach. Learn. Knowl. Extr.*, 6, 464–505. https://doi.org/10.3390/make6010024

Alzheimer's Association. (2015). 2015 Alzheimer's disease facts and figures. *Alzheimers Dement.*, 11(3), 332-84. doi: 10.1016/j.jalz.2015.02.003. PMID: 25984581.

Arfat, A. K., et al. (2024). Dual-3DM 3-AD: mixed transformer-based semantic segmentation and triplet pre-processing for early multi-class Alzheimer's diagnosis. *IEEE Trans Neural Syst Rehab Eng*, 32, 696–707.

Avinash Chandra, et al. (2019). Resonance imaging in Alzheimer's disease and mild cognitive impairment. *Journal of Neurology*, 266. doi: 10.1007/s00415-018-9016-3

Balamurugan, D., Aravinth, S. S., Reddy, P. C. S., Rupani, A., & Manikandan, A. (2022). Multiview objects recognition using deep learning based wrap-CNN with voting scheme. *Neural Process. Lett.*, 54, 1495–1521.

Burns, A., & Iliffe, S. (2009). Alzheimer's disease. *BMJ*, 338.

Blaikie, L et al. Current and emerging therapeutic targets of Alzheimer's disease for the design of multi-target directed ligands. MedChemComm 2019, 10, 2052–2072.

C. Morris, N. Cairns, L. A. Beckett, et al., "The alzheimer's disease neu roimaging initiative: progress report and future plans," Alzheimer's &

dementia: the journal of the Alzheimer's Association, vol. 6, no. 3, pp. 202–211, 2010

Cardarilli, G. C., Di Nunzio, L., Fazzolari, R., Giardino, D., Nannarelli, A., Re, M., & Spanò, S. (2021). A pseudo-softmax function for hardware-based high speed image classification. Scientific reports, 11(1), 1-1

Caselli, Richard J., and Eric M. Reiman, "Characterizing the preclinical stages of Alzheimer's disease and the prospect of presymptomatic intervention," Journal of Alzheimer's Disease, vol.33, no. s1, pp. S405-S416, 2013. Österlind,

Chimakurthi VNSS(2020)Application of convolution neural network for digital image processing. Eng Int 8(2):149–158

Crous-Bou et al. Alzheimer's Research & Therapy (2017) 9:71 DOI 10.1186/s13195-017-0297-z

D. P. Perl, "Neuropathology of Alzheimer's disease," Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine, vol. 77, no. 1, pp. 32–42, 2010.

Daviglus ML, et al. National Institutes of Health State-of-the-Science Conference statement: preventing Alzheimer disease and cognitive decline. Ann Intern Med. 2010;153(3):176–81.

Dubois B, et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. Lancet Neurol. 2014;13

Dubois, B et al. Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS–ADRDA criteria. Lancet Neurol. 2007, 6, 734–746.

E.-W. Radue, M. Weigel, R. Wiest, and H. Urbach, "Introduction to magnetic res onance imaging for neurologists," CONTINUUM:Lifelong Learning in Neurology, vol. 22, no. 5, Neuroimaging, pp. 1379–1398, 2016.

Ebrahimi A et al (2021) Deep sequence modelling for Alzheimer's disease detection using MRI .Comput Biol Med 134:104537

Ebrahimi, A.; Luo, S.; Alzheimer's Disease Neuroimaging Initiative. Convolutional neural networks for Alzheimer's disease detection on MRI images. J. Med. Imaging 2021, 8, 024503.

Estudillo-Romero, A.; Haegelen, C.; Jannin, P.; Baxter, J.S.H. Voxel-based diktiometry: Combining convolutional neural networks with voxel-based analysis and its application in diffusion tensor imaging for Parkinson's disease.

Hum. Brain Mapp. 2022, 43, 4835–4851.
Fawcett, T. An introduction to ROC analysis. Pattern Recogn. Lett. 2006, 27, 861–874.

Ganesh D et al (2023) Implementation of convolutional neural networks for detection of Alzheimer's disease. BioGecko, A J N Z Herpetol 12(01)

Guo, Y., Li, Y., Wang, L., & Rosing, T. (2020, April). Adafilter: Adaptive filter fine-tuning for deep transfer learning. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 04, pp. 4060-4066).

Hcini G, Jdey I, Ltifi H (2022) Improving malaria detection using L1 regularization neural network. JUCS: J Univers Comput Sci 285(10):1087–1107. https://doi.org/10.3897/jucs.81681

Hcini G, Jdey I, Ltifi H (2023) HSV-Net: a custom cnn for malaria detection with enhanced color representation. In: 2023 International conference on cyberworlds (CW). IEEE (2023). Sousse, Tunisia, pp 337–340

Hcini, Ghazala, Imen Jdey, and Habib Dhahri. "Investigating Deep Learning for Early Detection and Decision-Making in Alzheimer's Disease: A Comprehensive Review." *Neural Processing Letters* 56 (2024): 153. doi:10.1007/s11063-024-11600-5. Accepted March 17, 2024.

Hcini Getal(2021)Hyperparameter optimization in customized convolutional neural network for blood cells classification. J Theoret Appl Inf Technol 99:5425–5435

Inan, Yavuz 2024, Adaptive Multiscale Retinal Diagnosis: A Hybrid Trio-

Model Approach for Comprehensive Fundus Multi-Disease Detection Leveraging Transfer Learning and Siamese Networks.

Iqbal H. Sarker. Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. SN Computer Science, 2:1–20, 11 2021.

J. Bernal et al., "Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review", Artif. Intell. Med., 2019.

J.H.; et al. Identification of Alzheimer's disease using a convolutional neural network model based on T1-weighted magnetic resonance imaging. Sci. Rep. 2020, 10, 22252.

Jiang, X.; Chang, L.; Zhang, Y.-D. Classification of Alzheimer's disease via eight-layer convolutional neural network with batch normalization and dropout techniques. J. Med. Imaging Health Inform. 2020, 10, 1040–1048.

Joseph Gaugler et al, Alzheimer's disease facts and figures, Alzheimer's & Dementia: The Journal of the Alzheimer's Association, vol. 13, pp. 325–373, April 2017.

Karikari, T.K.; Ashton, N.J.; Brinkmalm, G.; Brum, W.S.; Benedet, A.L.; Montoliu-Gaya, L.; Lantero-Rodriguez, J.; Pascoal, T.A.; Suárez-Calvet, M.; Rosa-Neto, P.; et al. Blood phospho-tau in Alzheimer disease: Analysis, interpretation, and clinical utility. Nat. Rev. Neurol. 2022, 18, 400–418.

Karlsson, et al, "Memory improvement at different disease," Neuropsychologia, vol. 27, no. 5, pp.737-742, 1989.

Kim, B., Yuvaraj, N., Sri Preethaa, K. R., & Arun Pandian, R. (2021). Surface crack detection using deep learning with shallow CNN architecture for enhanced computation. Neural Computing and Applications, 33(15), 9289-9305.

Klette, R. (2014). Concise Computer Vision: An Introduction into Theory and Algorithms. Springer.

Konur, O.; Kingma, D.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015; pp. 1–15.

Lerch, Jason P., Jens C. Pruessner, Alex Zijdenbos, Harald Hampel, Stefan J. Teipel, and Alan C Evans, "Focal decline of cortical thickness in Alzheimer's disease identified by computational neuroanatomy," Cerebral cortex, vol. 15, no. 7 , pp.995-1001,2005.

M. W. Weiner, P. S. Aisen, C. R. Jack, W. J. Jagust, J. Q. Trojanowski, L. Shaw, A. J. Saykin, J.

Martin Prince et al, International, A.s.D. The global impact of dementia: an analysis of prevalence, incidence, cost and trends, in World Alzheimer Report. London: International, A.s.D; 2015.

Matthews, P.M.; Peter, J. Functional magnetic resonance imaging. J. Neurol. Neurosurg. Psychiatry 2004, 75, 6–12.

Mohi ud din dar G, et al (2023) A novel framework for classification of different Alzheimer's disease stages using CNN model. Electronics 12(2):469

N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," Journal of Machine Learning Research, vol. 15, pp. 1929–1958, 2014.

Nancy Noella RS, Priyadarshini J (2020) Diagnosis of Alzheimer's and Parkinson's disease using artificial neural network. Int J Sci Technol Res 9(3):3659–3664

Nguyen, Julia Cat-Vy. *Vessel Recognition in Ultrasound Images Using Machine Learning Techniques.* Department of Physics and Technology, University of Bergen, June 2023. Supervisors: Nello Blaser, Robert Hurry, and Hakon Grothe.

Nirthika, R.; Manivannan, S.; Ramanan, A.; Wang, R. Pooling in convolutional neural networks for medical image analysis: A survey and an empirical study. Neural Comput. Appl. 2022, 34, 5321–5347.

O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, Karpathy, rej, A. Khosla, M. Bernstein and others, "ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision (IJCV), vol. 115, no. 3, pp. 211-252, 2015.

Ouali, A., Hudelot, C., & Tami, M. (2020). An overview of deep semi-supervised learning. arXiv preprint arXiv:2006.05278.

Pan, S.J.; Yang, Q. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 2010, 22, 1345–1359.

Passeri .E et al, Alzheimer's Disease: Treatment Strategies and Their Limitations. Int J Mol Sci. 2022 Nov 12;23(22):13954. doi: 10.3390/ijms232213954. PMID: 36430432; PMCID: PMC9697769.

Rajendiran M et al (2022) Detection of Alzheimer’s disease in MRI images using different transfer learning models and improving the classification accuracy. Int J Health Sci 6:11851–11869. https://doi. org/10.53730/ijhs.v6nS3.8944

Ronald Petersen, Glenn Smith, Steve Waring, Robert Ivnik, Eric Tangalos, and Emre Kokmen. Mild cognitive impairment: Clinical characterization and outcome. Archives of neurology, 56:303–8, 04 1999. doi: 10.1001/archneur.56.6.760

S.-H. Tsang, Review: Googlenet (inception v1)- winner of ilsvrc 2014 (im age classi cation), Aug. 2018. Available: https://medium.com/coinmonks/paper review of google net inception v1 winner of ilsvlc 2014 image-classi cation-c2b3565a64e7.

Schneider LS, et al. Clinical trials and late-stage drug development for Alzheimer’s disease: an appraisal from 1984 to 2014. J Intern Med. 2014; 275(3):251–83.

Shamrat FMJ Metal(2023)AlzheimerNet:an effective deep learning based proposition for Alzheimer’s disease stages classification from functional brain changes in magnetic resonance images. IEEE Access 11:16376–16395

Shanmugam JVetal(2022)Alzheimer's disease classification using pre-trained deep networks. Biomed Signal Process Control 71:103217

Shenton, et al. A review of magnetic resonance imaging and diffusion tensor imaging findings in mild traumatic brain injury. Brain Imaging Behav. 2012, 6, 137–192.

Sikka, Bharat. (2021). "Elements of Deep Learning for Computer Vision: Explore Deep Neural Network Architectures, PyTorch, Object Detection Algorithms, and Computer Vision Applications for Python Coders" (English Edition). BPB Publications. pp. 101-110. ISBN: 978-93-90684-687.

Sisodia PS et al (2023) A review of deep transfer learning approaches for class-wise prediction of Alzheimer's disease using MRI images. Arch Comput Methods Eng 30(4):2409–2429

Sreelakshmi, S., Malu, G., & Sherly, E. (2022). Alzheimer's disease classification from cross-sectional brain MRI using deep learning. 2022 IEEE International Conference on Image Processing, Informatics, Communication and Energy Systems(SPICES), 1, 401–405.

Sriramakrishnan Padmanaban, Kalaiselvi Thiruvenkadam, Padmapriya T., M. Thirumalaiselvi, and RAM KUMAR. A role of medical imaging techniques in human brain tumor treatment. 8:565–568, 01 2020. doi: 10.35940/ijrte.D1105.1284S219.

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning (ICML 2019) (pp. 6105-6114). PMLR.

Waldemar, G.; Dubois, B.; Emre, M.; Georges, J.; McKeith, I.G.; Rossor, M.; Scheltens, P.; Tariska, P.; Winblad, B. Recommendations for the diagnosis and management of Alzheimer's disease and other disorders associated with dementia: EFNS guideline. Eur. J. Neurol. 2007, 14, e1–e26.

Wang,N.; Wang, Y.; Er, M.J. Review on deep learning techniques for marine object recognition: Architectures and algorithms. Control Eng. Pract. 2020, 118, 104458.

Yamashita R et al (2018) Convolutional neural networks: an overview and application in radiology. Insights Imaging 9:611–629

Yingge, H., Ali, I., & Lee, K. Y. (2020, February). Deep neural networks on chip-A survey. In 2020 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 589 592). IEEE.

Yinjun Zhang & Lingzhi Wang (2024) Early diagnosis of Alzheimer's disease using dual GAN model with pyramid attention networks, Connection Science, 36:1, 2321351, DOI: 10.1080/09540091.2024.2321351

Z. Breijyeh and R. Karaman, "Comprehensive Review on Alzheimer's Disease: Causes and Treatment," Molecules, vol. 25, no. 24, p. 5789, Dec. 2020.

Zaven et el,Alzheimer's Association Report, "2022 Alzheimer's disease facts and figures," Alzheimer's & dementia: the journal of the Alzheimer's Association, vol. 18, no. 4, pp. 700-789, 2022.

Websites

A. Rosebrock, Imagenet: Vggnet,resnet, inception, and xception with keras, Mar. 2017. Available :https://www.pyimagesearch.com/2017/03/20/imagenet-vggnet-resnet-inception-xception-keras/

ADNI : Alzheimer's Disease Neuroimaging Initiative database retrieved at january 3 , 2024 from http://adni.loni.usc.edu

Al Shehri W (2022) Alzheimer's disease diagnosis and classification using deep learning techniques. PeerJ Comput Sci 8:e1177. https://doi.org/10.7717/peerj-cs.1177

Amani, Keaun. "Understanding the Differences between AI, Machine Learning, and Deep Learning." *Neurosnap*, 1 Apr. 2023, https://neurosnap.ai/blog/post/64279cadfeb3e5ca5ba0904a. Accessed 24 May 2024.

B. Raj, A simple guide to the versions of the inception network, May 2018. [Online]. Available: https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202

Backès, Lucas. "Diagnosis of Neurodegenerative Diseases with Deep Learning: The Case of Alzheimer's Disease." Master's thesis, Faculty of Applied Sciences, University of Liège, 2021-2022. Promoters: Christophe Phillips, Gilles Louppe. URI: http://hdl.handle.net/2268.2/1598. Accessed May 16, 2024.

Balaji Petal (2023) Hybrid deep learning approach for detecting Alzheimer's disease. Biomedicines 11(1):149. https://doi.org/10.3390/biomedicines11010149

Bejnordi, Babak Ehteshami, Mitko Veta, Paul Johannes van Diest, et al. "Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer." *JAMA* 318, no. 22 (2017): 2199–2210. doi:10.1001/jama.2017.14585. PMCID: PMC5820737. PMID: 29234806. Published online December 12, 2017. Accessed May 16, 2024. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5820737/.

Brownlee, J. (2022). Softmax Activation Function with Python. Retrieved 8 June 2022, from https://machinelearningmastery.com/softmax-activation-function-with-python/

Ekuma, Godwin O. *An Explainable Deep Learning Prediction Model for Severity of Alzheimer's Disease From Brain Images.* MSU Graduate Theses, Summer 2023, Missouri State University, goe948s@missouristate.edu.

Huang, J., Malachowsky, C., & Priem, C. (2024). Convolutional Neural Network. NVIDIA. Retrieved april 27, 2024, from https://www.nvidia.com/en-us/glossary/convolutional-neural-network/

I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016. http: //www.deeplearningbook.org.

M. R. Hasan, "Deep Learning," School of Computing, University of Nebraska-Lincoln, 2021. Available: https://engineering.unl.edu/hasan/deep-learning/.

MahimSMetal(2024) Unlocking the potential of XAI for improved alzheimer's disease detection and classification using a ViT-GRU model. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3351809

National Institute on Aging. "Alzheimer's Disease Fact Sheet." National Institutes of Health, U.S. Department of Health and Human Services. Accessed May 16, 2024. https://www.nia.nih.gov/health/alzheimers-and-dementia/alzheimers-disease-fact-sheet.

National Institute on Aging. "What Are the Signs of Alzheimer's Disease?" National Institutes of Health, U.S. Department of Health and Human Services. Accessed May 16, 2024. https://www.nia.nih.gov/health/alzheimers-symptoms-and-diagnosis/what-are-signs-alzheimers-disease.

OASIS : Open Access Series Of Imaging Studies "OASIS database" Washington University in St. Louis. (2024). *Open Access Series of Imaging Studies (OASIS)*. Retrieved at january 2, 2024 from https://sites.wustl.edu/oasisbrains/

Odusami M et al (2021) Analysis of features of Alzheimer's disease: Detection of early stage from function albrain changes in magnetic resonanc eimages using a fine tuned ResNet18network.Diagnostics 11(6):1071. https://doi.org/10.3390/diagnostics11061071

Open Access Series Of Imaging Studies "OASIS database" http://adni.loni.usc.edu Washington University in St. Louis. (2024). *Open Access Series of Imaging Studies (OASIS)*. Retrieved at january 2, 2024 from https://sites.wustl.edu/oasisbrains/

Powers, D.M.W. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv 2020, arXiv:2010.16061.
PyTorch. (n.d.). torchvision.models.efficientnet_b3. Retrieved may 1, 2024, fromhttps://pytorch.org/vision/main/models/generated/torchvision.models.efficientnet_b3.html

Raza N et al (2023) Alzheimer disease classification through transfer learning approach. Diagnostics 13(4):801. https://doi.org/10.3390/diagnostics13040801

Tan, M., & Le, Q. V. (2020). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv preprint arXiv:1905.11946v5.Inception v3 — Netscope CNN Analyzer (dgschwend.github.io)

The Alzheimer's Disease Neuroimaging Initiative "ADNI " database retrieved at January 3 , 2024 from http://adni.loni.usc.edu

Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2021). Dive into Deep Learning. *arXiv preprint arXiv:2106.11342*.

Appendixes

```
Model: "cnn_01"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
 conv2d_2 (Conv2D)           (None, 200, 200, 16)      160

 conv2d_3 (Conv2D)           (None, 200, 200, 16)      2320

 Max_pooling2d_5 (MaxPoolin  (None, 100, 100, 16)      0
 g2D)

 sequential_5 (Sequential)   (None, 50, 50, 32)        2160

 sequential_6 (Sequential)   (None, 25, 25, 64)        7392

 sequential_7 (Sequential)   (None, 12, 12, 128)       27072

 dropout_2 (Dropout)         (None, 12, 12, 128)       0

 sequential_8 (Sequential)   (None, 6, 6, 256)         103296

 dropout_3 (Dropout)         (None, 6, 6, 256)         0

 flatten_1 (Flatten)         (None, 9216)              0

 sequential_9 (Sequential)   (None, 512)               4721152

 sequential_10 (Sequential)  (None, 128)               66176

 sequential_11 (Sequential)  (None, 64)                8512

 dense_3 (Dense)             (None, 4)                 260

=================================================================
Total params: 4938500 (18.84 MB)
Trainable params: 4936132 (18.83 MB)
Non-trainable params: 2368 (9.25 KB)
_________________________________________________________________
```

```
Model: "cnn_02"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d_4 (Conv2D)           (None, 200, 200, 32)      160

 conv2d_5 (Conv2D)           (None, 200, 200, 32)      4128

 batch_normalization_11                              (None,200,200,32)
128

 max_pooling2d_10 (MaxPooli  (None, 100, 100, 32)      0
 ng2D)

 dropout_7 (Dropout)         (None, 100, 100, 32)      0

 conv2d_6 (Conv2D)           (None, 100, 100, 64)      8256

 conv2d_7 (Conv2D)           (None, 100, 100, 64)      16448

 Batch_normalization                                 (None,100,100,64)
256

 Max_pooling2d_11 (MaxPooli  (None, 50, 50, 64)        0
 ng2D)

 dropout_8 (Dropout)         (None, 50, 50, 64)        0

 flatten_2 (Flatten)         (None, 160000)            0

 dense_4 (Dense)             (None, 512)               81920512

 dropout_9 (Dropout)         (None, 512)               0

 dense_5 (Dense)             (None, 4)                 2052

=================================================================
Total params: 81951940 (312.62 MB)
Trainable params: 81951748 (312.62 MB)
Non-trainable params: 192 (768.00 Byte)
_________________________________________________________________
None
```

```
Model: "cnn_03"
_________________________________________________________________
 Layer (type)                 Output Shape              Param #
=================================================================
 conv2d (Conv2D)              (None, 200, 200, 16)      448

 conv2d_1 (Conv2D)            (None, 200, 200, 16)      2320

 max_pooling2d(MaxPooling2D)  (None, 100, 100, 16)
0

 sequential (Sequential)      (None, 50, 50, 32)        14016

 sequential_1 (Sequential)    (None, 25, 25, 64)        55680

 sequential_2 (Sequential)    (None, 12, 12, 128)       221952

 dropout (Dropout)            (None, 12, 12, 128)       0

 sequential_3 (Sequential)    (None, 6, 6, 256)         886272

 dropout_1 (Dropout)          (None, 6, 6, 256)         0

 flatten (Flatten)            (None, 9216)              0

 sequential_4 (Sequential)    (None, 512)               4721152

 sequential_5 (Sequential)    (None, 128)               66176

 sequential_6 (Sequential)    (None, 64)                8512

 dense_3 (Dense)              (None, 4)                 260

=================================================================
Total params: 5,976,788
Trainable params: 5,974,420
Non-trainable params: 2,368
_________________________________________________________________
```

```
Model: "cnn_04"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 rescaling (Rescaling)       (None, 200, 200, 3)       0

 conv2d (Conv2D)             (None, 200, 200, 16)      448

 max_pooling2d (MaxPooling2D  (None, 100, 100, 16)     0
 )

 conv2d_1 (Conv2D)           (None, 100, 100, 32)      4640

 max_pooling2d_1 (MaxPooling  (None, 50, 50, 32)       0
 2D)

 conv2d_2 (Conv2D)           (None, 50, 50, 64)        18496

 max_pooling2d_2 (MaxPooling  (None, 25, 25, 64)       0
 2D)

 dropout (Dropout)           (None, 25, 25, 64)        0

 flatten (Flatten)           (None, 40000)             0

 dense (Dense)               (None, 128)               5120128

 dense_1 (Dense)             (None, 4)                 516

=================================================================
Total params: 5,144,228
Trainable params: 5,144,228
Non-trainable params: 0
_________________________________________________________________
```

```
Model: "inception_cnn_model"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 inception_v3 (Functional)   (None, 4, 4, 2048)        21802784

 dropout (Dropout)           (None, 4, 4, 2048)        0

 global_average_pooling2d (G  (None, 2048)             0
 lobalAveragePooling2D)

 flatten (Flatten)           (None, 2048)              0

 batch_normalization_94 (Bat  (None, 2048)             8192
 chNormalization)

 dense (Dense)               (None, 512)               1049088

 batch_normalization_95 (Bat  (None, 512)              2048
 chNormalization)

 dropout_1 (Dropout)         (None, 512)               0

 dense_1 (Dense)             (None, 256)               131328

 batch_normalization_96 (Bat  (None, 256)              1024
 chNormalization)

 dropout_2 (Dropout)         (None, 256)               0

 dense_2 (Dense)             (None, 128)               32896

 batch_normalization_97 (Bat  (None, 128)              512
 chNormalization)

 dropout_3 (Dropout)         (None, 128)               0

 dense_3 (Dense)             (None, 64)                8256

 dropout_4 (Dropout)         (None, 64)                0

 batch_normalization_98 (Bat  (None, 64)               256
 chNormalization)

 dense_4 (Dense)             (None, 4)                 260

=================================================================
Total params: 23,036,644
Trainable params: 1,227,844
Non-trainable params: 21,808,800
_________________________________________________________________
```

```
Model: "EfficientNetB3"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 efficientnetb3 (Functional)  (None, 1536)             10783535

 batch_normalization (BatchN  (None, 1536)             6144
 ormalization)

 dense (Dense)               (None, 256)               393472

 dropout (Dropout)           (None, 256)               0

 dense_1 (Dense)             (None, 4)                 1028

=================================================================
Total params: 11,184,179
Trainable params: 11,093,804
Non-trainable params: 90,375
_________________________________________________________________
```