

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Deep Learning Methods to Detect Alzheimer's Disease from MRI

Mariana Filipa Macedo Coelho

DISSERTAÇÃO

U. PORTO

FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Mestrado em Engenharia Biomédica

Supervisor: João Manuel R. S. Tavares

July 18, 2023

Abstract

Alzheimer's disease (AD) is a progressive and irreversible neurodegenerative condition in the brain that affects memory, thinking, and behaviour. To overcome this problem, which according to the World Health Organisation is on the rise, the creation of strategies is essential to identify and predict the disease in its early stages before clinical manifestation. In addition to cognitive and mental tests, neuroimaging plays a promising role in this field, especially in the assessment of brain matter loss. Therefore, computer-aided diagnosis (CAD) systems have been imposed as fundamental tools to help imaging technicians, as the diagnosis becomes less subjective and time consuming.

Thus, machine learning and especially deep learning techniques have come into play. In recent years, articles addressing the topic of Alzheimer's diagnosis by means of deep learning models have become increasingly popular, with an exponential increase from year to year, with increasingly higher accuracy values. However, the disease classification remains a challenging and progressing issue, not only in distinguishing between healthy controls and AD patients but mainly in the differentiation of intermediate stages such as Mild Cognitive Impairment (MCI). Therefore, there is a need to develop more valuable and innovative techniques. This document presents an overview of this pathology, the selected imaging method (Magnetic Resonance Imaging - MRI), and the theory behind the deep learning algorithms used in this area. Regarding the subject of this study, it is evident that this is a current issue, with numerous papers proposing different approaches to reach an accurate classification. Hence, this document also presents an up-to-date systematic review of deep models to detect AD and its intermediate phase by evaluating MR images.

Finally, the development of a deep learning-based transfer learning methodology to diagnose Alzheimer's disease, performing the following classifications, is also added: binary classification (CN vs. AD; CN vs. EMCI; EMCI vs. LMCI; LMCI vs. AD) and multiclassification (CN vs. EMCI vs. LMCI vs. AD), where CN stands for cognitively normal, EMCI for early MCI, and for LMCI late MCI. To reach the best final model, several intermediate tests were performed, namely different data augmentation techniques, different pre-trained architectures, different numbers of unfrozen layers, and different brain regions under study. So, the proposed model is a fine-tuned 80-layer SEResNet152 with the Elastic Transformation data augmentation technique, whose inputs are skull-stripped MR images. The effectiveness of the presented tests is evaluated through the ADNI dataset with multiple metrics. The results show promising performance, having reached a value of 92.98% in terms of accuracy in the CN vs. AD classification.

Keywords: Alzheimer's disease, 3D MRI brain image, Deep Learning, CAD, Transfer Learning

Resumo

A doença de Alzheimer é uma condição neurodegenerativa progressiva e irreversível do cérebro que afeta a memória, o pensamento e o comportamento. Para ultrapassar este problema, que, segundo a Organização Mundial de Saúde, está a aumentar, é essencial criar estratégias para identificar e prever a doença nas suas fases iniciais, antes da manifestação clínica. Para além dos testes cognitivos e mentais, a neuroimagem desempenha um papel promissor neste domínio, nomeadamente na avaliação da perda de matéria cerebral. Por conseguinte, os sistemas de diagnóstico assistido por computador impuseram-se como ferramentas fundamentais para ajudar os técnicos de imagiologia, uma vez que o diagnóstico se torna menos subjectivo e demorado.

Assim, as técnicas *Machine Learning* e especialmente de *Deep Learning* entraram em ação. Nos últimos anos, os artigos que abordam o tema do diagnóstico da doença de Alzheimer através de modelos de *Deep Learning* tornaram-se cada vez mais populares, com um aumento exponencial de ano para ano, com resultados de precisão cada vez mais elevados. No entanto, a classificação da doença continua a ser um desafio e uma questão progressiva, não só na distinção entre controlos saudáveis e doentes de Alzheimer, mas principalmente na diferenciação de fases intermédias, como *Mild Cognitive Impairment*. Por conseguinte, é necessário desenvolver técnicas mais valiosas e inovadoras. Este documento apresenta uma panorâmica geral desta patologia, o método de imagiologia selecionado (Ressonância Magnética), e a teoria subjacente aos algoritmos de *Deep Learning* utilizados neste domínio. Relativamente ao tema deste estudo, é evidente que se trata de uma questão atual, com inúmeros artigos que propõem diferentes abordagens para alcançar uma classificação precisa. Assim, este documento apresenta também uma revisão sistemática atualizada de modelos de *Deep Learning* para detetar a doença de Alzheimer e a sua fase intermédia através da avaliação de imagens de ressonância magnética.

Por fim, o desenvolvimento de uma metodologia de *transfer learning* baseada em *Deep Learning* para diagnosticar a doença de Alzheimer é também apresentada, realizando as seguintes classificações: classificação binária (CN vs. AD; CN vs. EMCI; EMCI vs. LMCI; LMCI vs. AD) e multiclassificação (CN vs. EMCI vs. LMCI vs. AD). Para alcançar o melhor modelo final, foram realizados vários testes intermédios, nomeadamente diferentes técnicas de *data augmentation*, diferentes arquitecturas pré-treinadas, diferentes números de *unfrozen layers* e diferentes regiões cerebrais em estudo. Assim, o modelo proposto é uma SEResNet152 de 80 camadas com a técnica *Elastic Transformation*, cujos dados de entrada são imagens de ressonância magnética do cérebro sem crânio. A eficiência é avaliada utilizando a base de dados ADNI com várias métricas. Os resultados mostram um desempenho promissor, tendo atingido um valor de 92,98% em termos de exactidão na classificação CN vs. AD.

Keywords: Doença de Alzheimer, Imagem cerebral por ressonância magnética 3D, *Deep Learning*, Sistemas de diagnóstico assistido por computador, *Transfer learning*

Acknowledgements

I would like to express my deepest gratitude to all the people who helped and supported me during the realization of this project and in some way enriched its entire content.

First, a special thanks to my supervisor, professor João Manuel R. S. Tavares from FEUP, for the help, the reviews, the suggestions, the time spent, and for accepting the orientation of my work. I would also like to thank professor Mark Cerny for his guidance in the preparation of this dissertation while I was on Erasmus in Czechia.

I also want to single out PhD student Selim Reza for the fundamental support and decisive guidance he provided me in the early stages of the process.

Last but not least, a reference to my family, who were an extremely important pillar, not only for their support but also for the financial contribution they provided me throughout my academic career.

Mariana Coelho

*“Predicting the future is not magic,
it is artificial intelligence.”*

Dave Waters

Contents

1	Introduction	1
1.1	Context	1
1.2	Motivation and Goals	2
1.3	Main Contributions	3
1.4	Document Structure	3
2	Background	5
2.1	Alzheimer's Disease	5
2.1.1	Epidemiology	5
2.1.2	Pathophysiology	6
2.1.3	Clinical Presentation	8
2.1.4	Risk Factors	9
2.1.5	Staging	10
2.1.6	Treatment	10
2.2	Magnetic Resonance Imaging	11
2.2.1	Physical principles	12
2.2.2	Hardware	14
2.2.3	Utility of Structural MRI in the Study of AD	16
2.3	Summary	17
3	Deep Learning	19
3.1	Definition	20
3.2	Types of Deep Learning Models	21
3.2.1	Convolutional Neural Networks	22
3.3	Hardware	26
3.4	Data	26
3.4.1	Overfitting	27
3.5	Transfer Learning	29
3.6	Performance Evaluation	32
3.7	Summary	34
4	Literature Review - Deep Learning Methods to Detect AD from MRI	35
4.1	Articles Selection Methodology	35
4.2	Review Results	36
4.2.1	Reviews	36
4.2.2	Used Datasets	36
4.2.3	Data Preprocessing	37
4.2.4	Data Analysis	39

4.2.5	Deep Learning based Approaches	40
4.2.6	Data Augmentation and Transfer Learning	43
4.3	Discussion	44
4.4	Summary	58
5	Methodology	59
5.1	Overview	59
5.2	Used Image Dataset	60
5.3	Image Preprocessing	60
5.3.1	Data Augmentation	61
5.4	Predictiton Model	62
5.5	Computational Infrastructure and Implemented Resources	67
5.5.1	Computational Infrastructure	67
5.5.2	Development language	67
5.5.3	Visual Studio Code and Google Colaboratory	68
5.5.4	TensorFlow and Keras	68
5.5.5	ROIEditor	68
5.6	Summary	68
6	Results and Discussion	71
6.1	Experimental Settings	71
6.2	Effects of Different Dataset Sizes	72
6.3	Effects of Different Data Augmentation Methods	72
6.4	Effects of Different Model Architectures	75
6.5	Effects of Different Numbers of Unfrozen Layers	78
6.6	Effects of Different Study Regions	80
6.7	Proposed Model	81
6.8	Comparison Results With State of the Art Methods	84
6.9	Summary	86
7	Conclusion and Future Work	87
7.1	Conclusion	87
7.2	Future Work	88
References		91

List of Figures

2.1	Physiological structure of the brain and neurons in (a) healthy brain and (b) brain with AD.	6
2.2	APP cleavage pathways.	7
2.3	NINCDS/ADRDA clinical criteria for AD.	8
2.4	Risk factors for Alzheimer's disease.	9
2.5	Examples of MRI images of the brain of an Alzheimer's patient and a healthy person taken from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.	11
2.6	Spin and the magnetic moment of hydrogen.	12
2.7	Hydrogen protons under the action of an external magnetic field.	13
2.8	Coordinate axes used in MRI and the magnetic moment vector (M) associated with the hydrogen proton.	13
2.9	Example of MRI sequence based on different relaxation times.	14
2.10	Diagram of the components of an MRI scanner.	15
2.11	Series of T1-weighted studies, in coronal view. These images were acquired 1 year apart and demonstrate the progressive atrophy in the hippocampus as the individual's clinical status worsens: t=0 - memory complaints, t=1y - MCI and t=2y - criteria corresponds to the diagnosis of AD.	17
3.1	Articles from the Scopus database that were published within the previous five years, according to Chapter 4.	19
3.2	Deep representations a digit-classification model has learned.	20
3.3	Training loop of deep learning model.	21
3.4	An example of convolution operation with a kernel size of 3×3.	23
3.5	Activation functions: (a) ReLU, (b) sigmoid, and (c) hyperbolic tangent (tanh). .	23
3.6	An example of max pooling operation with a filter size of 2×2.	24
3.7	Gradient descent approach.	25
3.8	Underfitted vs. Fitted vs. Overfitted model.	27
3.9	Signs of overfitting in the loss and accuracy curves.	28
3.10	Transfer learning strategies.	31
3.11	a- Size-Similarity matrix; b- Decision map for fine-tuning pre-trained models. .	32
3.12	Representation of a confusion matrix.	33
3.13	The AUC is indicated by the gray area below the ROC curve.	34
4.1	PRISMA diagram showing the performed literature search in the Scopus database.	35
4.2	Prevalence of the used datasets in the reviewed articles.	37
4.3	Prevalence of used data processing approaches.	40

5.1	Proposed CAD system for AD diagnosis. It is divided into 3 steps: definition of the image type - 3D MRI T1; pre-processing, which encompasses the 4 techniques represented plus the segmentation phase; classification process, which is characterized by the compilation of the pre-trained network plus the classification block. The final output is the stage of AD.	59
5.2	Five-level hierarchical structures of the brain.	61
5.3	Data preprocessing techniques. The transverse section is shown in the top row, followed by the sagittal section and the coronal section. Each column represents the image after one processing step.	62
5.4	Data augmentation methods. The transverse section is shown in the top row, followed by the sagittal section and the coronal section. Each column represents an image that has undergone a type of data transformation.	63
5.5	Architecture of 3D DenseNet (Convolutional base). Input dimensions correspond to depth × height × width × channels, and a,b,c,d defines the architecture variant.	64
5.6	Architecture of 3D ResNet (Convolutional base). Input dimensions correspond to depth × height × width × channels, and a,b,c,d defines the architecture variant. . .	64
5.7	Architecture of 3D ResNeXt (Convolutional base). Input dimensions correspond to depth × height × width × channels, and a,b,c,d and C defines the architecture variant.	65
5.8	Architecture of 3D SEResNet (Convolutional base). Input dimensions correspond to depth × height × width × channels, and a,b,c,d defines the architecture variant.	66
5.9	Classification block.	66
5.10	Architecture of 3D SENet154 (Convolutional base). Input dimensions correspond to depth × height × width × channels.	67
5.11	ROIEditor interface.	69
6.1	Performance comparison of the two dataset sizes, namely the Confusion Matrix, the ROC curve, and the radar chart with the metrics Precision, Recall, F1-score, Sensitivity and Specificity for each class (CN - label 0; AD - label 1).	73
6.2	Loss and Accuracy curves for training and validation for the two dataset sizes. . .	73
6.3	Radar chart with the metrics Precision, Recall, F1-score, Sensitivity and Specificity for each class (CN - label 0; AD - label 1) for each data augmentation technique implemented.	74
6.4	Loss and Accuracy curves for training and validation for each data augmentation technique implemented.	75
6.5	Performance comparison of the DenseNet variants, namely the Confusion Matrix, the ROC curve, Loss and Accuracy curves for training and validation, and the radar chart with the metrics Precision, Recall, F1-score, Sensitivity and Specificity for each class (CN - label 0; AD - label 1).	76
6.6	Performance comparison between ResNet152 and ResNeXt101, namely the Confusion Matrix, the ROC curve, Loss and Accuracy curves for training and validation, and the radar chart with the metrics Precision, Recall, F1-score, Sensitivity and Specificity for each class (CN - label 0; AD - label 1).	77
6.7	Performance of SENet154, namely the Confusion Matrix, the ROC curve, Loss and Accuracy curves for training and validation, and the radar chart with the metrics Precision, Recall, F1-score, Sensitivity and Specificity for each class (CN - label 0; AD - label 1).	77

6.8	Performance comparison of the SEResNet variants, namely the Confusion Matrix, the ROC curve, Loss and Accuracy curves for training and validation, and the radar chart with the metrics Precision, Recall, F1-score, Sensitivity and Specificity for each class (CN - label 0; AD - label 1)	78
6.9	Confusion Matrix and radar chart with the metrics Precision, Recall, F1-score, Sensitivity and Specificity for each class (CN - label 0; AD - label 1) for different number of unfrozen layers.	79
6.10	ROC curves of the different model's performance with different number of unfrozen layers.	80
6.11	Confusion Matrix, radar chart with the metrics Precision, Recall, F1-score, Sensitivity and Specificity for each class (CN - label 0; AD - label 1), and Loss and Accuracy curves for training and validation for the regions Telencephalon and Hippocampus, respectively.	81
6.12	Grad-CAM visualizations of the features learned for each region, namely Skull-stripped, Telencephalon, and Hippocampus.	81
6.13	Loss and Accuracy curves for training and validation for the different classification tasks.	82
6.14	Performance comparison of the different classification tasks, namely the Confusion Matrix, the ROC curve, and the radar chart with the metrics Precision, Recall, F1-score, Sensitivity and Specificity for each class.	83

List of Tables

3.1	Different deep learning models that have been used to detect AD.	21
3.3	Choosing the right last-layer activation and loss function for a certain type of model.	25
3.5	Parameters and hyperparameters associated with each type of building block in a CNN.	26
3.7	Performance Evaluation Metrics.	33
4.1	Summary of the reviews found in the current study.	36
4.3	Summary of the approaches found in this review.	45
4.5	Summary of the best approaches found in this reviewed.	57
5.1	Summary of the studied subjects.	60
6.1	Hyperparameters adopted in the experiments.	72
6.3	Performance of different dataset sizes.	72
6.5	Performance of different data augmentation methods.	74
6.7	Performance of different model architectures.	76
6.9	Performance with different numbers of unfrozen layers.	78
6.11	Performance of different ROIs.	80
6.13	Performance of the proposed model for the different classification tasks.	82
6.15	Comparison results with state of art methods.	85

Abbreviations

ACC	Accuracy
AD	Alzheimer's disease
ADNI	Alzheimer's Disease Neuroimaging Initiative
ADRDA	Alzheimer's Disease and Related Disorders Association
AE	Autoencoding
AI	Artificial Intelligence
AIBL	Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing
APP	Transmembrane amyloid precursor
AUC	Area under the curve
$\text{A}\beta$	Amyloid plaques
BGRU	Bidirectional Gated Recurrent Unit
CAD	Computer-aided diagnosis
CN	Cognitively normal
CNN	Convolutional Neural Network
CPU	Central Process Unit
CSF	Cerebrospinal fluid
dA	Deep Autoencoder
DBM	Deep Boltzmann machine
DBN	Deep Belief Network
DCELM	Deep Conventional Extreme Learning Machines
DL	Deep Learning
DNN	Deep Neural Network
EMCI	Early Mild Cognitive Impairment
FDG	Fluorodeoxyglucose
fMRI	Functional MRI
FN	False Negatives
FP	False Positives
GM	Grey matter
GPU	Graphics Processing Unit
LMCI	Late Mild Cognitive Impairment
MCI	Mild cognitive impairment
MRI	Magnetic resonance imaging
NFT	Neurofibrillary Tangles
NINCDS	National Institute of Neurological and Communicative Disorders and Stroke
NMDA	N-methyl d-aspartate
OASIS	Open Access Imaging Studies Series
PET	Positron emission tomography
PSA	Pyramid Squeeze Attention
ReLU	Rectified linear unit
RF	Radio frequency
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
ROI	Region of interest
SA	Self-attention
SE	Squeeze and Excitation
SNR	Signal-to-noise ratio
SP	Senile Plaques
SVM	Support vector machine
TE	Time to Echo
TECD	Texture, Edge, Colour and density
TN	True Negatives
TP	True positives
TR	Repetition time
WM	White matter

Chapter 1

Introduction

This chapter presents a short summary of the dissertation. It begins with a brief contextualisation of the disease and respective problem, followed by the proposed objectives and, finally, the presentation of the structure of this report as well as a short summary of what is addressed in each chapter.

1.1 Context

One of the most common causes of dementia in the world today is Alzheimer's disease, which can be defined as a progressive and irreversible neurodegenerative disease characterised by abnormal deposition of neurofibrillary tangles and amyloid plaques in the brain, causing issues with memory, thinking, and behaviour [50]. According to the World Alzheimer Report (2018), the illness impacted around 50 million individuals in 2018, which is expected to triple by 2050 [82]. Currently, there is no treatment that can cure a patient who already has AD, but there are drugs and methods to slow down the progression of the disease [50]. The brain suffers structural and functional changes because of AD. Alzheimer's disease usually manifests its symptoms after the age of 60. Yet, some forms of AD develop relatively early (30–50 years) in people with a genetic mutation. As a result, developing techniques to identify AD before clinical manifestation is critical for timely treatment and slowing progression [50]. Patients first have mild cognitive impairment, which progresses to illness over time. Not all individuals with MCI convert into Alzheimer's patients, though. Therefore, it is necessary to understand the progressive alterations that take place in the brain when AD develops. It is also urgent to find treatments and/or solutions to slow down, stop, reduce the risk of, or completely prevent the onset of this disease [82].

In recent years, several computer-aided diagnosis systems have been created to aid in disease diagnosis and follow up. Between 1970 and 1990, the first rule-based models were created, and subsequently, supervised models were designed. To create these supervised models, it was necessary to extract pertinent features from the input data, which required the intervention of human experts being a tedious and time-consuming task. Nevertheless, with the emergence of deep learning models, it became possible to draw out the features used directly from the data without

recourse to human interaction. Therefore, researchers have focused their studies on creating deep learning models to accurately diagnose several diseases, including AD [78].

Mainly as to AD, neuroimaging has proved to be extremely important in identifying early changes in brain tissue that may be related to the disease. A wide range of brain imaging measurements continues to be developed for scientific research and clinical evaluation of Alzheimer's disease. Structural magnetic resonance imaging (MRI) is one of the best established procedures for early identification and monitoring of AD [75].

1.2 Motivation and Goals

Global demographic ageing is a phenomenon that highlights the development of better healthcare over the past century. There are more senior individuals in the world's population as a result of people living longer and healthier lives. Although there is increasing awareness of cases that begin before the age of 65, dementia primarily affects older people. Today, dementia ranks as the seventh most common cause of death worldwide and one of the diseases with the highest societal costs [6].

According to the World Health Organization, there are numerous varieties of dementia, with Alzheimer's disease accounting for between 60 and 70 percent of cases. Every three seconds, someone worldwide develops dementia. In 2020, 55 million people were living with dementia worldwide. By the year 2030, there will be 78 million people, and by the year 2050, there will be 139 million people [6]. All these numbers collected by Alzheimer's Disease International are worrisome, and therefore the disease is receiving increased attention in the medical community and is the subject of an explosion in research [6].

In the study of the disease, diagnosis remains a major challenge, with people seeking a diagnosis often experiencing long waiting periods, if they are able to receive one at all [6]. According to the World Alzheimer Report (2011), early detection and intervention are crucial strategies for closing the treatment gap. It is absolutely untrue that "early diagnosis is pointless" or "nothing can be done". If these therapies are implemented earlier in the disease's progression, they may be more successful. According to research, the majority of people who are now suffering from dementia have not been officially diagnosed. Only 20 to 50 percent of dementia cases in high-income countries are identified and recorded in primary care. In low- and middle-income countries, the "treatment gap" is undoubtedly considerably larger, as evidenced by a study in India, where 90 percent of cases are undiagnosed. If these figures are generalised to other nations, it is likely that almost 75 percent of those who have dementia have not been officially diagnosed and are consequently unable to obtain the care, support, and organised services that a formal diagnosis can provide [6].

Therefore, the goal of this project is to create an automated process that uses artificial intelligence to identify brain lesions brought on by AD. Prior to that, it will be necessary to treat and analyse the MR images in order to do the classification.

1.3 Main Contributions

There are two main goals for the contributions that this project has made. On the one hand, a theoretical goal, where a thorough background analysis of the literature surrounding the addressed issue was completed. On the other hand, it includes all the experiential tasks carried out from a practical standpoint. In this approach, it can be seen that the main advancements made with this effort are primarily related to:

- The contextualization of the stages and the methodology applied in the diagnosis of Alzheimer's disease;
- The study and application of different data augmentation techniques, pre-trained models, and regions of interest;
- The development of a framework to achieve binary classification (CN vs. AD; CN vs. EMCI; EMCI vs. LMCI; LMCI vs. AD) and multiclassification (CN vs. EMCI vs. LMCI vs. AD);
- The submission of a systematic review article about the application of deep learning methods to detect the disease from MRI to a Wiley Journal.

1.4 Document Structure

In addition to the Introduction chapter, this document is divided into seven separate chapters and a list of bibliographical references.

Chapter 2 explores the disease background, which defines the basic concepts of Alzheimer's disease, the epidemiology, the pathophysiology and clinical characteristics, followed by the risk factors and the disease's treatment. Also in this chapter, a brief description of magnetic resonance imaging is given, covering everything from the underlying physical concepts to the hardware. In addition, this chapter concludes with the usefulness of structural MRI in the study of Alzheimer's disease. In Chapter 3, the area of deep learning is explored, including a brief characterization, typology, and required hardware. The topics of data and transfer learning are also covered, as are methods for evaluating the performance of a model.

In Chapter 4, a systematic review of deep learning applications in Alzheimer's detection is presented. In this way, the literature review performed is divided into five points: used datasets; data preprocessing; data analysis; deep learning based approaches; and data augmentation and transfer learning.

The subsequent Chapter 5 starts the practical strategy from a theoretical standpoint, providing an overview of the adopted process, a description of the used image dataset, image preprocessing, and the prediction model selected for this project's development, and a exhibition of the computational infrastructure and some managed software resources. The experimental work and results are

described in Chapter 6, including the different studies performed and an analysis of the obtained results.

To conclude, the final considerations and future work perspectives are given in Chapter 7.

Chapter 2

Background

The most important subject of this dissertation, Alzheimer's disease, was briefly discussed in the current chapter. Therefore, in addition to a brief description of this neurological condition, the epidemiology, pathophysiology, clinical presentation, risk factors, phases, and treatment are presented. With a clearer understanding of the disease, the adopted imaging technique, magnetic resonance imaging, is presented. In this topic, a description of the technique, physical principles, and hardware is given. It is also described how MRI can be useful in the detection of AD, thus making a correlation between the two topics. Finally, a summary of the information revealed in this chapter is provided.

2.1 Alzheimer's Disease

The concept of dementia can be described as a decline in cognitive activity that causes disruption to a person's daily life [10]. One of the most common causes of dementia in the world today is Alzheimer's disease, which is a progressive and irreversible neurodegenerative disease causing problems with memory, thinking, and behaviour, and is characterised by an abnormal accumulation of amyloid plaques and neurofibrillary tangles in the brain, as represented in Figure 2.1 [50]. The first case of this pathology was reported in 1907, when Alois Alzheimer noted, in a 51-year-old woman, the presence of amyloid plaques and a massive loss of neurons during an analysis of her brain. This patient presented with memory loss and psychiatric disorders, which were described as severe disease of the cerebral cortex [25, 23].

2.1.1 Epidemiology

70% of all existing dementia cases are caused by Alzheimer's disease and, according to the World Alzheimer Report (2018), the illness impacted around 50 million individuals in 2018, which is expected to triple by 2050 [25, 82]. AD is a typical disease of ageing, so its incidence increases with age, doubling every five to ten years. Studies indicate that people aged 65-69 have an estimated

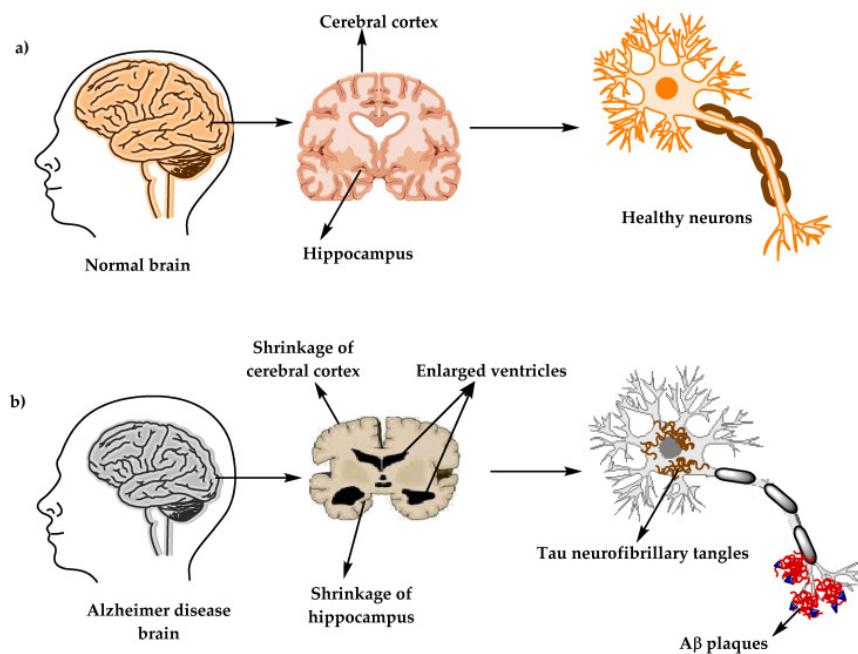


Figure 2.1: Physiological structure of the brain and neurons in (a) healthy brain and (b) brain with AD [23].

incidence of 0.6%, 70-74 of 1.0%, 75-79 of 2.0%, 80-84 of 3.3%, and over 85 of 8.4%. Apart from age, which is the most prominent risk factor, there are others such as family history, genetics, being female, cranial trauma, other diseases (such as cardiovascular diseases), and environmental factors that also contribute to the development of this condition [25]. In addition to the health consequences of AD, it also affects individuals, their families, and the economy at a monetary level, with estimated global costs of US\$1 trillion annually [23].

2.1.2 Pathophysiology

The pathology of Alzheimer's disease can be described by two types of changes that provide information about its progress and symptoms. Positive lesions are defined by the accumulation of amyloid plaques ($A\beta$), neurofibrillary tangles containing tau, neuropil threads, dystrophic neurites, and other deposits found in the brain. Or negative lesions that are characterised by a loss of synaptic homeostasis, neurons, or neuronal network integrity. There are still other factors, such as neuroinflammation or oxidative stress, that can cause neurodegeneration [25, 23].

- **Senile Plaques (SP)**

Amyloid plaques are deposits of extracellular fragments of beta-amyloid protein that can take different morphological forms such as neuritic, diffuse, dense-coloured plaques, or the classic, compact type. The beta-amyloid peptide comes from a membrane protein called the transmembrane amyloid precursor (APP), which is present especially in neuronal synapse zones. The

biosynthesis of APP is carried out by the action of secretases (alpha, beta, and gamma), which cleave APP into small fragments [49, 23]. As represented in Figure 2.2, APP can be cleaved in two ways: (a) APP is cleaved by α -secretase, leading to the formation of APP α and α CTF, which in turn are cleaved by γ -secretase producing the extracellular peptide p3 and the intracellular AICD fragment; (b) APP is cleaved by β -secretase into APP β and β CTF, which are cleaved by γ -secretase giving rise to $A\beta$ and AICD. Generally, APP is synthesised by alpha and beta secretases producing small fragments that are not toxic to neurons. However, when the synthesis is performed by beta and gamma secretases, beta-amyloid peptides are formed. A high concentration of this compound causes aggregation of amyloids, which in turn leads to neuronal toxicity [49].

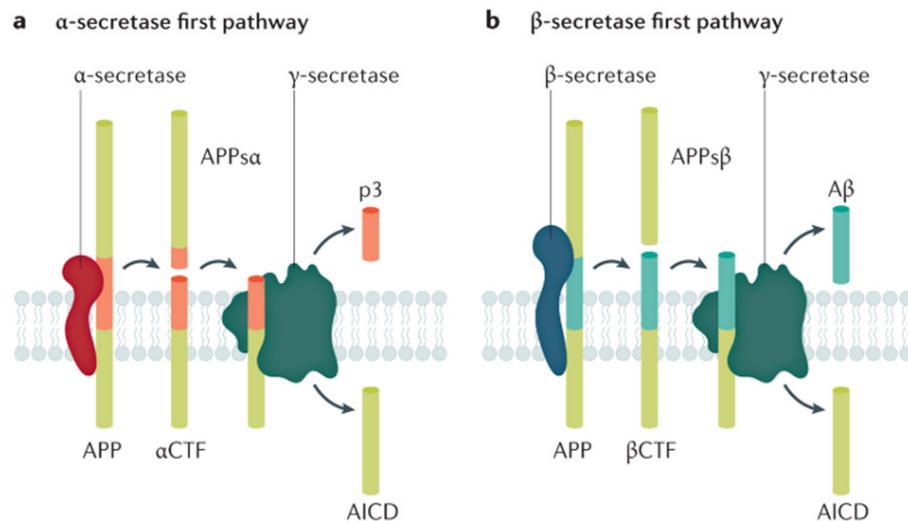


Figure 2.2: APP cleavage pathways [49].

- **Neurofibrillary Tangles (NFT)**

Neurofibrillary tangles are abnormal accumulations of a protein called tau that clump together inside neurons. The main function of the tau protein is to stabilise axon microtubules, which are internal structures in neurons that help transport nutrients and molecules from the cell body to the axon and dendrites. In Alzheimer's disease, given the aggregation of extracellular beta-amyloid that deploys abnormal chemical changes, the tau protein is hyperphosphorylated leading to tangle formation. These accumulations block transport between neurons, impairing synaptic communication. In AD, tangles are more strongly correlated with disease than plaques [10, 4].

- **Synaptic Loss**

Defects in axonal transport, oxidative stress, mitochondrial damage, accumulations of $A\beta$ and tau proteins, and other processes are mechanisms that result in synaptic loss. These processes translate into a loss of dendritic spines, presynaptic terminals, and axonal dystrophy, which affects normal communication between neurons [23].

2.1.3 Clinical Presentation

To diagnose a patient with AD, in addition to the analysis of the medical and family history, it is necessary to perform several tests, such as a neurological examination, magnetic resonance imaging for neurons, and laboratory tests (for example, vitamin B12) [23]. There are several types of tests that can be performed that allow conclusions to be drawn regarding the patient's condition. Normal laboratory tests do not reveal any difference, however a complete blood count, a complete metabolic panel, thyroid stimulating hormone, and B12 are used to exclude other pathologies. Computed tomography technique allows the detection of cerebral atrophy and enlargement of the third ventricle, which is suggestive but not specific for AD. Cerebrospinal fluid (CSF) analysis is useful in preclinical phase diagnosis when it detects low levels of beta-amyloid 42 and high levels of tau. EEG usually shows a generalised slowing with no focal features, which is also non-specific. Lately, volumetric MRI can be used to measure changes in brain volume. In AD, a retraction of the medial temporal lobe occurs, however these atrophies may be related to ageing itself. Other imaging techniques but in the functional component, such as positron emission tomography (PET), functional MRI (fMRI), and single-photon emission computed tomography, can be used to map patterns of dysfunction. These studies are useful for early detection and monitoring, however as a diagnostic method is not yet fully conceived. There are still several techniques that are under investigation [10]. Even after doing all these tests, it could still be impossible to definitively diagnose a particular type of dementia. Thus, given the subjectivity, criteria were developed to help physicians in their decision making [10].

In 1984, a group was formed by the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and the Alzheimer's Disease and Related Disorders Association (ADRDA) to establish criteria and describe the clinical diagnosis of Alzheimer's disease. The diagnosis requires clinical evidence of memory loss together with at least one disability in another of the cognitive domains, as shown in Figure 2.3 [25, 23].

NINCDS/ADRDA clinical criteria for AD.

-
- I. AD is characterized by progressive decline and ultimately loss of multiple cognitive functions, including both:
 - Memory impairment--impaired ability to learn new information or to recall previously learned information.
 - And at least one of the following:
 - Loss of word comprehension ability, for example, inability to respond to "Your daughter is on the phone." (aphasia);
 - Loss of ability to perform complex tasks involving muscle coordination, for example, bathing or dressing (apraxia);
 - Loss of ability to recognize and use familiar objects, for example, clothing (agnosia);
 - Loss of ability to plan, organize, and execute normal activities, for example, going shopping.

 - II. The problems in "I" represent a substantial decline from previous abilities, cause significant problems in everyday functioning, and begin slowly and gradually, becoming more severe.

 - III. The problems in "I" are not due to other conditions that cause progressive cognitive decline (e.g., stroke, Parkinson's disease, Huntington's chorea, brain tumor, etc.), and other conditions that cause dementia (e.g., hypothyroidism, HIV infection, syphilis, and deficiencies in niacin, vitamin B12, and folic acid), not caused by delirium, and not caused of another mental illness such as depression or schizophrenia.
-

Figure 2.3: NINCDS/ADRDA clinical criteria for AD [25].

27 years later (2011), the National Institute on Aging-Alzheimer's Association has updated the criteria, making some changes to the NINCDS-ADRDA assessment defined in 1984 with the aim of obtaining greater specificity and sensitivity in the diagnosis of AD. The new proposed criteria included the analysis of clinical biomarkers. These are divided into two categories: brain amyloid markers observed by PET or CSF; and markers of neuronal injury, such as the presence of cerebrospinal fluid tau, fluorodeoxyglucose (FDG) to analyse metabolic activity, and MRI to measure atrophy [23].

2.1.4 Risk Factors

AD is a multifactorial disease that is associated with multiple risk factors, as represented in Figure 2.4 [23].

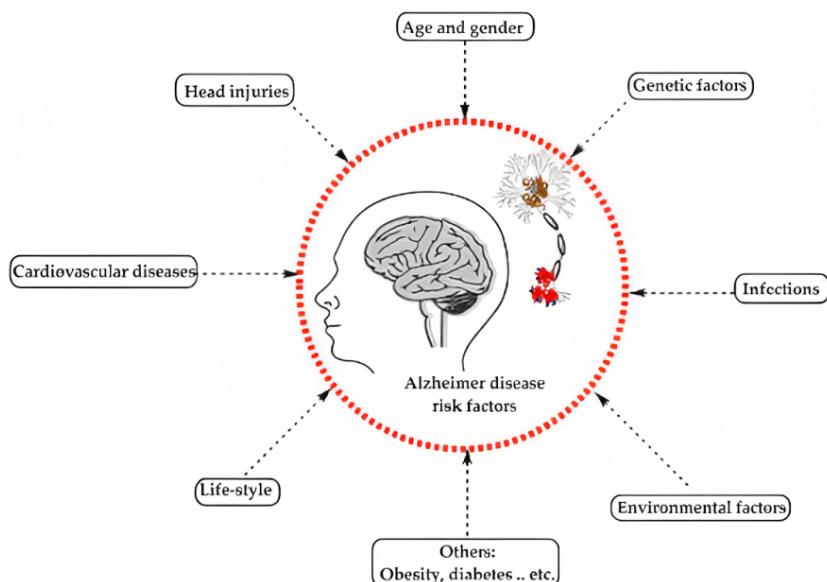


Figure 2.4: Risk factors for Alzheimer's disease [23].

- **Ageing**

The most prevalent risk factor for AD is ageing. This disease usually appears after 65 years of age, and cases in young people are very rare. Ageing is an irreversible and complex mechanism that occurs in multiple cellular systems causing, in the brain, weight and volume reduction, synaptic losses, and an increase in ventricles in specific areas accompanied by SP and NFT deposition [25].

- **Genetic Factors**

Genetic factors, which have been discovered and explored over the years, account for about 70 % of AD cases. Most are inherited by mutations occurring in autosomal dominant genes such as APP, presenilin-1, presenilin-2, and apolipoprotein E [25].

- **Environmental Factors**

Not all cases of AD can be explained by ageing and genetics. There are environmental conditions, such as air pollution, diet, exposure to metals, or infections that can induce the development of oxidative stress and inflammation that increase the risk of AD [25].

- **Medical Factors**

Older people are usually more prone to developing cardiovascular diseases, obesity, diabetes, among others, which also increase the risk of developing Alzheimer's [25].

2.1.5 Staging

Alzheimer's disease can be classified into three distinct clinical phases:

- **Preclinical or Presymptomatic**

This phase may last for several years and is characterised by early pathological changes in the cortex and hippocampus and a small loss of memory, however it does not present functional impairments in daily activities or clinical symptoms [23].

- **Mild Cognitive Impairment**

In this phase, the first symptoms of the disease appear, such as loss of memory and concentration, temporal and spatial disorientation, mood swings, among others. Each year, about 10% of patients with MCI progress to dementia [10, 23].

- **Dementia**

The dementia phase can be subdivided into two parts, considering its severity. In the moderate stage, the disease spreads to cerebral cortex areas resulting in increased memory loss, trouble recognising family and friends, communication difficulties (speaking, writing, and reading), and loss of impulse control. In the severe phase, the disease spreads throughout the cortex, with a severe accumulation of SP and NFT. As a result, there is a progressive functional and cognitive impairment where patients lose all their abilities, which can lead to their deaths [10, 23].

2.1.6 Treatment

Although AD is now a public health issue, there is no cure for it. There is only treatment for the symptoms, and only two drugs have been approved: inhibitors to cholinesterase enzyme (naturally derived, synthetic, and hybrid analogues) and antagonists of N-methyl d-aspartate (NMDA). Cholinesterase inhibitors increase the level of acetylcholine, which is a chemical that nerve cells use to communicate. This compound is important for cognitive functions, learning and memory. There are three drugs in this category approved by the Food and Drug Administration to treat

Alzheimer's: donepezil, rivastigmine, and galantamine. Partial NMDA antagonist memantine blocks NMDA receptors and slows down the intracellular accumulation of calcium. It is used to treat moderate and severe Alzheimer's disease and may be given concurrently with cholinesterase inhibitors [10, 23]. Although the two classes presented have therapeutic action, they act only in the treatment of symptoms, having no effect on the cure or prevention of the disease [10].

2.2 Magnetic Resonance Imaging

A wide range of brain imaging measurements for scientific research and clinical evaluation of AD continue to be developed. Magnetic resonance imaging is one of the most well-established procedures for detecting and tracking Alzheimer's disease [75]. Figure 2.5 represents the differences between a healthy patient and an Alzheimer's patient.

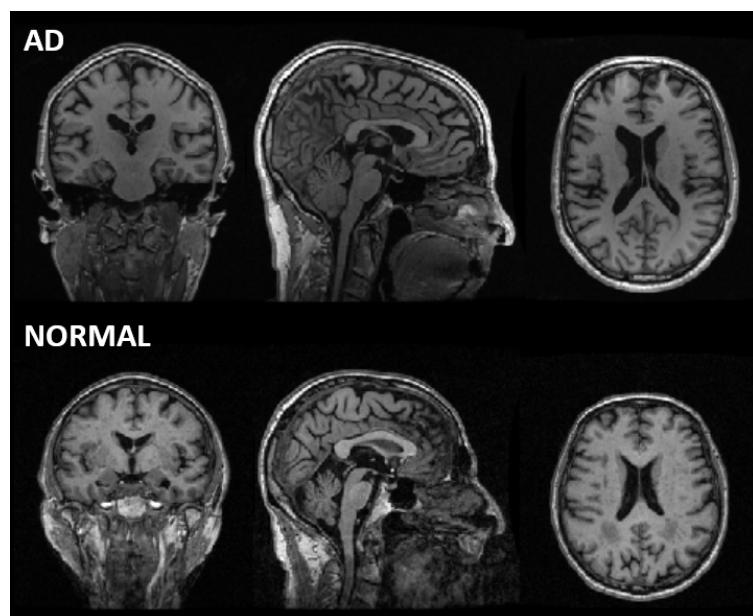


Figure 2.5: Examples of MRI images of the brain of an Alzheimer's patient and a healthy person taken from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [8].

MRI is a non-invasive imaging technique that generates accurate three-dimensional anatomical images, commonly used for disease detection, diagnosis, and therapy monitoring. This technology is based on a method that excites and detects the change in direction of the rotational axis of protons found in the water that constitutes living tissue [3]. In 1946, Bloch and Purcell, in independent articles in the Physics Review, were the first to describe the physical phenomenon of magnetic resonance, however, given the complexity of the technique only about 30 years later appeared the first images of the human body [62].

2.2.1 Physical principles

Powerful magnets are used in MRIs to produce a strong magnetic field that forces the body's protons to align with it. Thus, this technique uses the property of protons having a strong angular momentum that is polarised in a magnetic field. This means that a radiofrequency pulse can change the energy state of protons, and when the pulse is shut off, the protons will release a radiofrequency signal as they return to their original energy level. "Sequences" can be constructed to be sensitive to diverse tissue properties by combining different gradients and pulses [62]. Human tissue is mainly composed of these atoms: hydrogen, oxygen, carbon, phosphorus, calcium, fluorine, sodium, potassium, and nitrogen. Although both possess properties for use in MRI, hydrogen is chosen for three reasons [62]:

- it is the most abundant in the human body (about 10% of body weight);
- the MR characteristics are quite different between hydrogen present in normal tissue and in pathological tissue;
- the hydrogen proton has the highest magnetic moment and, therefore, the highest sensitivity to MR [62].

Protons are positively charged particles that have a property called spin or angular momentum. As represented in Figure 2.6, the hydrogen proton can be seen as a small sphere (1), which has a spin movement, or spin, around its own axis (2); because it is a positively charged particle (3), it will generate its own magnetic field around it (4), behaving like a small magnetic dipole (4) or like a magnet (5), with an associated magnetic moment (μ) [62].

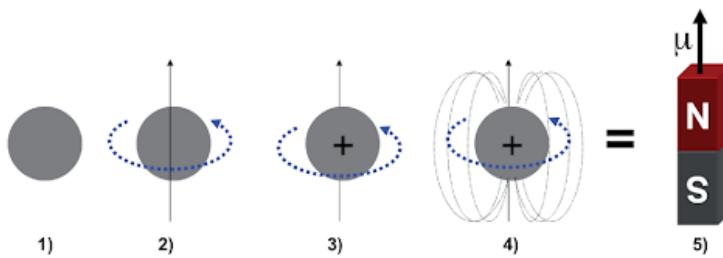


Figure 2.6: Spin and the magnetic moment of hydrogen [62].

Magnetic moments have no defined spatial orientation, being randomly distributed. This organisation results in a null total magnetization of the tissue volume. However, when the patient is subjected to an external magnetic field (B_0), the hydrogen protons will orient themselves in the direction of the applied field. There are two possible orientations: parallel to the field (low energy level) or anti-parallel (high energy level, excited state), as represented in Figure 2.7 [62].

In an attempt to align with the field and due to the existence of spin, a second movement called precession arises. Under the action of a magnetic field, hydrogen protons will precess at a frequency determined, w (Hz), by the Larmor equation (Equation 2.1) [62]:

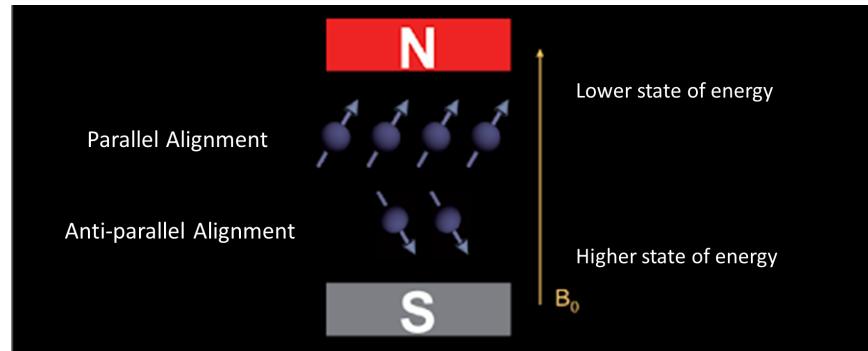


Figure 2.7: Hydrogen protons under the action of an external magnetic field [62].

$$\omega = \gamma B_0 \quad (2.1)$$

where, γ is gyromagnetic ratio and B_0 (T) is the value of the applied external magnetic field. Note that each magnetic field has its own specific precession frequency [62]. Figure 2.8 shows the coordinate axes (x , y , and z) and the vector representing the magnetic moment of a hydrogen proton performing precession motion around the z axis, as well as the same coordinates in a typical superconducting magnet. The z , or longitudinal, axis represents the direction of application of the main magnetic field (B_0). The xy plane is called the transverse B_0 plane [62].

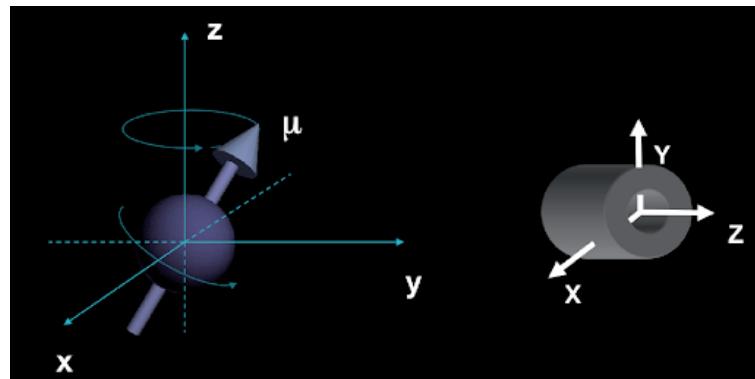


Figure 2.8: Coordinate axes used in MRI and the magnetic moment vector (M) associated with the hydrogen proton [62].

Although all the individual magnetic moments progress around B_0 at an angular frequency equal to ω , there is no phase coherence between them, and therefore there will be no magnetization component in the transverse plane. To reorient the magnetisation vector, a second magnetic field of short duration (pulse) must be applied. This B_1 field (radio frequency, or RF pulse) must be perpendicular to B_0 and be in phase with the precession frequency.

The effect on the magnetization vector (vector M) is to move it by a given offset angle (α), from alignment with B_0 . One of the most commonly used RF pulses is that which will result in a

90° angle of deviation [62]. With the application of a 90° RF pulse, for example, the magnetisation passes into the transverse plane inducing an electrical voltage in the coil with a frequency of w . When the RF pulse ceases, the signal gradually decays, which leads to the relaxation process or return of the magnetization vector to equilibrium, i.e., aligned with B_0 [62].

The amount of time between subsequent pulse sequences delivered to the same slice is known as repetition time (TR). The time between the delivery of the RF pulse and the reception of the echo signal is known as the Time to Echo (TE) [5]. Commonly, there are three types of MRI sequences based on different relaxation times, as indicated in Figure 2.9:

- T1-weighted images are produced by using short TE and TR times. The time constant T1 (longitudinal relaxation time) controls the rate at which stimulated protons return to equilibrium. It is the time it takes for spinning protons to realign themselves with the external magnetic field [5].
- T2-weighted images are produced by using longer TE and TR times. The rate at which stimulated protons reach equilibrium or diverge from one another is controlled by the time constant T2 (transverse relaxation time). It is the time it takes for spinning protons to lose phase coherence with nuclei spinning perpendicular to the main field [5].
- Fluid Attenuated Inversion Recovery (Flair) is similar to a T2-weighted image, but the TE and TR times are even longer [5].

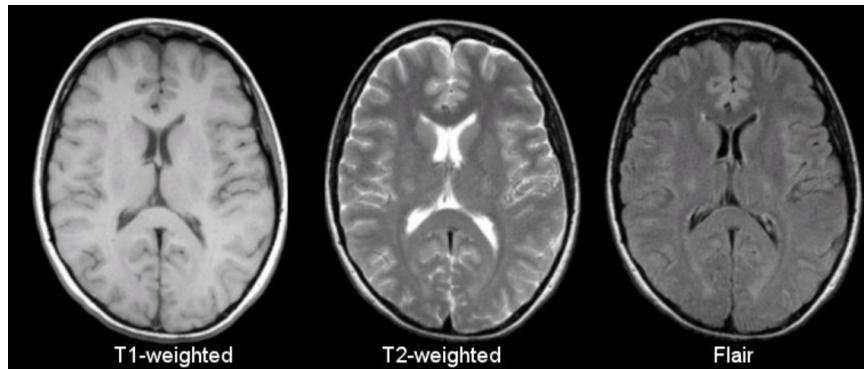


Figure 2.9: Example of MRI sequence based on different relaxation times [5].

2.2.2 Hardware

A typical MRI system consists of four sets of hardware components and accompanying parameters, as depicted in Figure 2.10 [21].

- **Primary Magnet**

The strength of the MR signal is mainly determined by the equilibrium magnetization, M_0 , which is proportional to the strength of the static magnetic field, B_0 . Hence, an extremely powerful

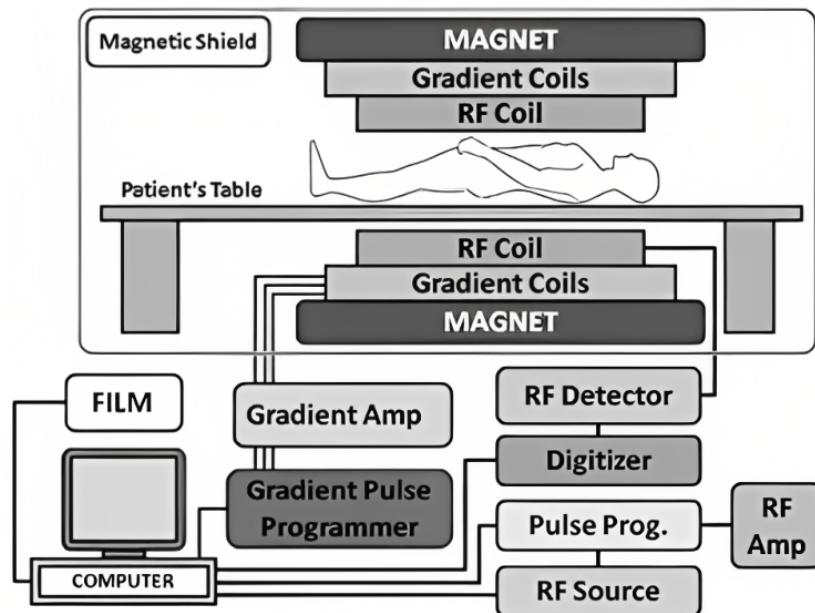


Figure 2.10: Diagram of the components of an MRI scanner [21].

magnetic field is required to achieve a sufficient signal-to-noise ratio (SNR). Gauss (G) or Tesla (T) units are used to express magnetic field intensity, with 1T being equivalent to 10,000 G. Gauss is the unit used in everyday life to measure magnetic fields, however as the fields used in scanners are 10000-100000 times stronger, they are specified in Tesla. The majority of clinical MR scanners in use today have 1.5T or 3.0T field strengths. In addition to being extremely powerful, the magnetic field of a MR system must be completely stable to prevent the Larmor frequency from drifting. In other words, it must ensure high signal uniformity and low image distortion. Thus, a solenoidal geometry - where the subject is positioned in the solenoid's centre, where the field is strongest and most uniform - is the most typical configuration [91].

- **Gradient System**

Maintaining homogeneity in the B_0 field is essential to minimise spin dephasing because any change can cause the spectral lines to broaden. Spatial variations in the magnetic field can also arise from susceptibility differences within the sample, so dynamic shim coils are used to adjust the field before scanning. Shim coils, which provide compensating magnetic fields, are incorporated into the MR system to account for linear and quadratic variations in the primary magnetic field. Moreover, gradient coils are utilised to encode spatial information into the signal and provide volume selectivity. Given three orthogonal directions (x, y, and z), three gradient coils are required in order to create magnetic fields that vary linearly along all three axes in terms of strength. Finer spatial resolution is made possible by higher gradient strengths, which are essential for imaging small animals and materials [91].

- **Radiofrequency System**

The MR system uses RF transmitter coils to create the $B_1(t)$ field, which uses RF energy to selectively excite nuclear spins. The RF system also includes RF receiver coils and data acquisition boards, which detect and process the signal emitted by the excited nuclei. The tissue volume of interest should be excited uniformly and with constant RF signal strength throughout the image. Surface coils and volume coils are the two types of RF coils. Surface coils, which are placed directly over the examination area and have a penetration depth about equal to the coil radius, are effective for picking up signals from relatively shallow locations of interest. Although their sensitivity range is limited, they offer significantly higher potential SNR than volume transmit coils. Volume coils encompass the tissue of interest and are designed to have a homogeneous sensitivity inside. To maximise SNR, the coil must be tailored to the size and shape of the anatomy because the signal is often captured from a small volume or tissue slice, while the noise is detected from all the tissue inside the coil's sensitivity range. Moreover, particular coils must be created for each field strength and nucleus being studied, such as hydrogen, since coils must resonate at the Larmor frequency and are only tunable within a constrained frequency range. To cover larger anatomical regions, several coil elements can be combined as phased-array surface coils, which optimise SNR while extending the region for signal reception. This sort of coil technology requires many RF receivers, which the coil elements can be attached to in batches or individually [91].

- **Computer System**

The scanner needs to be interfaced with a computer system in order to simplify the acquisition process and automate the reconstruction of images. The desired acquisition parameters, including the imaging plane, field of view, and spatial resolution, can be entered, and the results can be shown [91].

2.2.3 Utility of Structural MRI in the Study of AD

One of the features of neurodegeneration present in AD is progressive brain atrophy that can be observed by the MRI technique (with better results on T1-weighted volumetric sequences). The most prevalent factors causing this atrophy are dendritic and neuronal losses. AD is characterised by first manifesting in the medial temporal lobe. The earliest site of atrophy is the entorhinal cortex, followed by the hippocampus, amygdala, and parahippocampus, regions that constitute the limbic lobe [46].

Atrophy visible on MRI has been suggested as a biomarker of disease progression and a potential measure for clinical trials since this pathological condition manifests early (even before symptoms) and progresses continuously until individuals are severely affected, establishing a correlation with disease decline. The rate, distribution, and amount of brain atrophy are directly related to cognitive deficits. However, this technique has some disadvantages, namely the lack of molecular specificity. It does not allow the detection of histopathological markers characteristic of

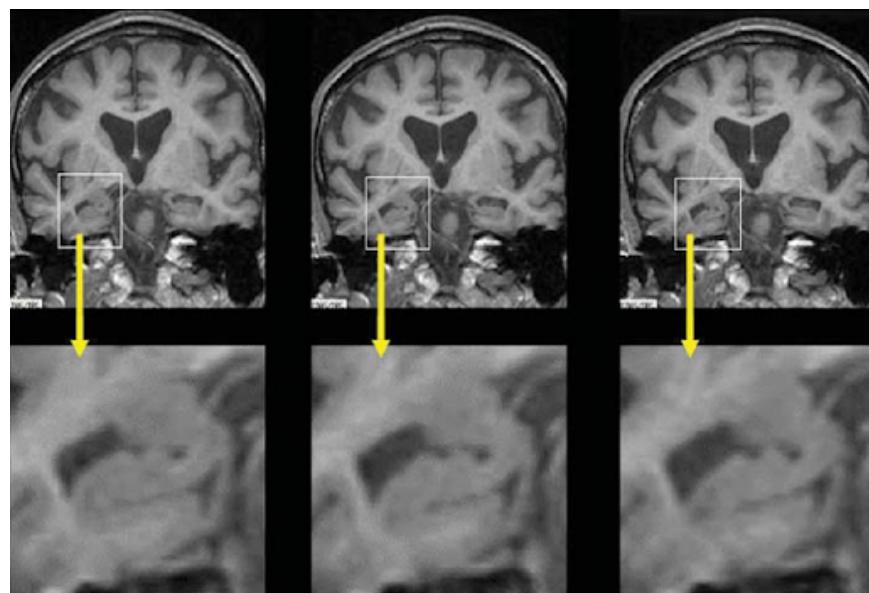


Figure 2.11: Series of T1-weighted studies, in coronal view. These images were acquired 1 year apart and demonstrate the progressive atrophy in the hippocampus as the individual's clinical status worsens: t=0 - memory complaints, t=1y - MCI and t=2y - criteria corresponds to the diagnosis of AD [46].

AD, such as amyloid plaques or neurofibrillary tangles. Furthermore, brain atrophy is not entirely specific neuronal damage [46].

2.3 Summary

The goal of this chapter was to make it easier to understand both the pathology and the medical imaging technique. Alzheimer's disease is a progressive neurodegenerative disorder marked by behaviour and cognitive impairment that eventually interferes with daily functional living activities. The pathology is incurable, and its rate of progression varies. Additionally, there are no specific laboratory or imaging tests to confirm the diagnosis of Alzheimer's disease in the early stages, making the disease difficult to diagnose. The medications available to treat the condition only work for the mild disease but also have a number of undesirable side effects. MRI is one of the most well-established procedures for detecting and tracking Alzheimer's disease. This non-invasive imaging technique generates accurate three-dimensional anatomical images that are frequently used for disease detection, diagnosis, and therapy monitoring.

Chapter 3

Deep Learning

Deep learning (DL) is a topic of great interest in many scientific fields, especially in medical image analysis. Currently, DL is the most effective machine learning (ML) method in many medical applications. By conducting a Scopus search, it is possible to determine that the use of CAD systems in AD diagnosis is proven due to the exponential increase in publications related to AD, medical imaging, and deep learning over the past five years, as represented in Figure 3.1.

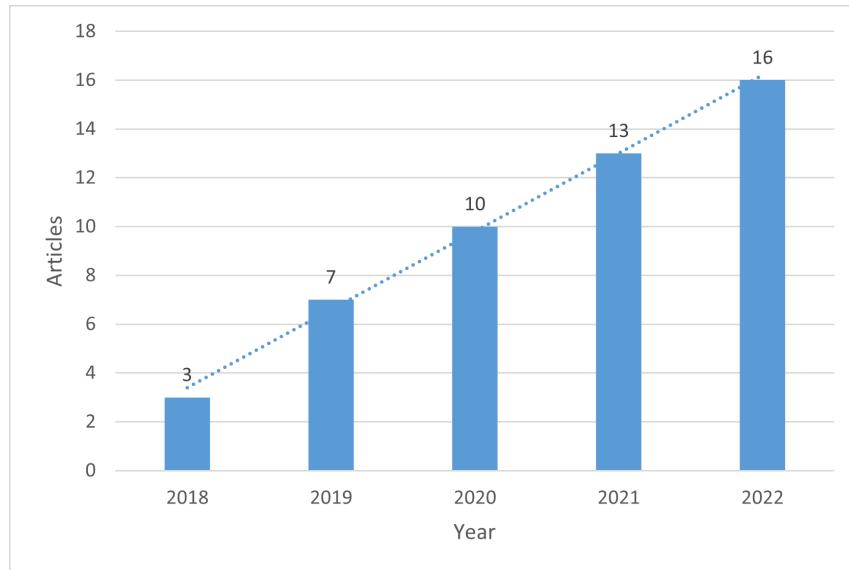


Figure 3.1: Articles from the Scopus database that were published within the previous five years, according to Chapter 4.

In this chapter, the pipeline of the deep learning approach is described in detail, along with the types of models that exist and the required hardware. The topics of data and transfer learning are also explored, as are metrics for evaluating the model's performance. At the end, a short summary of all the information described is presented.

3.1 Definition

Deep learning is a particular branch of machine learning that emphasises the learning of successive layers of increasingly meaningful representations. DL is a novel approach to learning representations from data. The deep in deep learning refers to successive layers of representations rather than any kind of deeper understanding that can be attained through the approach. The depth of the model refers to the number of layers that go into a data model. These layered representations in deep learning are (almost always) learned using neural network models, which are physically constructed as layers stacked on top of one another. Figure 3.2 shows how a multi-layered depth network transforms an image of a digit in order to recognise what digit it is [27].

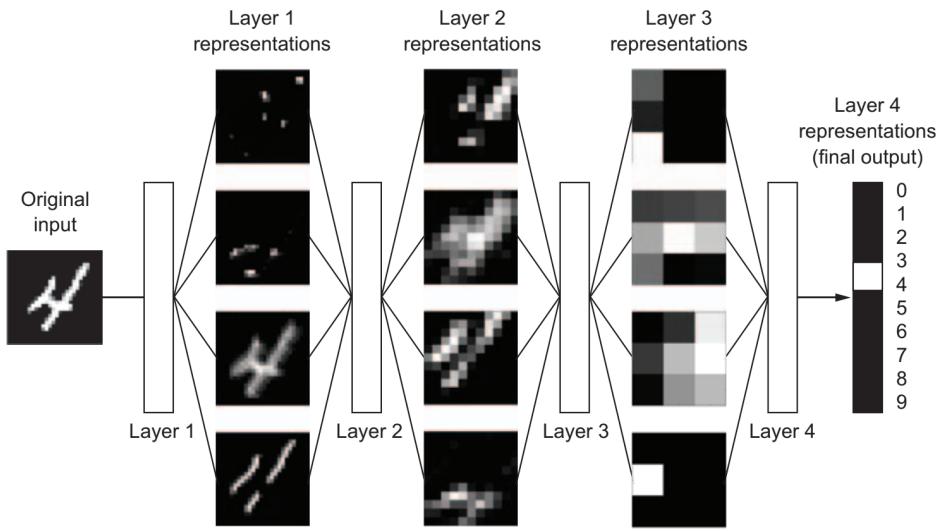


Figure 3.2: Deep representations a digit-classification model has learned [27].

A neural network is trained using the following items:

- The input information and associated targets;
- Layers that are incorporated into a network (or model);
- The loss function, which specifies the learning feedback signal;
- The optimizer, which controls the course of learning [27].

It is possible to picture how they interacted, as shown in Figure 3.3: the network converts the input data into predictions, via chained layers. In order to determine how closely the network's predictions match what was anticipated, the loss function compares these predictions to the targets and returns a loss value. This loss value is used by the optimizer to modify the weights of the network. This is the training loop, which, when carried out a sufficient number of times (normally tens of iterations over a large number of samples), produces weight values that minimise the loss function. A network with minimal loss is a network whose results are as close as possible to the targets: a trained net [27].

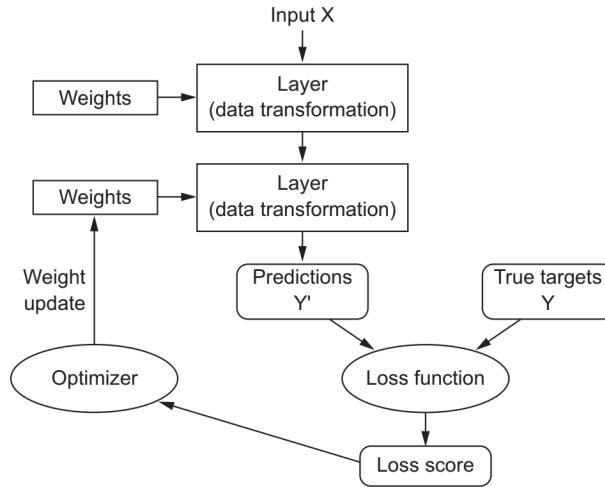


Figure 3.3: Training loop of deep learning model [27].

3.2 Types of Deep Learning Models

Researchers use various types of DL models, such as Convolutional Neural Network (CNN), Deep Neural Network (DNN), Deep Belief Network (DBN), Deep Autoencoder (dA), Deep Boltzmann machine (DBM), Deep Conventional Extreme Learning Machines (DCELM), Recurrent neural network (RNN) and their variants. A summary of the current DL models for AD diagnosis as to their main characteristics, advantages, and disadvantages is given in Table 3.1 [74, 24].

Table 3.1: Different deep learning models that have been used to detect AD (adapt from [74]).

DL model	Details	Pros	Cons
CNN	Three-layer network: a convolutional, a pooling and a fully connected layer. It is very interesting for two-dimensional (2D) data, since it applies convolutional filters that transform 2D into three-dimensional (3D) data. There are different types of CNN architectures, including: AlexNet, LeNet, faster Region-based Convolutional Neural Network (R-CNN), GoogleNET, Residual Network (ResNet), Visual Geometry Group Network (VGGNet), and ZFnet.	Excellent performance, quick learning.	A significant amount of labelled data is necessary.
DNN	Network with three or more layers, implementing complex non-linear connections. It can be applied either for classification or regression.	It is commonly employed and quite accurate.	The training procedure is challenging since the error is passed down to the previous individual layers and they get much smaller. The model's learning rate is likewise far too sluggish.
RNN	RNN uses sequential data or time series data. As sequential data is used, the weights are shared at all stages. They differ from other methods because they employ data from earlier inputs to alter current input and output. There are many variations, including: Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (BLSTM), Multi-dimensional Long Short-term Memory (MDLSTM), and Hierarchical Long Short-Term Memory (HLSTM).	It can model time dependencies.	Numerous problems arise as a result of a vanishing gradient and the demand for large datasets.

DL model	Details	Pros	Cons
DCELM	Convolutional neural networks' ability to abstract features and extreme learning machines' quick training speed are combined in DCELM. This network employs a Gaussian probability function for local connection sampling.	It is quite accurate and commonly utilised. It is a quick and efficient training technique in terms of computing. For random distortion, it works well.	Initialization might be successful if the learning function is fairly straightforward and there is a minimal amount of labelled data.
DBM	The Boltzmann family of machines serves as the basis for this model. It constructs the probability distribution using an energy function, then refines the parameters until the model understands the true distribution of the data. It is made up of one-way links among all hidden layers.	As top-down feed-back is applied, there is ambiguous data for more robust inference.	With large datasets, parameter optimization is not feasible.
DBN	DBN uses one-way connections and is applied to both supervised and unsupervised learning. Each subnetwork's hidden layers act as the following layer's visible layer.	The likelihood is maximised directly by the greedy approach employed in each layer and the tractable inference.	The training method requires a lot of processing power given the initialization.
dA	It is inserted in the unsupervised learning and was developed, mainly, to extract and/or reduce the dimensionality of the features. The quantity of inputs is equal to the quantity of outputs. There are numerous variants: De-noising Autoencoder, Sparse Autoencoder, Conventional Auto-Encoder for increased robustness.	It does not need labelled data.	It requires a step before training. Its training may not converge.

3.2.1 Convolutional Neural Networks

The convolutional neural network, a family of artificial neural networks that has dominated different computer vision tasks, is gaining interest across a range of areas, including neuroimaging. Convolution layers, pooling layers, and fully connected layers are some of the several building blocks that can be used in CNN to automatically and adaptively learn spatial hierarchies of information.

3.2.1.1 Convolution layer

The convolution layer performs feature extraction, which often entails combining linear and non-linear operations, such as the convolution operation and the activation function. This component is one of the most important layers of CNN. Convolution is a specific kind of linear operation used to extract features where a small set of numbers, called a kernel, is applied to the entire input [88].

As shown in Figure 3.4, to determine the output value in the corresponding place of the output tensor, an element-wise product between each element of the kernel and the input tensor is calculated at each point of the tensor and summed [88]. This process is repeated, applying multiple kernels to create an infinite number of feature maps reflecting various properties of the input tensors; this variety of kernels can be interpreted as multiple feature extractors. Size and the number of kernels are two essential hyperparameters that describe the convolution procedure. Beyond those, CNNs use padding as a solution to the problem of not being able to overlap the centre of each kernel with the outermost element of the input tensor, which reduces the height and width of the output feature map. Padding involves increasing the input tensor's sides with rows and columns of zeros, allowing a kernel's centre to overlap the outermost element, and keeping the convolution operation's in-plane dimension constant. Another element that characterises the convolution operation is the stride, which is the separation between two succeeding kernel positions. The most

typical value for stride is 1, but larger strides can also be used to achieve feature map downsampling. An alternative that is also used for downsampling is pooling, which will be presented next [88].

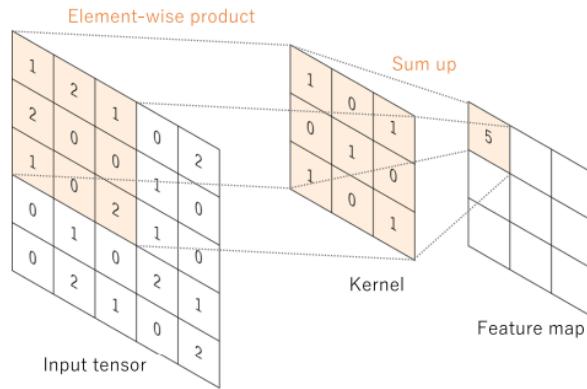


Figure 3.4: An example of convolution operation with a kernel size of 3×3 [88].

The objective of a convolutional neural network's training process is to determine the kernels that perform best for a specific task based on a training dataset. Only kernels are automatically learned during the convolution layer training process, so the hyperparameters size of the kernels, number of kernels, padding, and stride have to be set before training [88].

The outcomes of a linear operation like convolution are subsequently subjected to a nonlinear activation function. Despite the fact that smooth nonlinear functions such as the sigmoid or hyperbolic tangent (tanh) function have been used in the past because they are mathematical characterizations of biological neuron behaviour, the rectified linear unit (ReLU) is the most common nonlinear activation function, which computes the function: $f(x) = \max(0, x)$. Figure 3.5 represents the three types of functions mentioned above [88].

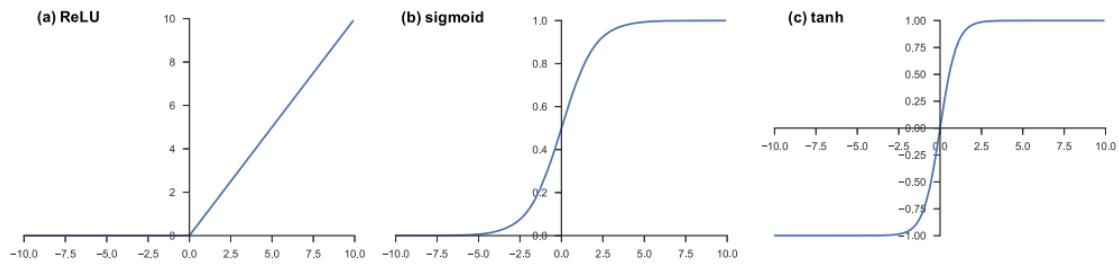


Figure 3.5: Activation functions: (a) ReLU, (b) sigmoid, and (c) hyperbolic tangent (tanh) [88].

3.2.1.2 Pooling layer

A pooling layer provides a common downsampling technique that lowers the feature maps' in-plane dimensions to introduce translation invariance to slight shifts and distortions, and limit the

number of ensuing learnable parameters. It is noteworthy that none of the pooling layers have a learnable parameter [88].

There are several pooling operations, but max pooling layer and global average pooling are the most commonly applied. The most popular kind of pooling operation is max pooling (Figure 3.6). It takes patches from the input feature maps, and only the greatest value in each patch is outputted, with the remaining values being ignored. A global average pooling algorithm executes a severe form of downsampling in which a feature map with dimensions of height \times width is downsampled into a 1×1 array by basically averaging all the elements in each feature map while keeping the depth of the feature map. Usually, only one application of this process is made prior to the fully connected layers [88].

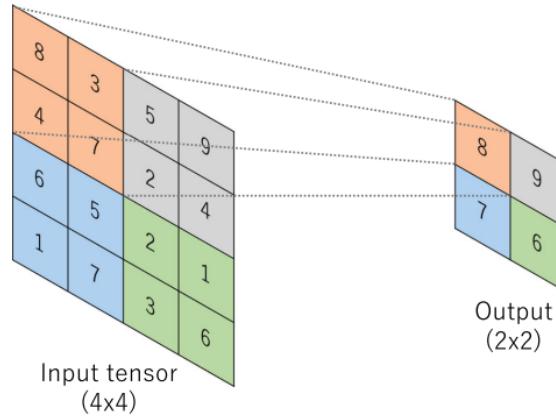


Figure 3.6: An example of max pooling operation with a filter size of 2×2 [88].

3.2.1.3 Fully connected layer

The output feature maps from the final convolution or pooling layer are typically flattened, or transformed into a 1D array of numbers, and linked to one or more dense layers, also known as fully connected layers, in which each input and each output are connected by a trainable weight. The characteristics obtained from the convolution layers and the downsampling layers are then mapped to the network's final outputs, by a subset of fully connected layers. A nonlinear function, like ReLU, follows each fully connected layer [88].

The final fully connected layer typically receives a different activation function than the others. Each task requires a different activation function, which must be chosen accordingly. Table 3.3 summarises the nonlinear functions to be applied for each type of problem [88].

3.2.1.4 Loss function

A loss function evaluates how well the network's output predictions, obtained by forward propagation, match the labels of the input ground truth data. One of the hyperparameters is the type of loss

function, and it must be chosen in accordance with the tasks provided. Table 3.3 also summarises the loss functions to be applied for each type of problem [88].

Table 3.3: Choosing the right last-layer activation and loss function for a certain type of model [88].

Problem Type	Last-layer Activation	Loss function
Binary Classification	sigmoid	binary_crossentropy
Multiclass, single-label classification	softmax	categorical_crossentropy
Multiclass, multilabel classification	sigmoid	binary_crossentropy
Regression to arbitrary values	none	mse
Regression to values between 0 and 1	sigmoid	mse or binary_crossentropy

3.2.1.5 Gradient descent

Gradient descent is a well-known optimisation technique that iteratively modifies the network's learnable parameters, such as its kernels and weights, in order to reduce loss. The direction in which the loss function increases at the sharpest rate is indicated by its gradient, and each learnable parameter is improved in the gradient's negative direction with a random step size dependent on a hyperparameter termed learning rate. It is noteworthy that, in practise, one of the most crucial hyperparameters to set before the training begins is the learning rate (Figure 3.7). There are several gradient descent algorithms that can be implemented, for example, stochastic gradient descent, RMSprop, or Adam [88].

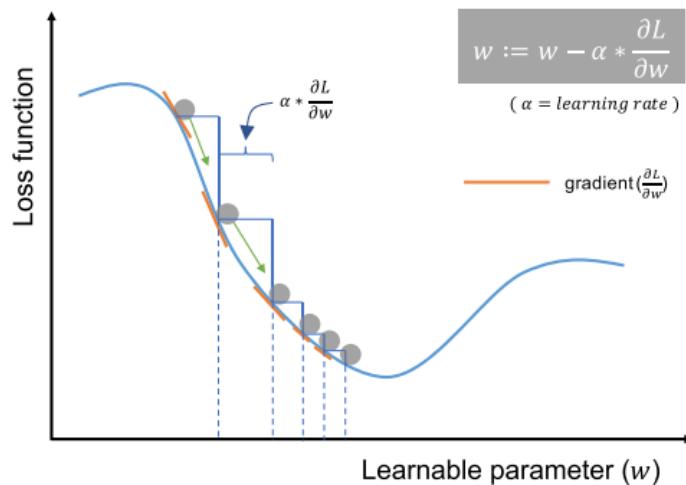


Figure 3.7: Gradient descent approach [88].

A list of the convolutional neural network's parameters and hyperparameters is defined in Table 3.5. It is important to emphasise that a hyperparameter is a variable that must be set before training, whereas a parameter is a variable that is automatically optimised during training [88].

Table 3.5: Parameters and hyperparameters associated with each type of building block in a CNN [88].

	Parameters	Hyperparameters
Convolution layer	Kernels	Kernel size, number of kernels, stride, padding, activation function
Pooling layer	None	Pooling method, filter size, stride, padding
Fully connected layer	Weights	Number of weights, activation function
Others		Model architecture, optimizer, learning rate, loss function, mini-batch size, epochs, regularization, weight initialization, dataset splitting

3.3 Hardware

Small deep-learning models could be run on personal computers between 1990 and 2010 thanks to the improvement in speed of off-the-shelf Central Process Units (CPUs). The conventional deep-learning models used in computer vision and speech recognition, however, demand a lot more computational power than what a laptop can offer. To power photorealistic video games, businesses like NVIDIA and AMD created quick, massively parallel Graphics Processing Units (GPUs). This investment in GPU technology benefited the scientific community when NVIDIA introduced CUDA in 2007. With the help of this programming interface, GPUs could take the place of CPUs in applications with a high degree of parallelization, such as physics modelling and neural networks. A gaming GPU from NVIDIA called the TITAN X has 350 times the processing capability of a current laptop and can train ImageNet models in just a few days. On clusters of hundreds of GPUs designed expressly for deep learning, large corporations train deep learning models [27]. So hardware like this is a must for developing projects like the one presented in this report.

3.4 Data

Artificial intelligence (AI) is sometimes hailed as the next industrial revolution. Data is the raw material that powers intelligent machines, without which nothing would be conceivable. The rise of the internet has changed the game when it comes to data, making it possible to gather and disseminate huge datasets for machine learning, in addition to the exponential growth in storage hardware over the past 20 years. The quality and quantity of the data are major issues. Deep learning algorithms are notoriously "data hungry," frequently needing millions of observations to work as expected. However, it is not uncommon to have to train an image classification algorithm

with very little data [27]. With just a few tens of examples, a convnet cannot be trained to handle a complex problem, but if the task is straightforward and the model is modest and highly regularised, a few hundred samples may be sufficient. In medical imaging, it is difficult to find large datasets, so this is a very present reality in this type of work. With this, some problems usually arise when implementing deep learning networks, even so, there are already several techniques that can be adopted to avoid them, as will be presented next.

3.4.1 Overfitting

Overfitting is a prevalent issue in machine learning that needs to be solved in order to become proficient. So, finding a balance between optimisation and generalisation is the main difficulty in ML. Generalisation is the ability of a model to perform well on data that it has not yet encountered, whereas optimisation refers to modifying a model to obtain the best performance possible on the training data. The model can only be adjusted based on its training data, despite the fact that excellent generalisation is the end goal. During the early stages of training, the model is underfit and can still be improved because optimisation and generalisation are strongly correlated. Generalisation, however, starts to deteriorate after a certain point, suggesting that the model is overfitting, that is, discovering patterns that are unique to the training data but irrelevant for new data [27]. The objective is to discover a good fit so that the model can recognise the patterns in the training data without memorising the smaller details. Figure 3.8 helps to understand the relationship between underfitting, overfitting, and a good fit [17].

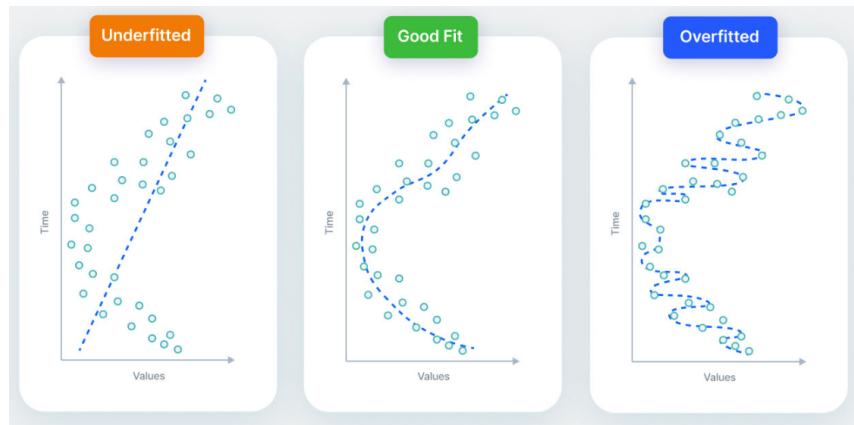


Figure 3.8: Underfitted vs. Fitted vs. Overfitted model [17].

Numerous factors contribute to overfitting, such as uncleansed data with noise values that prohibit the model from generalising, high model variance, a lack of training data that limits learning, and the complexity of deep neural networks that require extensive training time. Figure 3.9 characterises overfitting. The validation loss eventually stagnates and begins to rise while the training loss decreases over time. And in terms of the accuracy metric, the training accuracy rises linearly throughout the training until it approaches almost 100%, but the validation accuracy plateaus after a certain point [17].



Figure 3.9: Signs of overfitting in the loss and accuracy curves [17].

However, there are some techniques and strategies that help prevent overfitting and improve model performance:

- **Train with more data**

The easiest way to avoid overfitting is to collect more data. The important characteristics that must be extracted stand out as the training data grows. The connection between the input attributes and the output variable can be identified by the model. The only prerequisite for using this technique is that the data being fed into the model be clean; otherwise, overfitting would become a bigger issue [27, 17].

- **Reducing the network's complexity**

Another simple way to overcome this problem is to reduce the size of the model, or the amount of learnable parameters in the model (which is based on the number of layers and units per layer). In DL, a model's capacity is frequently referred to as the amount of learnable parameters in the model. A model with more parameters should theoretically have greater memory capacity and be able to memorise a perfect dictionary-like mapping between training samples and their labels [27].

- **Data Augmentation**

Data augmentation is an alternative approach to training with more data that is less costly and safer. Thus, the goal of this method is to increase the size and diversity of the training input, which can be thought of as a type of regularisation method to lower the generalisation error of the model. Data augmentation creates additional representative samples that mimic changes in the acquisition and anatomical variation of patients by applying modifications to the images of the training data. By avoiding learning features that are overly specific to the initial training data, the extra data should make the model more generalizable and eventually improve performance on the test set [27, 17, 26]. There are several types of data augmentation techniques. Basic augmentation methods

involve applying transformations to an image to create an augmented image, such as moving points to a different location or adjusting intensity values. Sizing, translation, rotation, reflection, and shearing are the most commonly used examples. Deformable augmentation methods can be used if geometric transformations do not offer enough variability to make the resulting model generalizable. This includes a randomised displacement field, spline interpolation, and deformable image registration. The last and most complex method is the application of deep learning-based approaches (e.g., generative adversarial networks - GAN), which can produce plausible images by using adversarial learning of a generator (G) and a discriminator (D). Data augmentation can also be the solution to another common problem in deep learning models - class imbalance. The algorithm may be biased towards the over-represented class if one or more of the classes to be predicted are under-represented in the dataset. By adding more data from the underrepresented class, data augmentation is a frequent technique to solve this problem [26].

- **Early stopping**

With this approach, the model's training is halted before random changes and noise from the data are remembered. The model could run the risk of underfitting if it stops training too quickly. The best duration and number of training iterations for the model must be determined [17].

- **Adding dropout layers**

Dropout is an efficient regularisation method for neural networks that involves arbitrarily removing (setting to zero) a few of the layer's output features during training. With this, the complexity of the model is reduced [27, 17].

- **Regularisation**

Regularisation techniques like Lasso (L1) can be helpful if it is difficult to be sure which features should be removed from the model. This approach imposes a "penalty" on the input parameters with higher coefficients, thereby reducing the variance of the model. In this way, overfitting is mitigated by adding constraints to the complexity of the model through a more regular arrangement of weights [27, 17].

- **Ensembling**

It is a ML method that creates an ideal predictive model by combining several base models. The forecasts are combined in ensemble learning to determine the most well-likely outcome [17].

3.5 Transfer Learning

Transfer learning is a well-liked technique in computer vision because it makes it possible to construct precise models quickly. With transfer learning, the learning process begins with patterns that have been learned from solving a different problem rather than from scratch. So it is a good deep

learning technique for small datasets. Transfer learning is typically expressed by using trained models. The hierarchy of spatial features learned by the pre-trained network can effectively act as a generic model for different problems in computer vision. Therefore, it is usual practise to import and employ models from published literature due to the computational cost of training such models (e.g., Inception, VGG, MobileNet) [27, 61].

Many pre-trained transfer learning models are built on massive convolutional neural networks. These deep learning models' ability to autonomously learn hierarchical feature representations is a crucial feature. This indicates that whereas features calculated by the last layer are specialised and rely on the specified dataset and assignment, features computed by the first layer are universal and can be applied to multiple problem fields. Because of this, the convolutional base of a standard CNN, particularly its lower layers (those closest to the inputs), refer to common characteristics, whereas the classifier component and some of the upper layers of the convolutional base refer to specialised features [27, 61].

A pre-trained model must first have its original classifier removed, then a new classifier that matches the new objectives must be added, and finally the model must be tuned using one of three strategies [27, 61]:

- 1. Train the all model.**

In this instance, the pre-trained model's architecture is utilised, and it is trained using the new dataset. Since the model learns from scratch, a sizable dataset and processing capacity are required [61].

- 2. Certain layers are trained while the rest are left frozen.**

As previously stated, lower layers correspond to general characteristics (problem independent), while upper layers correspond to specialised features (problem dependent). Here, it is possible to select how to update the net weights (a frozen layer remains unchanged during training). A crucial factor in the model's performance is determining how many layers to unfreeze. In general, it is advised to unfreeze a few layers for a short dataset and a large number of parameters to prevent overfitting. On the other hand, more layers can be unfrozen with a large dataset and few parameters [61].

- 3. Freeze the convolutional base.**

The train/freeze trade-off issue in this example is extreme. The key concept is to maintain the convolutional base in its original state and then feed the classifier with its results. If the computational capacity is limited, the dataset is tiny, and/or a pre-trained model already has a solution to a problem that is very similar to the current task, using the pre-trained model as a fixed feature extraction mechanism may be helpful [61].

These three tactics are shown schematically in Figure 3.10.

These three techniques—building a small model from scratch, extracting features from a pre-trained model, and fine-tuning a pre-trained model—will eventually make up a toolkit for dealing with the challenge of performing image classification with small datasets [27].

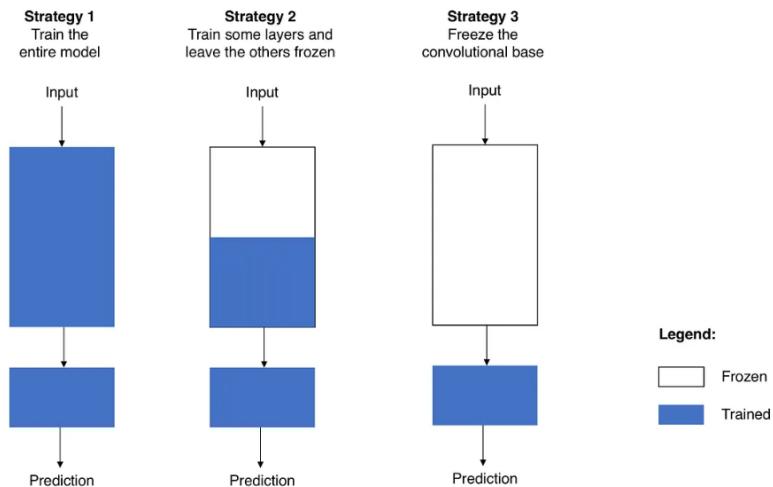


Figure 3.10: Transfer learning strategies [61].

The complete transfer learning process can be summed up as follows [61]:

1. **Choose a pre-trained model** - select a model that appears appropriate for the issue at hand [61];
2. **Use the Size-Similarity Matrix to categorise the task** - Figure 3.11a shows the matrix that directs the decisions. Taking into account the size of the dataset utilised and its resemblance to the dataset in which the pre-trained model chosen was developed, this matrix categorises the computer vision challenge [61].
3. **Fine-tune the model** - In this step, the Size Similarity Matrix helps to decide between the three possibilities described above, as shown in Figure 3.11b [61].

Note that in Figure 3.11, certain considerations must be taken into account for each quadrant [61]:

- **Quadrant 1** - Big dataset, yet distinct from the dataset used to train the model.

The result of this circumstance is Strategy 1. The availability of a sizable dataset makes it possible to train a model from scratch. In spite of the different datasets, initialising the model from a trained model and utilising its architecture and weights can be helpful in practise [61].

- **Quadrant 2** - Big dataset and comparable to the dataset used in the pre-trained model.

Any strategy is viable. The most effective choice is probably Strategy 2. Overfitting should not be a problem because a huge dataset is available, making it possible to learn as much as desired. Yet, since the datasets are comparable, it is possible to leverage prior knowledge to reduce training time significantly. Hence, training the classifier and the top layers of the convolutional base should be sufficient [61].

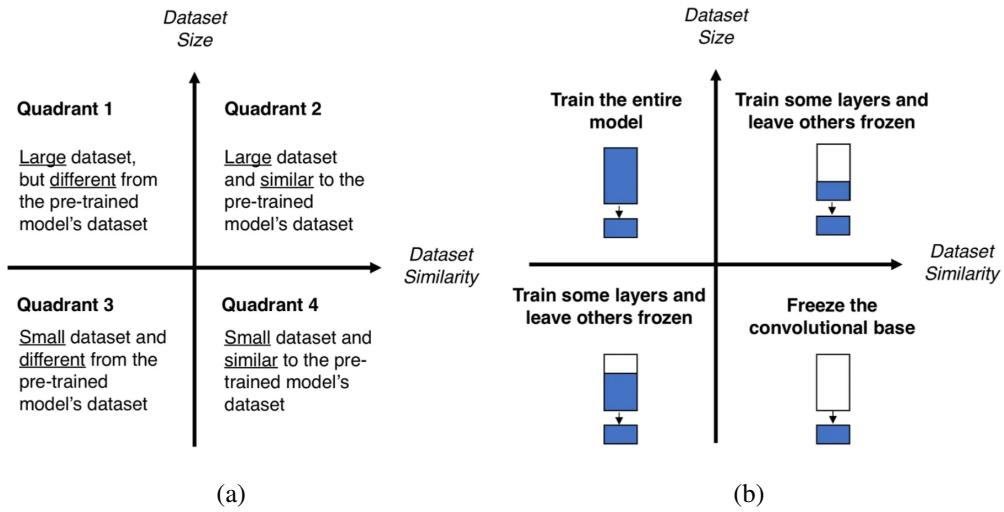


Figure 3.11: a- Size-Similarity matrix; b- Decision map for fine-tuning pre-trained models [61].

- **Quadrant 3** - Small dataset that differs from the dataset used to train the pre-trained model.

This is the worst case scenario, which happens in 2 out of 7 cases of computer vision problems. The only viable option is Strategy 2. The model will probably have to go deeper than in Quadrant 2 and will have to consider data augmentation techniques [61].

- **Quadrant 4** - Small dataset, but comparable to the dataset used to train the model.

In this case, the best strategy will be 3. The pre-trained model can be used as a fixed feature extractor, the final fully connected layer (output layer) can be eliminated, and the extracted features can then be used to train a new classifier [61].

When applying strategies 1 and 2 to pre-trained models built on CNN, it is crucial to be cautious with the learning rate. A high learning rate can cause the loss of prior knowledge because it determines how much the network weights are adjusted. So long as the pre-trained model has been properly trained, it is advised to use a small learning rate to prevent distorting the CNN weights too soon and too much. Contrarily, strategy 3 is simple and does not necessitate as much thought regarding the learning rate [61].

3.6 Performance Evaluation

An essential component of machine learning is performance assessment, but it is a difficult job. To evaluate ML models, it is essential to choose the appropriate measure. To better understand each metric and the applications they can be used for, an overview of the most common metrics are provided in Table 3.7. In some applications, one metric may not provide a full picture of the issue in question, so it may be useful to use a subset of the metrics instead [64].

Before introducing the metrics, it is important to grasp a crucial concept in classification performance - Confusion Matrix - which shows a tabular comparison of the model forecasts and the

truth labels. The instances in a predicted class are represented by each column in the confusion matrix, and the instances in an actual class are represented by each row. The diagonal elements of this matrix represent the samples that were correctly classified, whereas the off-diagonal elements represent the incorrect predictions for the various groups [64]. Figure 3.12 represents an example of a confusion matrix for a binary classification, where TP corresponds to True positives (prediction and reality positive), TN to True Negatives (prediction and reality negative), FP to False Positives (prediction positive, reality negative), and FN to False Negatives (prediction negative, reality positive) [67].

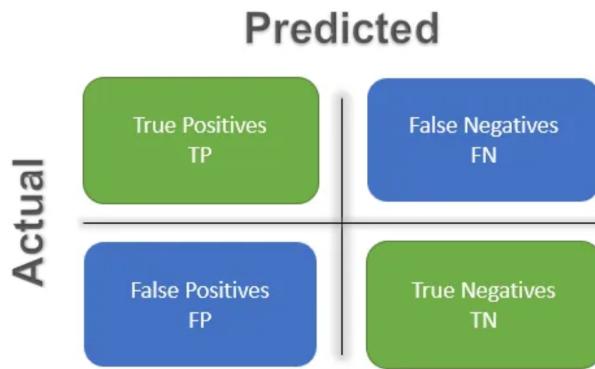


Figure 3.12: Representation of a confusion matrix [67].

Table 3.7: Performance Evaluation Metrics [47].

Metric	Description	Formula
Accuracy (ACC)	It is the number of correct predictions divided by the total number of predictions, which is perhaps the most straightforward metric.	$ACC = \frac{TP + TN}{TP + TN + FP + FN} * 100$
Precision	It is a class-specific performance metric, useful for evaluating a model performance when the class distribution is imbalanced.	$Precision = \frac{TP}{TP + FP}$
Recall	It is calculated as the percentage of samples from a class that the model correctly predicts.	$Recall = \frac{TP}{TP + FN}$
F1 Score	It is the harmonic mean of recall and precision, allowing a combination of this two metrics.	$F1score = \frac{2 * Precision * Recall}{Precision + Recall}$
Sensitivity	Determines how likely a test is to be positive when the condition is in fact positive. It also goes by the names false-negative rate and recall.	$Sensitivity = \frac{TP}{TP + FN} * 100(%)$
Specificity	Determines how likely a test is to be negative when the condition is in fact negative. It is also known as false-positive rate or precision.	$Specificity = \frac{TN}{TN + FP} * 100(%)$

In addition, there are two more metrics that can be useful for drawing conclusions about the model. The receiver operating characteristic curve (ROC curve) is a plot that displays how well a binary classifier performs in relation to its cut-off threshold. Basically, it compares the TP rate and FP rate for different threshold values. Area under the curve (AUC) is a combined indicator of how well a binary classifier performs across all potential threshold values (and therefore it is threshold invariant). AUC ranges from 0 to 1 because it measures the area under the ROC curve. AUC can be understood as the likelihood that the model values a randomly chosen positive example higher than a randomly chosen negative example [64]. In Figure 3.13, an example of these two measurements is visible.

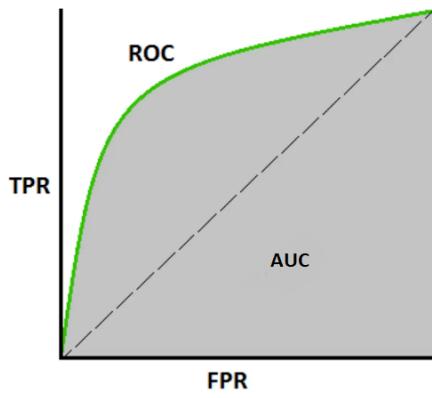


Figure 3.13: The AUC is indicated by the gray area below the ROC curve [64].

3.7 Summary

In many medical applications, deep learning is currently the most successful machine learning technique. These artificial neural networks with multiple layers can have different architectures, including CNNs for image and video, RNNs for sequential data, and transformers for natural language processing. Deep learning models have some requirements regarding hardware and data, i.e., they must be trained using large datasets and significant computational power (using GPUs). However, in the medical field, it is common to have a small amount of images, so some problems arise like overfitting. There are some techniques and strategies that help prevent overfitting and improve model performance, like training with more data, reducing the network's complexity, data augmentation, and more. Another good tool for small datasets is transfer learning, which is a technique that uses pre-trained models as a starting point for a new task. This way, the knowledge and features learned by the pre-trained model on a large dataset are leveraged to solve a new, related problem with a smaller dataset. At the end of the implementation of the model, it is necessary to evaluate its performance, so there are several metrics like ACC, Precision, Recall, and so on, to do so.

Chapter 4

Literature Review - Deep Learning Methods to Detect AD from MRI

This chapter presents an up-to-date systematic review of deep models to detect AD and its intermediate phase by evaluating magnetic resonance images. The deep learning models chosen by different authors are analysed, as well as their approaches regarding the used dataset, and the data preprocessing and analysis techniques.

4.1 Articles Selection Methodology

In the Scopus database, a systematic literature search was performed, Figure 4.1, using the following query: "ALL ("Alzheimer's Disease") AND ("image" OR "imaging") AND ("MR" OR "Magnetic Resonance") AND ("T1") AND ("Deep Learning") AND ("MCI" OR "mild cognitive impairment") AND ("3D" OR "3 dimensional") AND ("classification") AND (TITLE ("Alzheimer")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (PUBYEAR , 2023) OR LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018)) AND (LIMIT-TO (LANGUAGE , "English"))".

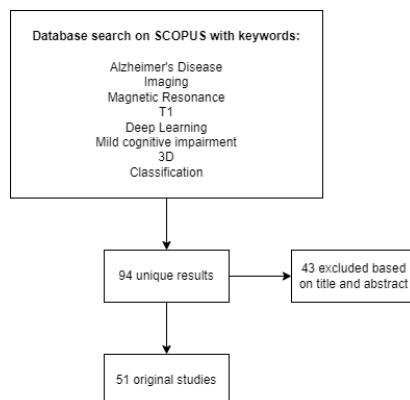


Figure 4.1: PRISMA diagram showing the performed literature search in the Scopus database.

In the performed search, a total of 94 articles were obtained, and of these, 51 were found to be useful for the topic under review. The remaining were excluded due to the following reasons: the proposed method does not classify the input data, does not include deep learning techniques or does not use 3D MR T1 images.

4.2 Review Results

4.2.1 Reviews

Two systematic reviews were also included in this review. Sethi et al. [78] explored different AD classification methods based on CNNs. In this study, topics such as data preprocessing, data processing, 2D and 3D CNNs and their variants, and data augmentation and transfer learning were addressed and compared. On the other hand, Tanveer et al. [82] discussed machine learning techniques used in AD diagnosis, namely support vector machine, artificial neural network, DL and ensemble models. Within their study, they also evaluated imaging modalities, feature extraction and selection techniques, and transfer learning. A summary of this information is presented in table 4.1.

Table 4.1: Summary of the reviews found in the current study.

A.	Years covered	Number of A.	Description
[78]	2012-2021	48	Analyse the efficacy of CNN classification approaches in AD using different types of datasets, neuroimaging modalities, data preprocessing and processing techniques.
[82]	2005-2019	165	The analysed machine learning methods are categorised into three main groups: support vector machine, artificial neural network, and deep learning and ensemble models.

In order to update and complement these reviews, the present analysis is proposed.

4.2.2 Used Datasets

In any machine learning algorithm, the used data plays a fundamental role in the final solution. The quantity of data as well as its quality, in terms of annotations and the way it is organised, must be evaluated with special attention before applying any model to it. In this section, a description of the datasets used in the reviewed articles is presented.

4.2.2.1 ADNI

The purpose of the Alzheimer's Disease Neuroimaging Initiative study is to analyse the brain's structure and function across the course of different disease states using biomarkers and clinical measurements to track the progression of the disease. The project was created in 2004 and consists of 4 studies - ADNI1, ADNIGO, ADNI2 and ADNI3, in each of which new participants are being added, while the original participants continue to be monitored. This dataset also has a wide variety of data types, namely: clinical information as to demographics, physical examinations,

and cognitive assessment data, genetic information, MR images, including Structural MRI, fMRI and Diffusion Tensor Imaging, PET images, including Amyloid PET, FDG PET, and Tau PET, and biospecimens, i.e., blood, urine, and CSF [8].

4.2.2.2 OASIS

The Open Access Imaging Studies Series (OASIS) is a project that makes available brain neuroimaging datasets to help researchers develop this field. Four studies have already been developed in this initiative: OASIS-1, OASIS-2, OASIS-3 and OASIS-4, and being a neuroimaging database, three imaging modalities are available: MRI, PET and Computed Tomography [9].

4.2.2.3 AIBL

The Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL) is a study whose major aim is to investigate which biomarkers, mental characteristics, and health and lifestyle aspects are determinants in the development of Alzheimer's disease. This dataset includes, neuroimaging such as MRI and PET, lifestyle data, i.e., questionnaires on diet, sleeping habits, etc., cognitive testing, family history and biomarkers: blood samples and CSF [7].

Several datasets are used in the reviewed articles, being the datasets presented in Figure 4.2 the most widely used and those with the greatest variety of information, in addition to being of public access.

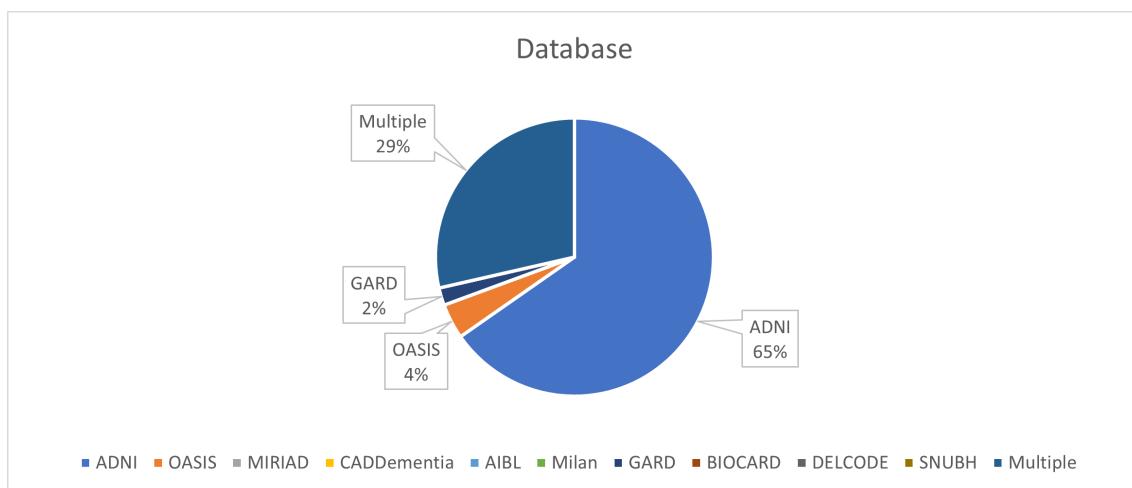


Figure 4.2: Prevalence of the used datasets in the reviewed articles.

4.2.3 Data Preprocessing

After the definition of the dataset and the acquisition of the neuroimages using the respective imaging modality, it is necessary to understand how these will be inputted into a DL architecture for the diagnosis of AD. However, before this process, the used images are usually submitted to several data preprocessing steps. Particularly, the preprocessing aims to enhance the input data

by ensuring that all the images have a degree of parity that in turn makes the following steps, for example, of image segmentation and feature extraction, more effective. That is, it entails removing artifacts, changing image resolution, and correcting contrast disparities caused by differing imaging acquisition hardware and parameters [84].

4.2.3.1 Bias field correction

The bias field, often known as intensity inhomogeneity, is a low frequency spatially changing MRI artifact that results in smooth fluctuations in signal intensity within tissue with similar physical qualities. The magnetic field strength is directly proportional to the bias field. The produced bias field is almost invisible when images are acquired at 0.5T and can be discarded. However, the bias field is powerful enough to cause issues and significantly impact the MRI analysis when images are scanned with a field of 1.5T, 3T, or greater [31]. So, the bias filed correction is a technique that has been usually used in this field [45, 71, 85].

4.2.3.2 Normalisation

Another method of preprocessing that has been commonly used is normalisation [38, 36, 14], which aims to reduce the grey (or colour) values in an image to a single set of relative grey (or colour) values. This guarantees that differences in imaging acquisition parameters across different imaging scanners not affect strongly the further results since similar tissues show up in a consistent range of values throughout all image scans [84].

4.2.3.3 Skull stripping

Skull stripping, also called brain extraction, is a computational approach for removing non-essential tissues from a brain image, such as skull, fat, or skin. In this way, the amount of noise can be reduced, and subsequent feature extraction step can be easier [84]. This method is one of the most widely used in this field [70, 35, 57].

4.2.3.4 Spatial smoothing

This technique allows increasing the signal-to-noise ratio by removing the high-frequency spatial noise components and is applied in several imaging, including in MRI. Averaging data pixels (or voxels in 3D) with their neighbours is referred to as spatial smoothing. Sharp "edges" of the images are blurred as a result, and spatial correlation within the data becomes more prominent [60].

4.2.3.5 Registration

Image registration is characterised by a process of overlaying two or more images containing the same object of study but taken at various times, from different perspectives, and/or by different imaging modalities. This step has been commonly used in this field [54, 53, 37], and implies

discovering the transformation between the images to be registered, so that their main characteristics are spatially aligned. Mainly, this alignment can be rigid or affine. A rigid transformation is composed of six parameters with translation and rotation, and is usually used in intrasubject registration when the object of attention is not relatively deformed, for example, for images at the same stage of brain development. The affine transformation is often used if there is scaling and skewing involved, such as in the case of different subjects or different developmental brain stages [31].

4.2.4 Data Analysis

In order to, from the preprocessed brain images, classify subjects into different categories, such as AD and CN, it is usually necessary to extract and analyse features from the preprocessed images. Thus, this process commonly includes several steps like feature extraction, selection, and classification. Yet, for most researchers, the challenge is to know how to tackle and input the neuroimages into the classifier, and for that different data processing approaches can be used, such as whole image, slice, region of interest (ROI), and patch based, as described in the following sections and summarized in Figure 4.3.

- **Whole 3D image**

This kind of approach uses as input the whole 3D image which allows to preserve all the spatial information of the original image as well as all the details, and has been commonly adopted by most researchers [18, 68, 44], however, it entails higher computational costs.

- **Slice based**

In order to obtain data features, the slice based approach extracts 2D images from the original 3D image. This enables the removal of background areas that act as interference; however, part of the information is lost. The authors in [51, 39, 81] selected slices in the coronal plane that contained the region of the hippocampus, as it is very relevant in the diagnosis of AD. In [90], the investigator used the same strategy; however, instead of one plane, slices in the 3 planes: coronal, axial and sagittal, were used.

In order not to lose so much relevant information, the author in [34], only removes the first and last slices, as they did not contain any useful detail.

Two completely different approaches were adopted by Shaji et al. [79] and Angkoso et al. [13]. The first authors, of the 182 slices of each image, only used the mid-axial slice in the study, while the second authors, for each plane, the slice with the largest area and the ones immediately before (-1) and after (+1) it were used.

- **ROI based**

The approach based on regions of interest involves the segmentation of particular anatomical structures from the original images, such as white matter (WM), grey matter (GM), cerebrospinal

fluid, and/or hippocampus. This technique allows focusing the DL model on the most relevant areas linked to the disease; however, other relevant structures can not be considered in the model.

- **Patch based**

As the designation implies, the patch based approach splits the original image into small patches that are used as input to extract features. Like the two previous approaches, this one also allows reducing the amount of performed calculations required by the used DL classifier; however, there is a loss of spatial information between the patches. The authors in [52, 89] applied this technique in their study.

- **Multiple**

It is also possible to use more than one of the previous approaches simultaneously, for example: ROI based + Patch based or Slice based + ROI based. As to the first alternative, three works [53, 58, 29] used this combination and extracted patches centred in the region of interest of the hippocampus. On the other hand, as to the second case, basically the regions of interest were extracted from 2D images instead of the original 3D image. Examples of this approach can be found in [56, 44, 63].

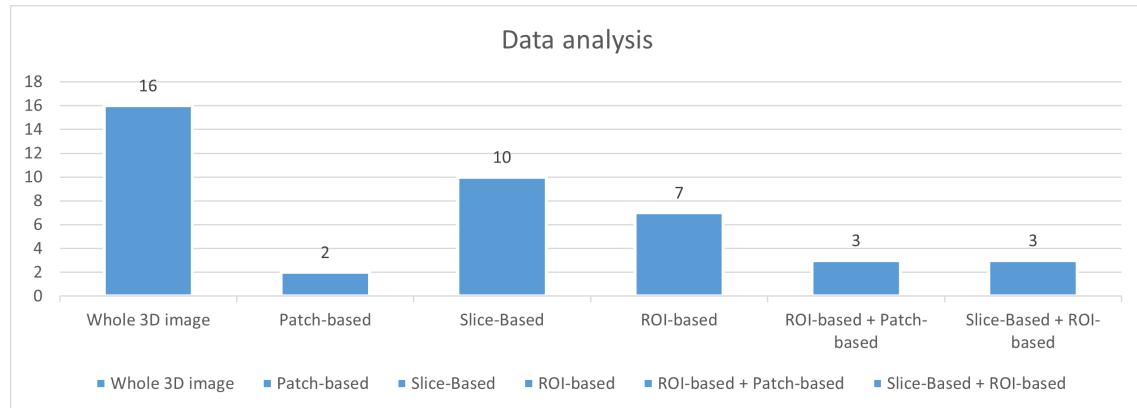


Figure 4.3: Prevalence of used data processing approaches.

4.2.5 Deep Learning based Approaches

The construction of the AD classification pipeline, namely the selection of the type of neural network to be implemented, the definition of its hyperparameters and the method for its validation and evaluation are the usual last steps in the development of computational systems to diagnose AD, and many different approaches have been adopted by the researchers. Nevertheless, two major paths can be taken in this process: either classify directly or perform a feature extraction process before the classification step. Both paths were found in the performed review. In addition to classification approaches, prediction approaches were also found, however in much smaller number.

4.2.5.1 Classification approaches

Most researchers chose to directly apply a classifier to the original or preprocessed images. In the field of image analysis, CNNs have attracted most of the attention due to being specially designed to recognize patterns. Basheera et al. [19], Faisal et al. [34] and Yigit et al. [90] applied 2D CNNs, while Basaia et al. [18] and Dyrba et al. [33] opted for a 3D CNN. An interesting study was taken by Wen et al. [86], who used a CNN to compare the influence of different data processing approaches. The authors conclude that the results show the superiority of 3D approaches over 2D approaches, but there was little difference in the outcomes of the different 3D approaches. In addition to CNNs, their variants have also been applied in this field, namely U-net [54, 35], DenseNet [73, 11, 29, 52, 58], ConvNet [37], ResNet [36, 51, 79, 66, 56, 77, 15], LeNet [36, 39], AlexNet [66], GoogleNet [36, 66], Inception [66, 79, 16], EfficientNet [11], VGG [36, 68, 44, 92], and MSCNet [59].

In these studies, Liu et al. [59] applied an MSCNet to segmented WM and GM in images, and proved that WM is more effective in AD diagnosing. Goenka et al. [37], on the other hand, applied a ConvNet to different types of data processing approaches, concluding that the whole 3D image approach, in contrast to the patch and slice based approaches, delivers the maximum accuracy. Folego et al. [36] compared multiple DL architectures, namely LeNet-5, VGG 512, GoogLeNet, ResNet, and obtained better results using the VGG 512 architecture. Another type of approach was adopted by some authors who instead of implementing only one DL network, used a combination of two or more networks to obtain better results. An example of this, was the case of Cui et al. [30] who combined a CNN with a RNN: the first model is used to extract the spatial features of each time point and obtain a single time classification result; then, the RNN model, which is based on cascaded Bidirectional Gated Recurrent Unit (BGRU), is used to tackle the temporal fluctuations and produce the longitudinal features used to enhanced the final classification. Liu et al. [58] also chose to combine two CNNs, using a multi-task deep CNN model to capture the multi-level features for hippocampal joint segmentation and disease classification, while the 3D deep DenseNet model picks up the characteristics from hippocampal image patches for the disease classification. Both the multi-task and the DenseNet model were trained individually, and the final classification was carried out by a fully connected layer followed by a softmax layer that was carefully calibrated. In [53], Li et al. used very similar reasoning: constructed hybrid convolutional and recurrent neural networks based on DenseNets and BGRU networks for the inner and outer hippo patches separately. 3D DenseNets were built on the patches to learn more specific image and shape features of the hippocampus for classification, while stacked BGRUs were utilised to record the high-level correlation and asymmetry features between the right and left hippocampus. In the end, two fully connected layers were attached to merge the features that the hybrid neural networks had learnt from the inner and outer patches of the hippocampus to improve the final classification.

Additionally to the common use of CNNs and RNNs, Mendoza-Léon et al. [63] developed Supervised Switching Autoencoders (SSAs) to conduct AD classification using just a sMRI slice.

4.2.5.2 Feature extraction + classification approaches

The other path taken by the reviewed authors consisted in applying a network that extracted features from the images and then applying a classifier that allowed obtaining a diagnosis for AD.

In this sense, several researchers have chosen to use deep learning models to extract the features that are then linked to simpler machine learning algorithms to perform the final classification. For example, Wang et al. [85] implemented a DenseNet with self-attention (SA) and auto-encoding (AE) to obtain the features that served as input to a linear support vector machine (SVM) classifier that performed the classification. Suh et al. [81] employed a deep CNN model, which divides each brain image into 82 areas. The used approach combines a target section with nearby slices in the channel dimension feeds this information into the XGBoost module that classifies the patients. Finally, Jiang et al. [45] implemented a Deep Learning Radiomics (DLR) model to extract features from MR images that are then combined with clinical information and classified using an SVM. For the DRL model, several architectures were tested: AlexNet, ZFNet, ResNet18, ResNet34, InceptionV3, and ResNet34 proved to be the best one.

DNN also proved to be a widely used model as a classifier followed by feature extractor. Raghavaiah et al. [71] applied a Gabor filter, Raghavaiah et al. [72] developed a hybrid Texture, Edge, Colour and density (TECD) feature extraction approach paired with clinical data, and Basher et al. [20] built a discrete volume estimator CNN model to identify positions of the left and right hippocampus.

Two groups of authors chose to use a deep Siamese neural network, Liu et al. [57] mapped several atlases using a Large Diphomorphic Deformation Metric Mapping (LDDMM) and then got the atlas plot labels for each imaging scan through the Multi-Atlas Likelihood Fusion (MALF) algorithm before applying the neural network; while Sampath et al. [76] utilised grey-level co-occurrence matrices (GLCM), Gabor, and wavelet features to extract the MR image's biomarker data and a Hilbert Schmidt independence criteria lasso (HSICL) algorithm to select the most preponderant features.

M. Cobbina et al. [28] implemented a convolutional adversary autoencoder (CAAE) to lessen the existing variations in multi-centre raw scans by storing them in an aligned common space. Subsequently, a convolutional residual soft attention network (CRAT) was also intended for AD classification. While in [14], a 3D Convolutional Autoencoder (3D-CAE) was developed to extract features followed by the application of a 3D Deeply Supervised Adaptive CNN (3D-DSA-CNN) to perform task-specific classification.

In [89], Yan et al. chose to implement attention mechanisms with the aim of significantly enhancing the performance of CNNs. Thus, the authors added a Squeeze and Excitation (SE) and Pyramid Squeeze Attention (PSA) mechanisms to the Fully Convolutional Network (FCN) model, in order to obtain the information from each image regarding the disease probability map. In addition, they also built a Multi-Layer Perceptron (MLP) classifier, combining a disease probability map's feature information with age, gender, and Mini-Mental State Examination (MMSE) of each sample to obtain the final classification. Finally, Odusami et al. [69] suggested a concatenation

of deep and random weight features taken from the ResNet18 and DenseNet121 networks, which simultaneously learnt DL features from MR images.

4.2.5.3 Time-to-event prognostic approaches

In addition to the classification from MR images at the different AD phases, some works are focused on predicting the evolution of the disease.

In this regard, Zhao et al. [93] proposed a new paradigm for predicting disease progression that embraces five cases of evolution: MCI-AD, MCI-MCI, MCI-CN, CN-MCI and CN-CN. In the process, a patch based 3D mi-GAN model was developed to produce high-quality images at future time points with two innovations: implementing a 3D U-Net based network conditioning on image patches and additional knowledge, i.e., age, academic level, gender, and APOE, in BL; modifying the final objective function by adding a gradient difference loss (GDL loss) and a mean square error loss in image space and frequency domain. The 3D U-Net based model can utilise the multi-scale characteristics of input image patches and the additional data. The generator can create less blurry and more reliable images due to the GDL loss. Then, the created whole-brain 3D images feed into a trained 3D ternary classification model to identify their phases and better track the development of the disease within four years.

In [68], it was developed a prediction model with two alternatives. A single time point CNN that for classification, a single 3D MRI with a full time point of patients diagnosed with AD or CN served as the input, and the output was CN vs AD binary classification, and for prediction, a single 3D MRI with a full time point of patients diagnosed with MCI served as the input and the outcome was a forecast of whether the patient had advanced (pMCI) or stayed stable (sMCI) 3 years later. A second approach was a dual time point CNN, where the input consisted of 3D MRI scans taken at both baseline and 12 months later, utilising the same study group and output classes used at a single time point for classification and prediction. Both types of models start with a sequence of convolutional blocks, flattened into at least one fully connected layer, concluding with a decision of classification or prediction.

4.2.6 Data Augmentation and Transfer Learning

Big data is widely utilised by deep neural networks to prevent overfitting. Regrettably, common neuroimaging datasets of AD patients are not so big. To overcome this problem, two strategies can be applied on imaging datasets with small samples: data augmentation and transfer learning [12].

Data augmentation is an effective strategy to enhance the performance of learning algorithms by increasing the samples. This strategy is characterised by performing transformations on already existing images, such as rotation, flipping, and cropping [12]. About 32% of the reviewed studies implement this strategies, namely [85, 53, 58, 19, 38, 90, 18, 79, 56, 68, 29, 33, 77, 73, 28, 30, 20]. As an example, this strategy was implemented by Dyrba et al. [33] who, by applying flipping along

the coronal (R/L) axis and also translation of +/- 10 voxels in each direction (x/y/z) to all original images, obtaining about fourteen times more samples.

Transfer learning is an ML technique where a model that has already been created for one task is used as a baseline for another task. In other words, this technique proposes storing information learned from solving one problem and using it to solve another, related challenge [12]. There are multiple studies that have opted for this procedure, namely in [38, 36, 52, 39, 34, 66, 86, 56, 68, 77, 15, 20]. A study that demonstrates the importance of this method was developed by Nanni et al. [66], who carried out a study about the effect of the transfer learning technique on several DL models. Thus, the study assessed the efficacy of transfer learning techniques on DL models trained on common images and then applied to sMR images of the brain. The models utilised were AlexNet, GoogleNet, ResNet50, ResNet101, InceptionV3. As opposed to a 3D CNN model trained from scratch on the MRI volumes, the results showed a performance increase of at least 4.7% in terms of accuracy.

4.3 Discussion

Aiming to classify/distinguish between the various Alzheimer's disease phases using MR imaging modality, 49 different approaches were found, which are summarised in Table 4.3.

Table 4.3: Summary of the approaches found in this review.

Ref.	Year	Dataset(s)	Image Type(s)	Subjects	Data Preprocessing	Data Processing	Feature Extraction	Classifier(s)	Accuracy metric(s)
[54]	2022	ADNI	T1 sMRI	928 subjects: 330 CN, 299 MCI, 299 AD	Resample Skull Stripping Intensity Correction Register Clip	Whole 3D image	U-Net		AD vs CN: 93.16% MCI vs CN: 80.44%
[76]	2022	ADNI AIBL OASIS	T1 sMRI	6400 subjects: 3200 CN, 2240 SMC, 896 MCI, 64 AD	Resizing Adaptive filtering for noise removal Adaptive histogram equalization	Whole 3D image	GLCM Gabor filter Wavelet features HSCL	FSODSNN based classifier	CN vs SMC vs MCI vs AD: 99.89% AIBL: CN vs SMC vs MCI vs AD: 99.67% OASIS:
[72]	2022	ADNI		955 subjects: 273 CN, 432 MCI, 250 AD	Skull-stripping Normalization TC-BWO-FCM for CSF, WM and GM (Segmentation)	ROI-based	TECD features + Clinical features	HRF-DNN	AD vs CN: 98.68% MCI vs AD: 95.88% MCI vs CN: 97.23%
[71]	2021	ADNI		280 subjects: 89 CN, 122 MCI, 69 AD	Bias field correction Normalization CSF, WM and GM (Segmentation) Spatial Smoothing	ROI-based	Gabor filter Optimal DNN		AD vs CN: 96.43% MCI vs AD: 94.64% MCI vs CN: 91.07%
[85]	2022	ADNI J-ADNI	T1 sMRI	Non-image information: Cognitive scores, Age and APOE type	Image shape normalization Image intensity normalization Skull-stripping hippocampi(L/R) and anterior temporal lobes(L/R) Segmentation	ROI-based	DenseNet-121 + AE + SA Linear SVM		87.80%
[53]	2019	ADNI	T1 sMRI	807 subjects: 216 CN, 233 sMCI, 164 pMCI, 194 AD	Data augmentation Intensity homogenization Affine registration Skull-stripping Hippocampus Segmentation Data augmentation	ROI-based + Patch-based	3D DenseNet for Feature Learning + Bidirectional gated recurrent unit (BGRU)		AD vs CN: 89.10% MCI vs CN: 75.00% pMCI vs sMCI: 72.50%

Ref.	Year	Dataset(s)	Image Type(s)	Subjects	Data Preprocessing	Data Processing	Feature Extraction	Classifier(s)	Accuracy metric(s)
[58]	2020	ADNI	T1 sMRI	449 subjects: 119 CN, 233 MCI, 97 AD	Skull-stripping N3 bias field correction Affine registration Hippocampus Segmentation Data augmentation	ROI-based + Patch-based	1. 1. Multi-task deep CNN 2.3D DenseNet 3. Multi-task deep CNN + 3D DenseNet 3. AD vs CN: 80.10% MCI vs CN: 71.50%		
[19]	2020	ADNI	T1 sMRI T2 sMRI - Results not presented	120 subjects: 20 CN, 25 MCI, 70 AD	Skull stripping Data augmentation	Slice-Based	1. 1. Multi-task deep CNN 2.3D DenseNet 3. Multi-task deep CNN + 3D DenseNet 3. AD vs CN: 88.90% MCI vs CN: 76.20%		
[38]	2021	ADNI	T1 sMRI	1340 subjects: 394 CN, 401 sMCI, 197 pMCI, 348 AD	Cropping Bias field correction Normalization Skull-stripping Data augmentation	Whole 3D image	pABN	AD vs NC: 90.70% pMCI vs sMCI: 79.30%	
[36]	2020	AIBL CADD MRIAD OASIS	T1 sMRI	831 subjects: 229 CN, 403 MCI, 199 AD	Intensity winsorizing Bias field correction Translation alignment Rigid transformation Affine transformation Skull-stripping Normalization	Whole 3D image	LeNet-5 VGG 5/2 GoogLeNet ResNet	VGG 5/2 + CADDEmentia. CN vs MCI vs AD: 52.30%	
[53]	2018	ADNI	T1 sMRI	8751 subjects: 832 CN, 419 AD	Intensity homogenization Skull-stripping Linear registration	Patch-based	DenseNet	AD vs CN: 89.50% MCI vs CN: 73.80%	
[55]	2022	ADNI	T1 sMRI	210 subjects: 70 CN, 70 MCI, 70 AD	Spatially normalization Skull-stripping Normalization	Whole 3D image	DMRNet	AD vs CN: 89.30%	
[14]	2018	ADNI	T1 sMRI			3D-CAE	3D-DSA-CNN	AD vs CN: 99.31% AD vs MCI: 100.00% MCI vs CN: 94.20% AD + MCI vs CN: 95.73% AD vs MCI vs CN: 94.80%	

Ref.	Year	Dataset(s)	Image Type(s)	Subjects	Data Preprocessing	Data Processing	Feature Extraction	Classifier(s)	Accuracy metric(s)
[37]	2022	ADNI	T1 sMRI	769 subjects: 475 CN, 224 MCI, 70 AD	1.Whole 3D image N4 Bias Field Correction Skull Stripping Rigid Registration 2.Patched based 3.Slice based	ConvNet	1.	AD vs CN: 97.83% AD vs MCI: 98.68% CN vs MCI: 99.10% CN vs MCI vs AD: 98.26%	
[51]	2022	ADNI	T1 sMRI	503 subjects: 116 CN, 187 MCI, 200 AD	Standardized voxels values Size scaling Skull Stripping Image registration		1.CCS-ResNet-18 2.CCS-ResNet-50	AD vs CN: 94.34% AD vs MCI: 95.72% CN vs MCI: 89.44% CN vs MCI vs AD: 88.31%	
[39]	2021	ADNI	T1 sMRI	210 subjects: 70 CN, 70 MCI, 70 AD	Size scaling Skull Stripping	Slice-Based	LeNet	AD vs CN: 95.00% AD vs MCI: 97.00% MCI vs CN: 97.00%	
[69]	2022	ADNI	T1 sMRI	138 subjects: 25 CN, 25 SMC, 25 EMCI, 13 MCI, 25 LMCI, 25 AD	Resizing Cropping Formalisation	Slice-Based	ResNet18 DenseNet121	(Weight Initialization-Kaiming) AD vs CN vs MCI: 98.21% AD vs MCI vs LMCI vs CN: 93.06% MCI vs LMCI vs EMCI vs AD vs CN: 98.86%	
[90]	2020	OASIS MIRIAD	T1 sMRI	485 subjects: 339 CN, 116 MCI, 30 AD	Skull stripping Data augmentation Size scaling CLAHE Smoothing	Slice-Based	CNN	AD vs MCI: 82.00%	

Ref.	Year	Dataset(s)	Image Type(s)	Subjects	Data Preprocessing	Data Processing	Feature Extraction	Classifier(s)	Accuracy metric(s)
[22]	2021	ADNI		578 subjects: 188 CN, 210 MCI, 180 AD	CSF, WM and GM Segmentation Skull stripping			1.CNN 2.DMI	1. AD vs CN: 81.00% MCI vs CN: 63.00% 2. AD vs CN: 83.00% MCI vs CN: 65.00%
[18]	2018	ADNI Milan	T1 sMRI	1638 subjects: 407 CN, 813 MCI, 418 AD	Registration CSF, WM and GM Segmentation Normalization Data augmentation	Whole 3D image	CNN		AD vs CN: 98.20% cMCI vs CN: 87.70% sMCI vs CN: 76.40% AD vs cMCI: 75.80% AD vs sMCI: 86.30% cMCI vs sMCI: 74.90%
[34]	2019	ADNI	T1 sMRI	489 subjects: 163 CN, 163 MCI, 163 AD	Slice-based Size scaling	Slice-Based	CNN		AD vs MCI vs CN: 96.12 %
[79]	2021	ADNI	T1 sMRI	200 subjects: 100 CN, 100 AD	Spatial normalization Skull-stripping Data augmentation	Slice-Based	Inception-ResNet		1.AD vs CN: 90.80% cMCI vs CN: 84.20% cMCI vs ncMCI: 69.10% 2.AD vs CN: 89.60% cMCI vs CN: 81.60% cMCI vs ncMCI: 70.00% 3.AD vs CN: 89.80% cMCI vs CN: 81.80% cMCI vs ncMCI: 70.40% 4.AD vs CN: 89.90% cMCI vs CN: 82.20% cMCI vs ncMCI: 71.20% 5.AD vs CN: 88.80% cMCI vs CN: 79.90% cMCI vs ncMCI: 69.80% 6.AD vs CN: 84.10% cMCI vs CN: 72.30% cMCI vs ncMCI: 61.10%
[66]	2020	ADNI	T1 sMRI	773 subjects: 162 CN, 234 ncMCI, 240 cMCI, 137 AD	Reorientation Cropping Skull-stripping Co-registration GM Segmentation	Slice-Based Whole 3D image		1.AlexNet 2.GoogleNet 3.ResNet50 4.ResNet101 5.InceptionV3 6.CNN	

Ref.	Year	Dataset(s)	Image Type(s)	Subjects	Data Preprocessing	Data Processing	Feature Extraction	Classifier(s)	Accuracy metric(s)
[86]	2020	ADNI AIBL OASIS	T1 sMRI	1951 subjects; 835 CN, 311 sMCI, 315 pMCI, 490 AD	Bias field correction Intensity rescaling Linear registration or Non-linear registration and skull-stripping	1.3D subject-based 2.3D ROI-based 3.3D patch-based 4.2D slice-based	CNN	1. AD vs CN: 77.33% (Baseline) sMCI vs pMCI: 60.50% (Baseline) AD vs CN: 79.66% (Longitudinal) sMCI vs pMCI: 61.50% (Longitudinal)	
[56]	2022	ADNI	T1 sMRI	819 subjects; 229 CN, 398 MCI, 192 AD	Non-uniform intensity correction CSF, WM and GM Segmentation Pixel values normalization	Slice-Based + ROI-based	1.CNN 2.VGG-16 3.ResNet-50	1. AD vs CN vs MCI: 72.70% 2.AD vs CN vs MCI: 78.57% 3.AD vs CN vs MCI: 75.71%	
[68]	2021	ADNI	T1 sMRI	557 subjects; 320 CN, 237 AD	Skull-stripping Linear registration Intensity normalization Data augmentation	Whole 3D image	1.Single time point CNN 2.Dual time point CNN(12 months)	1. AD vs CN: 86.00% sMCI vs pMCI: 2. AD vs CN: 88.70% sMCI vs pMCI:	
[81]	2020	ADNI OASIS Private databases	T1 sMRI	2727 subjects; 1432 CN, 743 MCI, 552 AD	Non-uniform intensity normalization (N3) correction Skull-stripping Pixel values normalization Data augmentation	Whole 3D image	deep CNN XGBoost	AD vs MCI: 79.53% MCI vs CN: 74.76% AD vs CN: 91.17%	GM.
[59]	2022	ADNI		520 subjects; 160 CN, 200 MCI, 160 AD	Slice timing Head movement correction Size scaling WM and GM Segmentation	Slice-Based + ROI-based	MSC-Net	AD vs MCI: 94.44% MCI vs CN: 90.74% WM. AD vs CN: 98.96% AD vs MCI: 95.37% MCI vs CN: 92.60%	

Ref.	Year	Dataset(s)	Image Type(s)	Subjects	Data Preprocessing	Data Processing	Feature Extraction	Classifier(s)	Accuracy metric(s)
[44]	2019	ADNI	T1 sMRI FDG-PET	1211 subjects: 731 CN, 441 sMCI, 326 pMCI, 647 AD	Reorientation Resample	Whole 3D image	VGG	MRI	CN vs AD: 81.19% Multi-modal CN vs AD: 90.10% CN vs pMCI: 82.38% sMCI vs pMCI: 72.22%
[11]	2022	ADNI	T1 sMRI	719 subjects: 245 CN, 229 sMCI, 245 AD	N4 bias correction Denoising Brain extraction Affine registration	Whole 3D image	1.DenseNet 2.EfficientNet	1. AD vs CN: 99.55% sMCI vs AD: 82.06% 2. AD vs CN: 91.55% sMCI vs AD: 81.38%	
[29]	2019	ADNI	T1 sMRI	811 subjects: 223 CN, 231 sMCI, 165 pMCI, 192 AD	Non-uniform intensity normalization (N3) correction Skull-stripping Rigid registration Hippocampus Segmentation Data augmentation	ROI-based + Patch-based	DenseNet	AD vs CN: 90.12% pMCI vs sMCI: 73.23% MCI vs CN: 73.02%	
[16]	2020	ADNI SNUBH	T1 sMRI	780 subjects: 390 CN, 390 AD	Skull-stripping Rigid transformation	Slice-Based	Inception V4	AD vs CN: 92.50%	ADNI-GO/2
[33]	2021	ADNI AIBL DELCODE	T1 sMRI Amyloid PET AV45-AIBL DELCODE	663 subjects (254 CN, 220 MCI, 189 AD)-ADNI-GO/2 575 subjects (326 CN, 187 MCI, 62 AD)-ADNI-3 606 subjects (448 CN, 96 MCI, 62 AD)-AIBL 474 subjects (215 CN, 155 MCI, 104 AD)-DELCODE	WM and GM Segmentation Spatial normalisation Data augmentation	ROI-based	CNN	MCI vs CN: 74.50% AD vs CN: 88.90% ADNI-3. MCI vs CN: 63.10% AD vs CN: 84.40% AIBL. MCI vs CN: 68.20% AD vs CN: 85.00% DELCODE. MCI vs CN: 71.00% AD vs CN: 85.50%	
[77]	2021	OASIS	T1 sMRI	741 subjects: 513 CN, 168 very-mild dementia, 55 mild dementia, 5 moderate dementia)	Normalisation Data augmentation	1.Slice-Based,d 2.Whole 3D image 3.Slice-Based	1.BrainNet2D 2.BrainNet3D 3.ResNet18	1.CN vs others: 86.50% 2.CN vs others: 82.00% 3.CN vs others: 87.00%	

Ref.	Year	Dataset(s)	Image Type(s)	Subjects	Data Preprocessing	Data Processing	Feature Extraction	Classifier(s)	Accuracy metric(s)
[92]	2019	ADNI Private database	T1 sMRI 18F-FDG PET	392 subjects: 101 CN, 200 MCI, 91 AD	Slice-Based		VGGNet-19	MRL	AD vs CN: 95.12% CN vs MCI: 83.24% AD vs MCI: 82.41% Multi-modal. AD vs CN: 95.89% CN vs MCI: 85.74% AD vs MCI: 88.20%
[13]	2021	ADNI	T1 sMRI	449 subjects: 500 CN, 500 MCI, 500 AD	Skull-stripping Spatially normalisation Linear registration	Slice-Based	Mp-CNN	CN vs MCI vs AD: 93.00%	
[73]	2022	ADNI	T1 sMRI	379 subjects: 146 CN, 138 MCI, 95 AD	Resizing Cropping Data Augmentation	ROI-based	1.DenseNet 2.PartialNet	1. AD vs CN vs MCI: 96.72% AD vs CN: 80.11% AD vs MCI: 96.65% MCI vs CN: 97.76% 2. AD vs CN vs MCI: 98.23% AD vs CN: 87.11% AD vs MCI: 99.26% MCI vs CN: 100.00%	
[93]	2021	ADNI OASIS	T1 sMRI	139 subjects: 55 CN, 84 AD	CSF, WM and GM Segmentation Linear registration Cropping	ROI-based + Patch-based	mi-GAN		
[89]	2022	ADNI	T1 sMRI	1125 subjects: 375 CN, 375 MCI, 375 AD	Template registration Skull-stripping Filter Denoising Intensity Normalisation	Patch-based	SE mechanism PSA mechanism +FCN	AD vs CN: 98.85%	AD vs CN: 98.85%
[28]	2022	ADNI	T1 sMRI T2 sMRI	830 subjects: 229 CN, 236 sMCI, 167 pMCI, 198 AD	Data augmentation	Whole 3D image	CAAE CRAT	AD vs CN: 91.90% AD vs MCI: 90.05% MCI vs CN: 88.10%	1. AD vs CN: 88.99% pMCI vs sMCI: 70.22% 2. AD vs CN: 85.01% pMCI vs sMCI: 68.49% 3. AD vs CN: 91.33% pMCI vs sMCI: 71.71%
[30]	2019	ADNI	T1 sMRI		Anterior Commissure (AC) - Posterior Commissure (PC) correction Skull-stripping Nonuniform intensity normalisation (N3)	ROI-based	1.CNN 2.RNN 3.CNN+RNN		
					CSF, WM and GM Segmentation Data augmentation				

Ref.	Year	Dataset(s)	Image Type(s)	Subjects	Data Preprocessing	Data Processing	Feature Extraction	Classifier(s)	Accuracy metric(s)
[63]	2020	OASIS	T1 sMRI	174 subjects: 87 CN, 87 AD	Field-bias intensity correction Spatial image normalisation Skull-stripping	Slice-based + Patch-based		SSA	Axial,CN vs AD: 87.50% Coronal, CN vs AD: 90.00% Sagittal, CN vs AD: 90.00%
[15]	2020	ADNI	T1 sMRI	3940 subjects: 2084 CN, 222 cMCI, 228 ncMCI, 1406 AD	Reorientation Cropping Padding Skull-stripping Correct intensity in homogeneity(N3)	Whole 3D image		ResNet29	ncMCI vs cMCI: 82.40%
[70]	2022	ADNI	T1 sMRI	238 subjects: 154 CN, 84 AD	ACPC alignment correction Intensity correction for uniform homogeneity Skull-stripping Registration	Whole 3D image	JD	CNN	AD vs CN: 96.61%
[35]	2021	ADNI AIBL	T1 sMRI	204 subjects: 65 CN, 56 EMCI, 40 LMCI, 43 AD	Skull-stripping Sampling Clipping Intensity normalisation	Whole 3D image		U-net	AD vs CN: 95.71% CN vs EMCI: 87.98% EMCI vs LMCI: 90.14% LMCI vs AD: 90.05%
[45]	2022	ADNI	T1 sMRI Amyloid PET Non-image information: Cognitive scores, Age and APOE type	417 subjects: 236 CN, 181 preAD	CSF, WM and GM Segmentation Normalisation Smoothing	Slice-Based + ROI-based	AlexNet ZFNet ResNet18 ResNet34 InceptionV3 Xception	SVM	AlexNet, CN vs preAD: 87.91% ZFNet, CN vs preAD: 87.91% ResNet18, CN vs preAD: 87.67% ResNet34, CN vs preAD: 89.53% InceptionV3, CN vs preAD: 84.88%
[57]	2019	ADNI BIOCARD	T1 sMRI	3304 subjects (At scan time - 1417 CN, 1146 MCI, 741 AD) 3304 subjects (Latest diagnosis - 1266 CN, 926 MCI, 1112 AD)	Skull-stripping Reorientation Intensity normalisation Inhomogeneity correction	Whole 3D image	LDDMM MALF	Siamse Net	At scan time,CN vs (MCI + AD): 91.03% Latest diagnosis,CN vs (MCI + AD): 92.20%
[48]	2021	ADNI GARD	T1 sMRI	188 subjects (74 CN, 40 MCI, 74 AD) - 123 subjects (42 CN, 39 MCI, 42 AD) -	Normalisation Registration CSF, WM and GM Segmentation Smoothing GARD			BR-FNN	ADNIR(ROI_CL). AD vs CN: 88.67% AD vs MCI: 72.86% ADNI (ROI_AAL). AD vs CN: 78.22% AD vs MCI: 80.57% GARD (ROI_CL). AD vs CN: 71.67% AD vs MCI: 59.75% GARD (ROI_AAL), AD vs CN: 75.60% AD vs MCI: 42.35%

Ref.	Year	Dataset(s)	Image Type(s)	Subjects	Data Preprocessing	Data Processing	Feature Extraction	Classifier(s)	Accuracy metric(s)
[20]	2021	GARD	T1 sMRI	252 subjects: 171 CN, 81 AD	Normalisation Data augmentation	Patch-based + Slice-based + ROI-based	DVE-CNN	DNN	L Hippocampus (L.H). AD vs CN: 76.26% Right Hippocampus (RH). AD vs CN: 77.34%
Legend:									

cMCI - Mild Cognitive Impaired converting to AD
ncMCI - Mild Cognitive Impaired not converting to AD
CN - Cognitively normal
EMCI - Early Mild Cognitive Impairment
LMCI - Late Mild Cognitive Impairment
MCI - Mild Cognitive Impairment
MID - Mild Demented
MOD - Moderate Demented
ND - Non-Demented
VMD - Very Mild Demented
preAD - Individuals at risk of Alzheimer's disease (according to standard uptake ratio >1.18 calculated by amyloid PET)
aAD - asymptotic Alzheimer's disease
SMC - significant memory concern
sMCI - stable MCI
pMCI - progressive MCI

It is very difficult to define which methods are best, given the different nuances that it is possible to create across the computational pipeline. However, in the reviewed articles, different evaluation metrics were used, such as ACC, sensitivity, F1 score, specificity, Precision, Recall, and AUC. In order to have a better view of the whole procedure adopted by the solutions with best results in terms of accuracy, since it is the metric adopted by all researchers, a more detailed description of the process applied by the authors of all articles found in this review whose prediction was higher than 95.00% is presented in the following and summarised in Table 4.5.

Sampath et al. [76] used 3D T1 weighted images acquired from 3 different databases: ADNI, OASIS and AIBL, which at the processing level underwent resizing in order to improve classification performance while using less memory, adaptive filtering to remove noise, adaptive histogram equalization for image enhancement and Voxel-based morphometry for segmentation of the ROI region, namely including grey matter, white matter, and cerebrospinal fluid. In this solution, the authors applied several feature extraction methods after the preprocessing phase, namely grey-level co-occurrence matrices, which extract the numerical features using spatial correlations of similar grey levels, Gabor filtering to extract energy-based texture features, and Wavelet ion order to extract the time-frequency representation. After that, a Hilbert Schmidt independence criteria loop method was introduced to reduce more irrelevant features among those extracted from the MR images. Finally, the new feature of investigating the food source direction of the fish shoal optimizer (FSO) is incorporated into the classification phase of the deep Siamese neural network (DSNN). The results obtained, with respect to Acc, were very promising, with 99.89% for the ADNI database.

Raghavaiah et al. [72] started from MR images and applied the following preprocessing techniques: skull removal, normalization, and GM, WM and CSF segmentation using a clustering method that combines a Temporally Consistent Black Widow Optimization (BWO) with a Fuzzy C-Means Clustering (FCM). Regarding feature extraction, the authors implemented a hybrid TECD approach merged with clinical data to add information concerning the patient's emotional condition. The TECD method characterises the input image features based on statistical properties of image colours (SPIC), grey level run length matrix (GLRLM), local discriminative powerful binary patterns (DPLBP) and tube density features by calculating each pixel's modified probability depending on its neighbours, and the clinical features include Functional Activities Questionnaire (FAQ), Neuropsychiatric Inventory (NPI) and geriatric depression scale (GDS). This feature vector is given as input for the proposed Deep Rotation Forest Neural Network to perform the classification task. Thus, the rotation forest creates the training data for the deep enhanced stacked auto encoder (DESAE) with back propagation learning classify algorithm.

Goenka et al. [37] used 3D T1 sMRI, on which the authors applied bias correction, skull-stripping and rigid registration. Given the reduced size of the used dataset, techniques for data augmentation were implemented, namely: - 5 and 5 degree angle rotations. Next, three ConvNets were created for three different input data processing approaches: hole 3D Image, 3D Patches and Slices. In the context of three-class categorization, a 14-layer architecture was employed with seven convolutional layers, four max-pooling layers, batch normalization, one Global Average

Pooling layer, and two dense layers. The same architecture with 14 layers was used in the 3D-Patch Based model, with a 72x72x72 patch. The torch unfold function, which extracts sliding blocks from a batched input array, was employed to create these patches through a non-overlapping strategy. And finally, the 3D-Slice level ConvNet uses only 13 layers, with slight differences to as other two networks. The best results were achieved for the methodology applied to the whole image, with an accuracy of 99.10% in the classification between CN and MCI.

Hazarika et al. [39] converted the 3D T1 weighted MR images into a group of 2D slices and identified the most suitable slices that can provide the regions of the hippocampus. Since the skull part is ignorable, skull-stripping and size scaling were applied. In this way, these images served as input for a LeNet network; however, instead of the traditional MaxPooling layer, the authors implemented a MinPooling layer to also evaluate the low intensity pixels. This allowed the model to achieve high performances, with an average accuracy of 96.64%.

Faisal et al. [34] started by decreasing the size of the 3D T1 weighted MR images. The images were then divided into the 3 slices: axial, coronal and sagittal, with those at the beginning and end omitted as they contained no useful information. In addition, the slices were normalized with a mean and standard deviation of zero and one, respectively. From this, the authors proposed a straightforward but powerful convolutional method (2D-CNN) that at once carries out standard convolution, deep convolution, and point convolution, followed by a jumping convolution layer to learn multi-level characteristics from brain MRI data. The great feature of this model is the reduced amount of parameters and consequently the computational weight and speed of the model, while achieving good results, namely 96.12% classification accuracy between CN, MCI and AD classes.

Qasim Abbas et al. [70] adopted four standard operations for preprocessing: ACPC alignment correction for identical orientation, skull removal to eliminate non-brain tissue, intensity correction for uniform homogeneity and image registration for geometric alignment, in 3D T1 weighted sMRI. The Jacobian domain is then applied to the previously preprocessed images to create a Jacobian determinant map, which is then utilised to train a CNN model. The proposed architecture uses a sequential model, consisting of one input, three convolutional, three max-pooling, one flattened, one fully connected and one output layer/s. The number of layers is determined by the effectiveness of the validation process. Thus, the results obtained by this domain Jacobean convolutional neural network (JD-CNN), in terms of accuracy, it was equal to 96.61% in the binary classification between CN and AD.

Yan et al. [89] used MR-weighted 3D T1 weighted images, which in the processing underwent affine transformation, skull removal, filter denoising, including median filtering, Gaussian blur filtering and anisotropic diffusion filtering, which was the one that showed the best results, and intensity normalization. After that, the SE and the PSA mechanisms were added to significantly improve the performance of the CNN. The SE mechanism obtains global data for each feature map by clustering the global average and then a fully connected layer is implemented to identify the feature maps' global dependencies. Whereas, the PSA mechanism can manipulate spatial data of input multi-scale feature maps, and can successfully create long-term dependence among multi-

scale channels' attention. Next, a Fully Convolutional Network (FCN) model was implemented, which consists of four convolutional blocks and two fully connected layers. A 3D convolutional layer, 3D maxpool layer, 3D batch normalization, Leaky ReLU, and Dropout are the convolutional block's components. The final fully connected layers play a vital part in increasing the model's effectiveness, and the model is trained by random initialization of weights. The authors adopt a patch (47x47x47 in size) random sampling method from MRI scans to train the FCN model. In addition, we also construct the MLP model structure, which consists of two fully connected layers, batch normalization, Leaky ReLU, and Dropout. In the MLP model's image classification test, the authors pick the probability value of AD from disease probability map data, choose the ROI according to the MCC heat map of FCN model, and match with age, MMSE, and gender of the patients. The accuracy obtained by the used pipeline was 98.85% for the distinction between CN and AD.

Table 4.5: Summary of the best approaches found in this reviewed.

A.	Dataset(s)	Image Type	Data Preprocessing	Data Processing	Feature Extraction + Segmentation	Classifier	ACC
[76]	ADNI AIBL OASIS	T1 sMRI	Resizing Adaptive filtering for noise removal Adaptive histogram equalization	Whole 3D image	GLCM Gabor filter Wavelet features HSICL	ADNI. CN vs SMC vs MCI vs AD: 99.89% AIBL. FSODSNN based classifier CN vs SMC vs MCI vs AD: 99.67% OASIS.	ADNI. CN vs SMC vs MCI vs AD: 99.89% AIBL. FSODSNN based classifier CN vs SMC vs MCI vs AD: 99.67% OASIS.
[77]	ADNI		Skull-stripping Normalization TC-BWO-FCM for CSF, WM and GM Segmentation	ROI-based	TECD feature extraction + Clinical features	HRE-DNN	AD vs CN: 98.68% MCI vs AD: 95.88% MCI vs CN: 97.23%
[37]	ADNI	T1 sMRI	N4 Bias Field Correction Skull Stripping Rigid Registration	Whole 3D image		ConvNet	AD vs CN: 97.83% AD vs MCI: 98.68% CN vs MCI: 99.10% CN vs MCI vs AD: 98.26%
[39]	ADNI	T1 sMRI	Size scaling Skull Stripping	Slice-Based		LeNet	AD vs CN: 95.00% AD vs MCI: 97.00% MCI vs CN: 97.00%
[89]	ADNI	T1 sMRI	Template registration Skull-stripping Filter Denoising Intensity Normalisation	Patch-based	SE mechanism PSA mechanism + FCN	MLP	AD vs CN: 98.85%
[70]	ADNI	T1 sMRI	ACPC alignment correction Intensity correction for uniform homogeneity Skull-stripping Registration	Whole 3D image	ID	CNN	AD vs CN: 96.61%

4.4 Summary

From this literature review, it could be realized that the AD classification with neuroimaging is the subject of a significant number of studies based on different deep learning approaches. This review was focused on studies that at least used MRI data, which led to the exclusion of some studies included in the initial search. Nonetheless, in practice, other imaging techniques are less common in the clinical environment owing associated high costs.

As it was possible to perceive, there are several brain image datasets for the study of Alzheimer's, but almost unanimously the ADNI dataset is the most used, providing greater variety and quantity of samples.

Although, it is undisputed that classification between AD and CN is the most straightforward task leading to high accuracies, reaching almost 100% in some studies. Therefore, it is not so clinically relevant as opposed to the prediction of sMCI to CN and pMCI to AD that are more desirable, but further challenging to reach good results. Regarding the implemented models, the results obtained in 2D or 3D approaches are similar, not being possible to conclude which is better. Both present limitations and advantages. 3D models represent more arduous challenges due to the increased number of parameters and computational cost, while in 2D approaches, not all information can not be used as spatial relations are discarded.

Given the data approach adopted, most researchers opted for direct classification, with CNNs being the most predominant architecture. However, other methodologies that show potential to be explored, such as the RNN are used in the longitudinal data processing to extract the minimal differences between consecutive image scans on the same subject. When comparing this approach with feature extraction plus classification, it is difficult to draw many conclusions. However, it is possible to see that it does not depend so much on the approach adopted, but rather on the models, parameters, data treatment, etc. Time-to-Event Prognostic approaches are still little explored, with very few studies on the subject, but they have a promising future since they work in the area of prediction.

Data augmentation and transfer learning techniques are widely used by researchers, and allow to obtain promising performances even using small amounts of data.

Finally, by evaluating the best results presented in the reviewed articles, it was not possible to find a common characteristic to all, proving that in this area several paths can be adopted and still produce promising results.

Chapter 5

Methodology

Following an overview of the background information regarding Alzheimer's disease, magnetic resonance imaging, and the deep learning techniques used to achieve the disease's classification into various stages, this chapter outlines the implemented methodology from the initial data collection to the final output. The theoretical framework of the DL algorithms implemented is also explored in greater detail. Then, some managed computational resources are introduced, along with the computational infrastructure.

5.1 Overview

The objective of this experimental study is to develop a computational platform for Alzheimer's disease diagnosis based on deep recurrent networks. In this sense, the development of a CAD system that incorporates the steps of image dataset, preprocessing, and classification was analysed and developed based on MR images. The scheme in Figure 5.1 provides an overview of the steps taken, which are then discussed in more detail in the following sections.

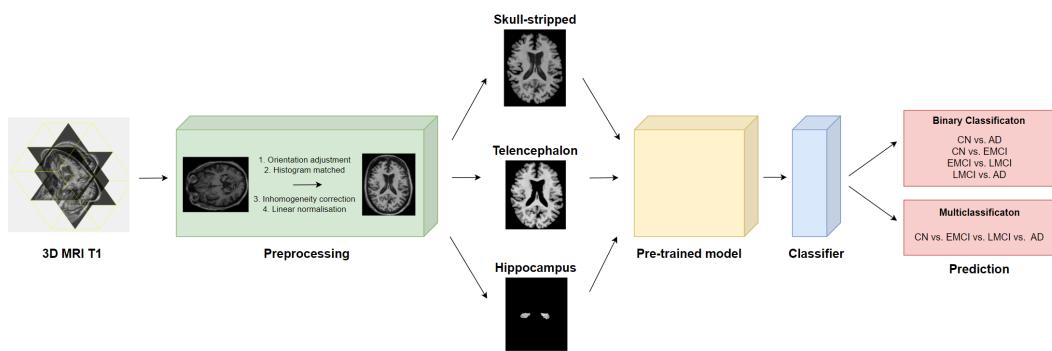


Figure 5.1: Proposed CAD system for AD diagnosis. It is divided into 3 steps: definition of the image type - 3D MRI T1; pre-processing, which encompasses the 4 techniques represented plus the segmentation phase; classification process, which is characterized by the compilation of the pre-trained network plus the classification block. The final output is the stage of AD.

5.2 Used Image Dataset

The dataset adopted in the project was gathered from the Alzheimer’s Disease Neuroimaging Initiative. The ADNI study’s purpose is to analyse the brain’s structure and function across the course of different disease states using biomarkers and clinical measurements to track illness progression [8]. This dataset was the most reported in the state of the art by far, and it is described in Section 4.2.2.1. From the phase ADNI2, 1140 T1-weighted MR images were acquired, including 286 CNs, 283 EMCIs, 287 LMCI, and 284 ADs. Table 5.1 illustrates the demographic and clinical information about the subjects used.

Table 5.1: Summary of the studied subjects.

Diagnosis	Number	Age	Gender(F/M)
CN	286	74.01 ± 20.01	146/140
EMCI	283	70.52 ± 20.48	134/149
LMCI	287	71.42 ± 20.58	150/137
AD	284	74.46 ± 18.46	118/166

In terms of acquisition procedure, these images were acquired in the sagittal plane using volumetric 3D MPRAGE with $1.00 \times 1.00 \text{ mm}^2$ spatial resolution and 1.2 mm of slice thickness. The scanners used had a field strength of 3T. The ADNI website has more detailed information about MR acquisition protocols.

5.3 Image Preprocessing

Regarding the preprocessing step, all the images acquired were processed and parcellated by MRI-Cloud. The MRICloud is a completely automated cloud service based on the Multiple-Atlas Likelihood Fusion algorithm, JHU multi-atlas inventories with 286 defined structures, and Ontology Level Control technology [83, 65, 32]. This software allows brain segmentation of MPRAGE images, namely T1-weighted sMR images. In the pipeline, the raw images were automatically preprocessed using five techniques: (1) orientation adjustment, (2) histogram matched, (3) inhomogeneity correction, (4) linear normalisation in the MNI space, and (5) skull-stripping, before the segmentation step. The output is an image with a matrix size of $181 \times 217 \times 181 / 1 \times 1 \times 1 \text{ mm}$. In this study, a set of 31 atlases with an age range of 50-90 years old were used to parcellate 287 brain structures (so-called Adult50_90yrs_287labels_31atlases_V10B). All the segmented structures are hierarchically organised into 5 levels, where each structure is also divided into right and left sides, as shown in Figure 5.2.

For this work, in addition to the skull-stripped image, an image with the region of interest (ROI) of the Telencephalon and Hippocampus were also studied. Figure 5.3 represents an example of the various steps performed in this phase, with the following images illustrated: original image,

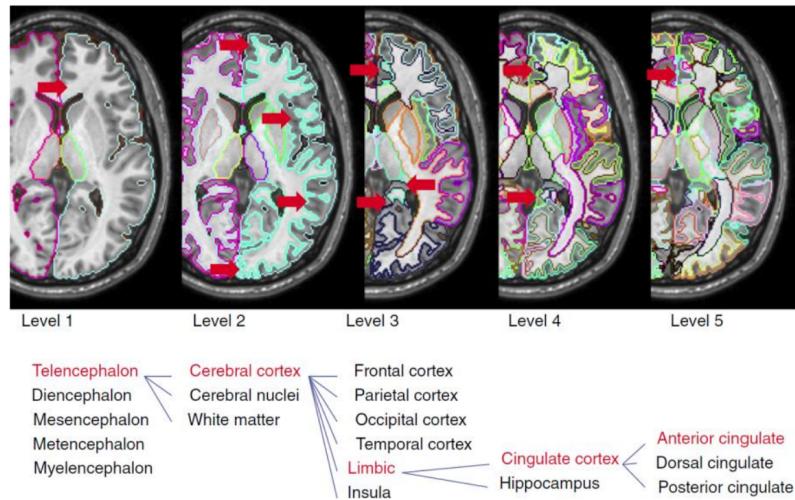


Figure 5.2: Five-level hierarchical structures of the brain [83, 65, 32].

processed image (techniques 1 to 4 mentioned above), skull-stripped image, telencephalon ROI image, and hippocampus ROI image. After segmentation, the images were further normalised in the range [0, 1] and resized to 96x128x128. While MRIs store data in 3-dimensional single-channel volumes, pre-trained models are made to operate on three channels. To fully explore the potential of transfer learning via pre-trained 3D architectures, the MR images were converted from one to three channels.

Finally, the dataset was split into an 80%/20% ratio for training and testing, respectively, and then the training set was again split into 80%/20% for training and validation, respectively. Thus, 64% of a dataset was used as the training set, 16% as the validation set, and 20% as the testing set by random sampling.

5.3.1 Data Augmentation

In order to reduce the general issue of the small dataset, which is overfitting during training, the data augmentation technique was applied. Thus, the detailed data transformation techniques, such as rotation, flip, elastic transformation, and adding Gaussian noise are introduced in this part. Figure 5.4 shows an example of each of these transformations.

- **Rotation** - The image is rotated to the right or left along an axis between 1° and 359° . The rotation degree parameter plays an important role in determining the safety of rotation augmentations. Thus, random rotations between -15° and 15° along the z axis were applied.
- **Flip** - This is a spatial transformation technique that inverts an image along a certain axis. In this work, this technique was performed along the x axis.

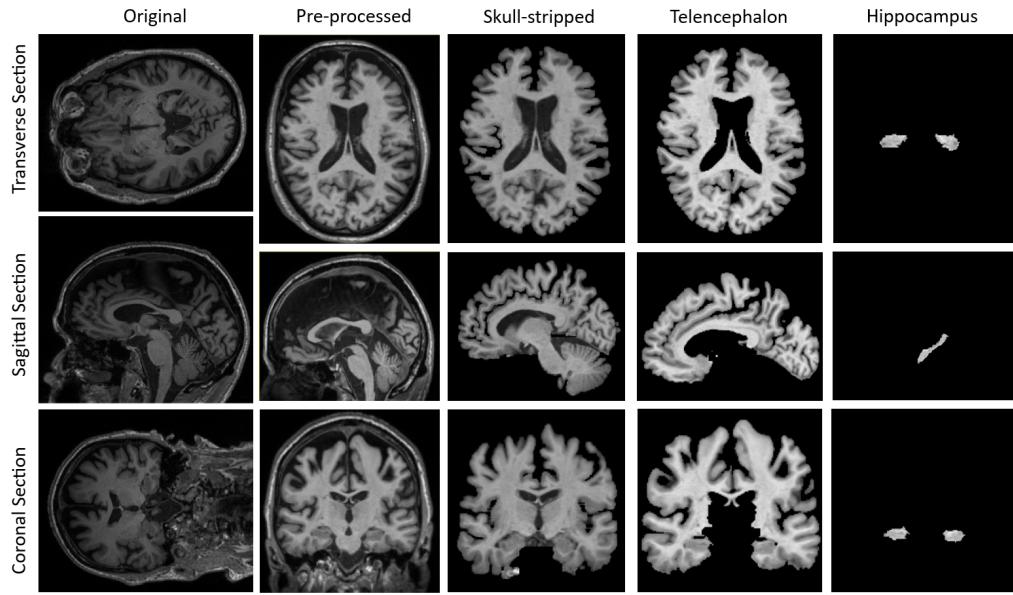


Figure 5.3: Data preprocessing techniques. The transverse section is shown in the top row, followed by the sagittal section and the coronal section. Each column represents the image after one processing step.

- **Elastic transformation** - This technique focuses on random elastic deformations in the shape, geometry, and size of an object. Here the deformation parameters were set between 0 and 0.25 with an interpolation of 1.
- **Gaussian Noise** - The process of "noise injection" entails injecting a matrix of random values, often selected using a Gaussian distribution. The added Gaussian noise had as parameters a mean of 0 and a variance of 0 to 1.

5.4 Predictiton Model

As laid out in Section 3.4.1, a whole deep-learning architecture from scratch (with random initialization) could not be trained due to the very little amount of data provided in this study compared to the number of samples required to train a CNN model that is stable, unbiased, and not overfitted. So, as was described in Section 3.5, transfer learning can be used to combat this problem.

Pre-trained 2D models in the ImageNet dataset have previously been demonstrated in several studies to have great potential for transfer learning to different domains with higher accuracy and/or faster convergence, including applications for medical image analysis. However, the problem addressed in this paper requires 3D inputs, so the study developed by R. Solovyev et al. was adopted, which took advantage of pre-training on large-scale 2D image datasets for 3D image analysis [80]. Thus, R. Solovyev et al. developed a method for transferring weights from pre-trained 2D CNNs in the ImageNet dataset to 3D CNNs. The ImageNet dataset, which contains

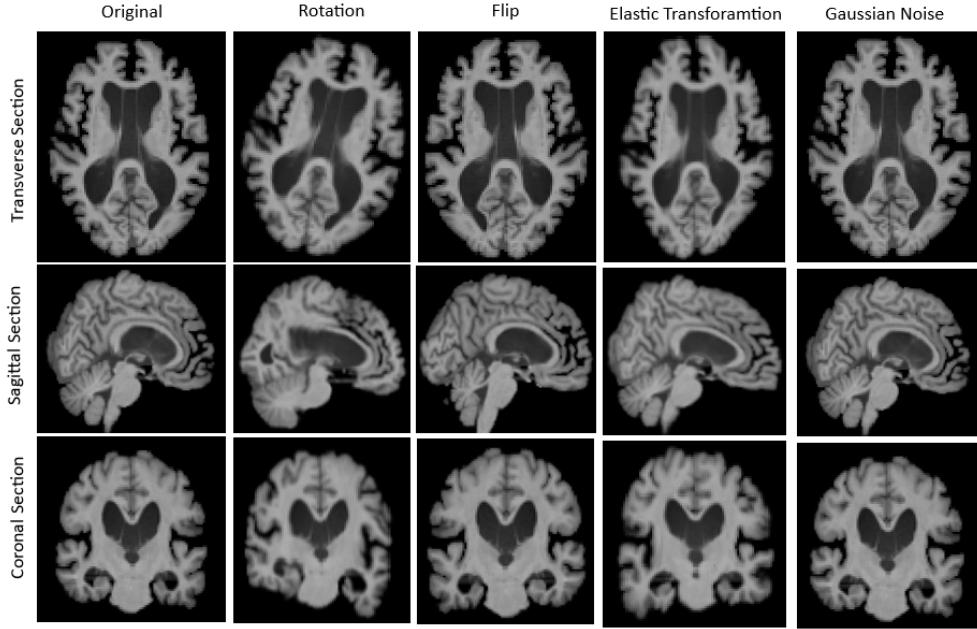


Figure 5.4: Data augmentation methods. The transverse section is shown in the top row, followed by the sagittal section and the coronal section. Each column represents an image that has undergone a type of data transformation.

more than a million images, were used to train these 2D models to categorise 1000 different image classes [80]. The following pre-trained networks were used:

- **DenseNet** - A Densely Connected Convolutional Network is a type of convolutional neural network that makes use of dense connections between layers by connecting all layers (with matching feature-map sizes) directly with one another using Dense Blocks. Each layer receives additional inputs from all earlier layers and transmits its own feature-maps to all later layers in order to maintain the feed-forward nature of the system. It is used for various tasks, however, considering published articles, it predominates in the area of image classification [43]. Supplementary Figure 5.5 shows the detailed architecture of the DenseNet implemented in the project, which is basically the same as the one proposed by Huang et al., where the 2D convolution and pooling layers were modified to 3D ones [80]. Three variants of this architecture were explored in this study, namely DenseNet121, DenseNet169 and DenseNet201. The differences between the three are apparent in the number of Conv_Blocks implemented within each Dense_Block, as represented in Figure 5.5. Thus, (a,b,c,d) corresponds to (6, 12, 24, 16) for DenseNet121, to (6, 12, 24, 16) for DenseNet169, and to (6, 12, 48, 32) for DenseNet201.
- **ResNet** - A Residual Network is a type of convolutional neural network that uses residual connections. These connections allow information to move straight from one layer to

DenseNet

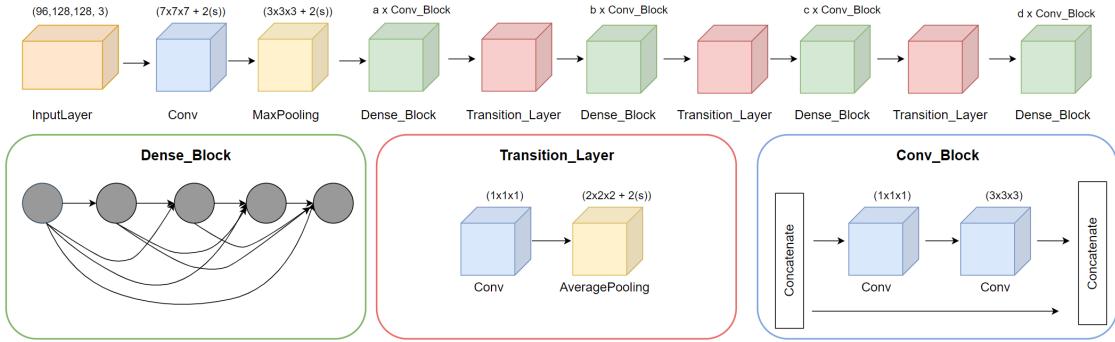


Figure 5.5: Architecture of 3D DenseNet (Convolutional base). Input dimensions correspond to depth \times height \times width \times channels, and a,b,c,d defines the architecture variant.

another, avoiding one or more intermediate levels. This makes it easier to train very deep neural networks with hundreds or even thousands of layers and helps to solve the vanishing gradient problem. In numerous computer vision applications, including image classification, ResNets have attained state of the art performance [40]. Figure 5.6 shows the detailed architecture of the ResNet implemented in the project. There are several variants within ResNets, such as ResNet50, ResNet101, ResNet152 and others. The difference between them is determined by the number of residual blocks they use. Thus, as shown in Figure 5.6, variables (a,b,c,d) correspond to (3, 8, 36, 3) for ResNet152, which was implemented in this work.

ResNet

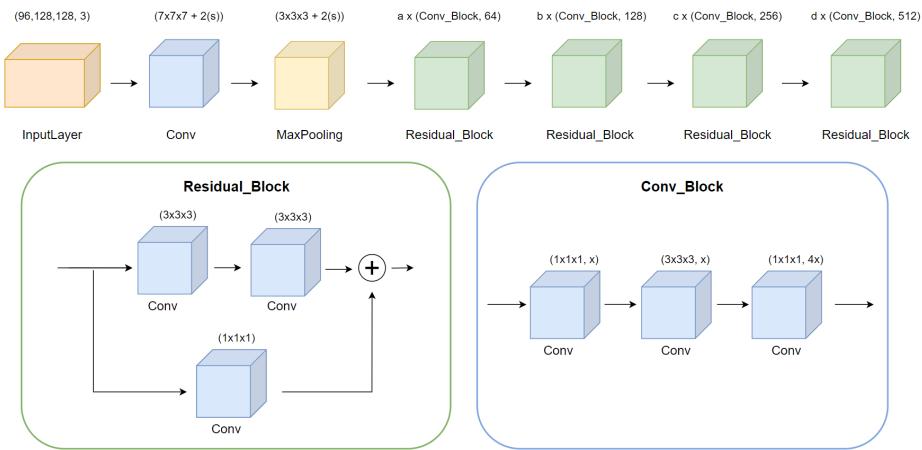


Figure 5.6: Architecture of 3D ResNet (Convolutional base). Input dimensions correspond to depth \times height \times width \times channels, and a,b,c,d defines the architecture variant.

- **ResNeXt** - A ResNeXt is a type of deep neural network that is based on the ResNet architecture and uses a novel block design that enhances the network's capacity to capture complex features in the input data to increase the representational power of the network. A "cardinality" block, which is a type of construction block in the ResNeXt architecture, is made up of parallel convolutional layer sets. The outputs of these layers are merged using concatenation or summation to create the block's final output. Each of these layers learns a unique collection of features from the input. The number of parallel convolutional layers (C) and the width of the output channel dimension are both determined by the block's cardinality [87]. Figure 5.7 exemplifies the architecture of the ResNeXt implemented in the project. Like the ResNets, this model typology also has several variants, such as ResNeXt50, ResNeXt101, and ResNeXt152. The ResNeXt101 model, where (a,b,c,d) correspond to (3, 4, 23, 3) and C equals 32, was implemented in the current work.

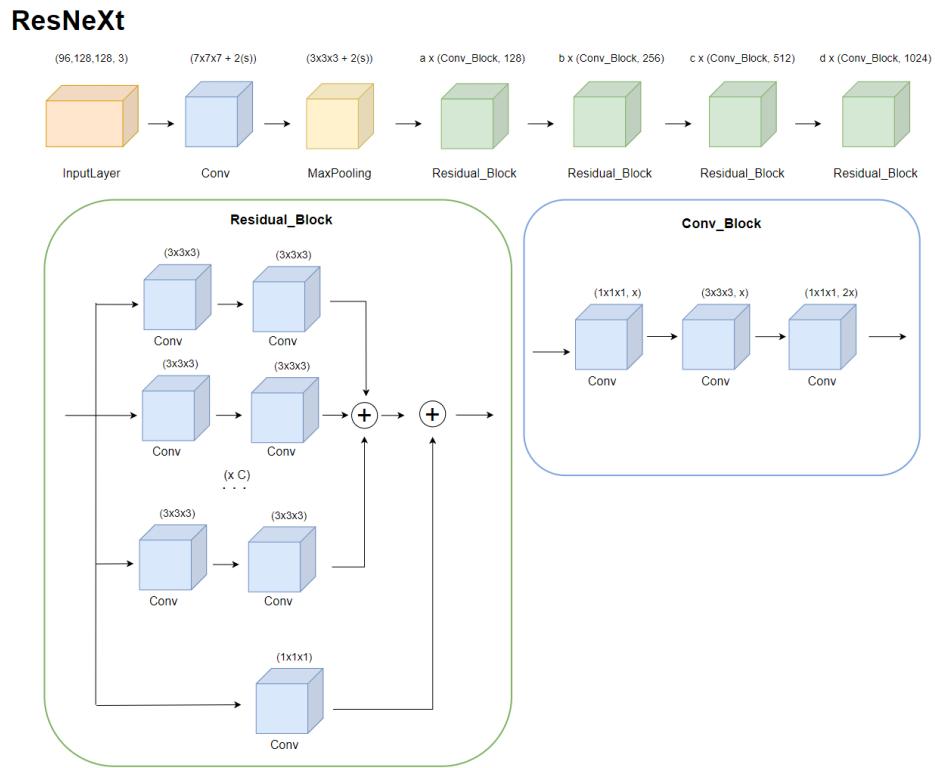


Figure 5.7: Architecture of 3D ResNeXt (Convolutional base). Input dimensions correspond to depth × height × width × channels, and a,b,c,d and C defines the architecture variant.

- **SENet** - A Squeeze and Excitation Network is a type of convolutional neural network that enables the network to dynamically recalibrate features channel-wise via squeeze and excitation blocks. The SENet model consists of a series of convolutional layers, followed by a squeeze and excitation block. The squeeze operation aggregates information from multiple channels and reduces the feature maps' spatial dimensions. The feature maps are

then rescaled using a set of channel-wise weighting factors that the excitation operation learns. The network's subsequent layer receives the generated feature maps. As a result, the network is better able to perform well on a variety of image classification tasks by concentrating on the most informative channels and suppressing noisy or irrelevant ones [42]. The Squeeze-and-Excitation (SE) block can be implemented in multiple CNN architectures like ResNet or ResNeXt. Therefore, Figure 5.8 represents the architecture of the SEResnet model, where a, b, c, d represent the variant. From this, SEResNet101 with (a, b, c, d) equal to (3, 4, 23, 3), and SEResNet152 with (a,b,c,d) equal to (3, 8, 36, 3) were implemented. Hu et al. [42] also developed a modified version of ResNeXt152 called SENet154, whose architecture is depicted in Figure 5.10.

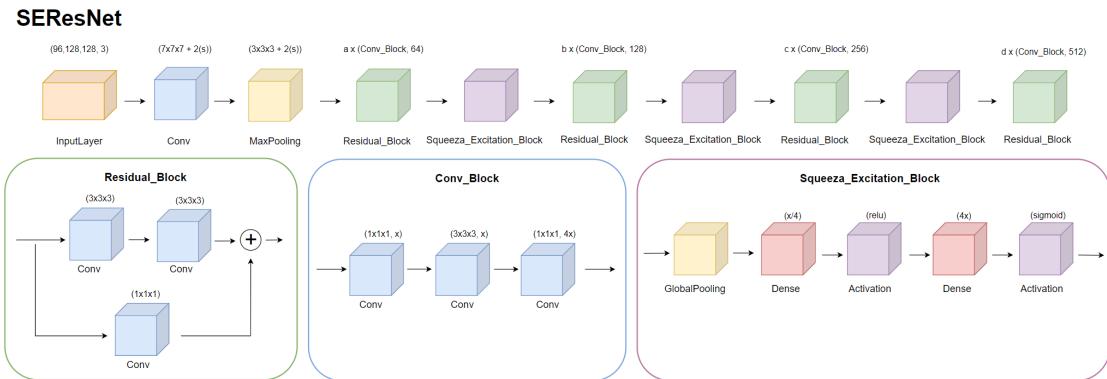


Figure 5.8: Architecture of 3D SEResNet (Convolutional base). Input dimensions correspond to depth \times height \times width \times channels, and a,b,c,d defines the architecture variant.

The defined classification block was implemented for all the presented architectures and was added to the end of them. Figure 5.9 represents the layers used: a GlobalAveragePooling, a Dropout of 0.5, a Dense and a sigmoid Activation. As a result, a binary or multiclassification prediction was obtained.

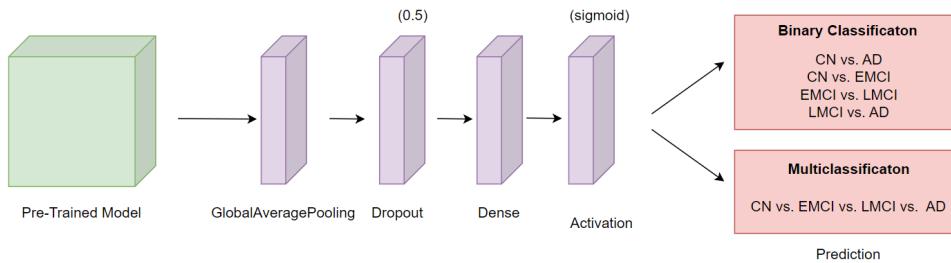


Figure 5.9: Classification block.

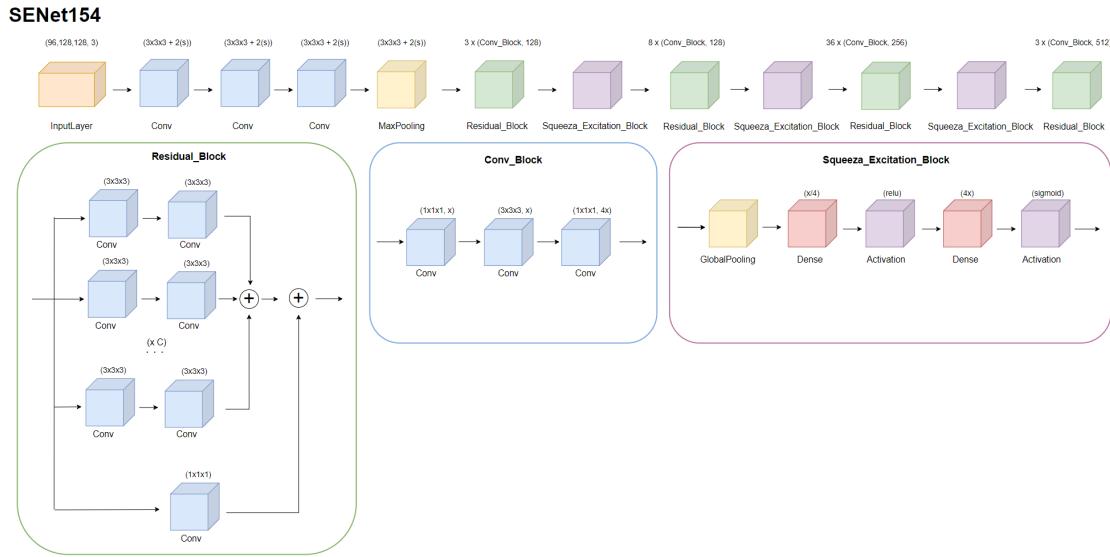


Figure 5.10: Architecture of 3D SENet154 (Convolutional base). Input dimensions correspond to depth × height × width × channels.

5.5 Computational Infrastructure and Implemented Resources

The hardware and software tools used throughout the creation of this project are described in the following section.

5.5.1 Computational Infrastructure

For the development of this work, a computer Surface Laptop 3 with a processor i5-1035G7 CPU at 1.20 GHz and 8 GB RAM was used. Microsoft Windows 11 Home is the 64-bit operating system installed on this computer and is compatible with all computational tools used during the preparation phase. In addition, the training of the DL network was carried out on a Linux computer equipped with a NVIDIA DGX with 4 Teslas V100 (32GB) with Intel Xeon (40 cores) and 256GB of memory.

5.5.2 Development language

Python is a high-level programming language whose design philosophy prioritises code readability through the use of off-side rule-based significant indentation. It supports a variety of programming paradigms, including object-oriented, structured, and functional programming. It is one of the most widely used programming languages.

Python is a useful tool in the medical field that can be incorporated into a number of steps, including medical imaging processing and segmentation, data extraction, statistical analysis, and a gateway to ML and DL. In addition, it has multiple libraries that can be used in this area and are very easy to integrate, such as Nibabel, SimpleITK, Pydicom, and more.

5.5.3 Visual Studio Code and Google Colaboratory

Visual Studio Code, or VS Code, is a source code editor created by Microsoft that includes multiple features like debugging, syntax highlighting, intelligent code completion, and more. It is a very versatile software and can be used with different programming languages, including C, C++, Python, Java, and so on. This tool was incorporated into the project to create and edit the script uploaded to the Linux machine.

Furthermore, to evaluate the results, the Google Colaboratory ("Colab" for short) was also adopted. It is a data analysis and machine learning tool that enables the incorporation of rich text, charts, images, executable Python code, HTML, LaTeX, and more into a single Google Drive document. Colab is a hosted Jupyter notebook service that offers free access to computing resources, including GPUs, and requires no setup to use. For this reason, it was also used at an early stage for small tests with simple networks.

5.5.4 TensorFlow and Keras

TensorFlow is an open-source software library for ML. It can be applied to many tasks, although it focuses mostly on deep neural network training and inference. This tool was developed by the Google Brain team, with the latest version being TensorFlow 2.0, released in September 2019. Python, JavaScript, C++, and Java are just a few of the programming languages that support TensorFlow. Due to its versatility, it can be used in a wide variety of sectors [2].

On top of the machine learning framework TensorFlow, Keras is a deep learning API created in Python. Keras is an interface to the TensorFlow library, designed to allow rapid experimentation with deep neural networks. Its main goals are to be easy to use, modular, and extensible. It was created as a component of the ONEIROS (Open-ended Neuro-Electronic Intelligent Robot Operating System) research project, and François Chollet, a Google engineer, is its main author and maintainer. To make working with image and text data easier and to streamline the coding required to create deep neural networks, Keras includes numerous implementations of widely used neural-network building blocks like layers, objectives, activation functions, and optimizers [1].

5.5.5 ROIEditor

ROIEditor is one module of the MRI Studio, which is an image processing programme. This module comprises a set of tools that allow the user to define and/or manipulate regions of interest on various types of MR images. Some of the tools available are the Image viewer (3D visualisation), ROI drawing, selection and manipulation, and ROI statistics. The interface of this programme is shown in Figure 5.11 [41].

5.6 Summary

The description of the used image dataset, image preprocessing, and prediction model were the three processes that are highlighted in this chapter's description of the project's pipeline. The

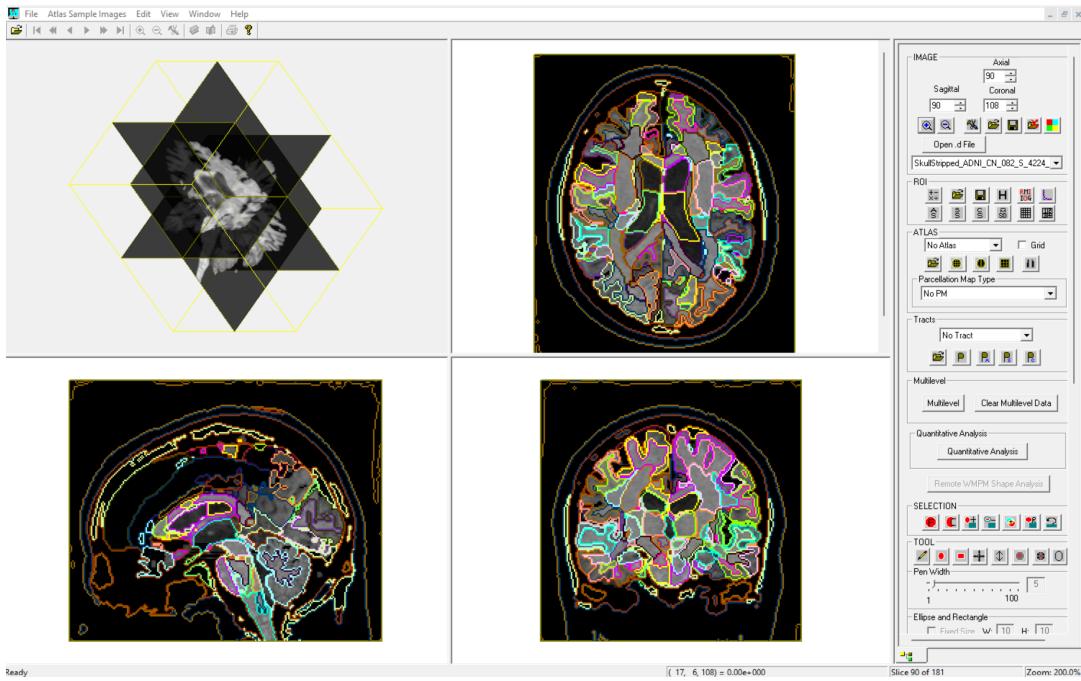


Figure 5.11: ROIEditor interface.

dataset adopted in the project was gathered from the Alzheimer's Disease Neuroimaging Initiative, and from the ADNI2 phase, 1140 T1-weighted MR images were acquired. The preprocessing was achieved through the MRICloud software with five techniques: orientation adjustment, histogram matched, inhomogeneity correction, linear normalisation in the MNI space, and skull-stripping. After that, the segmentation of two more regions, namely the telencephalon and hippocampus, was performed. And finally, the images were normalised, resized, and converted to three channels. Still in the preprocessing step, and in order to combat the issue of a small dataset, multiple data transformation techniques were experimented with. In the final step, the prediction model, multiple 3D pre-trained architectures were explored, such as DenseNet, ResNet, ResNeXt, SENet, and SEResNet. The differences and details of each type are presented in the sections above. In addition, and as an important component of the study, the computational resources used throughout the development were described.

Chapter 6

Results and Discussion

In this chapter, a significant number of tests were implemented to evaluate the efficacy of the key elements of the framework adopted, including the size of the dataset, the augmentation techniques, the number of layers fine-tuned in the pre-trained model, the architecture, and the region studied. To do this, all the metrics described in Section 3.6 were checked. Finally, after the conclusions drawn from these studies, the proposed model was performed in binary classifications for AD versus CN, CN versus EMCI, EMCI versus LMCI, and LMCI versus AD, and in a multiclassification for CN versus EMCI versus LMCI versus AD, and then compared with methods already described in the state of the art.

6.1 Experimental Settings

Due to the significance of hyperparameters in neural network training, Table 6.1 includes a list of all the hyperparameters used in the models implemented in this study. It is important to note that, as mentioned in Section 3.5, the first step for transfer learning is to freeze all layers (convolutional base) and train only the classifier for the new dataset. So in this process, 20 epochs and a learning rate of 0.0001 were set. The second step is to unfreeze some layers and fit the model using a smaller learning rate, so 100 epochs and a learning rate of 0.00001 were set.

Besides that, to facilitate model training, three types of "callbacks" in Keras were implemented during training: *EarlyStopping*, *ModelCheckpoint* and *ReduceLROnPlateau*. *EarlyStopping* allowed the model to stop training if its performance did not improve after fifteen iterations by tracking validation loss. In this way, it is possible to control/avoid overfitting. *ModelCheckpoint* made sure that, in order to prevent loss of progress, the model always saved the best weights while training. Finally, *ReduceLROnPlateau* allows reducing the learning rate when a metric has stopped improving. So the validation loss is monitored, and if it does not improve after three epochs, the learning rate is reduced by a factor of 0.95. All the networks used the same training strategy and hyperparameter settings.

Table 6.1: Hyperparameters adopted in the experiments.

Hyperparameter	
Number of epochs	20 (freeze) + 100 (fine-tuned)
Batch size	3
Optimizer	Adam
Learning Rate	0.0001 (freeze) and 0.00001 (fine-tuned)
Loss function	Cross-Entropy Loss
Metrics	Accuracy

6.2 Effects of Different Dataset Sizes

This experiment discussed the impact of the dataset's sample size on classification performance. For this, the CN versus AD classification task was run for the dataset presented in Section 5.2 that includes longitudinal data, and for a dataset with fewer samples, i.e., baseline data (103 AD and 140 CN). For this, both were tested on a pre-trained DenseNet201 model and the results are visible in Table 6.3 (in terms of ACC, loss and AUC), and in Figure 6.1 (in terms of Confusion Matrix, ROC curve, and radar chart with the metrics Precision, Recall, F1-score, Sensitivity and Specificity).

Table 6.3: Performance of different dataset sizes.

Dataset	Size (per class)	Size (per folder)	ACC (%)	Loss	AUC (%)
Baseline Data	AD: 103 CN: 140	Train: 155	71.42	0.6280	70.51
		Validation: 39			
		Test: 49			
Longitudinal Data	AD: 284 CN: 286	Train: 364	83.33	0.4430	89.48
		Validation: 92			
		Test: 114			

As expected, there is a significant improvement (+11.91 % in ACC) in test performance when increasing the number of samples used. Directly related and as a complementary consequence, in Figure 6.2 a reduction in overfitting is also noticeable with increasing data. Therefore, in all subsequent studies, the longitudinal dataset was used.

6.3 Effects of Different Data Augmentation Methods

The performance of four different data augmentation techniques was analysed, as presented in Figure 5.4, through the AD versus CN classification task. In this experiment, the model used

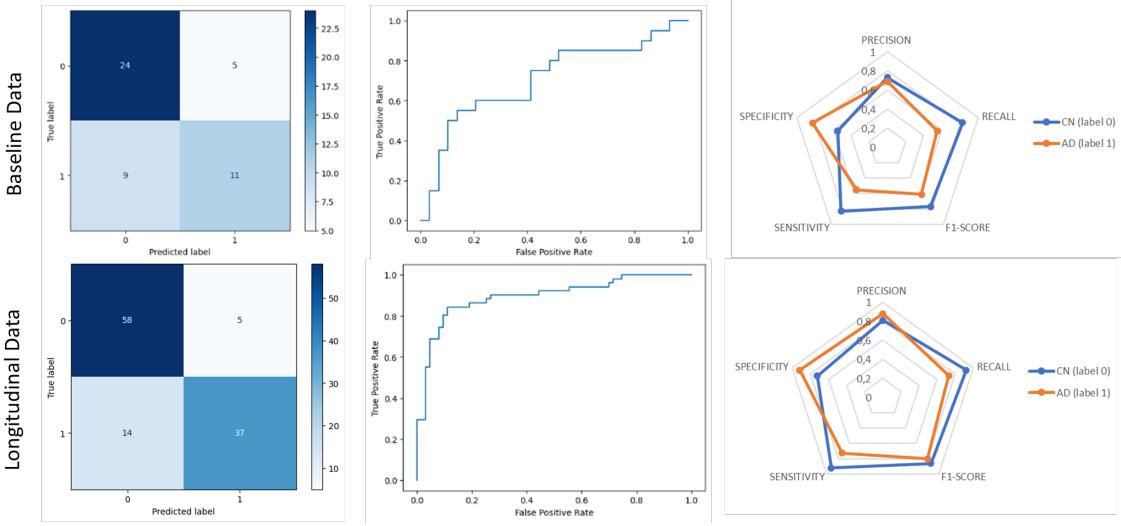


Figure 6.1: Performance comparison of the two dataset sizes, namely the Confusion Matrix, the ROC curve, and the radar chart with the metrics Precision, Recall, F1-score, Sensitivity and Specificity for each class (CN - label 0; AD - label 1).

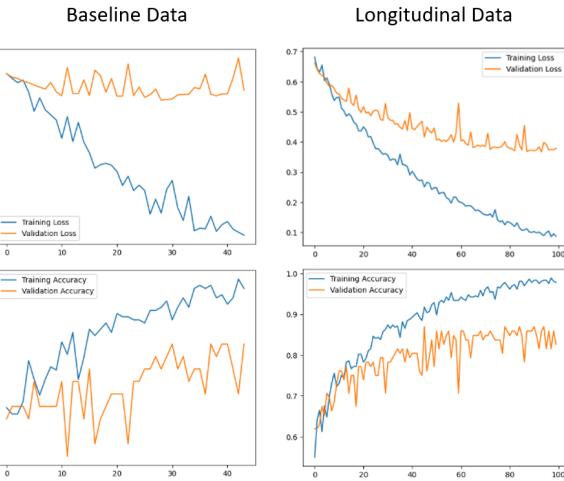


Figure 6.2: Loss and Accuracy curves for training and validation for the two dataset sizes.

was a pre-trained DenseNet201. Table 6.5 shows the results obtained for the classification given different transformations. To compare these techniques, Table 6.5 shows the ACC, loss, and AUC of each method; Figure 6.4 represents the loss and accuracy curves for training and validation; and Figure 6.3 shows the metrics Precision, Recall, F1-score, Sensitivity and Specificity for each class. Note that in Table 6.5, the values in parentheses represent the variation of the respective technique with the baseline, where the green values for ACC and AUC show an increase, i.e., an improvement in performance, while for loss, the green represents a decrease that also translates into an improvement in model performance. The red colour highlights the opposite situations for

each metric.

Table 6.5: Performance of different data augmentation methods.

Technique	ACC (%)	Loss	AUC (%)
Baseline (No techniques)	83.33	0.4430	89.48
Rotation	84.21 (+ 0.88)	0.4102 (- 0.033)	89.79 (+ 0.31)
Flip	85.08 (+ 1.75)	0.4205 (- 0.023)	89.44 (- 0.04)
Elastic Transformation	86.84 (+ 3.51)	0.4075 (- 0.036)	90.00 (+ 0.52)
Gaussian Noise	82.45 (- 0.88)	0.4297 (- 0.013)	88.29 (- 1.19)
Combination (all techniques)	81.57 (- 1.76)	0.4260 (- 0.017)	88.23 (- 1.25)
Combination (Rotation + Flip + Elastic Transformation)	84.21 (+ 0.88)	0.4040 (- 0.039)	90.07 (+ 0.59)

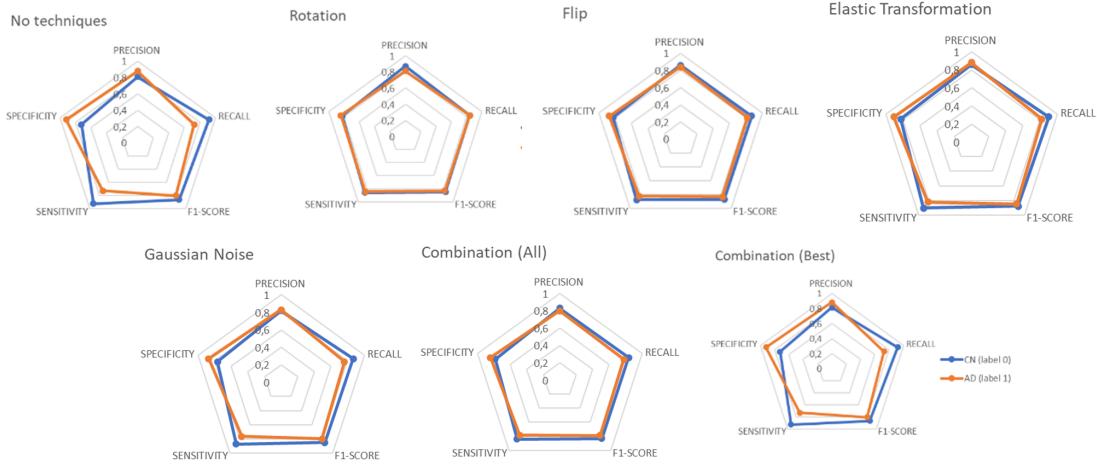


Figure 6.3: Radar chart with the metrics Precision, Recall, F1-score, Sensitivity and Specificity for each class (CN - label 0; AD - label 1) for each data augmentation technique implemented.

Thus, as can be seen from the table, several conclusions can be drawn: 1) elastic transformation is the most crucial data augmentation method; 2) rotation and flipping can also significantly enhance the performance of the model; 3) however, the model's performance suffers as a result of the Gaussian Noise technique; 4) furthermore, applying a combination of all techniques has a negative impact on the model; 5) combining only the three techniques that showed positive results (Rotation, Flip and Elastic Transformation) gives a better performance than when no technique is applied, yet is worse or equal to both of the three methods applied individually in terms of ACC. It is thought that this may be due to the fact that the Gaussian Noise method leads to the loss of useful information for classification, while the combinations are believed to add so much variability in the images that instead of facilitating the learning of the model they only hinder it.

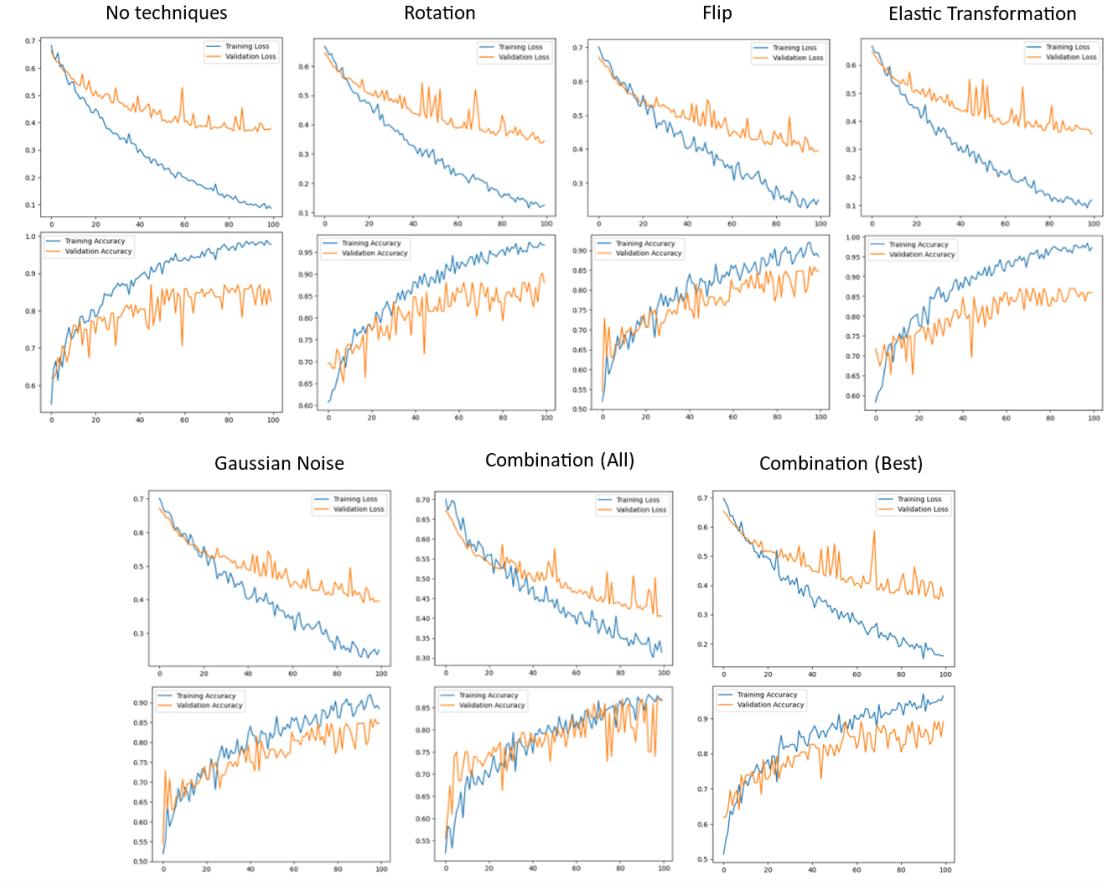


Figure 6.4: Loss and Accuracy curves for training and validation for each data augmentation technique implemented.

6.4 Effects of Different Model Architectures

In this experiment, eight fine-tuned 3D CNNs pre-trained on generic images (transfer learning) were implemented, namely DenseNet121, DenseNet169, DenseNet021, ResNet152, ResNeXt101, SENet154, SEResNet101, and SEResNet152. The respective architectures are explained in detail in Section 5.4. The performance of these different models was evaluated using the AD versus CN classification task, the Elastic Transformation technique for data augmentation, and 100-layer fine-tuning. To analyse the results, Table 6.7 shows the ACC, loss, and AUC of each method; Figure 6.5 displays the performance of the DenseNet variants; Figure 6.6 exhibits the results of the ResNet152 and ResNeXt101; Figure 6.7 displays the performance of the SENet154; and finally, Figure 6.8 contains the results obtained by the SEResNet variants.

Table 6.7: Performance of different model architectures.

Model	ACC (%)	Loss	AUC (%)	Train time (H:min)	Number of epochs
DenseNet121	84.21	0.4903	89.10	03:10	58
DenseNet169	80.70	0.4543	90.53	06:47	86
DenseNet201	86.84	0.4075	90.00	08:05	100
ResNet152	44.73	0.6946	61.90	03:17	16
ResNeXt101	44.73	0.6966	62.34	05:08	40
SENet154	77.19	0.6615	84.03	13:10	83
SEResNet101	85.96	0.5946	96.20	03:28	23
SEResNet152	91.22	0.4321	93.52	04:36	26

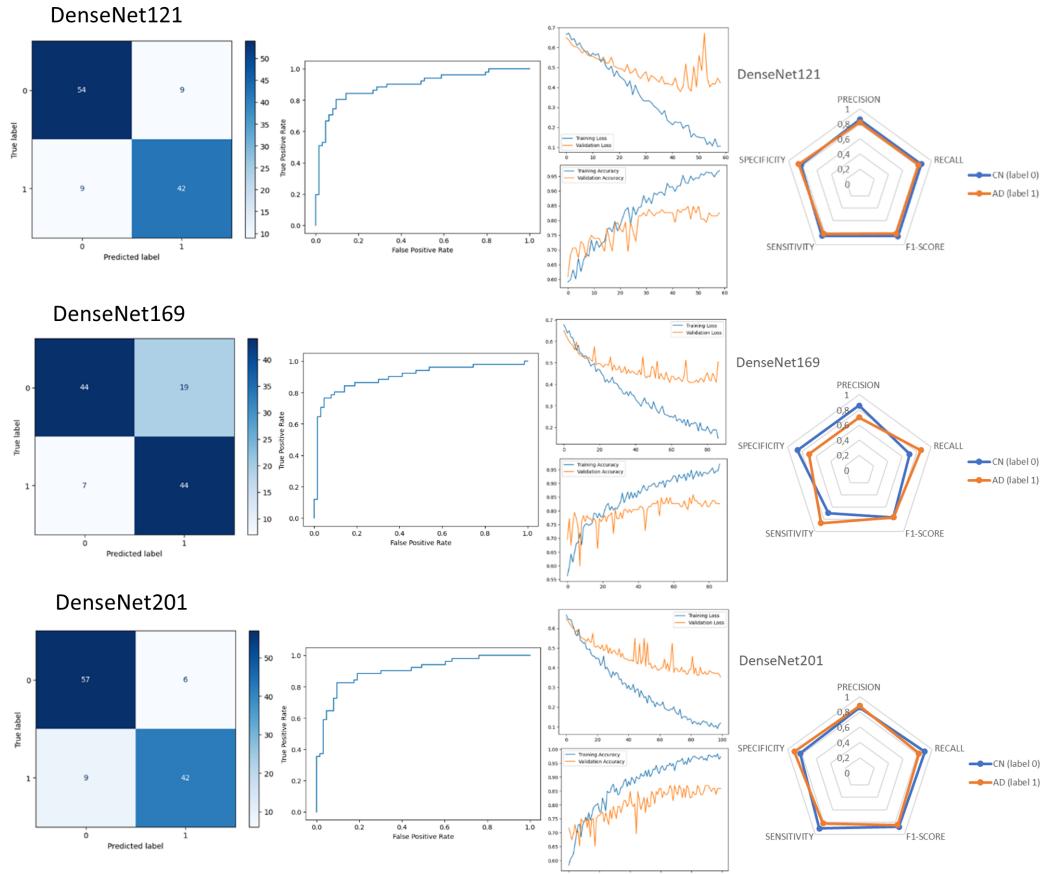


Figure 6.5: Performance comparison of the DenseNet variants, namely the Confusion Matrix, the ROC curve, Loss and Accuracy curves for training and validation, and the radar chart with the metrics Precision, Recall, F1-score, Sensitivity and Specificity for each class (CN - label 0; AD - label 1).

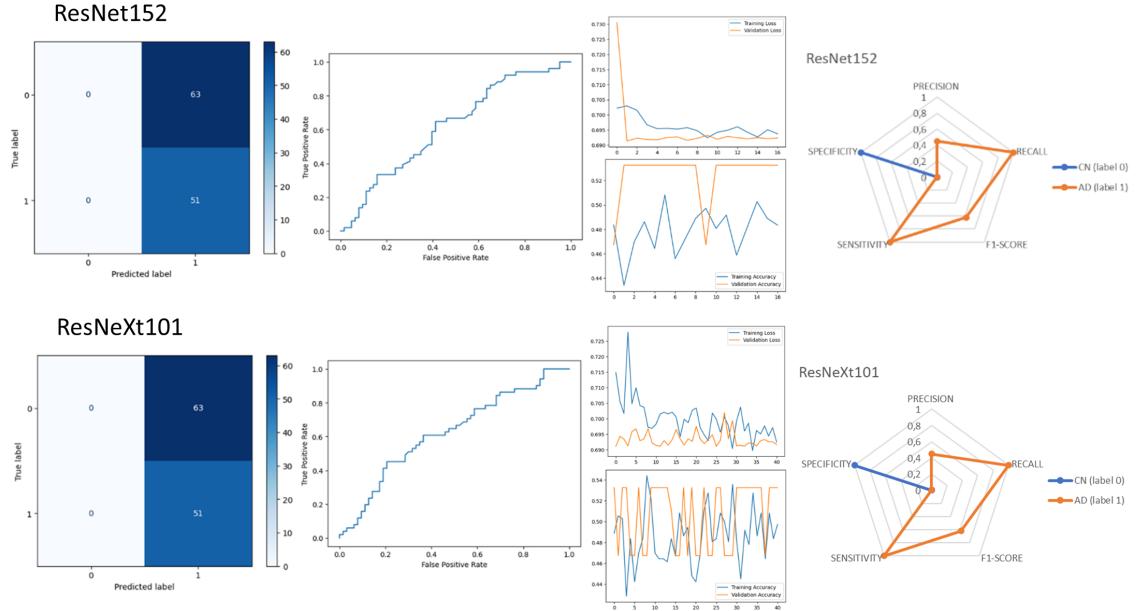


Figure 6.6: Performance comparison between ResNet152 and ResNeXt101, namely the Confusion Matrix, the ROC curve, Loss and Accuracy curves for training and validation, and the radar chart with the metrics Precision, Recall, F1-score, Sensitivity and Specificity for each class (CN - label 0; AD - label 1).

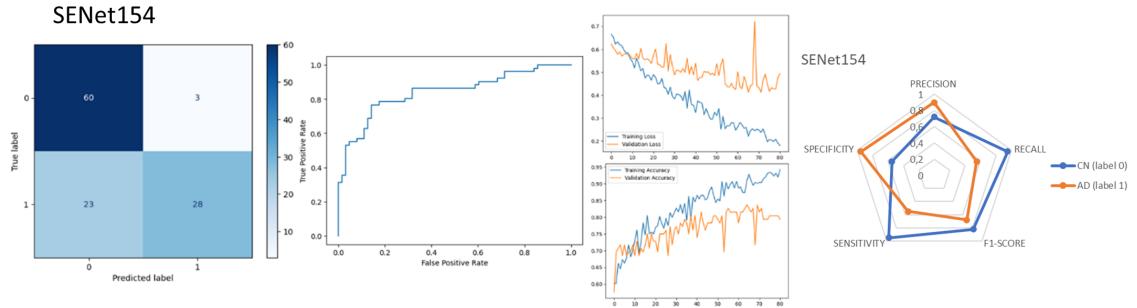


Figure 6.7: Performance of SENet154, namely the Confusion Matrix, the ROC curve, Loss and Accuracy curves for training and validation, and the radar chart with the metrics Precision, Recall, F1-score, Sensitivity and Specificity for each class (CN - label 0; AD - label 1).

Several conclusions can be drawn from these results: 1) the ResNet152 and ResNeXt101 models failed to learn anything from the task at hand, as is visible in Figure 6.6 since they always predicted the same class; 2) the DenseNet, SENet, and SEResNet architectures proved to be suitable for the task at hand, with all their variants obtaining an ACC greater than 77.00%, a loss smaller than 0.6700, and an AUC greater than 84.00%; 3) the SEResNet models converge quickly, needing less than 27 epochs and therefore relatively little training time when compared to the others; 4) the SEResNet152 model has the best testing performance in terms of ACC with

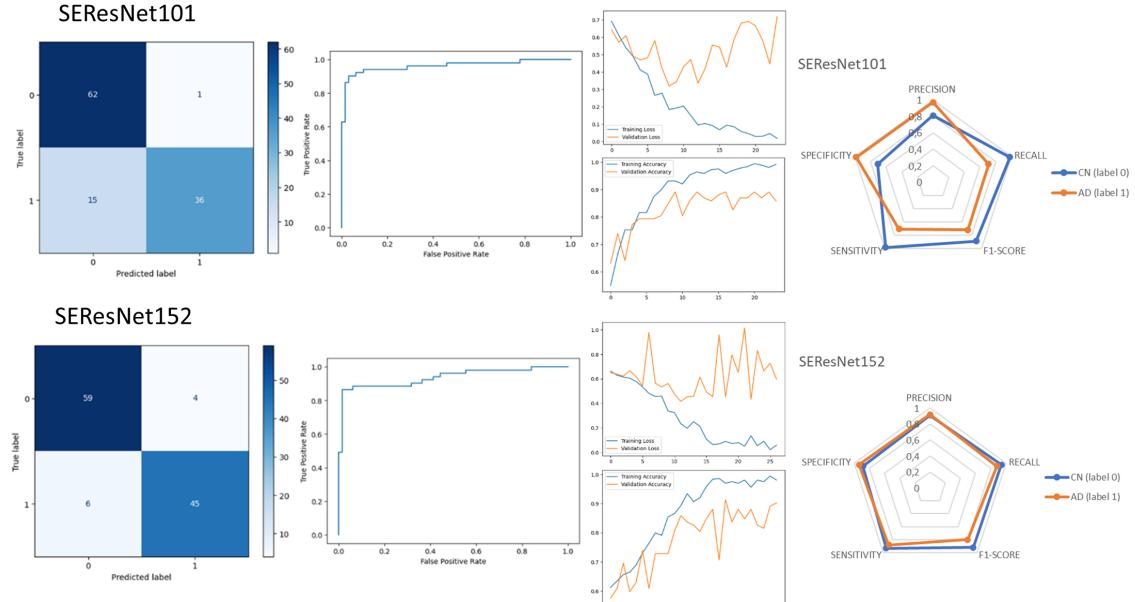


Figure 6.8: Performance comparison of the SEResNet variants, namely the Confusion Matrix, the ROC curve, Loss and Accuracy curves for training and validation, and the radar chart with the metrics Precision, Recall, F1-score, Sensitivity and Specificity for each class (CN - label 0; AD - label 1).

91.22%. All the results can be justified by multiple factors, namely, the complexity of the model, the complexity of the problem, the type of the problem, the size of the dataset, and the chosen settings.

6.5 Effects of Different Numbers of Unfrozen Layers

As discussed in Section 3.5, finding the optimal number of layers to unfreeze for a given problem is a challenge. Within this, and based on the size of the dataset in use and its similarity to the ImageNet images, the adopted strategy was the 2. However to understand how many layers to unfreeze, five tests were performed using the SEResNet152 model in the AD versus CN classification task and with the Elastic transformation technique as a data augmentation method. The results are presented in Table 6.9, Figure 6.9, and Figure 6.10.

Table 6.9: Performance with different numbers of unfrozen layers.

Unfrozen Layers	ACC (%)	Loss	AUC (%)
20	87.71	0.3245	93.09
50	87.71	0.3470	93.83

Unfrozen Layers	ACC (%)	Loss	AUC (%)
80	92.98	0.3840	94.33
100	91.22	0.4321	93.52
150	90.35	0.7654	93.74

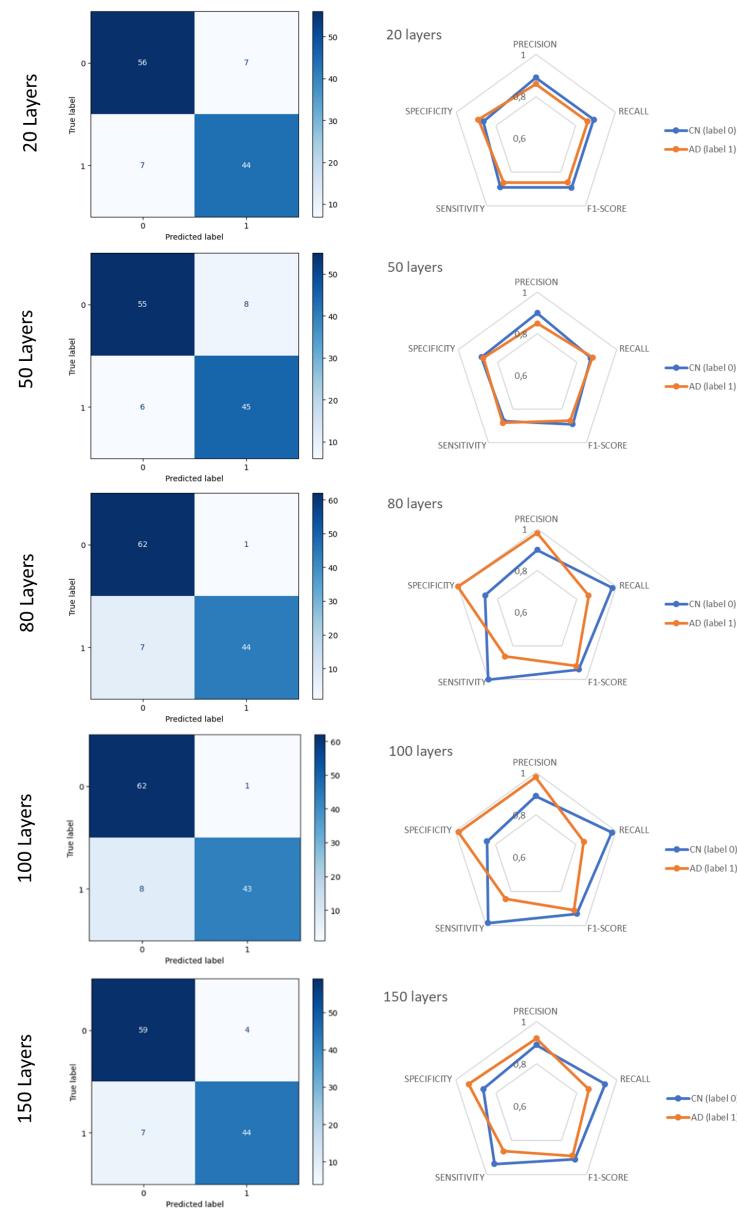


Figure 6.9: Confusion Matrix and radar chart with the metrics Precision, Recall, F1-score, Sensitivity and Specificity for each class (CN - label 0; AD - label 1) for different number of unfrozen layers.

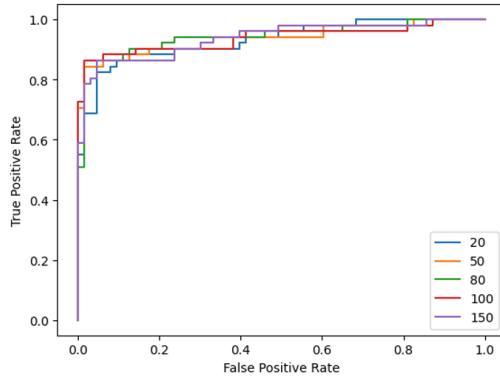


Figure 6.10: ROC curves of the different model's performance with different number of unfrozen layers.

As a result, it can be concluded that for these settings, with 80 layers the best test performance is obtained in both ACC (92.98 %) and AUC (94.33 %).

6.6 Effects of Different Study Regions

The last experiment focused on three regions of interest, namely the skull-stripped image, the telencephalon, and the hippocampus. Figure 5.3 demonstrates an example of each of these areas. The results were obtained in the classification of AD versus CN using the SEResNet152 model fine-tuned to 80 layers and with the Elastic Transformation technique. The results are shown in Table 6.11 which lists the ACC, loss, and AUC for each region, and in Figure 6.11, which includes the remaining metrics gathered. Beyond that, Figure 6.12 provided visualisations of the features discovered by the method using Grad-CAM to demonstrate the efficacy of the suggested method in feature extraction. The colour denotes how significant a region is for classification. Note that the results for the skull-stripped image are already presented in the test above for the 80 layers.

Table 6.11: Performance of different ROIs.

Region of interest	ACC (%)	Loss	AUC (%)
Skull-stripped	92.98	0.3840	94.33
Telencephalon	87.71	0.7596	89.73
Hippocampus	85.84	0.3611	93.77

The results obtained suggest that the image with more information (skull-stripped) allows for better performance in the test, with 92.98 % of ACC and 94.33 % of AUC. Thus, despite the fact that the hippocampus is one of the regions most affected by the disease, the proposed model behaves better with a more extensive image of the brain.

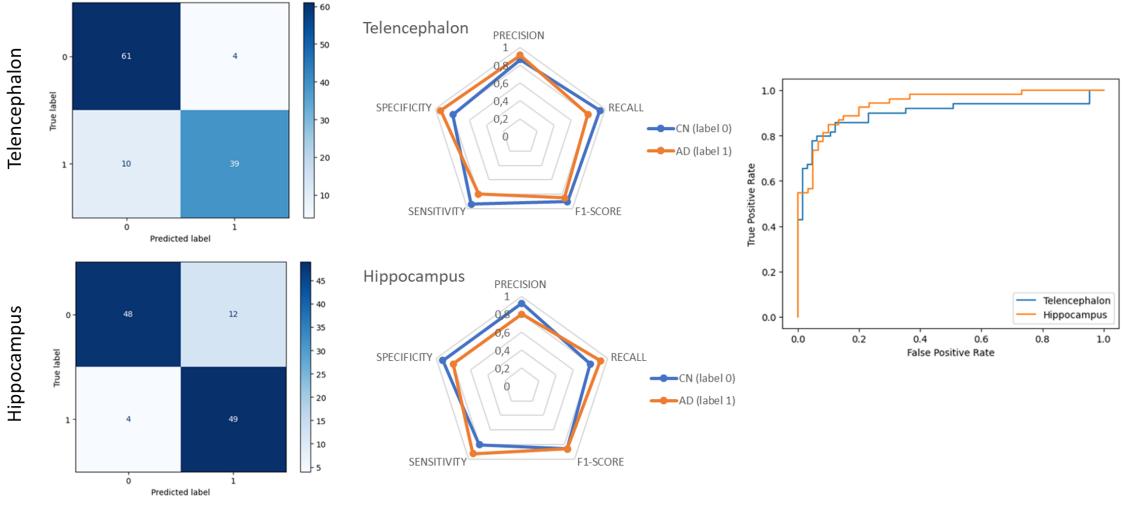


Figure 6.11: Confusion Matrix, radar chart with the metrics Precision, Recall, F1-score, Sensitivity and Specificity for each class (CN - label 0; AD - label 1), and Loss and Accuracy curves for training and validation for the regions Telencephalon and Hippocampus, respectively.

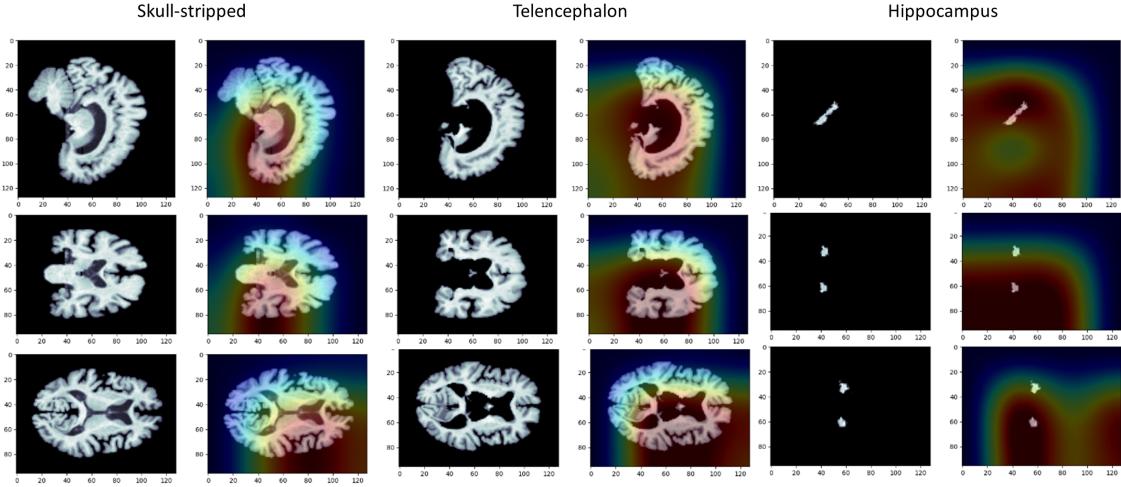


Figure 6.12: Grad-CAM visualizations of the features learned for each region, namely Skull-stripped, Telencephalon, and Hippocampus.

6.7 Proposed Model

Summing up all the experiments performed and their conclusions, the proposed model is a fine-tuned 80-layer SEResNet152 with the Elastic Transformation data augmentation technique, whose inputs are skull-stripped MR images. Thus, given these settings, binary classifications were performed for AD versus CN, CN versus EMCI, EMCI versus LMCI, and LMCI versus AD, and a multiclassification for CN versus EMCI versus LMCI versus AD. Table 6.13, Figure 6.14 and

Figure 6.13 demonstrate the results obtained for the various tasks described.

Table 6.13: Performance of the proposed model for the different classification tasks.

Task	ACC (%)	Loss	AUC (%)
AD vs. CN	92.98	0.3840	94.33
CN vs. EMCI	75.43	1.3839	80.43
EMCI vs. LMCI	78.94	1.0007	84.23
LMCI vs. AD	85.08	0.5212	90.60
CN vs. EMCI vs. LMCI vs. AD	67.98	1.091	-

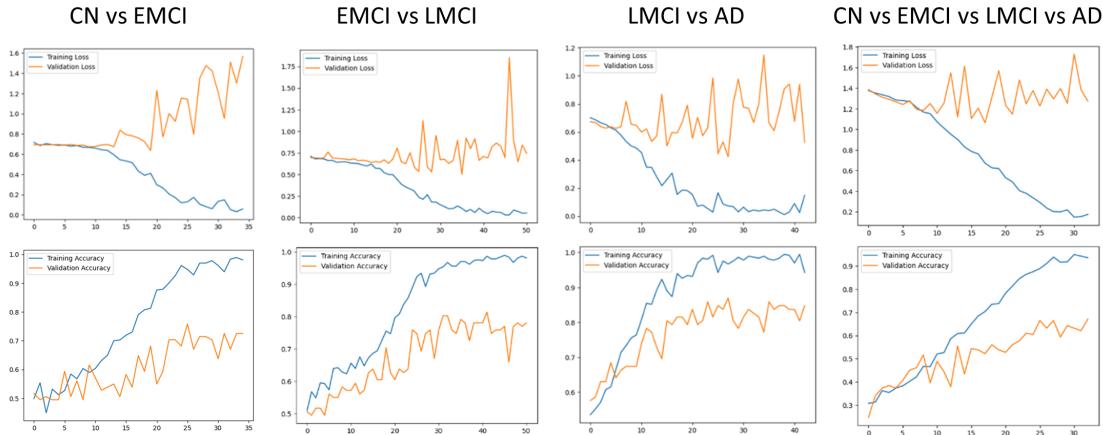


Figure 6.13: Loss and Accuracy curves for training and validation for the different classification tasks.

As expected, and since it is the least complex task, the binary classification CN vs. AD obtained better results with an ACC of 92.98%. The results also show that it was easier for the model to classify LMCI vs. AD than EMCI vs. LMCI or CN vs. EMCI, this may be linked to the progression of the disease, where the more advanced the stage, the more manifestations are visible. Also, as one would expect given the complexity, multiclassification was the most complicated task for the model. Even though SEResNet152 performed well in learning the representations, it nevertheless experienced the standard overfitting issue due to the short dataset used, as is visible in Figure 6.13. This is true even with the application of various regularisation methods, as described in Section 3.4.1, such as dropout, data augmentation, early stopping, and even the transfer learning technique itself. This might be a result of the classification task's high level of complexity. The subtle discrepancies between the CN and EMCI, and EMCI and LMCI images require a lot of data to learn the differences and perform the correct classification. Another reason might be that the

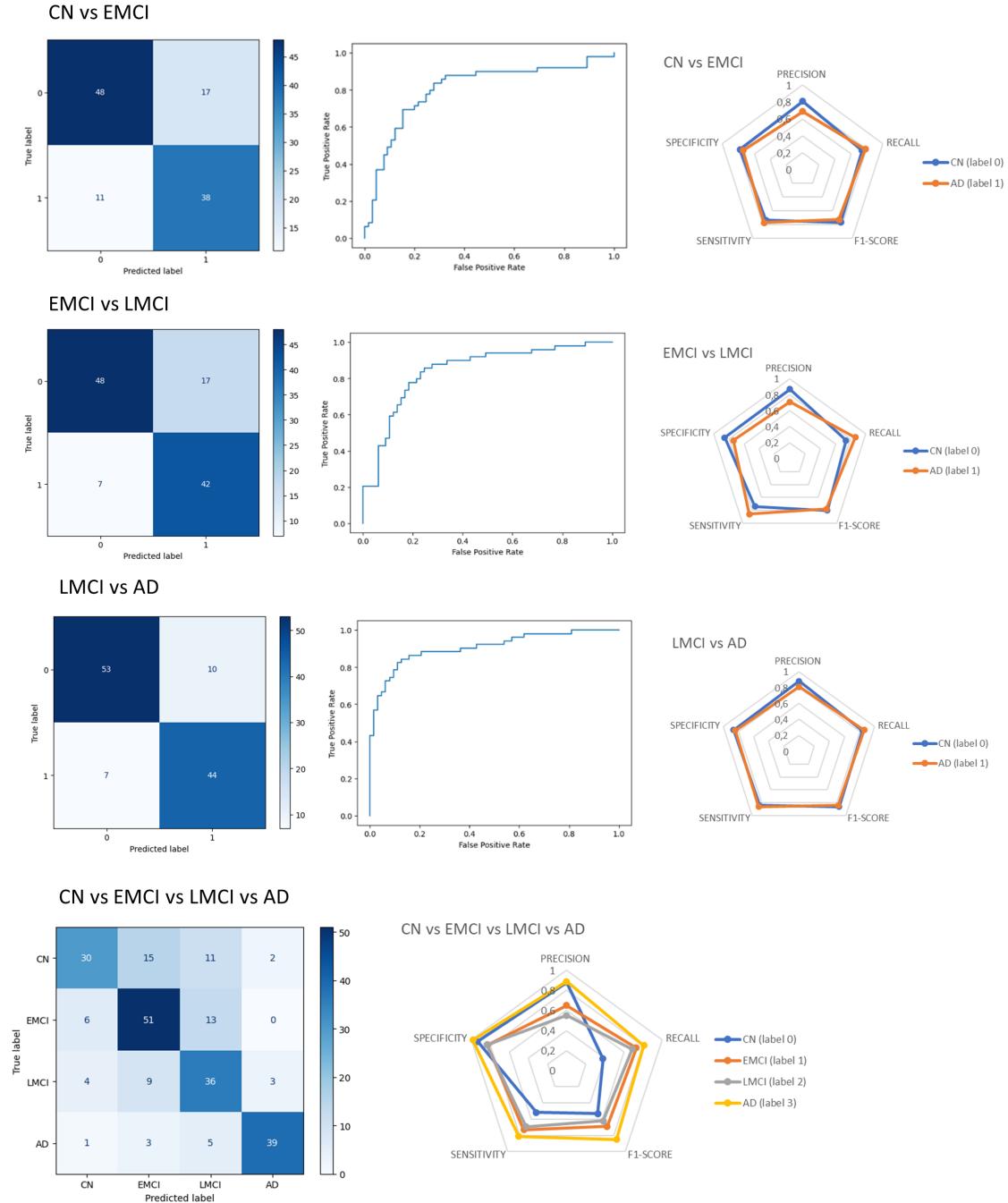


Figure 6.14: Performance comparison of the different classification tasks, namely the Confusion Matrix, the ROC curve, and the radar chart with the metrics Precision, Recall, F1-score, Sensitivity and Specificity for each class.

dataset being used in this project differs significantly from the ImageNet dataset. The SEResNet was pretrained on common images from the ImageNet without including medical images. Therefore, the high-level features picked up by the model's higher layers are insufficient to distinguish

the classes in this study. Thus, the results suggest that it is more difficult to detect this pathology in its earlier stages, as the very problem defined by this work anticipated.

6.8 Comparison Results With State of the Art Methods

The proposed model is compared with the state of the art algorithms using MR images in Table 6.15. The outcomes of these approaches should not be directly compared because they use different data. So, these techniques should only be used as general references.

Table 6.15: Comparison results with state of art methods.

Method	Year	Subjects	Data Pre-processing	Data Processing	Feature Extraction + Segmentation	Classifier	ACC
[76]	2022	6400 subjects: 3200 CN, 2240 MCI, AD, SMC, 896MCI, 64	Resizing Adaptive filtering for noise removal Adaptive histogram equalization	Whole 3D image	GLCM Gabor filter Wavelet features HSICL	FSODSNN based classifier	ADNI: CN vs SMC vs MCI vs AD: 99.89% IBIL: CN vs SMC vs MCI vs AD: 99.67% OASIS: CN vs SMC vs MCI vs AD: 99.61%
[72]	2022	955 subjects: 273 CN, 432 MCI, 250 AD	Skull-stripping Normalization TC-BWO-FCM for CSF, WM and GM Segmentation	ROI-based	TECD feature extraction + Clinical features	HRF-DNN	AD vs CN: 98.68% MCI vs AD: 95.88% MCI vs CN: 97.23% AD vs CN: 97.83%
[37]	2022	769 subjects: 475 CN, 224 MCI, 70 AD	N4 Bias Field Correction Skull Stripping Rigid Registration	Whole 3D image		ConvNet	AD vs MCI: 98.68% CN vs MCI: 99.10% CN vs MCI vs AD: 98.26%
[39]	2021	210 subjects: 70 CN, 70 MCI, 70 AD	Size scaling Skull Stripping	Slice-Based		LeNet	AD vs CN: 95.00% AD vs MCI: 97.00% MCI vs CN: 97.00%
[89]	2022	139 subjects: 55 CN, 84 AD	Template registration Skull-stripping Filter Denoising	Patch-based	SE mechanism PSA mechanism + FCN	MLP	AD vs CN: 98.85%
[70]	2022	238 subjects: 154 CN, 84 AD	ACPC alignment correction Intensity correction for uniform homogeneity Skull-stripping Registration	Whole 3D image	JD	CNN	AD vs CN: 96.61%
Proposed		1140 subjects: 286 CN, 283 EMCI, 287 LMCI, 284 AD	Orientation adjustment Histogram matched Inhomogeneity correction Linear normalisation Skull-stripping	Whole 3D image			AD vs CN: 92.98% CN vs EMCI: 75.43% EMCI vs LMCI: 78.94% LMCI vs AD: 85.08% CN vs EMCI vs LMCI vs AD: 67.98%
						SEResNet152	

The results obtained are a little inferior to the methods explored in the state of the art present in the table, however, this is the best 5 articles found out of 51. This way, if the results obtained are compared with the Table 4.3, the model proposed in this work presents a very interesting performance. It is also interesting to note that the model presented is distinct from the others, which also becomes an interesting contribution to future work by other researchers.

Thus, it is possible to conclude that at the performance level the goals have not been fully achieved, but they are a very entrancing starting point, and there is still plenty of room for progress.

6.9 Summary

This chapter presented all the results and respective analysis of the experimental part implemented in this project. In the first phase of the experimental portion, some tests were carried out to evaluate some variations that were essential to developing the optimal model. From this, it was possible to draw some conclusions and define the best model: a fine-tuned 80-layer SEResNet152 with the Elastic Transformation data augmentation technique, whose inputs are skull-stripped MR images. In the second phase, this model was subjected to several different classification tasks, namely binary classification (CN vs. AD; CN vs. EMCI; EMCI vs. LMCI; LMCI vs. AD) and multiclassification (CN vs. EMCI vs. LMCI vs. AD). These results were compared with the best algorithms found in the systematic review performed, and the best result was found in the CN vs. AD classification with an ACC of 92.98%.

Chapter 7

Conclusion and Future Work

The project's progress and the last thoughts on the outcomes are discussed in this chapter, which is followed by potential advancements.

7.1 Conclusion

The central theme of this dissertation was the development of a computer-aided diagnosis system based on 3D magnetic resonance images to identify different stages of Alzheimer's disease. Therefore, the main goal was the creation of a computational tool able to help in the practical context of the disease diagnosis. The development of the state of the art made it possible to analyse numerous methods to detect the disease accurately. The various techniques relied on selection data, preprocessing, segmentation (some of them), and the main key, classification.

Emphasising the complexity of the disease, the distinction between normal cognition, mild cognitive impairment (and its intermediate phases, EMCI and LMCI), and an AD subject is not easy, and it is even harder in the early stages of the disease when the structural changes are less obvious. The development of automatic classification systems, which can detect more things than human eyes and save medical professionals time and effort, is intended to aid in early diagnosis. These systems can also spot image differences that are imperceptible to the human eye.

Pre-trained networks have grown in popularity across a variety of applications because they make training quicker thanks to early learning about basic features that can be used in multiple tasks, like edges, shapes, corners, and intensities. Hence, this work explored various networks with the application of ImageNet weights, followed by fine-tuning regarding the task.

The experimental work was executed on the publicly available dataset, the ADNI. The pipeline is composed of three main stages: preprocessing, pre-trained models, and the classification task. The preprocessing steps for MR images that were reviewed and discussed make it clear that such enhancement is necessary for optimising the following steps. Thus, the 3D T1 MR images were subjected to 5 processes, namely orientation adjustment, histogram matched, inhomogeneity correction, linear normalisation, and skull-stripping.

Some trials were conducted during the experimental phase to test some variants that were crucial to achieving the best model. Thus, the results suggested that: the number of images used in training plays an essential role in the results obtained, the more the better; as a general rule, data augmentation techniques improve model performance, particularly the Elastic Transformation technique that obtained the best results; SEResNet152 proved to be the best architecture for the task at hand; the number of fine-tuned layers in a pre-trained model also influences the achieved performance, with 80 being the best number for the tests performed; and although it is scientifically proven that certain regions undergo more changes during the progression of AD, as is the case of the hippocampus, more extensive (skull-stripped) imaging allows for better results. The best project results were obtained through the CN vs. AD classification task, as expected, where an accuracy of 93% in the test set was achieved. The proposed method still has some drawbacks, though. The first limitation of this study is that, despite using the ADNI dataset, which is the largest worldwide regarding AD, the sample size was still quite small. Second, the model settings may not be fully optimised, such as batch size and dropout value. The overall outcomes were encouraging and show that computed methods can help medical professionals. They can certainly spread farther into clinical facilities in the near future.

7.2 Future Work

One of the long-term objectives could be to conduct a network test using images provided by the hospital to evaluate the degree of generalisation and investigate potential applications in that setting.

In the present work, only the MR imaging modality was explored. This has its advantages, namely that this imaging modality provides more data when compared to others, and it is a medical test performed more regularly and in almost all hospitals, which does not happen with PET, for example. However, it is typically more advantageous to combine information from genetics, cognition, illness state, structural and functional imaging, and other phenotypic modalities. Thus, the information adopted is more detailed and covers several areas related to the development of the disease, which can translate into improved model performance. So it would be a future possibility to explore more data types and their combinations.

Furthermore, and related to the problem that few MR images were used it would be interesting to explore, besides the data augmentation techniques presented, generative adversarial networks (GAN) to create fake images derived from the existing ones, a method that has been growing in popularity among researchers.

As far as regions are concerned, and although the skull-stripped image obtained better results than, for example, the hippocampus region, it would be of potential interest to segment other areas also strictly linked to the progression of this disease, e.g. Amygdala, Entorhinal cortex, ect, and develop an ensemble of classifiers that merges all the information extracted from these influential regions.

As has been seen throughout this paper, this topic is very current and of great interest to the scientific community, and because of this, new methods and solutions emerge daily. Thus, above are presented some ideas that could be evaluated in the future as a continuation of the project developed so far, however, many more possibilities could arise.

References

- [1] Keras: Deep Learning for humans. <https://keras.io/>, (accessed Apr. 11, 2023).
- [2] TensorFlow. <https://www.tensorflow.org/>, (accessed Apr. 11, 2023).
- [3] Magnetic Resonance Imaging (MRI). <https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri>, (accessed Mar. 03, 2023).
- [4] What Happens to the Brain in Alzheimer's Disease? | National Institute on Aging. <https://www.nia.nih.gov/health/what-happens-brain-alzheimers-disease>, (accessed Mar. 03, 2023).
- [5] MRI Basics. <https://case.edu/med/neurology/NR/MRIBasics.htm>, (accessed Mar. 16, 2023).
- [6] Alzheimer's Disease International (ADI). <https://www.alzint.org/>, (accessed May 03, 2023).
- [7] AIBL. <https://aibl.csiro.au/>, (accessed Nov. 11, 2022).
- [8] ADNI | Alzheimer's Disease Neuroimaging Initiative. <https://adni.loni.usc.edu/>, (accessed Nov. 15, 2022).
- [9] OASIS Brains - Open Access Series of Imaging Studies. <https://www.oasis-brains.org/>, (accessed Nov. 15, 2022).
- [10] Kumar A, Sidhu J, Goyal A, and Tsao JW. Alzheimer Disease. *StatPearls*, pages 1–27, may 2018.
- [11] D. Agarwal, M.A. Berbis, T. Martín-Noguerol, A. Luna, S.C.P. Garcia, and I. de la Torre-Díez. End-to-End Deep Learning Architectures Using 3D Neuroimaging Biomarkers for Early Alzheimer's Diagnosis. *Mathematics*, 10(15), 2022.
- [12] Ahmad Al-Qerem, Amer Abu Salem, Issam Jebreen, Ahmad Nabot, and Ahmad Samhan. Comparison between transfer learning and data augmentation on medical images classification. *2021 22nd International Arab Conference on Information Technology, ACIT 2021*, 2021.
- [13] C.V. Angkoso, H.P.A. Tjahyaningtjas, M.H. Purnomo, and I.K.E. Purnama. Multiplane Convolutional Neural Network (Mp-CNN) for Alzheimer's Disease Classification. *International Journal of Intelligent Engineering and Systems*, 15(1):329–340, 2022.

- [14] Ehsan Hosseini Asl, Mohammed Ghazal, Ali Mahmoud, Ali Aslantas, Ahmed Shalaby, Man-ual Casanova, Gregory Barnes, Georgy Gimel’farb, Robert Keynton, and Ayman El Baz. Alzheimer’s disease diagnostics by a 3D deeply supervised adaptable convolutional network. *Frontiers in Bioscience - Landmark*, 23(3):584–596, 2018.
- [15] J. Bae, J. Stocks, A. Heywood, Y. Jung, L. Jenkins, V. Hill, A. Katsaggelos, K. Popuri, H. Rosen, M.F. Beg, M.F. Beg, and L. Wang. Transfer learning for predicting conversion from mild cognitive impairment to dementia of Alzheimer’s type based on a three-dimensional convolutional neural network. *Neurobiology of Aging*, 99:53–64, 2021.
- [16] J.B. Bae, S. Lee, W. Jung, S. Park, W. Kim, H. Oh, J.W. Han, G.E. Kim, J.S. Kim, J.H. Kim, J.H. Kim, and K.W. Kim. Identification of Alzheimer’s disease using a convolutional neural network model based on T1-weighted magnetic resonance imaging. *Scientific Reports*, 10(1), 2020.
- [17] Pragati Baheti. What is Overfitting in Deep Learning [+10 Ways to Avoid It]. <https://www.v7labs.com/blog/overfitting{#}h1>, 2023 (accessed Apr. 03, 2023).
- [18] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, R. Santangelo, and M. Filippi. Automated classification of Alzheimer’s disease and mild cognitive impairment using a single MRI and deep neural networks. *NeuroImage: Clinical*, 21, 2019.
- [19] S. Basheera and M. Satya Sai Ram. A novel CNN based Alzheimer’s disease classification using hybrid enhanced ICA segmented gray matter of MRI. *Computerized Medical Imaging and Graphics*, 81, 2020.
- [20] A. Basher, B.C. Kim, K.H. Lee, and H.Y. Jung. Volumetric Feature-Based Alzheimer’s Disease Diagnosis from sMRI Data Using a Convolutional Neural Network and a Deep Neural Network. *IEEE Access*, 9:29870–29882, 2021.
- [21] Tushar Kanti Bera. Noninvasive electromagnetic methods for brain monitoring: A technical review. *Intelligent Systems Reference Library*, 74:51–95, 2015.
- [22] X. Bi, W. Liu, H. Liu, and Q. Shang. Artificial Intelligence-based MRI Images for Brain in Prediction of Alzheimer’s Disease. *Journal of Healthcare Engineering*, 2021, 2021.
- [23] Zeinab Breijeh and Rafik Karaman. Comprehensive Review on Alzheimer’s Disease: Causes and Treatment. *Molecules (Basel, Switzerland)*, 25(24), dec 2020.
- [24] Chensi Cao, Feng Liu, Hai Tan, Deshou Song, Wenjie Shu, Weizhong Li, Yiming Zhou, Xiaochen Bo, and Zhi Xie. Deep Learning and Its Applications in Biomedicine. *Genomics, proteomics & bioinformatics*, 16(1):17–32, feb 2018.
- [25] Rudy J. Castellani, Raj K. Rolston, and Mark A. Smith. Alzheimer Disease. *Disease-a-Month*, 56(9):484–546, sep 2010.
- [26] Phillip Chlap, Hang Min, Nym Vandenberg, Jason Dowling, Lois Holloway, and Annette Haworth. A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology*, 65(5):545–563, aug 2021.
- [27] François Chollet. *Deep Learning with Python*. Manning Publications Co., 2017.

- [28] B.M. Cobbinah, C. Sorg, Q. Yang, A. Ternblom, C. Zheng, W. Han, L. Che, and J. Shao. Reducing variations in multi-center Alzheimer's disease classification with convolutional adversarial autoencoder. *Medical Image Analysis*, 82, 2022.
- [29] R. Cui and M. Liu. Hippocampus Analysis by Combination of 3-D DenseNet and Shapes for Alzheimer's Disease Diagnosis. *IEEE Journal of Biomedical and Health Informatics*, 23(5):2099–2107, 2019.
- [30] R. Cui and M. Liu. RNN-based longitudinal analysis for diagnosis of Alzheimer's disease. *Computerized Medical Imaging and Graphics*, 73:1–10, 2019.
- [31] Ivana Despotović, Bart Goossens, and Wilfried Philips. MRI Segmentation of the Human Brain: Challenges, Methods, and Applications. *Computational and Mathematical Methods in Medicine*, 2015, 2015.
- [32] Aigerim Djamanakova, Xiaoying Tang, Xin Li, Andreia V. Faria, Can Ceritoglu, Kenichi Oishi, Argye E. Hillis, Marilyn Albert, Constantine Lyketsos, Michael I. Miller, and Susumu Mori. Tools for multiple granularity analysis of brain MRI data for individualized image analysis. *NeuroImage*, 101:168–176, nov 2014.
- [33] M. Dyrba, M. Hanzig, S. Altenstein, S. Bader, T. Ballarini, F. Brosseron, K. Buerger, D. Cantré, P. Dechant, L. Dobisch, F. Jessen, and S.J. Teipel. Improving 3D convolutional neural network comprehensibility via interactive visualization of relevance maps: evaluation in Alzheimer's disease. *Alzheimer's Research and Therapy*, 13(1), 2021.
- [34] F.U.R. Faisal and G.-R. Kwon. Automated Detection of Alzheimer-s Disease and Mild Cognitive Impairment Using Whole Brain MRI. *IEEE Access*, 10:65055–65066, 2022.
- [35] Z. Fan, J. Li, L. Zhang, G. Zhu, P. Li, X. Lu, P. Shen, S.A.A. Shah, M. Bennamoun, T. Hua, T. Hua, and W. Wei. U-net based analysis of MRI for Alzheimer's disease diagnosis. *Neural Computing and Applications*, 33(20):13587–13599, 2021.
- [36] G. Folego, M. Weiler, R.F. Casseb, R. Pires, and A. Rocha. Alzheimer's Disease Detection Through Whole-Brain 3D-CNN MRI. *Frontiers in Bioengineering and Biotechnology*, 8, 2020.
- [37] N. Goenka and S. Tiwari. AlzVNet: A volumetric convolutional neural network for multiclass classification of Alzheimer's disease through multiple neuroimaging computational approaches. *Biomedical Signal Processing and Control*, 74, 2022.
- [38] H. Guan, C. Wang, J. Cheng, J. Jing, and T. Liu. A parallel attention-augmented bilinear network for early magnetic resonance imaging-based diagnosis of Alzheimer's disease. *Human Brain Mapping*, 43(2):760–772, 2022.
- [39] R.A. Hazarika, A. Abraham, D. Kandar, and A.K. Maji. An Improved LeNet-Deep Neural Network Model for Alzheimer's Disease Classification Using Brain Magnetic Resonance Images. *IEEE Access*, 9:161194–161207, 2021.
- [40] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-December:770–778, dec 2015.
- [41] Malini Hosakere. RoiEditor User Manual - MRI Studio. https://www.mristudio.org/user{_}manual/roieditor.html, apr 2007, (accessed Apr. 27, 2023).

- [42] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu. Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8):2011–2023, sep 2017.
- [43] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely Connected Convolutional Networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-January:2261–2269, aug 2016.
- [44] Y. Huang, J. Xu, Y. Zhou, T. Tong, and X. Zhuang. Diagnosis of Alzheimer’s disease via multi-modality 3D convolutional neural network. *Frontiers in Neuroscience*, 13(MAY), 2019.
- [45] J. Jiang, J. Zhang, Z. Li, L. Li, and B. Huang. Using Deep Learning Radiomics to Distinguish Cognitively Normal Adults at Risk of Alzheimer’s Disease From Normal Control: An Exploratory Study Based on Structural MRI. *Frontiers in Medicine*, 9, 2022.
- [46] Keith A. Johnson, Nick C. Fox, Reisa A. Sperling, and William E. Klunk. Brain Imaging in Alzheimer Disease. *Cold Spring Harbor Perspectives in Medicine*, 2(4), 2012.
- [47] Kwetishe Joro Danjuma. Performance Evaluation of Machine Learning Algorithms in Post-operative Life Expectancy in the Lung Cancer Patients. *International Journal of Computer Science Issues*, 2015.
- [48] B. Khagi, K.H. Lee, K.Y. Choi, J.J. Lee, G.-R. Kwon, and H.-D. Yang. Vbm-based alzheimer’s disease detection from the region of interest of t1 mri with supportive gaussian smoothing and a bayesian regularized neural network. *Applied Sciences (Switzerland)*, 11(13), 2021.
- [49] David S. Knopman, Helene Amieva, Ronald C. Petersen, G  el Ch  telat, David M. Holtzman, Bradley T. Hyman, Ralph A. Nixon, and David T. Jones. Alzheimer disease. *Nature Reviews Disease Primers* 2021 7:1, 7(1):1–21, may 2021.
- [50] Garam Lee, Kwangsik Nho, Byungkon Kang, Kyung Ah Sohn, Dokyoon Kim, Michael W. Weiner, Paul Aisen, Ronald Petersen, Clifford R. Jack, William Jagust, John Q. Trojanowski, Arthur W. Toga, Laurel Beckett, Robert C. Green, Andrew J. Saykin, John Morris, Leslie M. Shaw, Zaven Khachaturian, Greg Sorensen, Maria Carrillo, Lew Kuller, Marc Raichle, Steven Paul, Peter Davies, Howard Fillit, Franz Hefti, Davie Holtzman, M. Marcel Mesulam, William Potter, Peter Snyder, Tom Montine, Ronald G. Thomas, Michael Donohue, Sarah Walter, Tamie Sather, Gus Jiminez, Archana B. Balasubramanian, Jennifer Mason, Iris Sim, Danielle Harvey, Matthew Bernstein, Nick Fox, Paul Thompson, Norbert Schuff, Charles DeCarli, Bret Borowski, Jeff Gunter, Matt Senjem, Prashanthi Vemuri, David Jones, Kejal Kantarci, Chad Ward, Robert A. Koeppe, Norm Foster, Eric M. Reiman, Kewei Chen, Chet Mathis, Susan Landau, Nigel J. Cairns, Erin Householder, Lisa Taylor-Reinwald, Virginia Lee, Magdalena Korecka, Michal Figurski, Karen Crawford, Scott Neu, Tatiana M. Foroud, Steven Potkin, Li Shen, Kelley Faber, Sungyun Kim, Lean Tha, Richard Frank, John Hsiao, Jeffrey Kaye, Joseph Quinn, Lisa Silbert, Betty Lind, Raina Carter, Sara Dolen, Beau Ances, Maria Carroll, Mary L. Creech, Erin Franklin, Mark A. Mintun, Stacy Schneider, Angela Oliver, Lon S. Schneider, Sonia Pawluczyk, Mauricio Beccera, Liberty Teodoro, Bryan M. Spann, James Brewer, Helen Vanderswag, Adam Fleisher, Daniel Marson, Randall Griffith, David Clark, David Geldmacher, John Brockington, Erik Roberson, Marissa Natelson Love, Judith L. Heidebrink, Joanne L. Lord, Sara S. Mason, Colleen S. Albers, David

- Knopman, Kris Johnson, Hillel Grossman, Effie Mitsis, Raj C. Shah, Leyla DeToledo-Morrell, Rachelle S. Doody, Javier Villanueva-Meyer, Munir Chowdhury, Susan Rountree, Mimi Dang, Ranjan Duara, Daniel Varon, Maria T. Greig, Peggy Roberts, Yaakov Stern, Lawrence S. Honig, Karen L. Bell, Marilyn Albert, Chiadi Onyike, Daniel D'Agostino, Stephanie Kielb, James E. Galvin, Brittany Cerbone, Christina A. Michel, Dana M. Pogorelc, Henry Rusinek, Mony J. de Leon, Lidia Glodzik, Susan De Santi, Kyle Womack, Dana Mathews, Mary Quiceno, P. Murali Doraiswamy, Jeffrey R. Petrella, Salvador Borges-Neto, Terence Z. Wong, Edward Coleman, Allan I. Levey, James J. Lah, Janet S. Celli, Jeffrey M. Burns, Russell H. Swerdlow, William M. Brooks, Steven E. Arnold, Jason H. Karlawish, David Wolk, Christopher M. Clark, Liana Apostolova, Kathleen Tingus, Ellen Woo, Daniel H.S. Silverman, Po H. Lu, George Bartzokis, Charles D. Smith, Greg Jicha, Peter Hardy, Partha Sinha, Elizabeth Oates, Gary Conrad, Neill R. Graff-Radford, Francine Parfitt, Tracy Kendall, Heather Johnson, Oscar L. Lopez, Mary Ann Oakley, Donna M. Simpson, Martin R. Farlow, Ann Marie Hake, Brandy R. Matthews, Jared R. Brosch, Scott Herring, Cynthia Hunt, Anton P. Porsteinsson, Bonnie S. Goldstein, Kim Martin, Kelly M. Makino, M. Saleem Ismail, Connie Brand, Ruth A. Mulnard, Gaby Thai, Catherine Mc-Adams-Ortiz, Christopher H. van Dyck, Richard E. Carson, Martha G. MacAvoy, Pradeep Varma, Howard Chertkow, Howard Bergman, Chris Hosein, Sandra Black, Bojana Stefanovic, Curtis Caldwell, Ging Yuek Robin Hsiung, Howard Feldman, Benita Mudge, Michele Assaly, Elizabeth Finger, Stephen Pasternack, Irina Rachinsky, Dick Trost, Andrew Kertesz, Charles Bernick, Donna Munic, Kristine Lipowski, Masandra Weintraub, Borna Bonakdarpour, Diana Kerwin, Chuang Kuo Wu, Nancy Johnson, Carl Sadowsky, Teresa Villena, Raymond Scott Turner, Kathleen Johnson, Brigid Reynolds, Reisa A. Sperling, Keith A. Johnson, Gad Marshall, Jerome Yesavage, Joy L. Taylor, Barton Lane, Allyson Rosen, Jared Tinklenberg, Marwan N. Sabbagh, Christine M. Belden, Sandra A. Jacobson, Sherye A. Sirrel, Neil Kowall, Ronald Killiany, Andrew E. Budson, Alexander Norbash, Patricia Lynn Johnson, Thomas O. Obisesan, Saba Wolday, Joanne Allard, Alan Lerner, Paula Ogrocki, Curtis Tat-suoka, Parianne Fatica, Evan Fletcher, Pauline Maillard, John Olichney, Owen Carmichael, Smita Kittur, Michael Borrie, T. Y. Lee, Rob Bartha, Sterling Johnson, Sanjay Asthana, Cynthia M. Carlsson, Adrian Preda, Dana Nguyen, Pierre Tariot, Anna Burke, Nadira Trnicic, Adam Fleisher, Stephanie Reeder, Vernice Bates, Horacio Capote, Michelle Rainka, Douglas W. Scharre, Maria Kataki, Anahita Adeli, Earl A. Zimmerman, Dzintra Celmins, Alice D. Brown, Godfrey D. Pearson, Karen Blank, Karen Anderson, Laura A. Flashman, Marc Seltzer, Mary L. Hynes, Robert B. Santulli, Kaycee M. Sink, Leslie Gordineer, Jeff D. Williamson, Pradeep Garg, Franklin Watkins, Brian R. Ott, Henry Querfurth, Geoffrey Tremont, Stephen Salloway, Paul Malloy, Stephen Correia, Howard J. Rosen, Bruce L. Miller, David Perry, Jacobo Mintzer, Kenneth Spicer, David Bachman, Elizabeth Finger, Stephen Pasternak, Irina Rachinsky, John Rogers, Dick Drost, Nunzio Pomara, Raymundo Hernando, Antero Sarrael, Susan K. Schultz, Laura L. Boles Ponto, Hyungsub Shim, Karen Ekstam Smith, Norman Relkin, Gloria Chaing, Michael Lin, Lisa Ravdin, Amanda Smith, Balebail Ashok Raj, and Kristin Fargher. Predicting Alzheimer's disease progression using multi-modal deep learning approach. *Scientific Reports* 2019 9:1, 9(1):1–12, feb 2019.
- [51] C. Li, Q. Wang, X. Liu, and B. Hu. An Attention-Based CoT-ResNet With Channel Shuffle Mechanism for Classification of Alzheimer's Disease Levels. *Frontiers in Aging Neuroscience*, 14, 2022.
- [52] F. Li and M. Liu. Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks. *Computerized Medical Imaging and Graphics*, 70:101–110, 2018.

- [53] F. Li and M. Liu. A hybrid Convolutional and Recurrent Neural Network for Hippocampus Analysis in Alzheimer’s Disease. *Journal of Neuroscience Methods*, 323:108–118, 2019.
- [54] J. Li, Y. Wei, C. Wang, Q. Hu, Y. Liu, and L. Xu. 3-D CNN-Based Multichannel Contrastive Learning for Alzheimer’s Disease Automatic Diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 71, 2022.
- [55] X. Liang, Z. Wang, Z. Chen, and X. Song. Alzheimer’s disease classification using distilled multi-residual network. *Applied Intelligence*, 2022.
- [56] B.Y. Lim, K.W. Lai, K. Haiskin, K.A.S.H. Kulathilake, Z.C. Ong, Y.C. Hum, S. Dhanalakshmi, X. Wu, and X. Zuo. Deep Learning Model for Prediction of Progressive Mild Cognitive Impairment to Alzheimer’s Disease Using Structural MRI. *Frontiers in Aging Neuroscience*, 14, 2022.
- [57] C.-F. Liu, S. Padhy, S. Ramachandran, V.X. Wang, A. Efimov, A. Bernal, L. Shi, M. Vaillant, J.T. Ratnanather, A.V. Faria, M. Albert, and M.I. Miller. Using deep Siamese neural networks for detection of brain asymmetries associated with Alzheimer’s Disease and Mild Cognitive Impairment. *Magnetic Resonance Imaging*, 64:190–199, 2019.
- [58] M. Liu, F. Li, H. Yan, K. Wang, Y. Ma, L. Shen, and M. Xu. A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer’s disease. *NeuroImage*, 208, 2020.
- [59] Z. Liu, H. Lu, X. Pan, M. Xu, R. Lan, and X. Luo. Diagnosis of Alzheimer’s disease via an attention-based multi-scale convolutional neural network. *Knowledge-Based Systems*, 238, 2022.
- [60] Sina Mansour L, Caio Seguin, Robert E. Smith, and Andrew Zalesky. Connectome spatial smoothing (CSS): Concepts, methods, and evaluation. *NeuroImage*, 250, apr 2022.
- [61] Pedro Marcelino. Transfer learning from pre-trained models. <https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751>, oct 2018 (accessed Apr. 06, 2023).
- [62] Alessandro A. Mazzola. Ressonância magnética: princípios de formação da imagem e aplicações em imagem funcional. *Revista Brasileira de Física Médica*, 3(1):117–129, 2009.
- [63] R. Mendoza-Léon, J. Puentes, L.F. Uriza, and M. Hernández Hoyos. Single-slice Alzheimer’s disease classification and disease regional analysis with Supervised Switching Autoencoders. *Computers in Biology and Medicine*, 116, 2020.
- [64] Shervin Minaee. 20 Popular Machine Learning Metrics. Part 1: Classification & Regression Evaluation Metrics | by Shervin Minaee | Towards Data Science. <https://towardsdatascience.com/20-popular-machine-learning-metrics-part-1-classification-regression-evaluation-metrics-1ca3e282a2ce>, oct 2019 (accessed Apr. 04, 2023).
- [65] Susumu Mori, Dan Wu, Can Ceritoglu, Yue Li, Anthony Kolasny, Marc A. Vaillant, Andreia V. Faria, Kenichi Oishi, and Michael I. Miller. MRICloud: Delivering high-throughput MRI neuroinformatics as cloud-based software as a service. *Computing in Science and Engineering*, 18(5):21–35, sep 2016.

- [66] L. Nanni, M. Interlenghi, S. Brahnam, C. Salvatore, S. Papa, R. Nemni, and I. Castiglioni. Comparison of Transfer Learning and Conventional Machine Learning Applied to Structural Brain MRI for the Early Diagnosis and Prognosis of Alzheimer's Disease. *Frontiers in Neurology*, 11, 2020.
- [67] Kartik Nighania. Various ways to evaluate a machine learning model's performance | by Kartik Nighania | Towards Data Science. <https://towardsdatascience.com/Various-ways-to-evaluate-a-machine-learning-models-performance-230449055f15>, dec 2018 (accessed Apr. 04, 2023).
- [68] E. Ocasio and T.Q. Duong. Deep learning prediction of mild cognitive impairment conversion to Alzheimer's disease at 3 years after diagnosis using longitudinal and wholebrain 3D MRI. *PeerJ Computer Science*, 7:1–21, 2021.
- [69] M. Odusami, R. Maskeliūnas, and R. Damaševičius. An Intelligent System for Early Recognition of Alzheimer's Disease Using Neuroimaging. *Sensors*, 22(3), 2022.
- [70] S. Qasim Abbas, L. Chi, and Y.-P.P. Chen. Transformed domain convolutional neural network for Alzheimer's disease diagnosis using structural MRI. *Pattern Recognition*, 133, 2023.
- [71] P. Raghavaiah and S. Varadarajan. A CAD system design to diagnosize alzheimers disease from MRI brain images using optimal deep neural network. *Multimedia Tools and Applications*, 80(17):26411–26428, 2021.
- [72] P. Raghavaiah and S. Varadarajan. A CAD system design for Alzheimer's disease diagnosis using temporally consistent clustering and hybrid deep learning models. *Biomedical Signal Processing and Control*, 75, 2022.
- [73] I. Razzak, S. Naz, A. Ashraf, F. Khalifa, M.R. Bouadjenek, and S. Mumtaz. Mutliresolution ensemble PartialNet for Alzheimer detection using magnetic resonance imaging data. *International Journal of Intelligent Systems*, 37(10):6613–6630, 2022.
- [74] Muhammad Imran Razzak, Saeeda Naz, and Ahmad Zaib. Deep Learning for Medical Image Processing: Overview, Challenges and the Future BT - Classification in BioApps: Automation of Decision Making. *Springer*, 26:323–350, 2018.
- [75] Eric M. Reiman and William J. Jagust. Brain imaging in the study of Alzheimer's disease. *NeuroImage*, 61(2):505–516, jun 2012.
- [76] R. Sampath and M. Baskar. 3D brain image-based Alzheimer's disease detection techniques using fish swarm optimizer's deep convolution Siamese neural network. *Expert Systems*, 39(9), 2022.
- [77] C.L. Saratxaga, I. Moya, A. Picón, M. Acosta, A. Moreno-Fernandez-de leceta, E. Garrote, and A. Bereciartua-Perez. Mri deep learning-based solution for alzheimer's disease prediction. *Journal of Personalized Medicine*, 11(9), 2021.
- [78] M. Sethi, S. Ahuja, S. Rani, D. Koundal, A. Zagaria, and W. Enbeyle. An Exploration: Alzheimer's Disease Classification Based on Convolutional Neural Network. *Bio[1] M. Sethi, S. Ahuja, S. Rani, D. Koundal, A. Zagaria, and W. Enbeyle, "An Exploration: Alzheimer's Disease Classification Based on Convolutional Neural Network," Biomed Res. Int., vol. 2022, 2022, doi: 10.1155/2022/8739960.Med Research International*, 2022, 2022.

- [79] S. Shaji, N. Ganapathy, and R. Swaminathan. Classification of Alzheimer Condition using MR Brain Images and Inception-Residual Network Model. *Current Directions in Biomedical Engineering*, 7(2):763–766, 2021.
- [80] Roman Solovyev, Alexandr A. Kalinin, and Tatiana Gabruseva. 3D convolutional neural networks for stalled brain capillary detection. *Computers in Biology and Medicine*, 141:105089, feb 2022.
- [81] C.H. Suh, W.H. Shim, S.J. Kim, J.H. Roh, J.-H. Lee, M.-J. Kim, S. Park, W. Jung, J. Sung, and G.-H. Jahng. Development and validation of a deep learning-based automatic brain segmentation and classification algorithm for Alzheimer disease using 3D T1-weighted volumetric images. *American Journal of Neuroradiology*, 41(12):2227–2234, 2020.
- [82] M. Tanveer, B. Richhariya, R.U. Khan, A.H. Rashid, P. Khanna, M. Prasad, and C.T. Lin. Machine learning techniques for the diagnosis of alzheimer’s disease: A review. *ACM Transactions on Multimedia Computing, Communications and Applications*, 16(1s), 2020.
- [83] John Towns, Timothy Cockerill, Maytal Dahan, Ian Foster, Kelly Gaither, Andrew Grimshaw, Victor Hazlewood, Scott Lathrop, Dave Lifka, Gregory D. Peterson, Ralph Roskies, J. Ray Scott, and Nancy Wilkens-Diehr. XSEDE: Accelerating scientific discovery. *Computing in Science and Engineering*, 16(5):62–74, sep 2014.
- [84] Vachan Vadmal, Grant Junno, Chaitra Badve, William Huang, Kristin A. Waite, and Jill S. Barnholtz-Sloan. MRI image analysis methods and applications: an algorithmic perspective using brain tumors as an exemplar. *Neuro-Oncology Advances*, 2(1):1–13, jan 2020.
- [85] C. Wang, Y. Li, Y. Tsuboshita, T. Sakurai, T. Goto, H. Yamaguchi, Y. Yamashita, A. Sekiguchi, H. Tachimori, C. Wang, Y. Li, and T. Goto. A high-generalizability machine learning framework for predicting the progression of Alzheimer’s disease using limited data. *npj Digital Medicine*, 5(1), 2022.
- [86] J. Wen, E. Thibeau-Sutre, M. Diaz-Melo, J. Samper-González, A. Routier, S. Bottani, D. Dormont, S. Durrleman, N. Burgos, and O. Colliot. Convolutional neural networks for classification of Alzheimer’s disease: Overview and reproducible evaluation. *Medical Image Analysis*, 63, 2020.
- [87] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated Residual Transformations for Deep Neural Networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-January:5987–5995, nov 2016.
- [88] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, 9(4):611–629, aug 2018.
- [89] B. Yan, Y. Li, L. Li, X. Yang, T.-Q. Li, G. Yang, and M. Jiang. Quantifying the impact of Pyramid Squeeze Attention mechanism and filtering approaches on Alzheimer’s disease classification. *Computers in Biology and Medicine*, 148, 2022.
- [90] A. Yiğit and Z. İşik. Applying deep learning models to structural MRI for stage prediction of Alzheimer’s disease. *Turkish Journal of Electrical Engineering and Computer Sciences*, 28(1):196–210, 2020.

- [91] Tayyabah Yousaf, George Dervenoulas, and Marios Politis. Advances in MRI Methodology. *International review of neurobiology*, 141:31–76, jan 2018.
- [92] F. Zhang, Z. Li, B. Zhang, H. Du, B. Wang, and X. Zhang. Multi-modal deep learning model for auxiliary diagnosis of Alzheimer’s disease. *Neurocomputing*, 361:185–195, 2019.
- [93] Y. Zhao, B. Ma, P. Jiang, D. Zeng, X. Wang, and S. Li. Prediction of Alzheimer’s Disease Progression with Multi-Information Generative Adversarial Network. *IEEE Journal of Biomedical and Health Informatics*, 25(3):711–719, 2021.