FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Brain Morphometry from the Analysis of Magnetic Resonance Images: Computational Techniques and Applications in the Study of Dementia

Vanessa Alexandra Martins Ramos

Mestrado em Engenharia Biomédica

Supervisor: João Manuel R.S. Tavares

August 4, 2021

© Vanessa Alexandra Martins Ramos, 2020

Brain Morphometry from the Analysis of Magnetic Resonance Images: Computational Techniques and Applications in the Study of Dementia

Vanessa Alexandra Martins Ramos

Mestrado em Engenharia Biomédica

Faculdade de Engenharia da Universidade do Porto

August 4, 2021

Resumo

A doença de Alzheimer é uma doença neurodegenerativa progressiva. É o tipo de demência com maior incidência na população, sendo uma das principais causas de morte nos países desenvolvidos. A Alzheimer causa danos ou destrói células nervosas em partes do cérebro específicas que são responsáveis por funções cognitivas essenciais, o que leva à perda de capacidades como memória e raciocínio. Numa fase mais tardia da doença pode chegar a incapacitar a realização de atividades do quotidiano devido ao detrimento das funções motoras levando muitas vezes ao acamamento do paciente. Como não há cura, a única opção a fim de melhorar a qualidade de vida do doente e de seus cuidadores baseia-se na diminuição da intensidade dos sintomas.

Além dos testes cognitivos e mentais, as imagens cerebrais surgem de forma promissora neste área, avaliando a perda de massa cerebral. Assim, os sistemas de diagnóstico auxiliado por computador passam a ter um papel fundamental no auxílio aos especialistas na interpretação das imagens, tornando esta tarefa menos subjetiva e demorada. Deste modo, surgem as técnicas de *Machine Learning* e neste caso, mais especificamente, *Deep Learning*, de forma a que o sistema alcance uma melhor performance.

Nos últimos anos, artigos relacionados com o diagnóstico de Alzheimer aplicando técnicas de *Deep Learning* têm se tornado mais populares, havendo um aumento drástico a cada ano. Algoritmos com base neste tipo de abordagem têm alcançando uma precisão maior do que com técnicas gerais de *Machine Learning*. Apesar disso, o diagnóstico e a classificação da doença de Alzheimer continua a ser um foco no uso de técnicas computacionais nesta área, por ainda apresentar inúmeros desafios, requerendo o aperfeiçoamento de abordagens com o estudo e incorporação de técnicas inovadoras e que sejam mais benéficas tanto no resultado final como na perceção do modelo pelos especialistas.

Atendendo ao domínio da classificação da doença de Alzheimer através da aplicação de técnicas de *Deep Learning*, o presente documento exibe uma visão geral da doença, assim como da técnica de imagem escolhida, seguida de alguns conceitos relativos ao processamento de imagem médica, e a teoria dos algoritmos de *Machine Learning* utilizados nesta área. Quanto ao foco deste estudo, verifica-se que se trata de um problema atual, devido aos inúmeros artigos encontrados que propõem diferentes abordagens para se chegar a uma classificação mais rigorosa.

De forma a distinguir os diferentes estágios da doença, esta Dissertação apresentada diferentes abordagens, incluindo implementações em duas e três dimensões para realizar a tarefa, seguindo técnicas promissoras e atuais na área de *Deep Learning*. Por fim, a aplicação de *Convolutional Neural Networks* para o problema em questão revelou-se apropriada tal como comprovado com resultados obtidos acima de 88%.

Abstract

Alzheimer's disease is a progressive neurodegenerative disorder. It is the type of dementia with the highest incidence in the population, being one of the leading causes of death in developed countries. It causes damage or destruction of nerve cells in the brain parts involving essential cognitive functions, which leads to a loss of capabilities like memory and thinking. In a severe stage of the disease, the patient can be prevented from performing quotidian activities due to damage in the motor functions enforcing to bed-bound. As there is no cure, the only options are based on decreasing the strength of symptoms improving the patient and his caregivers' quality of life.

Besides the cognitive and mental tests, neuroimaging enrolls promisingly in this field, assessing brain matter loss. Thenceforth, the computer-aided diagnosis system employs a fundamental role in helping the experts in image examination, making this task less subjective and time-consuming. Thus arises machine learning and, specifically deep learning techniques, to improve these systems.

In recent years, papers reporting Alzheimer's diagnosis with deep models have become popular and have been increasing drastically every year, reaching a higher accuracy than general machine learning algorithms. Nevertheless, AD diagnosis and classification remains a focus in this area, still challenging and evolving. It requires the improvement of approaches with the study and incorporation of more beneficial and innovative techniques.

Attending the domain of Alzheimer's disease classification through deep models, the report presents an overview of this disorder, the chosen imaging technique, followed by the biomedical image processing, and the theory of machine learning algorithms used in this area. Regarding this study's focus, it is verified that this is a nowadays problem, with numerous papers proposing different approaches to reach an accurate classification.

In order to distinguish the different stages of the disease, this Dissertation presents different approaches, including implementations in two and three dimensions to accomplish the task, following promising and current techniques in the area of *Deep Learning*. Finally, the application of Convolutional Neural Networks for the problem in question proved to be appropriate, as shown with results obtained above 88%.

Contents

List of Figures

List of Tables

Abbreviations and Symbols

AD	Alzheimer's Disease
ADAS	Alzheimer's Disease Assessment Scale
ADAS-Cog	Alzheimer's Disease Assessment Scale-Cognition
ADNI	Alzheimer Disease Neuroimaging Initiative
AE	Auto-Encoder
AI	Artificial Intelligence
AIBL	Australian Imaging, Biomarker and Lifestyle
BGRU	Bidirectional Gated Recurrent Units
CAD	Computer Aided Diagnosis
CNN	Convolutional Neural Network
CSF	Cerebrospinal Fluid
CT	Computed Tomography
DBM	Deep Boltzmann Machine
DBN	Deep Belief Network
DL	Deep Learning
DNN	Deep Neural Network
DPN	Deep Polynomial Network
DTI	Diffusion Tensor Imaging
ELM	Extreme Learning Machine
FA	Fractional anisotropy
FCL	Fully - Connected Layers
fMRI	functional Magnetic Resonance Imaging
FOV	Field of View
FPR	False Positive Rate
GAN	Generative Adversarial Network
GM	Gray Matter
JLLR	Joint Linear and Logistic Regression
LASSO	Least Absolute Shrinkage and Selection Operator
LDMIL	Landmark-based Deep Multi-Instance Learning
LM	Logical Memory
LSTM	Long ShortTerm Memory
MCI	Mild Cognitive Impairment
MD	Mean Diffusivity
MIRIAD	Minimal Interval Resonance Imaging in Alzheimer's Disease
ML	Machine Learning
MLP	Multi Layer Perceptron
MM	Multi-Modality
MMSE	Mini-Mental State Exam

MoCA	Montreal Cognitive Assessment
MPRAGE	Magnetization Prepared Rapid Acquisition Gradient Echo
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
NC	Normal Control
NMR	Nuclear Magnetic Resonance
OASIS	Open Access Series of Imaging Studies
PCA	Principal Component Analysis
PET	Positron Emission Tomography
pMCI	progessive Mild Cognitive Impairment
RAVLT	Rey Auditory-Verbal Learning Test
RBM	Restricted Boltzmann Machine
ReLU	Rectified Linear Unit
RF	Radio Frequency
RL	Reinforcement Learning
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
ROI	Region of Interest
SAE	Stacked Auto-Encoder
SCFR	Spatially - Constrained Fisher Representation
SELU	Scaled Exponential Linear Unit
sMCI	stable Mild Cognitive Impairment
SNP	Single Nucleotide Polymorphism
SNR	Signal-to-Noise Ratio
SVM	Support Vector Machine
WHO	World Health Organization
WM	White Matter

Chapter 1

Introduction

This initial chapter contains a light context of the disease and the neuroimaging technique focused on the study. Afterward, it is presented the motivation and the main goals of this project, as well as the main contributions that arose. The document structure is also declared by the end of this chapter.

1.1 Context

The human brain is the command center of the individual nervous system controlling all our (voluntary or involuntary) actions and the entire body. It receives signals from the sensory receptors and outputs information to the muscles and glands, besides all the mental capability of thinking in a complex form. The brain and spinal cord are parts of the Central Nervous System (CNS), while the Peripheral Nervous System (PNS) is constituted by the nervous tissue outside the CNS (nerves and ganglia). The neuron is the fundamental unit of the nervous system, responsible for the communication within this system in the form of action potentials. Depending on the cerebral area, the neurons have specific functions. Due to the brain's complexity, it is still one of the main focuses of studies in the scientific community. [2].

Dementia is a syndrome that affects brain activity, usually of a chronic or progressive nature, where there is deterioration in cognitive functions, like memory, orientation, comprehension, thinking, judgment, and others, beyond what might be expected from normal aging [3].

Worldwide, around 50 million people have dementia and it is estimated around 10 million new cases every year, expecting that number to triple by 2050, Figure 1.1. There are many different forms of dementia, being Alzheimer's disease the most common with around 60-70% of cases according to the World Health Organization [3]. With the increasing average life expectancy, the number of people with dementia and particularly with Alzheimer also rises. This pathology begins long before the first symptoms, around 20 years or more of alterations at the brain level, not notable to the affected person. With the advance of time, the nerve cells begin to be dysfunctional or even destroyed, which are part of the brain required in cognitive functions. At this point, the

patient starts to feel the disease progression. This brain-damaged can affect parts that lead to the inhibition of primary body function like movements [4].

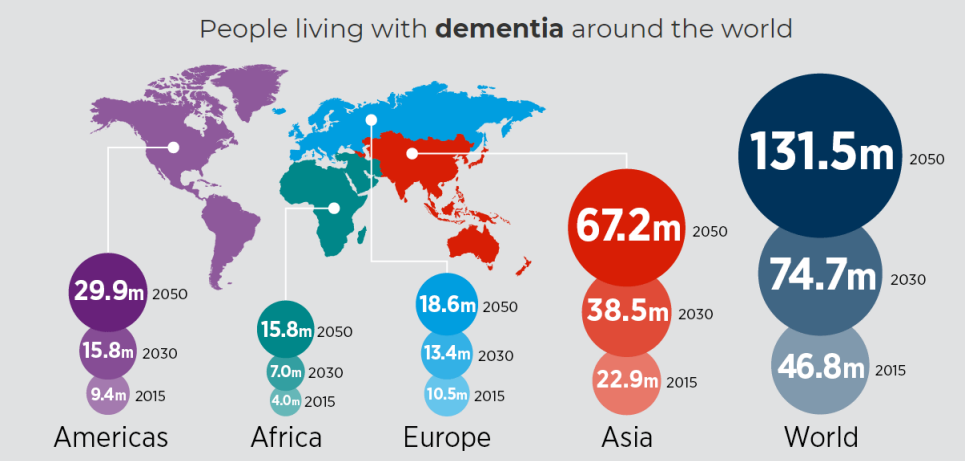


Figure 1.1: People living with dementia around the world (from [5]).

As there is no cure, there must be an early detection/diagnosis to promote early and optimal management, trying to keep everyday life as normal as possible, including the patient and his carers.

The first signal in the history of Alzheimer's disease was on November 3 of 1906, where Alois Alzheimer announced, "A peculiar severe disease process of the cerebral cortex" resulting from a long-term study performed on a female patient. The subject was observed and studied since she was admitted to the Frankfurt Psychiatric Hospital in November 1901 with 50-years-old showing symptoms such as sleep disorders, memory disturbances, and progressive confusion. During these years, the symptomatology growingly deteriorated until her death on April 8 of 1906, which allowed an investigation, through an autopsy, of the morphological and histological field. The results from the post-mortem examination, along with the clinical findings recorded, became the basis of Alzheimer's lecture reporting the changes later understood as plaques and neurofibrillary tangles [6].

Already, in 1911 Alzheimer described a presenile and senile dementing process in a male patient, which after autopsy were noticed only plaques, which was considered contradictory. These results were forgotten until 1995, when Maurer found the documentation and reinvestigated both cases with modern technology, concluding that the presence of only plaques or plaques with neurofibrillary tangles were different stages of the same disease [6].

The study of Alzheimer's disease begun with autopsied brains to observe the morphological changes. Afterward, the technology advance allowed the examination and disease investigation before the patients' death through neuroimaging.

The human brain imaging started with the evaluation of blood flow relating to brain function. It was only after the X-Ray Computed Tomography and Positron Emission Tomography that the Magnetic Resonance emerged.

The physical principles correlated to MRI, in that time known as Nuclear Magnetic Resonance, were disclosed separately by Bloch (Harvard University) and Purcell (Stanford University) in 1946. It was around thirty years later that Paul Lauterbur arose with a plan to create cross-sectional images like CT using NMR signals. Being a free ionizing radiation technology caused interest in the community and produced better images than CT, considering the sensitivity to soft tissues [7].

The analysis of dementia is predominantly based on clinical criteria, but neuroimaging was started to play an increasingly significant role over time. Magnetic Resonance Imaging remains at the core of differential diagnosis due to the valuable capability of tissue differentiation and a reliable morphometry presentation, exhibiting peculiar patterns of cortical and subcortical changes depending on the medical condition [8].

In the AD study, structural MRI can help detect progressive cerebral atrophy, one characteristic of the disease allowing an early detection notwithstanding the difficulty to visualize that feature in the initial stages.

MRI technology becomes an indispensable tool in the medical field due to its properties being the most used neuroimaging technique.

Depending on the imaging acquisition device, the image obtained will be different. The same happens within the MRI technology when its internal parameters vary or due to equipment diversity.

Thus, it may be necessary to apply specific techniques such as enhancement of the region of interest, remotion of image artifacts, or an image crop.

Of this need arose the Computer-Aided Diagnosis systems to assist specialists in the interpretation of medical images. Furthermore, in recent times, some developed systems implement artificial intelligence to help health professionals or make the system capable of deciding. In this way, it is avoided that professionals waste much time in routine tasks and can detect and identify morphology anomalies not visible in the first instance to the human eye.

1.2 Motivation and Goals

Alzheimer's disease is the most prominent neurodegenerative disorder. Therefore, it is a significant attention focus in medical and research areas generating an exponential growth of studies on the topic.

Since Alzheimer's disease gathers more and more patients, a lot relating to the increase of lifespan, early detection is crucial to retard the disease evolution and support the patients and their caregivers, usually direct family, to deal with the symptoms. By the WHO numbers, in 2019, Alzheimer's disease, jointly with other types of dementia, was the 7^{th} leading cause of death [9].

Thereby, this project aims to develop an automated procedure using artificial intelligence to distinguish brain lesions caused by the mentioned disorder. Before that, the treatment and analysis of the magnetic resonance images will be essential to perform the classification.

The application of the system in clinical practice is also a motivation reinforcement.

The main goals of this project are:

- The investigation of deep learning approaches to detect Alzheimer's disease and its different stages;
- The investigation and application of preprocessing imaging techniques in python;
- The development of a deep learning-based approach using the programming language - Python;
- The application of this system in clinical practice.

1.3 Main Contributions

The main contributions that arose from this project can be divided into two purposes. On the one hand, a theoretical purpose, where was done an extensive background analyzes of the literature around the adressed theme. On the other hand, in a practical aspect, it is included all the experimental tasks performed. In this way, it can be noted that foremost contributions achieved with this work are related essentially to:

- The contextualization of the stages and the methodology applied in the analysis of MR images of patients with Alzheimer's disease or mild cognitive impairment;
- A search for computational methods applied with the aim of disease detection;
- The study and application of generic methodologies in the processing of MR images and known intelligent classification systems applied in this field;
- The development of a network to achieve the classification between healthy subjects, mild cognitive impairment, and Alzheimer's patients;
- The analysis and comparison of different activation functions and the influence in the network;
- The submission of a review paper about the application of deep learning methods to detect the disease in the Neurocomputing Journal.

1.4 Document Structure

Besides the Introduction chapter, this document is organized into six distinct chapters, accompanied by a list of bibliographic references. Chapter 2 is explored the disease background, which defines the basic concepts of Alzheimer's disease, the epidemiological and clinical characteristics, followed by the diagnostic criteria and the disease treatment. Also, it is briefly addressed a description of Magnetic Resonance Imaging from the physical principles to image generation. The hardware, imaging parameters, artifacts, and considerations in neuroimaging scans are also

taken into account. Besides that, the present chapter reports the pipeline and the main steps in biomedical image processing and analysis, as well as preprocessing techniques usually applied in Alzheimer's challenge.

Chapter 3 refers to the typical Machine Learning process, highlighting multiple classifiers used in the problem in focus. Deep Learning and some worked architectures in this field are also briefly theoretically reported, along with some popular techniques explored in the last years. To finish the chapter, the most applied validation metrics are described. In Chapter 4 is presented a systematic review of Deep Learning applications in Alzheimer's and Mild Cognitive Impairment diagnosis. The papers presented in the literature are segregated into three categories. Articles that used only an imaging modality (MRI) in their methodologies, works that implemented multi-modality approaches, and projects where is done a time conversion prediction instead of pure classification.

The following Chapter 5 starts the practical plan from a theoretical point of view, giving an overview of the ensued procedure, a theoretical description of the Deep Learning algorithm chosen for this project development, and a description of the computational infrastructure and some managed software resources. The experimental work and results are described in Chapter 6 including the preprocessing steps, the two different implementations, and an analysis of the obtained results. To conclude, the final considerations and future work projection are given in Chapter 7.

Chapter 2

Background

The present chapter started to address the most crucial focus of this dissertation, Alzheimer's disease. Following a historical contextualization, proceeds to its description, causes, and disease stages. Succeding, there is a characterization of the diagnosis criteria and the treatment applied to the patients. Having a clarified idea about the disease, it is presented the technique adopted of image acquisition, Magnetic Resonance Imaging, where is described the basic principles of MR and expressed a brief description of the imaging characteristics of the disease and some conditioning factors in the acquisition process. By the end, it is characterized the habitual medical imaging processing and some preprocessing techniques that are usually applied in this area. After all, it is presented a summary relating to the information exposed in this chapter.

2.1 Alzheimer's Disease

Alzheimer's disease (AD) is a progressive neurodegenerative disorder being the most common cause of dementia. It is characterized by loss of capabilities like thinking, learning, and memory due to the damage or destruction of nerve cells in the brain parts involved in these cognitive functions.

2.1.1 Description of Alzheimer's Disease

Alzheimer's disease was reported for the first time by Alois Alzheimer more than 100 years ago. He identified as pathologic features the amyloid plaques and neurofibrillary tangles, which are nowadays still considered in the patients' evaluation. AD is a central nervous system disorder, being a progressive neurodegenerative disease, is usually considered a normal part of aging. It affects people aged 60 years and more, being considered a rare condition below that age [5]. In less than half of the events, AD appears to be pure, being in most cases a mix of dementia types. Among these, the most prevalent secondary diseases are Lewy body dementia and vascular dementia [10].

An AD brain often has cortical atrophy, involving a loss of neurons in the cerebral cortex. The frontal and temporal cortices are often the most affected regions, while primary motor and

somatosensory cortices most often appear unaffected. Frontal and temporal lobes play a crucial role in controlling voluntary motor functions, motivation, mood, memory, and abstract functions like thought and judgment [2].

Therefore, the early symptoms are based on memory loss events and apathy or depression. At an advanced stage, the symptoms can be impaired communication, confusion, weak judgment, and behavioral alterations. In the ultimate, the patient has difficulties in speaking, swallowing, and walking [4].

It is thought that this disorder starts 20 years or more before the onset of the first symptoms, with brain alterations unnoticeable to the individual affected. Only when the brain starts to be damaged is that the disease can be noticeable by the patient [4].

During AD's progression, with unnoticeable followed by noticeable brain changes, it can be distinguished three broad phases: preclinical AD, Mild Cognitive Impairment (MCI), and AD, which in turn can be divided into three stages of mild, moderate, and severe reflecting the symptoms implication in the patients daily, demonstrated in Figure 2.1. The duration of each stage varies according to age, hereditary characteristics, gender, and other factors [4].

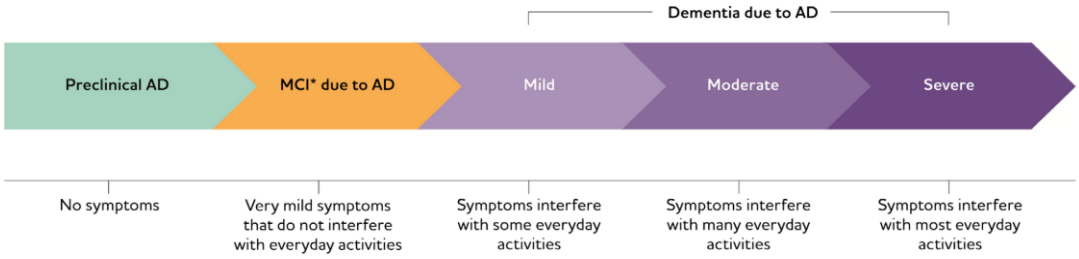


Figure 2.1: Alzheimer disease evolution (From [4]).

In the preclinical phase, the individuals possess measurable functional brain differences without structural differences neither developing symptoms that suggest the earliest manifestations of the disorder (biomarkers). These biomarkers stand out the unusual beta-amyloid levels (visible in PET or CSF analysis) and the decreased glucose metabolism (shown on PET scans).

Not all the individuals who fit in this stage with AD biomarkers can develop MCI or dementia symptoms. Furthermore, these differences can remain constant until death, making it more difficult for the AD early-stage study.

The MCI stage is the previous stage of AD, where the biomarkers are more visible than in the preclinical phase and begin to have slight symptoms like problems in memory and thinking. These mild symptoms occur due to the lack of brain capacity to compensate with new pathways for the actual damage. They do not interfere with the patient's daily life but can be noticeable to people close to him.

From longitudinal studies, MCI patients reported that after two years' follow-up, 15% of individuals older than 65 years developed AD, and in five years' follow-up 32%. However, in this phase, the individuals can also remain stable or even revert to normal cognition [4].

In the diagnostic AD, the patients have noticeable cognitive problems, and some daily abilities are compromised. The symptoms change over the years, revealing the severity of brain damage, making possible the discrimination in mild, moderate, and severe [4].

1. Mild - Most activities and functions can be realized independently, but assistance is needed to remain safe or to complete some activities.

2. Moderate - Often the longest stage, the patients may have problems communicating, performing daily activities and personality/behavioral changes.

3. Severe - The individuals need help and supervision all day due to the inability to perform the quotidian activities. The brain can be damaged in the movement areas, precluding simple activities, forcing the patient to be bedridden, and causing other issues like circulatory system problems or skin infections. They can also be affected in the swallowing process, making feeding through the conventional way impossible.

2.1.2 Epidemiological and Clinical Characteristics

It was predicted that more than 47 million people were directly affected by dementia in 2015 [5]. With the increase of lifespan, chronic diseases become more widespread due to lifestyle changes and behaviors that can make people more predisposed to them. Some studies demonstrate how AD is influenced by age, other diseases, habits, or activities, increasing or decreasing the risk of developing this disorder.

As already mentioned, AD affects people over 60 years. Previous studies reported that people under that age with some form of dementia represent 2 to 8% of all cases and are thus considered a rare condition [5]. The prevalence of dementia increased exponentially with age, Figure 2.2.

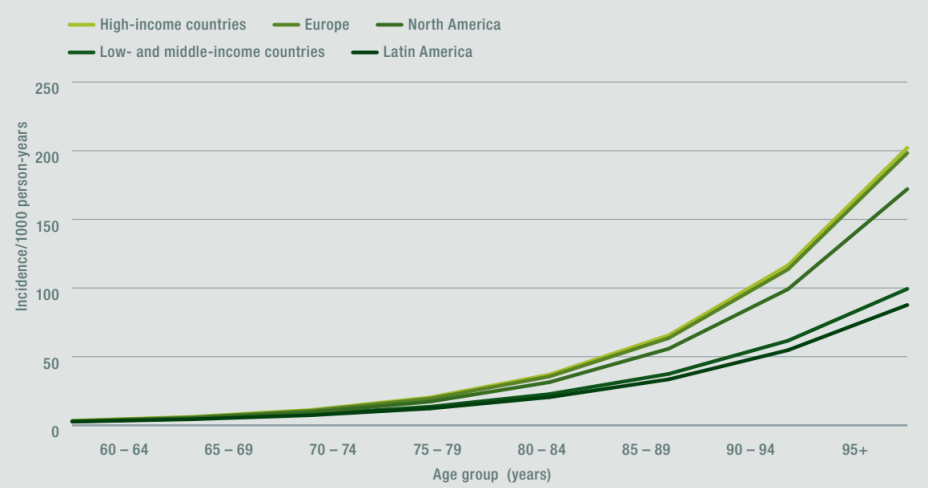


Figure 2.2: Estimation of the incidence of dementia by world region/development status (From [11]).

The incidence of dementia doubled every 5.9 years, from 3.1/100 person-years at age 60-64 to 175/1000 person-years at age 95+ shown in image 2.2. The high-income countries appear to have

more dementia cases but are deeply related to the health service available to almost all populations [12].

Gender is noted as a variable in some regions, such as Latin America, Western Europe, Caribbean, East Asia, and Asia South, where women's disease prevalence is higher 14% to 32% than for men [5].

Beyond age and gender, a genetic risk factor was identified. The APOE-e4 genotype's presence provides a substantial risk of developing AD. Individuals who carry that gene possess an odds ratio of 3 compared to non-carriers [10].

Other factors like cerebrovascular diseases, diabetes, hypertension, smoking, obesity, and dyslipidemia (abnormal amount of lipids in the blood) have been associated with an increase in AD risk [11].

The first biomarkers shown by Alois Alzheimer were plaques and neurofibrillary tangles. These two remain as microscopic features nowadays. Additionally, tau-positive neuropil threads, dystrophic neurites, activated microglia, reactive astrocytes, eosinophilic Hirano bodies, Granulo-Vacuolar Degeneration (GVD), and Cerebral Amyloid Angiopathy (CAA) can be detected in AD patients. These features lead to the loss of synapses and neurons in regions commonly associated with the disorder revealing symptoms. Evidence shows that amyloid deposition and tau pathology can precede by decades the structural changes [10].

Regarding the macroscopic features, the AD brain presents cortical atrophy. The frontal and temporal cortices have extended sulcal spaces with atrophy of the gyri. This atrophy results in an enlargement of the frontal and temporal horns of the lateral ventricles, shown in Figure 2.3. Medial temporal atrophy affecting amygdala and hippocampus is also a feature of AD. This volume loss essentially influences the white matter, a region that contains the axonal connections between the nerve cells allowing the brain cells to transmit and receive messages [10].

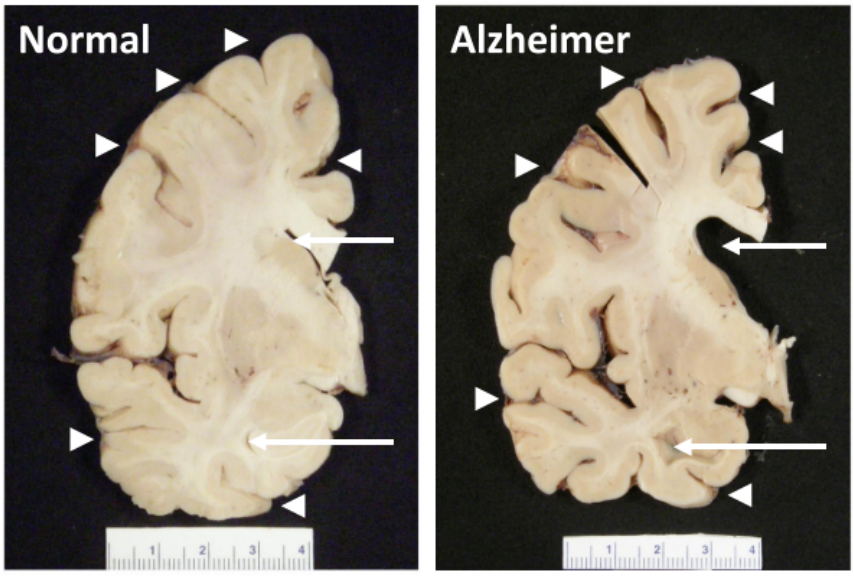


Figure 2.3: Differences between a normal and an AD brain (Adopted from [10]).

2.1.3 Diagnostic Criteria

To diagnose AD prior to autopsy, it is necessary to analyze the features previously mentioned. The amyloid and tau biomarkers are not accessible to clinicians.

The physicians use multiple approaches to diagnose Alzheimer's or dementia, which is easier to determine, being difficult to identify the cause. The most potent biomarker candidates for Alzheimer's disease involve neuroimaging studies using MRI or Positron Emission Tomography (PET) and proteins in CerebroSpinal Fluid (CSF).

Before the imaging tests, the clinicians can perform a neurological exam where reflex, coordination, speech, sensation, and eye movement are tested [13].

Moreover, there are two assessments commonly used the MMSE and the Mini-Cog test. The MMSE (Mini-Mental State Exam) is based on a series of questions to test a range of everyday mental skills. This test indicates a score that separates different levels of dementia. Additionally, the Mini-Cog test is a brief test where the patient needs to complete two tasks to examine the memory condition and the cognitive function with a clock's drawing and a time defined by the examiner [13].

The ADAS (Alzheimer's Disease Assessment Scale) and ADAS-Cog (Alzheimer's Disease Assessment Scale-Cognition) are based on clinical trials to assess AD's cognitive and non-cognitive symptoms being time-consuming [14].

Besides that, there is a growing research area to develop computer-based tests, cognitive tests, which evaluated thinking, learning, and memory capability [13].

A genetic test can be performed to examine the risk gene for AD, the APOE-e4, by blood analysis [13].

The brain imaging workup often includes structural imaging with MRI or CT, which can reveal in a first instance other conditions that can cause similar symptoms, such as tumors or strokes. A PET exam can also be executed to find high beta-amyloid levels, whereas normal levels suggest that AD is not the dementia cause [13], considering that the levels arise even before the MCI stage diagnosis and has a regularly high value in the AD stage, Figure 2.4. In this graph, it is visible that amyloid and functional markers' progression before the MCI stage is more significant than the evolution within the AD stage [15].

Analyzing the CSF proteins is also helpful to perform the diagnose. One example is the study about high CSF T-tau, which turned out valuable to discriminate MCI cases that will progress to AD from those that did not progress (with 90% sensitivity and 100% specificity) [16], being less relevant to track later disease progression. In Figure 2.4 is visible the earliest sites of tau deposition during the disease progression.

The AD diagnosis area is currently in research, looking for neuroimaging techniques and tools to better diagnose and track disease progress.

2.1.4 Disease Treatment

The AD symptoms are multiple and affect both cognitive and motor functions.

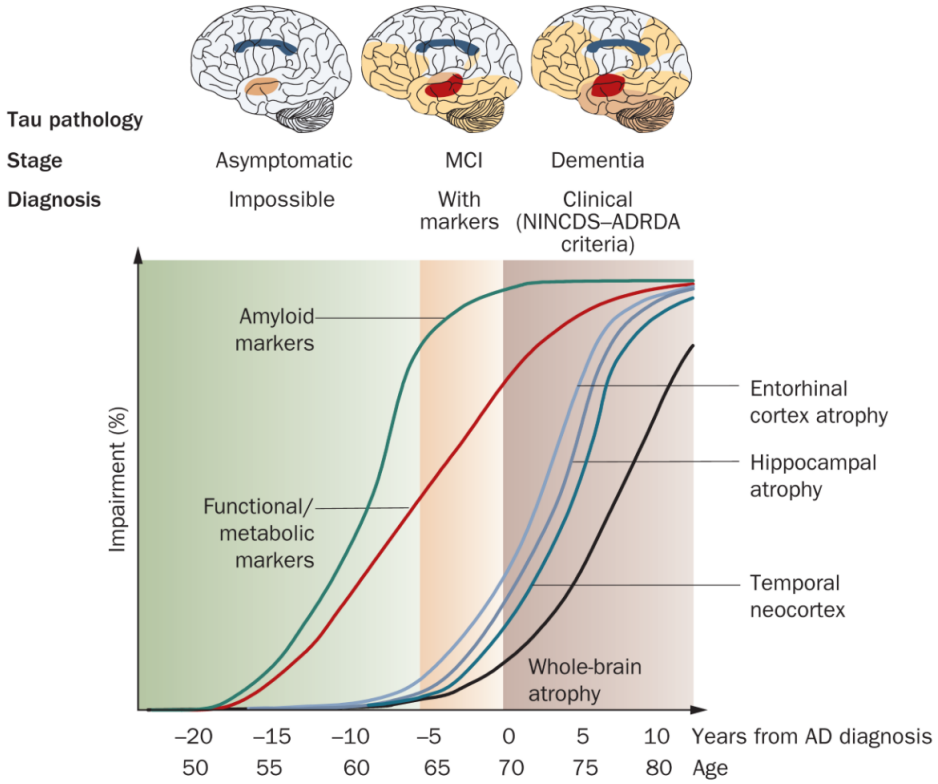


Figure 2.4: Markers impairment during the disease progression, (Adopted from [15]).

Nowadays, there is no cure for this disorder or a way to stop or slow the progression.

Therefore, there are options to help treat the symptoms improving the patient and his caregivers' quality of life. Pharmacologic treatment and non-pharmacologic therapy are adopted depending on the finality [5].

As AD causes the damage or destruction of neurons, some drugs can increase the number of neurotransmitters to improve cognitive symptoms, but for a limited duration, and the effectiveness varies with the individual. To other symptoms like depression or aggression can be applied very carefully and taking into account the risks, antipsychotics [4].

On the other side, it has been proven that non-pharmacologic treatments for agitation and aggression seemed to be more efficient in reducing them. These therapies include computerized memory training, listening to favorite music to stir recall and the use of special lighting to help in sleep disorders. Additionally, despite the difficulty to choose the therapies more valuable, researchers found that aerobic and non-aerobic exercise had positive effects on overall cognitive capacity and may reduce the speed of cognitive decline. Also, cognitive stimulation, music-based therapies, and psychological treatment can increase patients' quality of life [4].

Hence, it is essential to adopt strategies to reduce AD's risk like education, leisure activities, a Mediterranean diet, and physical activity [11].

2.2 Magnetic Resonance Imaging

The acquisition of images by Magnetic Resonance (MR) has been increasingly used as a non-ionizing radiation imaging technique with high sensitivity in differentiating human body tissues. Notice the usefulness of MRI to assist in analyzing and diagnosing diseases, particularly Alzheimer's disease, potentiated a rapid expansion and dissemination of studies and applications using this imaging technique.

2.2.1 Physical principles

The basic principles of Magnetic Resonance based on Nuclear Magnetic Resonance (NMR) have been understood since the early 1900s, where atomic nuclei were studied. Thence, it was discovered that they have the property, within a magnetic field, of absorbing a determined Radio-Frequency (RF) energy and then release it at a specific frequency. Subsequently, in 1938, it was demonstrated that particles have intrinsic quantum properties examined into nuclei's magnetic properties, the NMR phenomenon. Some years later, the procedure was redefined, allowing the measurement of NMR signals from liquids and solids [1].

Nevertheless, the first signals of Magnetic Resonance Imaging were given in 1973 by Paul Lauterbur. He revolutionized this field by developing a system for spatially encoding the released signal employing linear magnetic field gradients. Simultaneously, Peter Mansfield disclosed a way to demonstrate the solids spatial structure by submitting a linear gradient across the object. These discoveries allowed the receipt of a Nobel Prize in 2003 by the two scientists [1, 17].

The principle of MR is based on energy transfer manipulating the net magnetization. It is applied an excitation pulse of radiofrequency energy, which will be absorbed by the protons at a specific frequency proportionally to the magnetic field. Subsequently, the protons reemit that energy being then detected [18].

2.2.1.1 The nucleus and its properties

Solely the nuclei that have a non-null magnetic moment, a non-null spin, are observable by NMR. The most used nuclei in biological applications with a high natural abundance are Hydrogen (H), Carbon (C), Fluorine (F), Sodium (Na), Phosphorus (P), and Potassium (K), Table 2.1 [1]. A nucleus has no spin when it has an even atomic number and an even number of atomic weights, thereby not interact with an external magnetic field [18].

Table 2.1: Common nuclei for MR (Adopted from [1]).

Nucleus	Atomic Number	Relative abundance (%)	Spin (I)	Giromagnetic ratio (MHz/T)	Relative sensitivity	Abundance in the Human body (% of atoms)
H	1	99.98	1/2	42.58	1	63
C	13	1.11	1/2	10.71	0.016	0.13
F	19	100	1/2	40.05	0.83	0.0012
Na	23	100	3/2	11.26	0.093	0.037
P	31	100	1/2	17.23	0.066	0.14
K	39	93.1	3/2	1.99	5.08×10^{-4}	0.031

Within these, MRI comprises imaging the H atom's nucleus due to the body constitution of tissues that contain principally water and fat, both forms containing hydrogen.

The H nucleus consists of a single proton having a more significant value of the gyromagnetic ratio, which produces a better visualization. It has a spin of $\frac{1}{2}$, an intrinsic property, influencing angular momentum and a local magnetic field, and being fundamental to obtain an MR image [18].

An H nucleus has an axis of rotation, like a vector with a determined orientation and magnitude being its magnetic moment parallel to the rotation axis. The spin orientation and the variations induced in it present the basis for the MR signal [18].

With the omission of an external magnetic field, the spins will be randomly aligned in space resulting in zero net magnetization. Whereas, in the magnetic field presence, each spin will align in parallel (lower energy) or antiparallel (higher energy) to the applied field direction, becoming polarized, shown in Figure 2.5 [1].

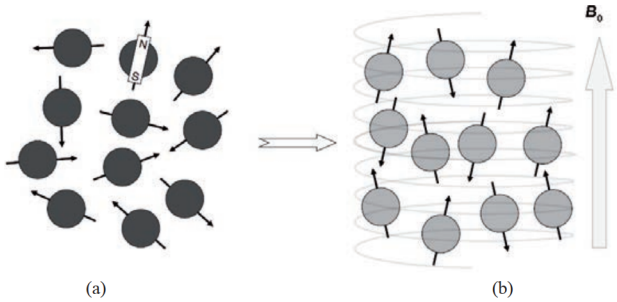


Figure 2.5: (a) The spins randomly oriented in space, (b) when applied a magnetic field, the spins become aligned (From [1]).

Beyond the polarization changes, the tissue will also become magnetized with an M_0 value, known as net magnetization, equation 2.1. The orientation of net magnetization will be the same as the magnetic field (B_0) and proportional to it [18].

$$M_0 = \chi \times B_0 \tag{2.1}$$

Where χ is the magnetic susceptibility.

The net magnetization is commonly treated as a single spin using a rotating frame in the z-axis, parallel to B_0 and with a proportional value. Thereby, the x and y axes rotate at the Larmor frequency, w_0 in Hz, equation 2.2 [18].

$$w_0 = \gamma \times B_0 \tag{2.2}$$

Where γ is the gyromagnetic ratio.

The manipulation of M_0 is the basis of MR. During the application of a short pulse of radiofrequency energy, the protons exposed will absorb a part of this energy at a specific frequency. Then, the protons will reemit energy at the same frequency, defined by Larmor, and consequently received and analyzed by the equipment [18].

2.2.1.2 Relaxation

Relaxation is the process by which protons discharge the energy that was absorbed from the pulse. It is a fundamental process in this imaging technique, like the own energy absorption, because it provides the essential mechanism for image contrast [18].

Following the excitation, the protons return to the original configuration releasing the energy. Relaxation times are measured for a complete sample and are averaged or statistically estimated. It can be measured two different relaxation times, T1 and T2, which differ in the final energy settlement [18].

Relaxation time T1, also known as the spin-lattice relaxation time or longitudinal relaxation time, is the time needed for the z-axis, longitudinal direction of magnetization back 63% to its equilibrium state (minimum energy) after the RF pulse perturbation. Some part of the spins' absorbed energy is missed to its surroundings ("lattice"), thence the spin-lattice relaxation. After the pulse, M_0, which is initially parallel to B_0, will rotate into the transverse plane due to energy absorption.

Beyond the interaction with the lattice, the spins also interact with other spins, producing spin-spin relaxation. A spin generates a particular magnetic field, which lightly changes the others in its surroundings, including their precessional frequency. There is an energy transfer from an excited proton to a different neighboring [1]. Therefore, the T2 time, also known as spin-spin time or transversal relaxation time, corresponds to the time after the 90º pulse that runs until the transverse magnetization is 37% of its equilibrium value [18].

This proton excitation process followed by signal detection is the basis for image creation and is repeated several times until a satisfactory amount of data have been acquired. Hence, it is called the repetition time (Tr) to the interval between this process recurrence.

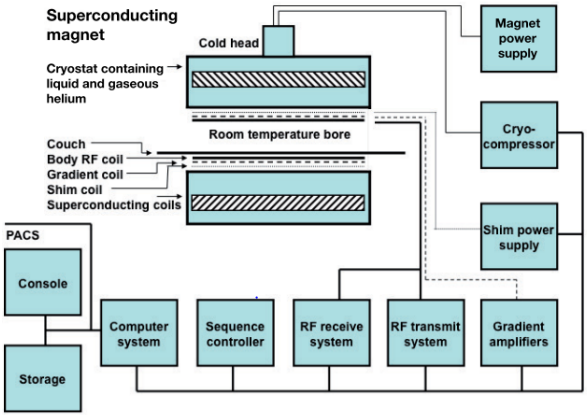
The image characteristics will change with a variation of Tr. To a shorter value, tissues with a smaller T1 value, that is, with a faster relaxation process, will appear brighter compared to tissues with a longer time, emphasizing the differences. On the other hand, when adjusted the echo time (Te), the interval between excitation and data acquisition, the image contrast is managed due to T2 differences. With more significant echo time, tissues with short T2 values will present a darker color than those with higher values.

Table 2.2 presents a list of T1 and T2 values for different tissues to a specific magnetic field of 1.5 Tesla (T), noting that the relaxation times vary with the strength.

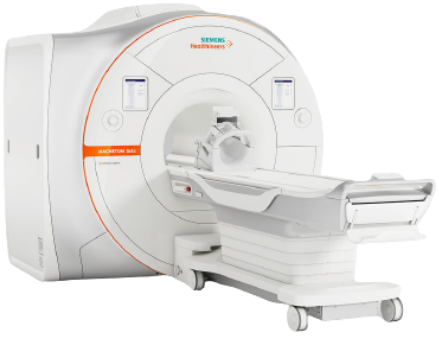
Table 2.2: Tissues and their T1 and T2 values at 1.5T (adopted from [1]).

Tissue	**T1 (ms)**	**T2 (ms)**
Blood (oxygenated)	1200	200
Blood (deoxygenated)	1200	50
White Matter	790	90
Grey Matter	920	100
Cerebrospinal fluid	4000	2000

2.2.2 Equipment



(a) Components of MRI scanner (from [1]).



(b) MRI scanner 1.5T (from Siemens Healthineers [19]).

Figure 2.6: MRI hardware.

The MRI system includes many components, Figure 2.6a, beginning with at least one workstation or PC, providing the interface to the operator, who can control and supervise the leading equipment, plan and execute the image acquisition. Moreover, to examine and analyze the results through particular software directed to special clinical questions [1].

Presently, medical field MRI scanners use cryogenic superconducting magnets in the range of 0.5 T to 1.5 T (example present in 2.6b). However, 3T systems are being studied in the research area due to higher energy advantages, such as the improvement of signal-to-noise ratio (SNR) and higher resolution. Despite that, there are also some disadvantages like magnetic susceptibility, instability, and artifacts [20].

An MRI system's main requirement is to produce a uniform and strong magnetic field to guarantee the Larmor resonance. This can be achieved by the central magnet or together with static and electronic shims (gradient system), which attend to fix the imperfections in magnet design, static effects from neighboring steel structures, and the patient's intrinsic magnetic susceptibility [1].

Besides the magnet, shims, and the computer system, another hardware group is essential, the radiofrequency system, which includes the RF coils necessary to the RF pulse transmission and signal reception, adapted to the understudy body part and designed considering the equipment characteristics [21].

2.2.3 Image acquisition and reconstruction

2.2.3.1 Pulse Sequences

According to a particular image application, it is mandatory to optimize the MR sequences, providing an optimum signal contrast. *Sequence* in this sense means a mixture of well-defined repetitive

RF pulses and signal acquisition, which must be synchronized with the gradients of the magnetic field. The most used in the general medical field are gradient-echo and spin-echo sequences. Within these two groups, there are several variations [21]. These two differ mainly in two points: the production of transverse magnetization using gradient fields and flip angles of less than 90º.

- **Gradient-Echo Sequences**

  In this type of sequence, a unique RF pulse is employed during particular Tr periods. The data is acquired simultaneously as the free induction decay, which is the attenuation of the transverse magnetization caused by the RF excitation [21].

  One disadvantage of the gradient echo sequence is its sensibility to magnetic field inhomogeneities resulting in a signal reduction, which can be avoided when used the spin-echo sequence [1].

- **Spin-Echo Sequences**

  In a spin-echo sequence is used a second RF pulse of 180º flip angle per period. The data is collected along the spin-echo once the spins become refocused. Due to the large flip angles, it is needed long TR times to allow a satisfactory recovery of the longitudinal magnetization [21].

- **Inversion-Recovery Sequences**

  Besides the two principal sequences, this third occurs when is used a 180º inversion pulse before the initial excitation. Thus, the data is obtained while recovering the longitudinal magnetization toward its equilibrium M_0. The time separating the inversion pulse and the first excitation is called the inversion time (TI) [21].

2.2.4 Imaging features

2.2.4.1 Imaging parameters

There is a range of parameters for any given pulse series that empowers the flexibility of MR imaging. To control the signal contrast is used the flip angle and timing parameters. Concerning image resolution, it is determined by spatial resolutions, and the number of signal averages commands the total SNR [21].

2.2.5 T1-weighted images

To produce this type of image is commonly used standard single echo sequence when acquired with relatively short Tr and Te, using a 180° inversion RF pulse before the excitation pulse.

The T1-weighted images enlarge the fat tissues' signal, appear whiter than other tissues like muscles or fluids, being the muscle hyperintense (whiter) compared to fluids. Concerning brain image, the grey matter presents a darker color compared with the white matter [18].

2.2.5.1 T2-weighted images

Relating to the T2-weighted images, to produce a spin echo is used a 90°–180° pair of pulses long Te when Tr is long enough to enable the total T1 relaxation for most tissues.

This sequence is useful for visualizing soft tissues, being fluids whiter than muscle tissues. Applying that in neuroimaging, the white matter is hypointense, darker, contrasted to grey matter [18].

2.2.5.2 Protocol Considerations for Brain

Considering neuroimaging, MR examinations' norm remains in sagittal T1-weighted spin-echo and transverse T1- and T2-weighted standard spin-echo.

For performing MR when there is suspicion of Alzheimer's, the subjects are scanned at 1.5T or 3T MRI scans. It is also selected the 3D Magnetization Prepared Rapid Acquisition Gradient Echo (MPRAGE) imaging sequence [22].

This type of gradient-echo sequence was developed to minimize the equipment vendor-to-vendor differences and to present a superior contrast gray/white matter, producing a superior performance in some projects demanding cortical segmentation and imaging times below 10 minutes to both field strengths [23].

2.2.6 Safety

The MRI presents potential risks in three major magnetic fields [24]:

- The static magnetic field may attract and accelerated ferromagnetic structures towards an MRI scanner's bore and interfere with implanted devices;
- The radiofrequency field can cause tissue heating, peculiarly in the presence of implants;
- The quick switching of magnetic gradient fields can provoke peripheral nerve stimulation and implant heating.

Beyond the security and the proper functioning of the equipment guaranteeing all the exam room conditions, it is crucial to ensuring the patient's well-being. For example, claustrophobic patients, children, or subjects with pain may require sedation or anesthesia during the exam due to the long duration and the motion sensibility, which can cause image artifacts resulting in useless results [24].

2.2.7 Alzheimer's Disease Imaging Features

The structural MRI had sequences that are particularly sensitive to different tissue characteristics and are applied to AD.

In AD, the imaging features can be divided into evaluating the overall atrophy (or volume) and changes in specific tissues that cause signal alterations on specific sequences, such as white matter in T1-weighted imaging [25].

A neurodegeneration feature is the progressive cerebral atrophy, visualized in life with MRI, preferentially with 3D T1-weighted. The earliest atrophy site is usually the entorhinal cortex, followed by the hippocampus, showed in Figure 2.7, amygdala, and parahippocampus. The atrophy in the medial temporal lobe is also considered a biomarker in MR images to AD diagnosis [25].

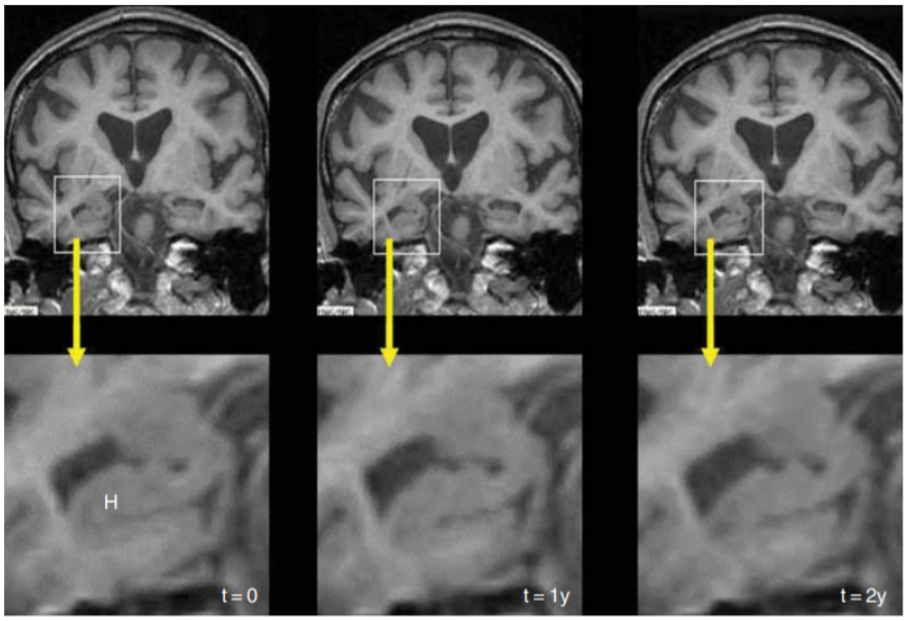


Figure 2.7: Three coronal T1-weighted images were acquired with an interval of 1 year of the same AD patient, which show progressive atrophy in hippocampal area, (from [25]).

2.2.8 Artifacts

In MR images, artifacts correspond to pixels that do not loyally reproduce the study's anatomy. Artifacts can be categorized into multiple forms. One manner is separating them into three groups, according to the signal misregistration's cause [18].

The first group comes from the patient's motion during the scan, the second group results from specific measurement techniques or parameters, and the last one is independent of the patient or measurement technique being provoked by an MR scanner malfunction or an external source [18].

Many artifacts may be removed or significantly decreased with attention to quality assurance, tuning, and optimizing sequences or parameters [1].

2.2.8.1 Motion Artifacts

Motion artifacts result from tissue movement at the time of the data acquisition. Common effects are relocated, ghosted, or low-intensity images of fat in the body. The most common motion artifact is the blood flow, which can be suppressed by inflowing blood out of the plane, and sequences involving gradient lobes to rephase motion. Signal averaging, navigator-induced acquisitions or rotation of the phase encode direction are all forms for reducing this type of problem [1].

2.2.8.2 Sequence/Protocol-Related Artifacts

This artifacts class derive from specific process measurements used to acquire images, more relating to technical aspects like the pulse sequence and the method used in data collection. The result in the image is relatively constant throughout the measurement and easy to recognize. The aliasing, chemical shift, phase cancellation, truncation, Coherence, and the magnetic-susceptibility difference can be highlighted within this group [18] and are presented in Table 2.3.

Table 2.3: Types of Sequence/Protocol-Related Artifacts.

Artifact	**Description**
Aliasing or "wrap around"	Aliasing happens when the sample extends outside of the field of view (FOV). Then, the discrimination is not possible between tissues within or on the edge of the FOV and outside [1].
Chemical Shift	This effect occurs due to the different frequencies resonate of water and fat. Usually, the systems have their reference to water. Consequently, where localization depends on frequency, slice selection or frequency encoded reading, the fat signal is spatially relocated from the water signal, resulting in a shifted image. The overlay of water and fat signals in the readout direction can produce a bright signal or void [1].
Phase Cancellation	The chemical shift differences also promote phase cancellation artifact, perceived in out-of-phase gradient-echo images. In voxels containing both fat and water, the signal intensity variations are more prominent in addition to those due to relaxation. Thus, for some choices of TE, out-of-phase TE, a minimal signal will be caught, signal cancellation, which will appear as a dark band [18].
Truncation	Truncation artifacts are caused by insufficient digital sampling of the echo, occurring when data acquisition is terminated, at the same time that an important signal is still emitted. Resulting in ringing at the image's sharp edges, typically seen in frequency or phase, or both directions, as parallel bands [1].
Coherence	Coherence is an artifact class, that depending on the used measurement technique and how it is applied to the scanner, the appearance will be variable. They are generated by RF pulses, which induce undesirable transverse magnetization that leads to the signal observed. Depending on the cause, it can produce a line in the final image or an additional image [18].
Magnetic-Susceptibility Difference	Magnetic-Susceptibility Difference artifacts occur mainly at the interface between soft tissues and different susceptibility areas. At the interface between these two areas, a significant change in the local magnetic field is present over a short distance, resulting in signal loss [18].

2.2.8.3 External Artifacts

External artifacts are produced from sources external to the patient and can derive from an MRI hardware malfunction or miscalibration or other reasons like metal implants or foreign electrical sources. The imaging results can be spoiled in different ways, depending on the artifact source [18].

2.3 Biomedical Image Processing and Analysis

Medical imaging is one of the essential tools used in disease diagnosis. These images have different characteristics depending on the functionality, equipment, and internal variables. It is possible to highlight within the image acquisition modalities, the MRI, the functional MRI (fMRI), a variation of the structural MRI with the variable time, Positron Emission Tomography (PET), Diffusion Tensor Imaging (DTI), Computed Tomography (CT), Ultrasound (US), and others.

After the acquisition and depending on the main goal, it is necessary to apply specific techniques to enhance the region of interest and make a final decision, solely by the system or a union between the system and health professionals. In this way, it is avoided human subjectivity and time-consuming tasks.

2.3.1 Main steps in biomedical image processing

In biomedical image processing, it can be highlighted the main steps present in Figure 2.8.

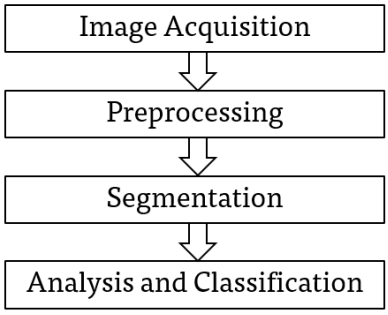


Figure 2.8: Main steps in medical imaging processing.

2.3.1.1 Acquisition

The input image is a scan result, a slice of the understudy body, through that it is possible to do a 3D reconstruction by a set of images overlapped in the z-axis.

A set of pixels forms an image in the digital field. Each pixel has its coordinates and intensity, the value present in the space x,y coordinates. A digital image can be binary, with values 0 or 1, monochromatic or chromatic. In a grayscale image, the most used in the medical field, the pixel value varies depending on the chosen scale, from 2 to 256 different intensities. The binary image, two intensities, has less information, but the memory requirement is also less. On the other side, for a 256 values image, exist more details with good differentiation between tissues. However, it requires more memory space and the processing time is also bigger. When it comes to an RGB image, a matrix with three dimensions where each channel corresponds to each color, the processing requirements are even more.

In the overlapping process to obtain the 3D object, the closer the scan slices are, i.e., the minimum distance, the more accurate will be the spatial reconstruction. Placing pixels and slices

together obtains a three-dimensional partition of the image in space. In volume elements, the unit is a voxel, forming a 3D scalar field [26].

2.3.1.2 Preprocessing

For easier image processing regardless of the goal, it is usually applied preprocessing techniques that aim to enhance the interest points or reduce the information to process, excluding irrelevant data. The original image is manipulated by a modification in pixel values, resulting in a more suitable image for the intended application. The techniques are problem-oriented, that is, not all go through the same process.

The preprocessing can be divided into two categories, image enhancement and image restoration. The first one is used to increase the quality or give more relevance to a particular structure. For a given problem, image restoration applied mathematical models of image degradation to perform the inverse process, recovering the image [27].

2.3.1.3 Segmentation

The image segmentation is implemented in nearly all biomedical image processing applications to segregate different regions or structures. It is considered the most sensitive step in the process [28].

To create masks, for example, to use as ground-truth, in segmentation step, is usually designed the set of pixels with useful information as an object, and the remaining pixels, as the background where the pixels are not profitable [27].

The segmentation can be classified into two general categories. The first category is based on the discontinuity of the points across two regions, while the second group takes advantage of the similarities among the points in the same region. In the first group stands out approaches founded on gray-level discontinuities (detection of point, lines, or edges) or thresholding (selection by gray-level) [28].

A robust segmentation can define the process's success when required objects to perform an analysis or classification. An image with less noise and good contrast between object and background will result in a more accessible and accurate segmentation [27].

Therefore, image segmentation is not always a straightforward task. The applied methodology will depend on the input, image characteristics, and final goal. This area has many different approaches presented in the literature and is never a static area with regularly new approaches.

2.3.1.4 Analysis and Classification

The last step, also called recognition in some applications, is based on a label's assignment to an object. For that, image features are extracted to be further submitted to a classifier.

These classifiers can follow supervised or unsupervised learning. On one side, in a supervised procedure, the classification labels are given during the training process. This method is widely used in the biomedical image when using a health professional classification to target the system.

On the other side, there are no labels in an unsupervised procedure, unknown objects are submitted in the classifier, and the learning process to perform the classification is based on similar patterns that group them into clusters [28].

2.3.2 Preprocessing techniques used in AD diagnosis problem

As mentioned before, in Section 2.3.1.2, the study of the preprocessing techniques requires previous knowledge to know which techniques are chosen for the AD diagnosis problem.

In the literature, it was found the main applied preprocessing steps before the AD diagnosis.

- **Intensity Normalization**: This process establishes all pixels or voxels into a reference scale. Consequently, similar structures will have similar intensities. The correction most applied in the literature is the N3, a non-parametric method for automatic correction of intensity non-uniformity in MRI data [29];
- **Registration**: This technique is widely applied to register the same objects in all the images, from different imaging modalities, or even in the same machine when the images are taken at different times. The process is based on image scans' spatial alignment to a specific reference anatomical space, being the most used MNI space. In this way, it is possible to compare the same pixel in a particular structure through different images. For instance, it often occurs that imaged objects are tilted, being necessary to compensate for these distortions. Thus, registration is one of the most used routines in the preprocessing step, primarily in multimodal approaches to align all imaging modalities of the same patient's images [28];
- **Skull Stripping**: It is defined by the remotion of the skull bone from images. It can be implemented individually or mutually with cerebellum or neck removal [30]. Consequently, there is less information to process, increasing the probability of system success;
- **Tissue Segmentation**: In approaches using only the gray matter (GM), it requires a previous segmentation due to the GM being affected by neurodegeneration in the initial stage [30]. Some systems also use GM + WM as input through 2D slices, or previously extracted volumes, or even with features from those particular structures;
- **AC-PC alignment**: Anterior Commissure- Posterior Commissure line is a neuroimaging standard also used to align the image, adjusting the two major anatomical brain landmarks (AC and PC) in the same axial plane [30].

2.3.3 Classification techniques used in AD problem

In the detection of AD, the diagnosis is based on observing cerebral magnetic resonance by health professionals. This process, besides being time-consuming, also results from a subjective decision. Artificial intelligence comes to this field to play a significant role in assisting professionals in their interpretations of medical imaging. Therefore the Computer-Aided Diagnosis (CAD) systems have exponentially increased [31].

Since applying the image enhancement techniques to the automatic decision with Deep Learning methods, engineering science has revolutionized the health field. This automatic classification falls into the next chapter, which explores Machine Learning and Deep Learning algorithms.

2.4 Summary

This chapter covers the project's neuro disease focus to understand the base principles and the characteristics that revealed Alzheimer's disease in image scans or medical analysis. Being the leading cause of dementia, and the number of patients continuing to increase exponentially, its study and early diagnosis are imperative. The symptoms can progress or stabilize from the prior stage of the disease to a severe stage, being crucial a continuous monitorization and complementary exams to help clinicians produce an accurate detection resorting to computational techniques.

From these imaging exams is highlighted the MRI, which is an imaging technique widely used when required a structural image, being the initial exam performed to detect Alzheimer's disease. This acquisition system presents a relatively higher resolution than other imaging modalities, beyond the fact that it can be performed in any arbitrary slice orientations. Various sequences can be chosen to this end, although the T1-weighted images are considered the most beneficial due to manifest better results and allow easy detection of cerebral atrophy, Alzheimer's disease feature. The scan's image outcome can be affected by several factors, from the patient's involuntary movements to hardware problems. Therefore, it is crucial to evaluate all the parameters before the scan execution and applied preprocessing techniques to enhance the image.

In the end, it is focused the biomedical image processing from the acquisition until its analysis. The digital image is formed by multiple individual units that have coordinates and a value. This set undergoes preprocessing steps, segmentation, analysis, and classification. To the last step can be required a Computer-aided system to facilitate the clinicians' work, falling in the artificial intelligence area. The most used techniques in Alzheimer's disease problem are also presented, like skull stripping, alignment, and intensity normalization.

Chapter 3

Machine Learning and Deep Learning

Machine learning (ML) is one of the principal subsets of Artificial Intelligence (AI). Nowadays, it is widely used in CAD systems studies to develop diagnostic tools to support the medical staff in their decisions. Performing a search in Scopus, the success of CAD systems application in AD diagnosis is proven due to the exponential increase of publications related to AD, neuroimaging, or medical imaging and machine learning views until 2020, reaching over 200 papers in the last year (Figure 3.1).

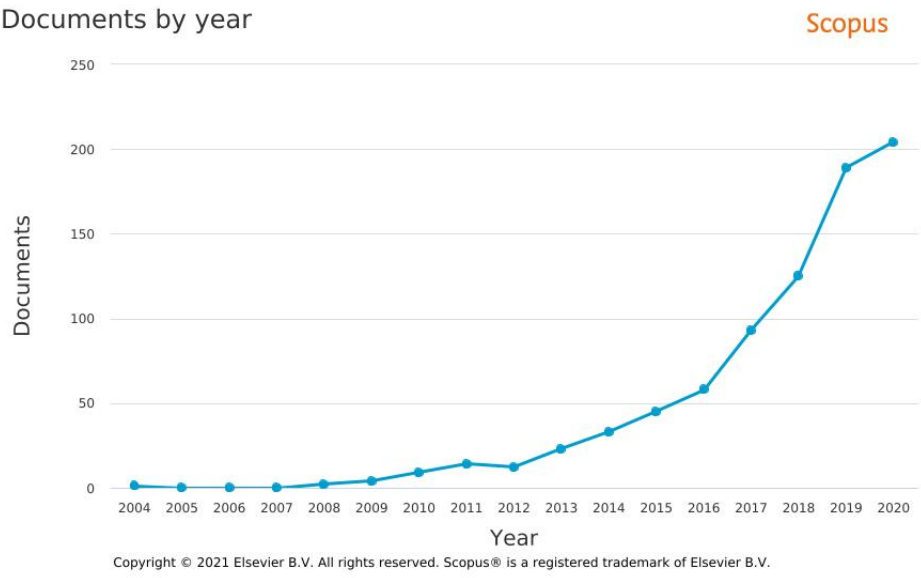


Figure 3.1: Machine Learning publications related to AD diagnosis based on imaging in Scopus Database.

This chapter presents an overview of the Machine Learning approach pipeline and a description of the fundamental steps. Afterward, a summary of the main algorithms applied nowadays in the literature is presented. It is highlighted the characteristics of Deep Learning and some popular techniques applied in these systems, mainly in image analysis and classification. In the end, it is described the standard measures used to assess AD diagnostic approaches.

3.1 Machine Learning

Machine learning techniques using neuroimaging data are popular within studies for automated segmentation or/and classification where stands out the MRI data [32]. The classification process is based on the identification of a specific brain disorder in a patient.

From an ML standpoint, the prevailing framework in the analysis of neuroimaging data for diagnosis passes by a long process, Figure 3.2, although can be summarized in 3 main steps: preprocessing, feature extraction/selection, and classifier learning [33].

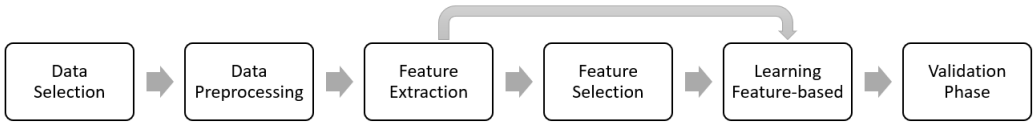


Figure 3.2: The representative process of an ML approach.

The preprocessing stage aims to minimize the noise or standardize the image data to reduce possible error sources in the following steps or even reduce the digital data to process. It follows the essential process regarding feature extraction in which the outcome is inserted into ML models such as Support Vector Machine and Logistic Regression for further analysis/classification. The intermediary step is the most crucial aspect in conventional ML approaches conditioning the learning algorithm's performance. Therefore, it required a lengthy study of possible features and several tests in different methods before finding the most suitable set for the task [34]. Herein, human experts play a critical role in these complex structures beyond the need for a ground truth to verify the performance [32].

On the one hand, the use of ML conventional classifiers requires a previous feature extraction and selection (in some cases). On the other hand, Deep Learning (DL) approaches can perform from feature extraction to classification.

Within the ML domain, some algorithms that are frequently used in the medical field to perform the classification are presented in Figure 3.3.

3.1.1 Single Layer Learning

In ML classifier algorithms stand out SVM, Linear Regression, and K-Means clustering, which are commonly used in AD classification.

3.1.1.1 SVM

The Support Vector Machine (SVM) is a supervised model, very stable, and widely used for pattern recognition, classification, and regression analysis [37]. This algorithm uses the maximum margin principle to classify the data points of the input. In non-linearly divisible data, it is implemented kernel functions to convert the dimensions [38]. Vapnik's method is based on training, testing, and performance evaluation, being more discriminative as it focuses on the points that are challenging

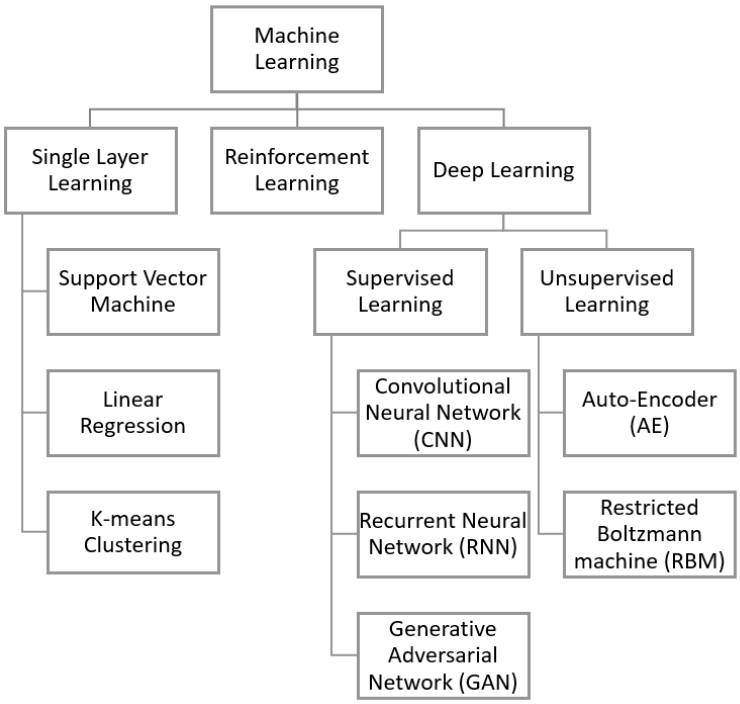


Figure 3.3: Breakdown of algorithm types in the machine learning classification, (adopted from [35] and [36]), used in AD problems.

to analyze [39]. The SVM provides adequate results and good generalization performance over the test data compared to other DL approaches in small datasets [37].

3.1.1.2 Logistic Regression

Logistic Regression algorithm, against the categorical classification labels, provides a quantitive probability outcome in order to classify [40]. It is a mathematical model approach based on the logistic function that varies between 0 and 1, so as the output [41].

3.1.1.3 K-Means

The K-Means is a clustering algorithm where the classification is performed by creating k groups departing from a sample set, where k is a number defined by the user. Initially, k centroids are initialized (usually randomly), and the data points are allocated to the nearest centroid. These centroids are recalculated depending on the belonging data points. The algorithm achieves convergence when the centroids do not move [42].

3.2 Reinforcement Learning

Reinforcement learning (RL) is simultaneously the problem, the solution, and the field that studies them. This machine learning paradigm involves learning to achieve a maximum numerical reward

signal instead of trying to find hidden patterns or abstract features. They are closed-loop problems, where the system performed actions influencing its later inputs in order to identify which actions generate the most considerable reward by trying them out. In this process, one of the challenges arises, the trade-off between exploration and exploitation. The system has to exploit what it already recognizes to reach a reward but also has to explore in order to execute more beneficial action choices in the future. The central dilemma is that neither exploration nor exploitation can be tracked without sometimes failing in the task. The agent must attempt a diversity of actions and progressively promote those that seem to be most profitable [43].

This form of learning also demonstrates applications in the medical field in personalized medicine with dynamic treatment regimes or personalizing drug dosage or chemo/radiotherapy for cancer treatments. The clinical professionals aim to achieve the same that RL, attempting a sequence of treatments, they want to reach the best conceivable outcome [44].

3.3 Deep Learning

Thus, the major drawback of the conventional ML algorithms is the need for handcrafted features. To overcome this problem arose Deep Learning (DL) techniques. DL is a recent branch of ML, inspired by the human brain. This method has been developed based on complex algorithms that trace high-level features and obtain those abstractions from data using neural network architecture similar to brain neuron connections[45].

Most DL applications use feedforward neural network architectures that learn to map a fixed-size input to a fixed-size output. When these outputs labels are available, the learning process is called supervised learning, where internally is done recursive adjustments of parameters to achieve the intended objective. Between these applications stand out image segmentation, speech recognition, document reading, and classification [46].

In this field, methods implementing unsupervised learning are essential when faced with data scarcity and high dimensionality. Furthermore, it is called unsupervised due to the absence of output labels. The primary goal is to reduce dimensionality and information retrieval.

Recently, DL implementations in automated detection of brain disorders have shown early success, outperforming statistical methods and predicting diseases like AD better than clinicians [47].

3.3.1 Deep Learning Architectures

This topic explores the architectural layouts of deep learning used in the AD classification or prognostic prediction presented in Figure 3.4.

3.3.1.1 AE

A DL architecture used in AD/MCI problems is the Auto-Encoders, which is an unsupervised learning method defined by three layers: the input layer, hidden layer, and output layer [49]. This

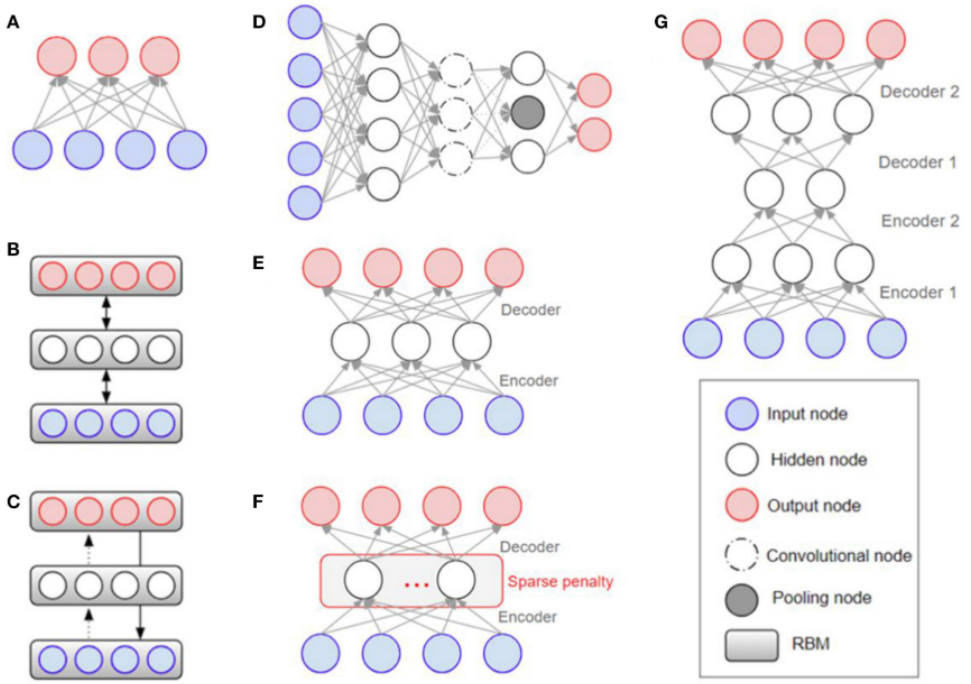


Figure 3.4: Architectures structures (from [48]) where: (A) RBM; (B) DBM; (C) DBN; (D) CNN; (E) AE; (F) Sparse AE; (G) Stacked AE.

feedforward network involves two components, the encoder, to learn to generate a latent representation of the input data and a second component, the decoder, where the learned representations are used to restore the input data as nearby as possible to the original. AE uses deterministic functions to perform the automatic feature extraction [50].

3.3.1.2 SAE

A Stacked Auto-Encoder is a configuration where multiple AEs are stacked to intensify the representational power by using as AE's input the values of hidden units of the previous one. Higher-level representations are learned due to the increased model depth improving the ability to discover intricate patterns [30].

3.3.1.3 Sparse AE

A Sparse Auto-Encoder is based on the simplest AE where the most hidden units are enforced to be close to zero. This network can be used to learn valuable filters for convolution operations being advantageous to control the variability of MRI images [51].

3.3.1.4 RBM

A different unsupervised method, similar to learning and extracting features from the data like AE, is the Restricted Boltzmann Machine (RBM), composed of only two layers (a visible and a hidden). The difference that separates these two methods is the training process. RBM is usually

trained using the maximum-likelihood estimation to find a stochastic representation of the input in its hidden layer (latent features) [50].

3.3.1.5 DBM

A Deep Boltzmann Machine (DBM) is an undirected graphical model based on a hierarchically multiple RBMs stack. The DBM captures deeply non-linear and intricate patterns or statistics such as the relations between input values due to the hierarchical nature [52]. The pre-training in this network can be done separately in each layer to accomplish better weight matrix initialization [34].

3.3.1.6 DBN

Deep Belief Networks (DBNs) are similar to Stacked AEs comprising stacked RBMs, which are trained using the greedy layer-wise algorithm allowing one RBM layer training at a time. This approach starts the training process with the input data inserted into the first layer and pursues training higher-level RBMs handling the hidden layer's current state of the previous level. With this method, the network acquired different feature levels. Low-level features are in the first layers and the higher abstraction features are in higher network levels [53].

3.3.1.7 CNN

Within DL models, the Convolutional Neural Network (CNN) is highlighted due to a clear train's capability and a better generalization than networks with full connectivity between adjacent layers. It is a type of feedforward network popularly embraced by the computer vision community.

CNN's are constructed to handle data that comes in the form of multiple arrays, 1D to signals or arrays of values, 2D in greyscale pictures or a set of 3 2D matrices to RGB data, and 3D for volumetric images or video [46].

Due to the relevance of this algorithm in the development of the present work, CNN will be explored in a forward chapter.

3.3.1.8 RNN

For temporal dynamics of longitudinal cognitive measures, like the disease evolution, Recurrent Neural Networks (RNN) are used to learn informative representation for each data analysis time point. Besides that, it is applied in tasks that involve sequential inputs such as speech and language.

RNN's process one element once a time, saving in their hidden units a "state vector" that implicitly holds information concerning all the past sequence elements' history. Considering that, the RNN reveals to be remarkable in predicting characters in words, words in sentences, or even in more complex tasks. Thereby analyzing the hidden units at different discrete time steps, their output will be different in the same component.

Noting that the primary purpose is to get long-term dependencies, it has been shown the difficulty of learning to store information for a long time. To counteract, it was created the Long Short-Term Memory (LSTM) networks that apply peculiar units to increase memory over time [46].

3.3.1.9 GAN

Generative Adversarial Networks (GAN) is an algorithm that generates new data statistically mimicking the input data through a latent distribution. This method has two "adversarial" networks. The first one is the generator, which produces the synthetic data, while the second is a discriminator that aims to distinguish the real from synthetic instances [35].

3.3.1.10 DNN

Also exists an architecture based on the fundamental components of DL, the Multi-Layer Perceptron, called Deep Neural Networks (DNN), a neural network with more than two hidden layers [54].

3.3.2 Currently Popular Techniques in DL Systems

In the more recent studies employing DL approaches in the AD problem, some strategies to improve the algorithm have been explored and deserve to be highlighted, like data augmentation and transfer learning.

3.3.2.1 Data augmentation

In the computer vision field, the data augmentation technique is commonly applied, which has also been explored in the neuroimaging context over the years. This technique is based on increasing the number of samples by applying transformations to the input data, such as mirror, flipping, scaling, small magnitude translations, or even soft Gaussian blur to the existent dataset images [50]. An example of data augmentation applications is presented in Figure 3.5.

It is used to add data to the neuroimaging dataset in DL approaches in order to reduce overfitting due to fewer images available. Then the original and the artificial data are injected into the system in the learning process [55].

Another way to perform data augmentation is by applying a Generative Adversarial Network, which produces synthetic medical images from a raw image. An example of GAN's application is the reproduction of PET images from MRI data of the same subject, having as basis other subjects with the complete data [48].

3.3.2.2 Transfer Learning

Regarding the limited data for AD or other disorder problem classification, transfer learning is implemented as a possible solution similar to the previous method mentioned [57]. This technique

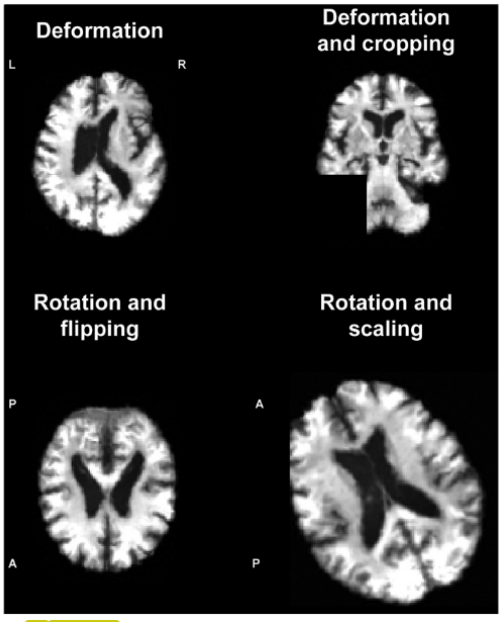


Figure 3.5: Example of data augmentation technique applied to neuroimaging (from [56]).

uses previously learned features from another sample with a different target domain or a related domain [50].

Thereby, there are multiple ways to apply transfer learning. A typical application implements a model pre-trained in a large dataset as ImageNet, which was developed to use in visual object recognition with millions of images. The system learned how to distinguish images and then is trained in the focus problem. The network that learns the simples features to distinguish images will merely need fine-tuning to recognize the different neuroimaging stages. Another way to employ transfer learning is to train the network in a task and later carry the weights to other tasks within the same area of expertise.

It is suggested that transfer learning could be beneficial when deeper models are employed, allowing accurate model design in a time-saving way [50].

3.4 Validation metrics to ML classifiers

To validate or quantify the machine learning algorithms' performance is used different evaluation metrics on classifiers. Depending on the goal of the medical approach, these metrics can evaluate tissue segmentation or class discrimination.

Therefore, it is implemented, commonly, two evaluation approaches:

- Subjective evaluation - the result is obtained by experts who provide their opinion on the quality of the images. The result is usually employed as ground truth to segmentation or classification methods;
- Objective evaluation - the outcome is based on mathematical algorithms.

The evaluation conducted by professionals is time-consuming and complex, hence the application of artificial intelligence in this field.

The classification quality measures are built from a confusion matrix that records the correct and incorrect recognized examples for each class, a binary class example is given in Table 3.1.

Table 3.1: A confusion matrix for binary classification.

Class \ Prediction	Positive	Negative
Positive	TP	FN
Negative	FP	TN

Where,

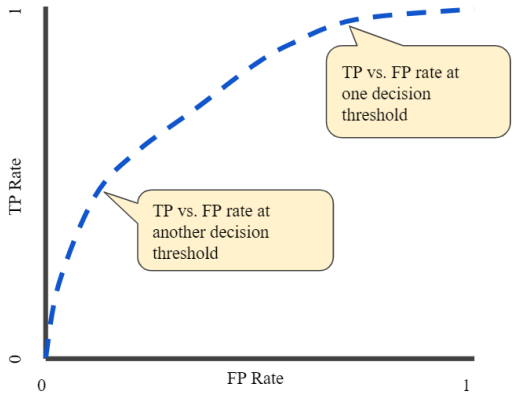
- TP - true positive, when the prediction is positive, and the actual value is positive;
- FP – false positive, when the prediction is positive, and the actual value is negative;
- FN – false negative, when the prediction is negative, and the actual value is positive;
- TN – true negative, when the prediction is negative, and the actual value is positive.

Based on the confusion matrix, it is computed the Accuracy, Precision, Recall, False Positive Rate (FPR), Balanced Accuracy, and F-Score or F1-Score ($\beta = 1$), presented in Table 3.2, and ROC Analysis, commonly used in the biomedical image field.

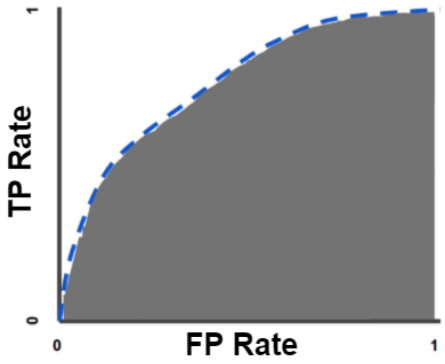
Table 3.2: Measures commonly used to assess AD diagnostic approaches.

Metric	**Formula**
Accuracy	$\frac{TP+TN}{TP+FP+FN+TN}$
Sensitivity or Recall	$\frac{TP}{TP+FN}$
Specificity	$\frac{TN}{FP+TN}$
Precision	$\frac{TP}{TP+FP}$
False Positive Rate (FPR)	$\frac{FP}{TN+FP}$
Balanced Accuracy	$\frac{Sensitivity+Specificity}{2}$
F-measure	$\frac{(\beta^2+1)*Precision*Recall}{\beta^2*Precision+Recall}$

Receiver Operating Characteristic (ROC) curve, Figure 3.6a, is a graph that plots the system's classification performance as the discrimination threshold varies. The parameters, Recall and FPR, define the curve. Area Under the Curve (AUC) is the measurement of the entire area under the ROC curve, Figure 3.6b, representing the separability degree [58].



(a) Receiver Operating Characteristic (ROC) curve (from [59]).



(b) Area Under Curve (AUC) (from [59]).

Figure 3.6: ROC curve and AUC.

3.5 Summary

This chapter proffers a theoretical base of machine learning algorithms, including all the fundamental steps to achieve in the practical implementation. The crucial action to accomplish is the feature extraction, being the basis of the classification influencing the algorithm performance.

Deep Learning and its most used architectures in the medical field are introduced, followed by popular techniques applied nowadays to eliminate the lack of data problems. To evaluate the classifier performance is adopted metrics to comparison purposes.

Chapter 4

Deep Learning in AD diagnosis - State-of-the-Art

This chapter focus on the current approaches for Mild Cognitive Impairment and Alzheimer's Disease diagnosis using Deep Learning architectures to perform these tasks. The input data provided to the assessment is from solely MRI or multimodal imaging data combining MRI with other techniques.

4.1 Method

In the Scopus database, a systematic literature search was implemented, Figure 4.1, with the following keywords: "Deep Learning" AND Alzheimer AND (Image OR Imaging) AND (MRI OR "Magnetic Resonance") AND (MCI OR "Mild Cognitive Impairment").

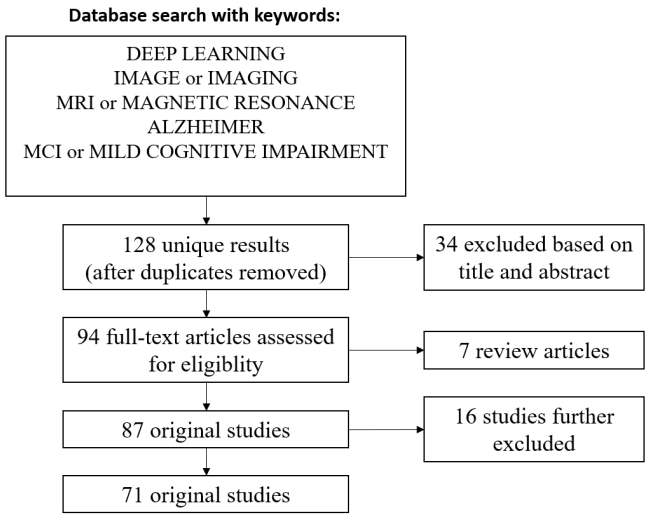


Figure 4.1: PRISMA diagram showing the results of the performed literature search.

This search provided a total of 149 results. From those, 71 articles were revealed to be helpful and within the subject. The remaining were excluded for the following reasons:

- The article was not entirely available;
- MRI was not one of the addressed imaging modalities;
- The goal was not the AD or MCI classification/prediction.

For further analysis, 54 articles of the previously selected used only data provided by MR neuroimaging. While the remaining 17 used multimodality data, where there is a higher predominance of PET images to support the MRI data. Beyond this division concerning the type of data used to construct the classification algorithm, an additional separation was performed, highlighting the articles that forecast the time elapsed until the conversion for a different AD stage.

4.1.1 MRI Approaches

With the purpose to perform classification or distinction between the different stages of the disease employing only the MRI technique, 52 unique approaches were found, Table 4.1.

In the field of imaging processing, CNN attracts most of the attention due to being specially designed to recognize patterns. Architectures already developed or created from scratch are presented and as input can be utilized slices from the whole brain or a segmented region. Although the ADNI dataset is almost a universal choice in the training and testing steps, even in similar architectures, the Authors present different results. It is noteworthy that softmax activation is the most usual activation function to proceed with the final classification.

In the reviewed articles, models based on 2D CNN appear in greater numbers. Within the slice approach, some Authors chose to perform image segmentation before inputted them into the CNN, for instance, to use only GM data, while others preferred an entire brain slice.

A known architecture trained from scratch, GoogleNet, obtained a high value of accuracy in NC vs. pMCI vs. sMCI vs. AD using only GM slices and applying data augmentation [60]. ResNet is conjointly evaluated but achieved a lower performance in this article. Following the same idea with data augmentation, in [61], three 2D CNN were applied separately to axial, sagittal and coronal views. They conclude that the sagittal view reached better performance in the binary classifications. In contrast, the Authors of [62], resorting to data augmentation with the whole slice matter, use coronal slices in an end-to-end 3-way CNN classification, considering previous tests where the axial view resulted in an inferior performance. Going deeper, [63] used two different datasets, MIRIAD and OASIS, and implemented three models to perform binary classification. These models were trained with OASIS and tested with MIRIAD datasets previously submitted to data augmentation. It was evaluated the contribution of the different brain slice planes, concluding that in AD vs. CN, a better result was achieved when the network was trained with all views and testing with axial images. For MCI vs. CN, the best results were obtained with training and testing with axial images. Following the last study, most of the found articles use the axial view.

Table 4.1: Summary of MRI modality approaches found in the state of the art.

A.	Year	Data Source + Size	Appr. type	D.A.[1]	Feature extraction + Feature Selection + Classification	T.L.[2] [4]	Accuracy	Sensitivity	Specificity	Other metrics
[60]	2017	ADNI: AD:33; NC:45; pMCI:22; sMCI:49	slices (GM)	YES	GoogleNet/ ResNet18/ ResNet152	NO	GoogleNet: 3way: 98.88%; ResNet-18: 3way: 98.01%; ResNet-152: 3way: 98.14%			
[61]	2019	ADNI: CN:200; sMCI: 200; pMCI:200	ROI	YES	CNN	NO	CN/pMCI:94.54%; CN/sMCI:93.96%; sMCI/pMCI:93%	CN/pMCI:91.7%; CN/sMCI:90.46%; sMCI/pMCI:91.48%	CN/pMCI:97.96%; CN/sMCI:98.19%; sMCI/pMCI:94.82%	**AUC**: CN/pMCI: 0.994; CN/sMCI: 0.988; sMCI/pMCI:0.981
[62]	2021	ADNI: AD:110; NC:51; MCI:105	slices	YES	DCNN	NO	3way: 98.57%(train); 87.72%(validation)			
[63]	2020	OASIS: AD:30; NC:316; MCI:70; MIRIAD: AD:46; NC:23	slices	YES	CNN	NO	AD/CN: 83%, MCI/CN: 82%	MCI/CN:93%; AD/CN:72%	MCI/CN: 81%; AD/CN:94%	**AUC**: MCI/CN:0.6; AD/CN:0.85
[64]	2020	ADNI: CN:159; AD:153; MCI:157	slices	NO	ACNN / ResNet	NO	ResNet: MCI+AD/NC:85.07% ACNN: NCI+AD/NC:80.17%			
[65]	2019	ADNI:[3] AD:211; NC:91; MCI:774	slices (GM)	NO	MCADNNet (CNN)	NO	3way:100%; AD/NC: 99.09%; AD/MCI:99.7%; NC/MCI:100%			
[66]	2020	ADNI: AD:137; NC:162; pMCI:76; sMCI:134	slices	YES	CNN (ensemble learning)	NO	AD/NC: 84%, pMCI/NC: 79%; pMCI/sMCI: 62%			**AUC**: AD/NC: 0.92; pMCI/NC: 0.83; pMCI/sMCI:0.59
[67]	2020	ADNI: AD:243; NC:307; MCI:525	slices	NO	PCANet (CNN) + K-Means clustering	NO	(one slice/TOP data) AD/MCI: 95.52%/97.01%; AD/NC: 76.32%/89.15%; MCI/NC: 90.63%/92.6%; 3way:89.38%/91.25%			

Table 4.1, continued

A.	Year	Data Source + Size	Appr. type	D.A.[1]	Feature extraction + Feature Selection + Classification	T.L.[2 4]	Accuracy	Sensitivity	Specificity	Other metrics
[68]	2017	ADNI: AD:188; NC:228; MCI:399	2D patches	YES	Deep CNN	NO	AD/NC:91.41%; NC/MCI: 65.62%; AD/MCI:69.53%	AD/NC:93.75%; NC/MCI: 65%; AD/MCI:71.88%	AD/NC:89.06%; NC/MCI: 66.25%; AD/MCI:67.19%	
[69]	2020	ADNI: sMCI:100; pMCI:100	2D ROI	NO	CNN	NO	pMCI/sMCI: 94%	pMCI/sMCI: 92%	pMCI/sMCI: 96%	
[70]	2017	ADNI: AD: 188; NC:228; MCI:399	2D+£ patches	YES	CNN	NO	AD/NC:82.8%; MCI/NC: 66.0%; AD/MCI: 62.5%	AD/NC:79.6%; MCI/NC: 73.7%; AD/MCI 60%	AD/NC: 85.9%; MCI/NC:58.7%; AD/MCI: 64%	
[71]	2018	ADNI: AD:188; NC:229; pMCI:169; sMCI:139; MCIun:93	2.5D -> 2D patches	YES	CNN / FreeSurfer + LASSO / PCA + ELM	NO	sMCI/pMCI:81.4%			**AUC**: sMCI/pMCI:0.878
[72]	2020	ADNI: AD:143; NC:144; MCI:203	slices (GM + WM)	NO	SKA-50	NO	AD/NC: 98.59% ; AD/MCI: 98.82%; MCI/NC: 89.53%	AD/NC: 97.22% ; AD/MCI: 98%; MCI/NC: 86.11%	AD/NC: 100% ; AD/MCI: 100%; MCI/NC: 92%	
[73]	2019	ADNI: AD:50; NC:50; MCI:50	slices	NO	VGG-19	YES ()	AD/NC: 99.36; AD/MCI: 99.2%; MCI/NC: 99.04%; 3way:95.19%	AD/NC: 98.7%; AD/MCI: 98.9%; MCI/NC: 99.5%	AD/NC: 100%; AD/MCI: 99.5%; MCI/NC: 98.6%	
[74]	2018	ADNI: AD:347; CN:537; MCI:806	slices (coronal view)	YES	CNN (based on DenseNet-121)	YES ()	AD/CN: 94.97%; AD/MCI: 91.98%; MCI/CN:74.7%	AD/CN: 94.33%; AD/MCI: 90.47%; MCI/CN:70.96%	AD/CN: 95.89%; AD/MCI: 95.38%; MCI/CN:78.2%	
[75]	2019	OASIS: AD:100; NC:316; ADNI: AD:192; NC:229; MCI:398	slices	YES	AlexNet	YES ()	OASIS: AD/NC:95.35%; ADNI(1.5T/3T): NC/AD: 98.74%/98.97%; 3way:98.06%/98.76%	OASIS: AD/NC:90.44%; ADNI(1.5T/3T): NC/AD: 96.32%/96.68% 3way: 95.77%/97.69%	OASIS: AD/NC: 90.39%; ADNI(1.5T/3T): NC/AD: 97.78%/98.53%; 3way: 95.19%/96.73%	**Precision**: OASIS: AD/NC: 96.91%; ADNI(1.5T/3T): NC/AD:99.4%/99.59%; 3way:98.69%/99.06%
[76]	2020	HNR: [3] aMCI:61; CN:59; ADNI-1: aMCI:397; NC:227	slices	NO	DeCAF + LSTM (based RRN)	YES ()	ADNI: aMCI/CN: 77% HNR: 90%	ADNI: aMCI/CN: 85% HNR: 90%	ADNI: aMCI/CN: 65% HNR: 91%	

Table 4.1, continued

A.	Year	Data Source + Size	Appr. type	D.A.[1]	Feature extraction + Feature Selection + Classification	T.L.[2] [4]	Accuracy	Sensitivity	Specificity	Other metrics
[77]	2018	ADNI: NC: 150; sMCI:150; pMCI:157	slices (RGB)	YES	GoogleNet/ CaffeNet	YES ()	GoogleNet: 3way : 83.23%; CaffeNet: 3way: 87.78%;			
[78]	2018	NACC: [3] NC:303; MCI:83	slices	NO	VGG-11 + MLP	YES ()	MCI/NC: 90.9%	MCI/NC: 96.3%		**Precision**: MCI/NC: 92.6%
[79]	2019	ADNI: AD:179; NC:182; MCI:254	slices (GM+ WM)	NO	CNN (eResNet50+ eNASNet + eMobileNet) (ensemble learning)	YES ()	AD/MCI: 97.65%; MCI/NC: 88.37%	AD/MCI: 96%; MCI/NC: 80.56%	AD/MCI: 100%; MCI/NC: 94%	**AUC**: AD/MCI: 1; MCI/NC: 0.96
[80]	2018	ADNI: AD:111; NC:154; MCI:150	slices	NO	ResNet18 / BoW + RNN	YES ()	AD/NC: 89.5%; NC/MCI: 81.7%			**BACC**: AD/NC: 88.1%; NC/MCI: 81.4%
[81]	2019	ADNI: AD:176; NC:416 MCI:304	slices	NO	Inception	YES	3way: 85.71%; AD/NC:100%; MCI/NC: 89.47%			
[82]	2020	ADNI: sMCI:70; NC:50	slices	NO	VGG16 + LASSO + SVM	YES	sMCI/NC: 89.4%	sMCI/NC: 92.9%	sMCI/NC: 89.3%	
[83]	2017	ADNI: AD:50; NC:61; pMCI:43; sMCI:77	3D ROI	NO	VoxCNN / VoxResNet	NO	(VOX/Res) pMCI/sMCI: 56%/52%; AD/NC:79%/80%; AD/sMCI:64%/63%; AD/pMCI:62%/59%; pMCI/NC:63%/61%; sMCI/NC: 54%/56%			**AUC**: (VOX/Res) pMCI/sMCI:0.47/0.52; AD/NC:0.88/0.87; AD/sMCI:0.66/0.67; AD/pMCI:0.61/0.62; pMCI/NC:0.67/0.65; sMCI/NC:0.57/0.58
[84]	2019	ADNI: AD:50; NC:62; sMCI:91; pMCI:39	3D ROI	NO	VoxCNN	NO	AD/NC: 76.8%; AD/sMCI 67.5%; AD/pMCI 66.7%; pMCI/NC 58.1%; pMCI/sMCI 56.8%; sMCI/NC 53.5%			**AUC**: AD/NC: 0.863; AD/sMCI: 0.729; AD/pMCI: 0.743; pMCI/NC: 0.611; pMCI/sMCI: 0.56; sMCI/NC: 0.535
[85]	2019	ADNI: AD:198; NC:229; pMCI:167; sMCI:236	3D ROI (GM)	YES	CNN - BGRU	NO	AD/NC: 91.33%; pMCI/sMCI:71.71%	AD/NC:86.87%; pMCI/sMCI: 65.27%	AD/NC:95.20%; pMCI/sMCI: 76.27%	

Table 4.1, continued

A.	Year	Data Source + Size	Appr. type	D.A.[1]	Feature extraction + Feature Selection + Classification	T.L.[2 4]	Accuracy	Sensitivity	Specificity	Other metrics
[33]	2017	ADNI: AD:186; NC:226; pMCI:167; sMCI:226	3D ROI (GM)	NO	DeepESRNet	greedy layer -wise	AD/NC: 91.02%; MCI/NC:73.02%; pMCI/sMCI: 74.82%	AD/NC: 92.72%; MCI/NC: 77.6%; pMCI/sMCI: 70.93%	AD/NC: 89.94%; MCI/NC: 68.22%; pMCI/sMCI: 78.82%	**BACC**: AD/NC: 91.33%; MCI/NC:72.91%; pMCI/sMCI:74.87%; **AUC**: AD/NC: 0.9272; MCI/NC: 0.7361; pMCI/sMCI:0.7539
[86]	2018	ADNI: AD:199; NC:229; MCI:403	3D patches	NO	DenseNet	NO	AD/NC: 89.5%; MCI/NC: 73.8%	AD/NC: 87.9%; MCI/NC: 86.6%	AD/NC: 90.8%; MCI/NC: 51.5%	**AUC**: AD/NC: 0.924; MCI/NC: 0.775
[87]	2020	ADNI: AD:97; NC:119; MCI: 233	3D patches	YES	V-Net (CNN) + 3D DenseNet	NO	AD/NC: 88.9%; MCI/NC:76.2%	AD/NC:86.6%; MCI/NC:79.5%	AD/NC:90.8%; MCI/NC:69.8%	**AUC**: AD/NC:0.925; MCI/NC:0.775
[88]	2019	ADNI: AD:192; NC:223; pMCI:165; sMCI:231	3D patches + voxel	NO	3D DenseNet / SPHARM-MAT + MLP	NO	AD/NC: 92.29%, pMCI/sMCI: 75%, MCI/NC: 74.64%	AD/NC: 90.63%; pMCI/sMCI:73.33%; MCI/NC: 77.27%	AD/NC: 93.72%; pMCI/sMCI:76.19%; MCI/NC: 69.96%	**AUC**: AD/NC: 0.9695; pMCI/sMCI: 0.797; MCI/NC: 0.777
[56]	2019	ADNI: AD:294; NC:352; MCI:763; MILAN: AD:124; NC:55; MCI:50	3D ROI	YES	CNN	YES (⊙)	ADNI: AD/NC:99.2%; pMCI/NC:87.1; sMCI/NC:76.1%; AD/pMCI:75.4%; AD/sMCI:85.9%; pMCI/sMCI:75.1%; ADNI + MILAN: AD/NC:98.2%; pMCI/NC:87.7; sMCI/NC:76.4; AD/pMCI:75.8%; AD/sMCI:86.3%; pMCI/sMCI:74.9%	ADNI: AD/NC:98.9%; pMCI/NC:87.8; sMCI/NC:75.1; AD/pMCI:74.5%; AD/sMCI:83.6%; pMCI/sMCI:74.8%; ADNI + MILAN: AD/NC:98.1%; pMCI/NC:87.3%; sMCI/NC:75.1%; AD/pMCI:74.8%; AD/sMCI:84%; pMCI/sMCI:75.8%	ADNI: AD/NC:99.5%; pMCI/NC:86.5%; sMCI/NC:77.1%; AD/pMCI:76.4%; AD/sMCI:88.3%; pMCI/sMCI:75.3%; ADNI + MILAN: AD/NC:98.3%; pMCI/NC:88.1%; sMCI/NC:77.8%; AD/pMCI:77.1%; AD/sMCI:88.7%; pMCI/sMCI:74.1%	
[89]	2018	ADNI:[3] AD:200; NC:230; MCI:411	3D ROI	YES	3D-CNN	YES (⊙)	AD/NC:94.1%; 3way: 61.1%	AD/NC:94%; 3way: 63%		**Precision**: AD/NC:92%; AD/NC/MCI: 59%
[90]	2019	ADNI: AD:150; NC:112; MCI:129	3D ROI	YES	3D sVoxCNN	YES (⊙)	3way:95.05%; AD/NC:99.47%; MCI/NC:98.32%; MCI/AD:93.01%	AD/NC:99.32%; MCI/NC:97.7%; MCI/AD:95.03%	AD/NC:99.59%; MCI/NC:98.72%; MCI/AD:90.6%	**AUC**: AD/NC:0.9945; MCI/NC:0.9821; MCI/AD:0.9281

Table 4.1, continued

A.	Year	Data Source + Size	Appr. type	D.A.[1]	Feature extraction + Feature Selection + Classification	T.L.[2] [4]	Accuracy	Sensitivity	Specificity	Other metrics
[91]	2019	ADNI: AD:198; NC:230; pMCI:166; sMCI:101	3D ROI	YES	ICAE	YES (⊙)	pMCI/sMCI:73.95%; AD/NC: 86.6%; pMCI/NC: 77.37%; sMCI/NC:63.04%; pMCI/AD:60.97%; sMCI/AD:75.06%	pMCI/sMCI:77.46%; AD/NC: 88.55%; pMCI/NC: 81.03; sMCI/NC:59.02%; pMCI/AD:64.53%; sMCI/AD:76.55%	pMCI/sMCI :70.71%; AD/NC: 84.54%; pMCI/NC: 74.07%; sMCI/NC:57.11%; pMCI/AD:56.13%; sMCI/AD:73.39%	
[92]	2020	ADNI+ IXI: [3] pMCI:168; sMCI:129	3D ROI	NO	ADNET	YES (⊙)	pMCI/sMCI: 76%	pMCI/sMCI: 77%	pMCI/sMCI: 76%	**AUC**: pMCI/sMCI: 0.81
[93]	2020	CADD, ADNI, AIBL, MIRIAD, OASIS	3D ROI	NO	ADNet + logistic regression	YES	AD/MCI/NC:52.3%			
[94]	2019	ADNI-1: AD:355; CN:242; MCI:415; ADNI-2: AD:261; CN:300; sMCI:314; pMCI:208; ASIAN: CN:176; AD:43; MCI:128		NO	Graph-CNN	YES (⊙)	ADNI-2: AD/CN: 85.8%, CN/pMCI:69.3%; CN/sMCI:51.8%; sMCI/pMCI: 60.9%; sMCI/AD:79.2%; pMCI/AD: 65.2%; ADNI-1: CN/AD:81%; CN/MCI: 67.6%; MCI/AD:65.4%	ADNI-2: AD/CN: 83.5%; CN/pMCI:65.6%; CN/sMCI:55.3%; sMCI/pMCI: 52.5%; sMCI/AD: 70.4%; pMCI/AD: 62.6%; ADNI-1: CN/AD:85.5%; CN/MCI: 71.3%; MCI/AD:77.5%	ADNI-2: AD/CN: 87.5%; CN/pMCI:72%; CN/sMCI:48.6%; sMCI/pMCI: 67.8%; sMCI/AD: 85.8%; pMCI/AD: 68%; ADNI-1: CN/AD:74.5%; CN/MCI: 60.7%; MCI/AD:54.6%	
[95]	2019	ADNI: AD: 198; NC:229; sMCI:214; pMCI:160	3D ROI	NO	rDNNS + SVM/-	YES	AD/CN:100%; AD/MCI: 87.32%; MCI/CN:100%; pMCI/sMCI:95.24%; 3way: 77%	AD/CN:100%; AD/MCI: 66.67%; MCI/CN:100%; pMCI/sMCI:89.29%	AD/CN:100%; AD/MCI: 97.87%; MCI/CN:100%; pMCI/sMCI:100%	**AUC**: AD/CN:1; AD/MCI: 0.9202; MCI/CN:1; pMCI/sMCI:0.9949;
[96]	2019	ADNI: AD:221; NC:315; MCI:297	3D	NO	3D-DenseNet	NO	3way:97.52%			
[97]	2020	ADNI: AD:120; CN:120; sMCI:120; pMCI:120	3D ROI	NO	3D-DenseNet-121	NO	4way: 83.33%			

Table 4.1, continued

A.	Year	Data Source + Size	Appr. type	D.A.[1]	Feature extraction + Feature Selection + Classification	T.L.[2 4]	Accuracy	Sensitivity	Specificity	Other metrics
[98]	2019	ADNI: AD:221; NC:315; MCI:297	3D ROI	NO	3D-DenseNet (ensemble learning)	NO	3way: 97.19%			
[99]	2020	ADNI-1: AD:199; CN:229; pMCI:167; sMCI:226; ADNI-2: AD:159; CN:200; sMCI:239; pMCI:38; MIRIAD: CN:23; AD:46	3D voxel + patches	NO	LDMIL	NO	ADNI-2: AD/CN: 91.1%; pMCI/sMCI: 76.9%; MIRIAD: AD/CN: 92.8%	ADNI-2: AD/CN: 88.1%; pMCI/sMCI: 42.1%; MIRIAD: AD/CN:93.5%	ADNI-2: AD/CN: 93.5%; pMCI/sMCI: 82.4%; MIRIAD:91.3%	**AUC**: ADNI-2: AD/CN: 0.959; pMCI/sMCI: 0.776; MIRIAD:0.972
[100]	2020	ADNI: AD:175; NC:214; MCI:260	slices / 3D ROI (GM)	NO	DRN(2D) / DAIN(3D)	NO	MCI/NC:87.5%; MCI/AD: 83.18%	MCI/NC:89.62%; MCI/AD: 91.92%	MCI/NC:84.99%; MCI/AD: 70.56%	**AUC**: MCI/NC:0.8731; MCI/AD: 0.8124
[101]	2018	ADNI: AD:150; NC:112; MCI:129	slices / 3D ROI	NO	3D-FCNN / 2D-CNN	NO	(3D)AD/NC:96.81%; AD/MCI:88.43%; MCI/NC: 92.62%; 3way:91.32%; (2D)AD/NC:95.91%; AD/MCI:86.84%; MCI/NC:90.11%; 3way:89.76%	(3D) AD/NC:96.3%; AD/MCI:88%; MCI/NC:92.2%; (2D) AD/NC:95.1%; AD/MCI:85.7%; MCI/NC:88.9%	(3D) AD/NC:97.2%; AD/MCI:88.9%; MCI/NC:93.1%; (2D) AD/NC:95.6%; AD/MCI:87.6%; MCI/NC:91.7%	
[102]	2020	ADNI:[3] NC:228; AD:187; MCI:396	slices / 3D ROI	NO	2D/3DCNN	YES (⊙)	3way: 71.2% (3DPET); 64.817%(3DMRI); 60.536%(2DPET); 63.25% (2DMRI)			
[103]	2020	ADNI:[3] MCI:396; AD:187	slices / 3D ROI	NO	2D/3DCNN	YES (⊙)	MCI/AD: 71.728%(3DPET); 67.581%(3DMRI); 56.901%(2DPET); 65.708% (2DMRI)			
[104]	2020	ADNI: NC:159; AD:153; MCI:157	3D ROI (GM)	NO	3D-CNN + SVM	NO	3way: 95.74%; NC/MCI:98.90%; NC/AD:99.1%; MCI/AD: 89.4%	NC/MCI:98.90%; NC/AD:99.8%; MCI/AD:86.7%	NC/MCI:98.80%; NC/AD:98.4%; MCI/AD:84%	

Table 4.1, continued

A.	Year	Data Source + Size	Appr. type	D.A.[1]	Feature extraction + Feature Selection + Classification	T.L.[2] [4]	Accuracy	Sensitivity	Specificity	Other metrics
[105]	2019	ADNI: [3] AD:192; NC:184; pMCI:181; sMCI:228	3D + Jacobian ROI	NO	CNN	NO	AD/NC: ≈ 100%; sMCI/pMCI: 86%	AD/NC: ≈ 100%; sMCI/pMCI: 87.5%	AD/NC: ≈ 100%; sMCI/pMCI: 85%	**AUC**: sMCI/pMCI: 0.925
[106]	2018	ADNI: [3] sMCI:532; pMCI:327	3D ROI	NO	ResNet3D + XGBoost	NO	sMCI/pMCI: 76%	sMCI/pMCI: 88%	sMCI/pMCI: 70%	**AUC**: sMCI/pMCI: 0.86
[107]	2020	ADNI: 1116; in-house: 716	3D ROI (GM)	NO	3DAN	NO	ADNI: AD/CN: 92.1%; pMCI/sMCI: 71.7%	ADNI: AD/CN: 89%; pMCI/sMCI: 74%	ADNI: AD/CN: 94.4%; pMCI/sMCI: 70.1%	**AUC**: ADNI: AD/CN: 0.941; pMCI/sMCI: 0.721
[108]	2019	ADNI: [3] CN:653; AD:238; MCI: 1002	3D ROI	NO	Freesurfer + Multinomial Logistic classifier	NO	3way :98.78%			
[54]	2018	ADNI:[3] AD:60; NC:60; pMCI:60; sMCI:60	3D ROI	NO	FreeSurfer + Random forest/- + fuzzy classes / DNN	NO	DNN:53.7%; fuzzy: 50.9%			

[1] D.A. - Data augmentation; [2] T.L. -Transfer Learning

[3] Use of clinical, demographic, or cognitive exams data in the approach.

- ImageNet; ⊙ - internal

Like in [64], where was implemented an end-to-end CNN, comparing two known architectures, ACNN and ResNet, to perform a 3-way classification having better results with ResNet.

Considering only the MRI data pipeline presented in [65], many GM slices from the same subject were evaluated with MCADNet. Afterward, for an adequate patient class allocation, the slices classification gave the final patient's decision by a voting majority process which improved and stabilized the outcome of the network, presenting a high three-way accuracy classification.

This final classification scheme was also used by [66]. However, it was applied to combine the predictions from the ensemble learning implemented between the three projections views, i. e., a set of slices per patient from each view were submitted on multiple 2D CNN. The five best classifiers produced the individual view result from every collection of slices to further achieve the final decision. Another example of ensemble learning was presented in [67], where was adopted a more complex ensemble model divided into two stages. In the first stage, they used an unsupervised CNN model based on PCANet to learn the features from MRI images of each view (implementation of 3 different PCANet). The final features were concatenated, resulting in a subject feature vector. Then, that vector was inputted into an unsupervised classification method based on K-Means clustering to achieve the classification task. The main conclusion was that using more input data was achieved a higher classification accuracy.

Similar to the initial approach, in [68] was used patches from hippocampal ROI (3 slices from each projection) with data augmentation to feed 3 DeepCNN separating the planes. An individual classification is collected from each network through a voting process where the results were obtained hinge on most votes. This data restriction was based on the fact that the hippocampus area is a strong biomarker of AD since its atrophy is visible in the earliest stages of the disease. Considering this information and the fact that a smaller examination area leads to a reduction in the computational cost, the Authors of [69] created an algorithm based on a CNN to classify MCI converted or non-converted. The left hippocampus was extracted from MRI images using an ROI mask mapping technique. Also in [70], it was presented a classification approach based on the same segmented region. A set of slices (5 slices) was selected in the sagittal view centered in the hippocampus (the technique of 2D+£, formed by extracting three patches from the three planes centered at the same point) to take into account the inter-subject morphological differences.

To a reliable approach, techniques like data augmentation, blurring the data (to augment the dataset too), and balancing data were applied. Finally, patches from the hippocampus ROI were extracted and fed to the 2D CNN to perform the classification. Nevertheless, in [71], the features to obtain the final classification were processed through two steps. The first one was with FreeSurfer software, and the second was with a pre-trained CNN with AD vs. NC images. The network input was 2.5D patches, like in [70], which produced an RGB image where each color corresponded to a particular axis. The two paths go by an age correction procedure that removed age-related effects as atrophy by estimating a pixel regression model fitting. Afterward, the features undergo a reduction by a PCA algorithm followed by a LASSO to select the most informative features. They were later inserted in an Extreme Learning Machine, a feed-forward neural network that learns much faster than regular networks performing the classification between sMCI and pMCI.

Joining the WM to GM slices, the Authors in [72] introduced a model based on the ResNeXt with Selective Kernel (SK) convolution and a spatial attention mechanism. SK adaptively adjusted the receptive field sizes to capture more useful information through a dynamic selection mechanism, while the spatial attention map defined how to highlight or suppress the feature maps. They obtained the best results with SKA-50, a fifty-layer network.

Despite the many studies where the network was trained from scratch, due to the small datasets in neuroimaging problems and the long training process, the use of pre-trained networks (transfer learning technique) has recently become more attractive. The weights of connections were obtained in most cases from ImageNet (a database with millions of natural images) with many different outputs. In addition, these networks are applied and adjusted to AD problems using a fine-tuning procedure to perform further classification (binary or multiclass).

Among the reviewed articles with MRI as the only neuroimaging technique, stand out 10 with transfer learning on 2D CNN [73, 74, 75, 76, 77, 78, 79, 80, 81, 82].

In [73] and [74], both implemented a pre-trained network using 2D MRI brain slices as input. The first used the axial view, while the second one defended that the coronal view is the most reliable to AD's diagnosis. In [74], the method followed to the classification step was based on a Dense-121 architecture working with data augmentation to increase the number of slices to train, avoiding overfitting. The Authors of [73] want to prove that with a small dataset, applying transfer learning, it is possible to achieve good results. They used only thirty-two slices from each subject, selected by the highest entropy, to train a VGG19 network to accomplish the goal. Also, in [75], the entropy technique was adopted to select the most discriminative slices from each plan, with careful consideration of the outliers pixels, which were excluded enabling the computation of more reliable values. Thereby, the best ten selected from each axis were inputted into a modified version of AlexNet to perform classification.

In [76], to supplement the information in the images injected into the network, the Authors generated a single RGB image joining a branded image (with sociodemographic data) in the Red channel, a blurred slice in the Green channel, and a CLAHE image in the Blue channel. The produced image is then inserted, and deep convolutional activation features were extracted from the average pooling layer of the DL system Inception-v3 (pre-trained also with ImageNet). Regarding the dependencies between predecessor and successor slices, the classifier is in the form of an LSTM architecture based on RNN. They concluded that visually fusing sociodemographic and genetic data in MRI slices enrich the extracted knowledge. Following this data augmentation manner, [77] presented the creation of an RGB image by the aggregation of different brain slices within a random interval of the same subject. They compared two different architectures, GoogleNet and CaffeNet (transform learning), to realize a 3way classification and the conversion time prediction, which will be further explored in a future section. CaffeNet proved to be the most successful in both tasks.

The Authors of [78] started with three individual DL models based on VGG-11 using MRI scans. They selected, per unit, one slice that was previously reported with anatomical areas correlated with AD and MCI. On the outputs of the three VGG-11 models was carried out max, mean,

and majority voting. The predictions from these three voting methods were then combined by majority voting to generate a fused MRI model. Jointly, they developed two MLP models, one using MMSE results and the other handling Logical Memory (LM) test data. Later, the predictions of these models were also combined by majority voting. Another approach employing ensemble learning was presented in [79], where a classifier was created based on three different classifiers models: eResNet50, eNASNet, and eMobileNet. In each model, were inserted 20 slices from each patient, the produced outcome resulted from the ensembling of the three models.

Following the use of transfer learning, [80] worked with a ResNet18 to extract slice-level features in top-50 large intracranial portions from each scan. After, through the BoW model, the features were aggregated by scan individually to integrate longitudinal-level features. As the last step, the data was evaluated with an RNN taking the LSTM model to capture disease pathological changes. With a different network, [81] described an approach to extract additional information from the most discriminative regions by evaluating the heat map inserted in the "Channel Attention Mechanism" consequently embedded into the Inception network. Thereby, during the convolutional phase, the kernel used to create the feature maps will acquire a lower dimension in the most relevant spots.

An implementation of a hybrid system that used an ML conventional classifier, adopting a transfer learning technique, was detailed in [82]. A VGG16 network was pre-trained in ImageNet and trained to classify sMCI vs. NC through different brain slices. Afterward, the VGG16 was supplied with brain slices from patients to a subject-level approach. In the feature extraction, from each patient were chosen 32 brain slices resulting in 32x256 features, which was reduced applying LASSO, a feature selection method, and then used an SVM classifier.

Some authors defend that 2D approaches are not enough to acquire all the pertinent information. Thereby, 3D structures are being developed to extract more spatial information relating to the slices and hierarchically distinctive features. From the most straightforward approaches like [83], where the Authors applied two pre-known procedures as ResNet and VoxCNN, including few steps in image preprocessing, and [84], where was used a 3D VoxCNN to reduce the training time without a decreasing of the prediction performance, to more complex methods, like [85], 3D CNN is widely used in the reviewed literature. Concerning [85], the Authors implemented a 3D CNN to learn spatial features of 3D GM density maps at different time points scan of the same patient. Thereupon, stacked bidirectional gated recurrent units (BGRU) were trained to capture the multiple time points' longitudinal features through CNN's outputs. These units can process the varying length image sequences to alleviate the problem of incomplete longitudinal data. Nevertheless, in [33], the Authors proposed a novel approach that ensembles multiple sparse regression models, with different values of sparse control and a DeepCNN. The sparse regression model is Joint Linear and Logistic Regression (JLLR) with L2, L1 norm regularization to select the most discriminative volumetric features from GM tissue volumes of 93 ROIs. After, a CNN (DeepESRNet) took as input the target-level representations acquired from the JLLR model and discovered non-linear relations to enable the final brain disease diagnosis. However, with this sparse regression model, the system needed a fine-tune of parameters to control the regularization,

which hindered the system development.

Still using 3D data, [86, 87, 88] applied a patch or a voxel + patch way to handling the input data. In [86], the total MRI volume was incorporated, being partitioned into several local regions, and consequently, 3D patches were selected from each image region. These patches were grouped into different clusters by the K-Means clustering method, and each group was individually injected into a 3D-DenseNet to proceed to feature extraction. The features learned are then concatenated to perform the final classification. In contrast, and attending the discriminative power of the hippocampus region, [87] and [88] presented 2 different approaches. [87] implemented Vnet architecture derived from CNN to do the hippocampal segmentation. The features used in the bottleneck phase to process the referred segmentation were used jointly with the features extracted by the 3D DenseNet realized with the input of the segmented hippocampal region. Those were concatenated to proceed with the classification through a softmax layer. On the other side, in [88], the Authors started with the segmented hippocampus by a mask computed from the centroid. After, it was simultaneously applied two networks, one for each hippocampus (left and right hippocampus), an MLP to extract shape features, and a 3D-DenseNet to obtain image features. Later, the two feature types from both sides were concatenated to perform the final classification step. Thus, the method can use the visual and shape features to reduce the computational complexity due to the decrease of data size. Both approaches achieved competitive classification accuracy comparing to other methods that use the whole brain.

Similar to 2D CNNs, some 3D architectures enable internal knowledge transfer by carrying the network weights from a task to another which provides a faster training step. The articles [56, 89, 90, 91, 92, 93, 94, 95] are examples of the application of this technique.

In applications end-to-end, there are works [56, 89, 90] which applied this transfer method and also adopted data augmentation. In [56], it was used the training information of AD vs. CN to pMCI vs. sMCI, as well as [91]. There was expressed the idea that firstly is necessary to learn how to achieve a more manageable problem and then use that information to process a higher challenging question. This argument was also acclaimed by the Authors of [89], who used the weights obtained through the network in AD vs. CN classification to a three-class task. Beyond the MR brain volume input, they added two features from demographic variables (age and gender) in the first fully connected layer to complement the data. Then, following the argument above, [91] pre-trained a Convolutional Auto-Encoder-based unsupervised learning in AD vs. NC task. Subsequently, the learned knowledge was applied to classify sMCI vs. pMCI using a Scaled Exponential Linear Unit (SELU) activation to accelerate the process and took a normalization effect before the softmax layer. Contrasting, in [90], the developers used sVoxCNN, a modified version of the known architecture VoxCNN with fewer layers, applying the trained weights from the three-category network to initialize the binary-category network weights, which was reported to quicken the convergence and consequently shorten the training time.

Without applying data augmentation, ADNET [92] implemented two ways of transfer learning, extending the traditional knowledge transfer capability used in the other articles. Therefore, data from healthy subjects was used in a network to obtain an age prediction, which was then

employed as a pre-trained network to classify pMCI vs. sMCI. Finally, the age information and the classification network data were fused and adjusted to achieve the final prediction system. The Authors of [93] also applied an ADNET with a slight variation, ADNET-DA, which had domain adaptation creating a system that was enhanced for the target domain. Besides, logistic regression was used instead of softmax using the one vs. rest technique to perform the classification in the CADDementia dataset (pre-trained and optimized in ADNI). So, and concerning a clinical environment scenario, the CAD system was trained and tested in different datasets with the target in CADDementia.

Exploring a different form of handling a CNN, [94] presented an unusual method to extract features, a Graph-CNN, where the convolution and pooling operations of this architecture were designed on an irregular grid, instead of a defined matrix. The cortical thickness graph (geometry of cortical values) was injected into the network to extract features. The proposed framework can capture the brain's morphological changes essential to AD/MCI detection regardless of the population's cohort. It was applied transfer learning between different datasets proving the appropriateness of this approach in small datasets.

A novel method to proceed with the classification was proposed by [95], exploring a system that systematically integrated different approach types into a combined structure. First, ROI volumes from the GM were obtained based on predefined anatomical atlas. Afterward, for each region, it was estimated a representation of the voxels in a subspace. Consequently, for an individual sector was compounded a set of randomized DNNs. Each of these was constructed based on volumetric measurements of voxels within the respective region, resulting in multiple outcomes from the same brain. The collected results were fed into an SVM classifier to perform the brain-wise classification. Subsequently, these Authors applied the transfer learning technique to pre-train the sMCI vs. pMCI with the weights obtained in AD vs. CN.

The ensemble learning technique was employed in the following articles adopting volume data, splitting up between an ensemble learning with probability-based fusion [96, 97], a weighted-based fusion method [98] or majority voting method [99]. In [96] was presented a model of ensemble 3D-DenseNet with brain volumes input, where was computed different hyperparameters to obtained distinctive networks. Each system will reach a result based on probabilities given by the softmax layer. After that, these probabilities will be merged to obtain the outcome. The final result was acquired by probability normalization and not by the majority voting method as usual. This architecture, including the indicated fusion method, achieved a high accuracy to distinguish AD vs. MCI vs. CN. With a similar approach, [97] introduced a system with three distinct 3D-DenseNet121 with different hyperparameters. The Authors performed the final classification separating four classes (AD vs. pMCI vs. sMCI vs. NC) with high accuracy. With the weighted-based fusion method, in [98], it was developed an ensemble of 3D-DenseNet to recognize AD and MCI. They joined five networks with different hyperparameters to execute the classification. The network with the best individual results had a higher weight in the final decision. Another system, adopting a majority vote process with a voxel + patches approach was presented in [99]. A Landmark-based Deep Multi-Instance Learning (LDMIL) compared the two extremist classes

(AD vs. NC) and identified the brain locations with significant statistical differences. With the selection of these points, 3D patches were extracted and classified. Therefore, the final subject decision was made by a majority voting strategy encompassing the outcome from all patches. This approach integrated the advantages of 3D without overloading the system with data.

Using a 2D and 3D approach, the Authors of [100] applied a Multi-View based Multi-Model (MVMM) learning framework to join the advantages of these approaches, selecting features from GM in these two ways. In 2D, they extracted local and in 3D global/spatial information. For the first procedure, the slices most informative obtained through entropy selection algorithm were selected and injected in a Dual Attention Inception Network (DAIN), which was based on an Inception network with a dual attention mechanism to learn extra local information. On the other side, the 3D features were extracted by DRN (Dilated Residual Network) derived from ResNet. The join of that info was done by a concatenation fusion method to achieve higher performance. Comparing these two dimensional ways of handling data, the articles [101, 102, 103, 104] presented different systems. On one hand, in [101] was created in parallel two classification systems with 2D and 3D GM input. They used an architecture based on VGGNet, an FCNN (Fine-tuning CNN) with an extra shortcut to merge low-level and high-level feature information. The brain volume scan has shown better performance compared with 2D CNNs. Already in [104], the Authors implemented three different approaches from 2D CNN, 3D CNN with a softmax classification, and a 3D CNN with SVM classification, concluding that the hybrid method achieved the best efficiency.

On the other hand, despite the articles [102, 103] referred to multimodalities approaches, the neuroimaging techniques were processed individually. They computed four architectures to compare MRI and PET and 2D / 3D CNN using the whole brain's image without segmentation or brain extraction. The best performance was obtained in 3D over 2D, as expected. Comparing imaging techniques with the presented approach in an AD vs. MC vs. CN classification, the PET modality achieved better results. The binary classification between AD vs. MCI demonstrated in [103] using a similar approach accomplished the same outcome. In summary, it was notable that the general report pointed to a better performance of 3D systems over 2D due to the importance of spatial features.

In [105, 106, 107] was exploited the use of clinical or demographic information, like cognitive tests, CSF measurements, genetic risk, and others in addition to MRI brain volume. On the one hand, the Authors of [105] proposed a CNN model that merged the two type of data in the final step of the network achieving the best classification. On the other hand, [106] defended that the clinical data allowed a better prediction quality than the neuroimaging, and this information can be used for conversion prediction tasks. In the executed tests with a ResNet3D, they achieved a slightly higher value in the outcome of an embedding method than using only clinical data. In the attempt to encompasses the two types of data, [107] used a 3DAN (3D attention network) based on ResNet architecture to extract features from GM volume. The attention mechanism layer into a deep CNN algorithm will understand which regions had more discriminative power, creating a subject-level attention score map, which was compared with neurobiological support, concluding

that GM atrophy and MMSE score highly correlated, influencing the final classification.

Freesurfer is a tool widely adopted in brain image processing, used in [108] and in [54] to extract features. While in [108], it was extracted 88 features, such as volume, surface area, and cortical thickness, in [54], it was obtained 431 features, among them, curvature, surface area, volume, or hippocampus volume. Considering the high number of acquired details, this article carried out an analysis, and the highly correlated features were removed. In both cases, features from the clinical forum were also used. In [108], all the extracted features were inserted in a Deep Sparse Autoencoder for feature selection. Then, the most discriminative ones were employed for supervised classification by a Multinomial Logistic Classifier. However, in [54], the Authors proceeding to feature selection, using a random forest classifier to identify the 20 most important features. The CSF volume, the lateral ventricle volume, the entorhinal cortex thickness, and the baseline MMSE were some of the identified features. The classification was done by a DNN, which included an optimal mixed configuration of activation functions and a decreasing number of neurons in each layer.

4.1.2 Multimodal Imaging Approaches

Of all reviewed articles, 17 adopted different image modalities, combining MRI and other neuroimaging techniques, Table 4.2. Regarding the research done, the most employed imaging modality is Positron Emission Tomography (PET), followed by Diffusion Tensor Imaging (DTI). To increase the variety of data features, other types of data such as CSF biomarkers and cognitive tests were also found. In all articles, the base data was from the public database ADNI. Yet, [119] also used the AIBL database to perform experiments.

Concerning DTI imaging to support MRI [121, 115], the reviewed articles showed a 2D plan on CNN. Focusing on [115], a CNN model was pre-trained on AD vs. NC classification using data augmentation to reduce overfitting. The Authors trained a Caffe network for each projection on the MRI dataset. Next, by applying transfer learning of learned weights, the network was trained with the MD map of the DTI dataset (with also the three projections). Classifiers were then fused by a majority vote, increasing binary classification accuracy between NC, MCI, and AD. Also adopting transfer learning technique, [121] implemented a pre-trained network in ImageNet to classify AD stages. Thereunto, RGB 2D image was raised from the overlay on different channels, R and G were respectively the FA and MD maps from DTI and B channel corresponded to the MRI slice. The generated images undergo the VGG16 architecture to extract features, followed by the LASSO algorithm to reduce feature dimension and redundant information. It was only considered the classification between sMCI and NC that used an SVM, an example of a hybrid approach: the feature extraction step was done by a DL technique followed by an ML classifier.

Following also a hybrid method in MRI+PET data, the [52] used MRI(GM) plus PET to manage a latent high-level feature representation in the application of a method based on the AD vs. NC comparison, finding the most statistically significant voxels in the process of class discrimination. Thus, these voxels were used to create patches to extricate local region information through a Deep Boltzmann Machine (DBM), a network constructed by hierarchically stacking multiple

Table 4.2: Resume table of multimodal imaging approaches in state of art.

Article	Year	Modal	Dataset source + size	Approach type	Data augment.	Feature extraction + feature selection + classification	Transfer Learning	Accuracy	Sensitivity	Specificity	Other metrics
[52]	2014	MRI+PET	ADNI AD: 93; NC:101; MCI:204 (pMCI:76; sMCI:128)	3D voxel + patches	NO	DBM + SVM	Greedy Layer-Wise	MRI+PET: AD/NC: 95,35%; MCI/NC:85.67%; pMCI/sMCI:75.92%; MRI: AD/NC: 92,38%; MCI/NC 84,24%; pMCI/sMCI:64,75%;	MRI+PET: AD/NC: 94,65%; MCI/NC 95,37%; pMCI/sMCI: 48,04%; MRI: AD/NC: 91,54%; MCI/NC 99,58%; pMCI/sMCI:36,70%;	MRI+PET: AD/NC: 95,22%; MCI/NC 65,87%; pMCI/sMCI:95,23%; MRI: AD/NC: 94,56%; MCI/NC 53,79%; pMCI/sMCI:90,98%;	**BACC**: MRI+PET: AD/NC: 94,93%; MCI/NC 80,62%; pMCI/sMCI: 71,63%; MRI: AD/NC:93,05%; MCI/NC: 76,69%; pMCI/sMCI:63,84%; **AUC**: MRI+PET: AD/NC: 0,9877; MCI/NC 0,8808; pMCI/sMCI:0,7466; MRI: AD/NC: 0,9697; MCI/NC:0,8478; pMCI/sMCI:0,7342;
[109]	2015	MRI + PET + CSF + clinical	ADNI AD:51; NC:52; MCI:99 (pMCI:43; sMCI:56)	3D ROI (93)	NO	SAE + LASSO + multi kernel SVM	Greedy Layer-Wise	AD/HC:98,8%, MCI/NC:90,7; AD/MCI: 83,7; pMCI/sMCI:83,3%			
[49]	2013	MRI + PET +CSF +MMSE +ADAS-Cog	ADNI AD:51; NC:52; MCI:99 (pMCI:43; sMCI:56)	3D ROI (93)	NO	SAE + multi kernel SVM	Greedy Layer-Wise	AD/HC:95,9%, MCI/HC:85%, pMCI/sMCI: 75,8%			
[110]	2018	MRI (GM) + PET	ADNI AD:51; NC:52; MCI:99 (pMCI:43; sMCI:56)	3D ROI (93)	NO	MM stacked - DPN + linear kernel SVM	NO	AD/NC: 97,13%; MCI/NC: 87,24%; pMCI/sMCI:78,88%; 4way:57%	AD/HC:95,93%, MCI/NC:97,91%; AD/MCI:83,7%; pMCI/sMCI:68,04%; 4way: 53,65%	AD/HC:98,53%, MCI/NC:67,04; AD/MCI: 83,7; pMCI/sMCI:86,81%; 4way: 85,05%	
[111]	2015	MRI + PET + CSF	ADNI AD:51; NC:52; MCI:99 (pMCI:43; sMCI:56)	3D ROI	NO	volumetric features + PCA-RBM + linear kernel SVM	NO	AD/HC: 91,4%, MCI/HC: 77,4%, AD/MCI: 70,1%, pMCI/sMCI : 57,4%			
[112]	2019	MRI+PET +genetic data	ADNI MRI: AD:190; NC:226; MCI:389 (pMCI:157; sMCI:205); PET: 360 (total); SNP: 737 (total)	3D ROI (93)	NO	3 DNN (one for each modality) + 3 for modality pairs + DNN final	NO	3way: 64,4%; 4way: 54%; NC/AD:90% ; NC/MCI:74%			
[113]	2018	MRI+PET	ADNI AD:358; NC:429; pMCI:205; sMCI:465	3D patches	NO	LM3IL	NO	AD/NC:92,5%; sMCI/pMCI: 79,06%	AD/NC:89,94%; sMCI/pMCI: 55,26%	AD/NC:94,53%; sMCI/pMCI:82,85%	**AUC**: AD/NC:0,9589; sMCI/pMCI: 0,7584
[114]	2019	MRI+PET	ADNI AD:93; NC:100; pMCI:76; sMCI:128;	3D ROI	NO	3D-CNN	NO	AD/NC:94.82%, pMCI/NC:86.36%, sMCI/NC: 65.35%	AD/NC:97,7%, pMCI/NC:83,33%; sMCI/NC:70,59%	AD/NC:92,45%, pMCI/NC:88,78%, sMCI/NC: 59,63%	**AUC**: AD/NC:0,9676; pMCI/NC:0,9111; sMCI/NC: 0,6917
[115]	2018	MRI + DTI (MD)	ADNI AD:188; NC:228; MCI:399 MM: AD:48; MCI:139; NC:58	2D ROI	YES	CNN	MRI model is applied in MD	AD/NC:92.5%, AD/MCI:85.0%; NC/MCI:80.0%	AD/NC:94,7%, AD/MCI:93,7%; NC/MCI:92,8%	AD/NC:90,4%, AD/MCI:79,1%; NC/MCI:73%	**BACC**: AD/NC:92,5%; AD/MCI:86,4%; NC/MCI:82,9%
[116]	2018	MRI(GM) +PET(GM) +CSF(3)	ADNI AD:51; NC:52; MCI: 99	3D ROI (93)	NO	sELM-AE + KELM	NO	AD/HC: 97,12%; MCI/HC:87,09%	AD/HC: 98,08%; MCI/HC:75%	AD/HC: 94,12%; MCI/HC:91,92%	**BACC**: AD/HC: 96,1%; MCI/HC:83,46%
[117]	2019	MRI + PET + RAVLT + MoCA + ECogT	ADNI AD:159; NC:248; pMCI:193; sMCI:296;	3D ROI	NO	Parameters +ANOVA + DNN	NO	sMCI/NC: 84%; pMCI/NC:84,1%; AD/CN:96,8%; sMCI/pMCI: 69,5%; sMCI/AD: 90,3; pMCI/AD:80,2%	sMCI/NC: 83,2%; pMCI/NC:80,4%; AD/CN:94,1%; sMCI/pMCI: 80,6%; sMCI/AD: 86,7%; pMCI/AD:86,8%	sMCI/NC: 84,4%; pMCI/NC:87,6%; AD/CN: 98,2%; sMCI/pMCI: 60,5; sMCI/AD: 92,2; pMCI/AD:71,9%	
[118]	2019	MRI+PET	ADNI AD:205; NC:231; pMCI:165; sMCI:147;	3D ROI	NO	DSNN	NO	AD/CN:92,25%; sMCI/pMCI: 80%	AD/CN:93,3%; sMCI/pMCI: 83,2%;	AD/CN:90,91%; sMCI/pMCI:70,79%	**AUC**: AD/CN:0,9702; sMCI/pMCI: 0,8394
[119]	2020	MRI+ PET	ADNI + AIBL total:355	3D ROI	NO	Spatially- Constrained Fisher Representation	NO	AD/CN: 93.58%; pMCI/sMCI: 77,44%	AD/CN: 91,52%; pMCI/sMCI:79,07%	AD/CN: 95,22%; pMCI/sMCI:77,22%	**AUC**: AD/CN: 0,9695; pMCI/sMCI: 0,8251
[120]	2019	MRI+PET	ADNI AD:114; NC:133; MCI:132	3D ROI	NO	3DResNet + DDNN	NO	AD/MCI: 79,17%; AD/NC: 93,55%; MCI/NC: 78,92%; 3way: 68,86%			**AUC**: AD/NC: 0,9514; MCI/NC: 0,7837
[121]	2020	MRI+DTI	ADNI sMCI: 70; NC: 50	slices	NO	VGG16 + LASSO + SVM	YES (ImageNet)	EMCI/NC: 94,2%	EMCI/NC: 97,3%	EMCI/NC: 92,9%	**AUC**: EMCI/NC: 0,953
[122]	2019	MRI+PET	ADNI AD:647; NC:731; pMCI:326; sMCI:441;	3D ROI	NO	VGG11	NO	CN/AD:90.10%; CN/pMCI:87.46%; pMCI/sMCI:76.90%	CN/AD:90.85%; CN/pMCI:90,73%; pMCI/sMCI:68,15%	CN/AD:89,21%; CN/pMCI:80,61%; pMCI/sMCI:83,93%	**AUC**: CN/AD:0,9084; CN/pMCI:0,8761; pMCI/sMCI:0,7961
[123]	2018	MRI + PET	ADNI sMCI:409; pMCI: 217	3D patches + ROI	NO	pre-training SAE + MMDNN (6 DNN) + DNN	NO	sMCI/pMCI:82,93%	sMCI/pMCI:79,69%	sMCI/pMCI:83,84%	

RBMs. So, it began to capture basic image features in a lower layer and evolve to a more abstract and higher level. The outcome was then inserted into an SVM. Finally, in the test set, it was constructed mega-patches following a greedy manner, a result of the patches selected with the highest classification accuracy from the training phase. The SVM classifier can be used subsequently to numerous feature extraction methods. In [110], from MRI (GM) along with PET, was employed a Stacked Deep Polynomial Network (DPN) consisting of two stages. In the first step, it was extracted characteristics in order to learn high-level features from each of the separate input modalities. Following the association between the two modalities, the information was fused, and the classification was performed. Furthermore, in [111], instead of two input modalities was added the CSF data and the features extraction is executed similarly to [116]. Those went by a selection where PCA was implemented to decide the most discriminative ones. This election was fed into Multi-task DL by a Restricted Boltzmann Machine (RBM) to create relations between modalities and added two additional clinical scores. The outcome of these associations was then adapted to classify.

The [109] and [49] articles presented similar algorithms with an improvement in the feature selection. The features were initially extracted from 93 3D ROIs, deriving from the GM volume and the mean intensity of PET rigidly registered. Three biomarkers from CSF complete the feature vector that will be inserted in SAE. Moreover, in [49], the concatenation of SAE-learned features and the original low-level features resulted in an augmented feature vector where multi-task learning was applied to select features that collectively realize the class label and the clinical scores. Whereas in [109], instead of multi-task learning, the Authors applied a sparse regression learning LASSO to select features that efficiently regress the target values, finding the discriminative features in a low dimension. To finalize, in both, this multimodal information was classified by multi-kernel SVM.

Initially, in [112, 123, 116], the features were extracted following the same method than [109] or [49]. While in [112, 116] these features were extorted from 93 3D ROIs, in [123] the data came from patches. Joining the neuroimaging data, [116] also used CSF data applying a Stacked Sparse Extreme Learning Auto-Encoder with three hidden layers to extract high-level representations individually. Then, another sELM-AE was implemented to fuse the learned features, resulting in a joint hierarchical feature representation.

To make the diagnosis, [112] presented a three-stage DNN with seven networks, where it is manipulated MRI, PET and Single Nucleotide Polymorphisms (SNP) data, the most common type of genetic variation among people. Firstly, the DNN learned the latent representations (high-level features) from each modality. Then, in stage 2, the modalities were fused in pairs in another DNN. Lastly, in stage 3, the three resulting features were injected in the last DNN, where they were repeatedly fused. In each DNN, a softmax layer provided the classification. A majority voting strategy was used to obtain the outcome through all the seven softmax output layers, producing a multiclass classification. A more complex network was presented in [123], with a multimodal and multiscale DNN composed of seven DNNs. The first step was based on the definition of three patch sizes (500, 1000, and 2000 voxels) and consequently varying the patches' quantity. For

each patch size and modality was implemented a single DNN. Suddenly, the features from these six DNNs were fused in the last DNN to generate a final probability output by a softmax layer. [117] also used a DNN to perform the final classification. The injected features were extracted by FreeSurfer, joining clinical information as neuropsychological test scores. The features were evaluated by a statistical analysis algorithm, the Analysis of Variance (ANOVA), where the highly correlated features were removed for dimensionality reduction purposes.

Following the most architecture used in image classification, it was introduced in [120, 114, 122] different known formations based on CNNs. The three applications commence with a 3D approach obtaining, separately, features from each imaging technique's volumes. After, they were combined and classified. The Authors defended that in 3D space, the networks will also capture spatial information enriching the information collected. Beginning with the article [120], it was employed two 3D ResNet for each image acquisition procedure. This info was fused in one feature vector and fed to a Dense DNN, giving a final binary and multiclass classification between NC, MCI, and AD. Succeeding the implementation of 3D-CNN in article [114], the learned features were injected in a Fully Stacked Bidirectional Long Short-Term Memory (FSBi-LSTM), instead of an FC layer which revealed to improve the extraction of high level semantic and all the spatial information from the feature maps. The extracted features by this technique were enhanced to boost diagnosis performance. [122] preferred the VGG11, another CNN architecture type to extract separately features from an ROI centered only in the hippocampus. Furthermore, they were concatenated to catalog in the end via the softmax layer.

In general, PET images are more challenging to obtain than MRI. In public datasets like ADNI, not all MRI subjects have PET, complicating these multimodal approaches. So, [113, 118, 119] developed three ways to generate synthetic PET images based on Generative Adversarial Networks (GAN), a 3D Cycleconsistent GAN, a Feature-consistent GAN model and a Hybrid GAN, respectively, capturing the underlying relationship from MRI data. In [113] was implemented a patch-based approach. These patches were centered at predefined disease-related landmarks on the two imaging techniques implemented by an LM3IL (Landmark-based Multimodal Multi-Instance Learning). The learned features from each patch were concatenated and learned by an FCL, followed by a softmax layer that gave the final output. Contrastly, [118] adopted a DNN, Disease-image Specific Neural Network (DSNN) to classify with features directly taken from the whole brain. While [119] applied a spatially-constrained Fisher representation (SCFR) network which generated some specific statistical descriptors to perform an end-to-end classification. The Authors of these articles prove that the usage of real and, when not available, synthetic PET data promotes brain disease classification with a network improvement.

4.1.3 Time-to-Event Prognostic Approaches

Besides the MR image classification into the different AD stages, some of the found articles focus on predicting the conversion of MCI to AD at a subject-level, i.e., predicting the time it takes the MCI patients to evolve the disease to an Alzheimer's diagnosis, Table 4.3.

Table 4.3: Summary table of time-to-event prognostic approaches found in the state of the art.

Article	Year	Modal	Dataset Source + Size	Approach	Data Augm.	Feature Extraction + Classification	Transfer Learning	Accuracy	Other metrics
[37]	2018	MRI	ADNI: MCI:165 (100 to AD within 12months; 65 to AD after 12 months)	slices (GM)	NO	CNN + SVM	NO	MCI to AD: within 1 year: 92.33%	MCI to AD: Sensitivity: 93.6%; Specificity: 92%; AUC: 0.9749
[124]	2019	MRI	ADNI:1711; AIBL:435	3D ROI (hippocampus)	NO	CNN	NO	prediction of MCI to AD within 1,2 and 3 years: 86.4%	
[77]	2018	MRI	ADNI NC: 150; sMCI:150; pMCI:157	slices (RGB)	YES	GoogleNet/ CaffeNet	YES (ImageNet)	MCI to AD with intervals of 6 months: GoogleNet: avg = 77.25%; CaffeNet: avg = 96.215%	
[123]	2018	MRI + PET	ADNI sMCI:409; pMCI: 217	3D patches + ROI	NO	pre-training SAE + MMDNN (6 DNN) + DNN	NO	MCI to AD: within 1 year: 90% within 2 years: 86.6% within 3 years:82.4%	

Most articles that perform this challenging task chose a CNN approach [37, 77, 124] similarly to what was previously concluded in the diagnosis assignment. Within a year, the study [37] used slices from the middle of GM, which were inputted in a CNN to feature extraction and further inserted on an SVM classifier. The SVM classification with RBF kernel was pointed due to the good results. Using the known architecture CaffeNet, [77] mentioned in Section 4.1.1 was also used to predict the conversion within three years. The pMCI data were divided into five groups according to the elapsed time until the AD diagnosis. As in the discrimination task, that network achieved the best result among those that the Authors have studied. Instead of using the whole brain, [124] worked with hippocampus area to extract in-depth imaging features to join subject clinical measures. In this way, the network was more specific and combined the advantages of 3D data. It was divided the MCI at three levels (low, middle, and high) of progression risk to AD through a variable distribution function. They implemented a CNN pre-trained in AD vs. CN classification to extract features and hereafter classify MCI subjects. According to the outcome, the DL imaging features had better results than shape and shape/texture features but worse compared with the use of clinical measures. Joining the DL hippocampal features with clinical resources was reached the best result (within a slight margin of clinical only).

The work developed in [123], previously mentioned in Section 4.1.2, also implemented a procedure to reach the time of conversion. The MMDNN classifier achieved high accuracy in converting pMCI individuals to AD due to the longitudinal information used.

4.2 Discussion

Within the studied articles, seven were review papers. The oldest refers to 2016, [125], where an extensive survey was presented for single subject prediction of several brain disorders, including AD and MCI. The biggest pitfall pointed by the Authors was the limited sample size that, therefore, provided a low ability to develop a robust predictor to discriminate the diseases and their subtypes accurately. A manner to counter that is using multi-site data-sharing initiatives like ADNI, where data is gathered using the same recruitment criteria and scanning protocols across sites, providing data like age, scores of cognitive functions test, genetic risk, and others [50].

Coinciding with the increasing accuracy and robustness in the prediction, machine learning techniques like deep learning also arise. Despite that, most of the studied articles did not report any DL architecture for feature extraction considering the publishing year 2016. Nonetheless, as advantages are denoting the ability to remove the selecting subjectivity and the depth of the models that augment the capability to model highly complex data patterns. To prove that, in the following year, 2017, DL was the focus of [50], compilating multiple psychiatric and neurological diseases with multimodal imaging techniques. The high-level features were also indicated as a successful way against noisy input data to complete the information already given. Since then, transfer learning and data augmentation were denoted as useful techniques in neuroimaging.

Regarding the MCI, it is crucial to highlight the difficulty of diagnosing this medium stage of the disease due to its relatively mild and nearly insignificant cognitive impairment symptoms. Even with the implementation of multimodal neuroimaging data, which improve the overall accuracy compared to the independent modality use [125]. A multimodal approach is valuable due to performing a metabolism interpretation besides the structural evaluation [126].

Some of the found reviews just focus on AD, such as [57, 38, 48]. Besides the DL approaches, the last one indicated the ensemble learning as successful, considering the promising results of highly complex data modeling with high precision [38]. In [35] and [48], it was highlighted the biggest challenge in implementing this type of system in a clinical environment since it required a consistent performance with good generalization and interpretability of models. Moreover, data privacy, accessibility, and ownership problems will also require attention and discussion between health professionals and society in general. A resolution is also needed to overcome the problem of heterogeneous data due to the differences in medical equipment of image acquisition.

In this literature review, it could be seen that the AD classification with neuroimaging is the subject of a significant number of studies with different DL approaches. Although the oldest article report to the year 2013, the most remarkable rise in this field occurred after 2017, with more than 87% of the considered articles published after that year. These results can be explained by the exponential increase of DL methods and CAD systems' need for diagnosing. Furthermore, DL methods prove to be a good system to apply in medical images in multiple tasks, which permit a faster analysis process with a relief on the workload of medical staff.

This review focuses on studies that at least used MRI data, which lead 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 high costs associated. Also, articles that only distinguish AD from NC were excluded due to the will to predict cognitive alterations before the condition is fully developed.

Concerning the multimodal articles, PET is the most used in association with MRI due to the properties to analyze the metabolism of structures bringing value to the structural evaluation already carried out.

There are several publicly available datasets for investigation purposes to analyze the different AD stages being ADNI the most adopted, providing multiple acquisition techniques and cognitive/clinical data.

Although, it is undisputed that classification between AD and NC is the most straightforward task obtaining a high accuracy reaching almost 100% in some studies. Therefore, it is not so clinically relevant as opposed to the prediction of CN to sMCI and pMCI to AD, which 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 possible to conclude which is better because of both present limitations and advantages. 3D models represent more arduous networks due to the increased number of parameters and computational cost. However, there is more information regarding the spatial characteristics, while in 2D approaches, not all information is acquired by discarding spatial connections, and consequently, the computational cost is lower. Attending the architecture type, CNNs are predominantly the most employed in image classification. However, recently, other architectures give good reasons to be further explored, like RNN in longitudinal data processing to extract the minimal differences between consecutive scans in the same subject.

The most prominent problem in the application of DL in neuroimaging is the lack of training data. Even though several datasets exist, reaching a well-trained model with no overfitting requires a large quantity of data. For this reason, several techniques have been developed to overcome this issue, being widely used in the reviewed articles. Data augmentation and transfer learning have been implemented, yielding satisfactory improvements.

A longitudinal conclusion in the review articles is the impossibility of comparing different systems performance reports due to the distinctions between the sample data despite the almost general use of ADNI, regarding the fact that the authors can do, within the same dataset, a selection of samples.

4.3 Summary

This chapter presents a systematic and extensive review of strategies for improving AD stages classification based on DL and neuroimaging data provided primarily by MRI, followed by PET and DTI, where other biomarkers or clinical information can be used. With the exponential increase of studies in this area due to the evolution of computational techniques, it is mandatory a recursive update of the advances related to DL applied to the disease under study due to be the leading cause of dementia.

Chapter 5

Methodology and Employed Resources

Following the background introduction about Alzheimer's disease and the processing of MR images, as well as the deep learning techniques applied to achieve the disease classification into different stages, this chapter presents the implemented methodology from the original data acquisition to the final output. Besides that, the theoretical framework of the DL algorithm is more deeply explored. Afterward, it is also introduced the computational infrastructure and some managed computational resources.

5.1 Overview

The objective of this experimental study is focused on testing preprocessing techniques and classification of brain MR images of patients with AD. In this sense, the steps of preprocessing, feature learning and classification of a typical image processing system were focused on and analyzed. The representative scheme in Figure 5.1 shows an overview of the steps adopted and described in detail in the following sections.

Initially, the images were studied and visualized through imaging software like ITK-SNAP. Some images did not show visible structural changes in the brain matter regarding the class where they were inserted, evidencing the value of an intelligent system application.

Concerning the preprocessing stage, the first step was the registration or the skull stripping of MR images, depending on the followed approach. For an atlas registration, firstly, the images needed to undergo skull stripping due to the atlas reference not presenting the skull. On the other hand, for an internal registration to facilitate and improve the process, the skull presence was required in the images and only after was done the brain extraction.

Therefore, for the registration step, where images are aligned between each other, a library specialized for effect was applied. Regarding the skull stripping task, the segmentation was evaluated in two algorithms. Afterward, and having an acceptable dimension to the process, a 3D-CNN was applied to perform the classification. For a 2D implementation, it was mandatory to cut and select slices by an entropy selection method to further inject in a pre-trained network, which, through a voting process, could perform a subject distinguishing disease class.

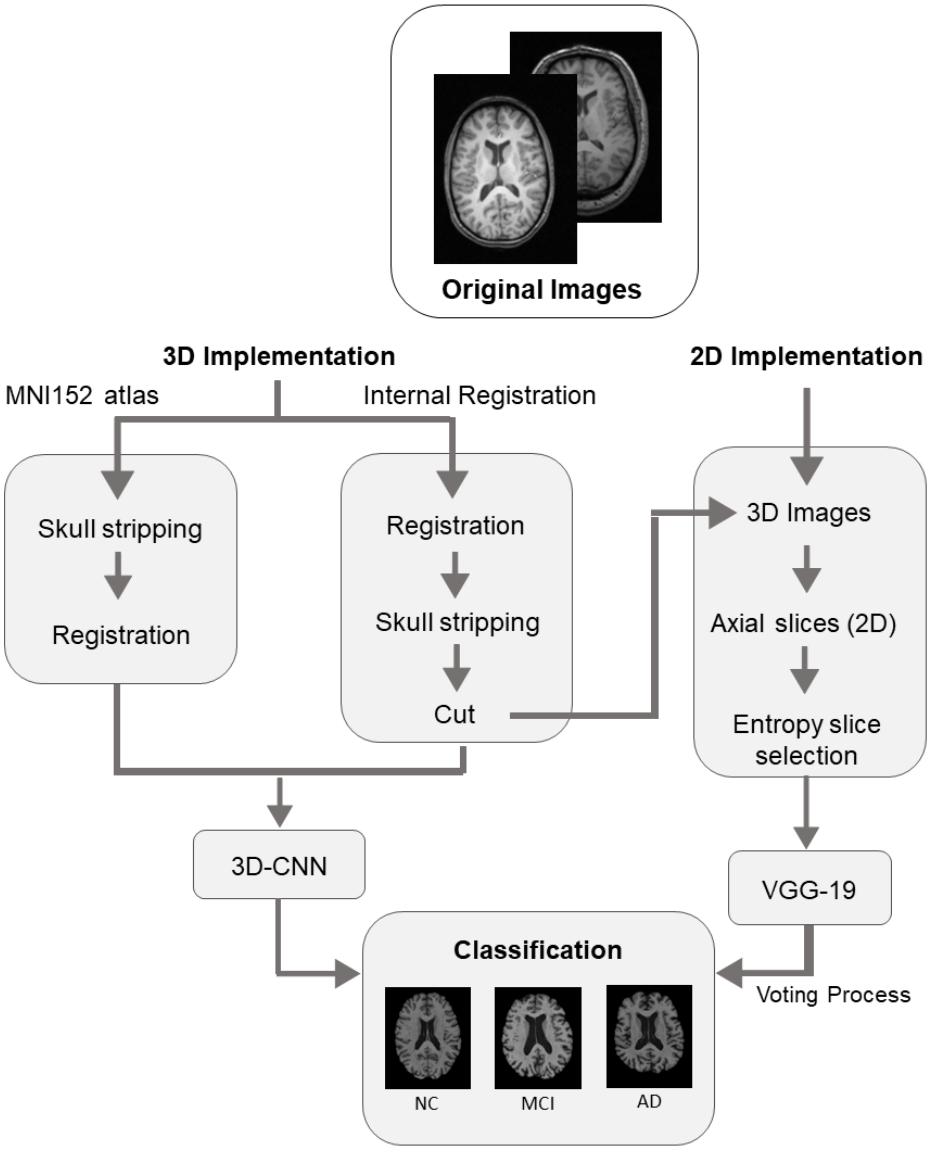


Figure 5.1: Schematic view of the steps implemented throughout this project, from the original image of input MR to classification.

5.2 Convolutional Neural Networks

Within DL models, the convolutional neural network (CNN) is highlighted due to a clear train's capability and a better generalization than networks with full connectivity between adjacent layers. It is a type of feedforward network, popularly embraced by the computer vision community.

CNN's are constructed to handle data that comes in the form of multiple arrays, 1D to signals or arrays of values, 2D in greyscale pictures or a set of three 2D matrices to RGB data, and 3D for volumetric images or video.

Deep networks can be formed by stacking convolutions, pooling operators, and non-linearity activation functions.

Ideally, the initial few layers should learn to extract low-level image features like lines, edges, and blobs while subsequent layers extract complex features. It is often followed by fully connected layers, which compile the data extracted by preceding layers and generates the outcome, transforming the features from spatially structured maps to, for example, a vector of estimated class probabilities [46]. The main significant distinction between CNNs and standard ANNs is that CNN's are mainly employed in image pattern recognition. This enables us to embed image-specific properties into the architecture, making the network better suited for image-focused tasks while also lowering the number of parameters needed to build up the model.

5.2.1 Network Layers

5.2.1.1 Convolutional Layers

Considering an image with a width and height of 32×32 pixels and a depth of 3, for an RGB image, a traditional ANN hidden layer will present a 32x32x3 weight connection. Alternatively, it will present 32x32xn for volume data with depth equal to n. By joining another neuron, the parameters will double. Two neurons may appear to be insufficient for any effective processing in an image classification application. The efficiency can be improved by connecting the input picture to the neurons in the next layer with the same height and width values, reducing the parameters.

Therefore, instead of a complete connection, analyzing small areas rather than the entire image is an excellent way to improve the learning skill. Even though the size of the connection has shrunk dramatically, there are still many aspects to work out. Another simplifying assumption is to keep the local connection weights for the entire layer of neurons unchanged. This will link the neighbor neurons in the next layer to the local region of the previous layer with the same weight. The size of this local region is typically called the receptive field. These basic assumptions have several advantages. Fixing the weights for the local connections is comparable to moving a window in the input neurons and mapping the resulting output to the appropriate spot. It allows the identification and recognition of features regardless of the place in the image, being this process called convolution.

Depending on the matrix set, the convolved result will be different, extracting much different information from the same input layer. These matrices, also called filters or kernels, are similar to the classical filters used in image processing. Adding further convolutional layers, the model will be more specialized [127]. Given an input feature map X with a size of $W^{in} * H^{in} * F^{in}$, where W is the width, H the height, and F the number of feature maps. A trained filters bank is implemented in line to

$$Y_j = B_j + \sum_i W_{ij} * X_i \tag{5.1}$$

Where X_i denotes the i-th feature map of X. The output feature map Y has a size of $W^{out} * H^{out} * F^{out}$. To mantain the same width and height as the input feature map, it can be applied padding. Optionally, a bias weight is learned for each output. The B_j indicates the bias resized to the width

and height of the output feature map. The filter size to produce each feature map varies depending on the application but is usually small in spatial dimensionality [127].

Three parameters control the size of an output volume: the depth, stride, and zero-padding size. The depth of a feature map results from the number of filters applied in the current layer. The stride is the step adopted in the "sliding" operation of the kernel. On the one hand, small strides can lead to overlapping receptive fields and larger output volumes. On the other hand, more considerable strides will result in fewer overlapping receptive fields and lower output volumes. The last one is the zero-padding based on the need to "pad" the borders of an image to retain the original image size. It is advantageous to train deep networks because, without this technique, the spatial dimensions of the input volume would reduce too quickly, and the network would not be able to do an efficient train. Considering these parameters, the output size will follow equation 5.2.

$$[\frac{n+2p-f}{s}+1] \tag{5.2}$$

Where, n stands out for image size, f,the size of the filter, p for padding, and s for stride.

5.2.1.2 Activation Function

The activation function is used to determine a neuron's output response. The total of the weighted input signal is combined with activation and added the bias parameter to decide whether it should be "fired" or not.

There are different groups to slip the numerous existent activation functions. A primary group is the linear functions which are helpful to handle simple tasks like interpretability, but it has limited power and capacity to work complexity. Therefore, it comes to the non-linear functions to allow the model to create complex mappings between the inputs and outputs of the network, which are essential for learning and modeling complex and non-linear or high- dimensionality data.

A non-linear activation function, such as ReLU, ELU, or any of the various Leaky ReLU versions, is applied after each convolutional layer in a CNN, being the RELU the most adopted. In this process, the data dimensions will remain the same [128].

Focusing on specific activation functions, the ReLU, PReLU, and ELU are further used in the convolutional layers of the model architecture while sigmoid and softmax, also non-linear activation functions, are adopted to perform the classification being part of the last layer of the model in a binary and a multiclass problem respectively.

1. **ReLU**

   Rectified Linear Unit (ReLU) is defined by equation 5.3, demonstrated in Figure 5.2 [128].

$$f(x) = max(0,x) \tag{5.3}$$

This function returns zero for negative inputs but then linearly increases for positive values, being not saturable and highly computationally efficient. However, a problem emerges in the presence of a zero value because the gradient cannot be computed.

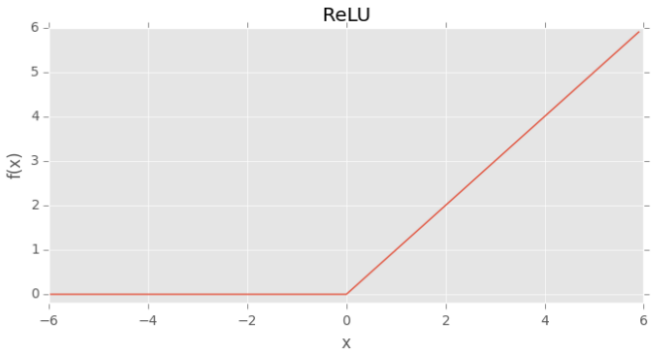


Figure 5.2: ReLU activation function (from [128]).

2. **Variants of ReLU**

A variant of ReLUs called Leaky ReLUs, Figure 5.3a, allow for a slight, non-zero gradient when the unit is not active, permitted to take on a negative value contrarily to the basic ReLU, equation 5.4.

$$f(x) = \begin{cases} x & \text{if } x \geq 0 \\ \alpha * x & \text{otherwise} \end{cases} \tag{5.4}$$

Parametric ReLUs, build on Leaky ReLUs, permits the parameter α to be learned on an activation-by-activation basis, indicating that each node in the network can learn a "leakage coefficient" different than the others.

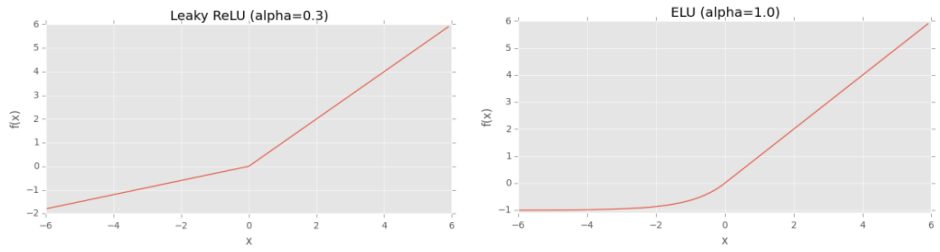
In 2015 arose the Exponential Linear Units (ELUs), Figure 5.3b, where the value of α contrarily to PReLU is constant and set when the network architecture is instantiated [129]. ELU with $\alpha > 0$ is is defined as,

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha * (e^{x_i} - 1) & \text{otherwise} \end{cases} \tag{5.5}$$

Through the work, ELUs rarely ever perform worse than the standard ReLU function [128].

3. **Sigmoid**

Sigmoid is a non-linear activation function that can be used to threshold predictions. It is a good choice since it is continuous and differentiable in its fullness. The most significant disadvantages are that it is not zero-centered, as visible in Figure 5.4, and saturated neurons



(a) Leaky ReLU function. (b) ELU function.

Figure 5.3: Variants of ReLU function (from [128]).

will kill the gradient due to the gradient delta be minimal. It is widely used in binary classification. Sigmoid function is defined by equation 5.6.

$$f(x_i) = \frac{1}{1+e^{-x_i}} \tag{5.6}$$

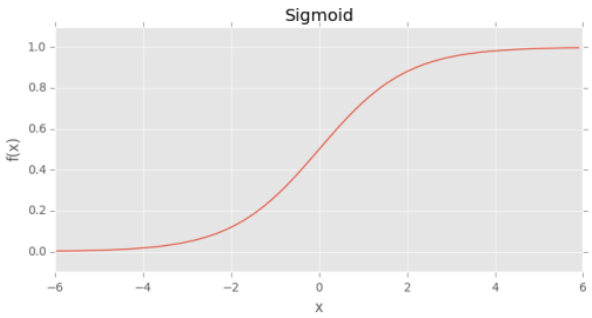


Figure 5.4: Sigmoid activation function (from [128]).

4. **Softmax**

   Although softmax is considered a non-linear activation function, it is the one that presents the most different characteristics due to it not only maps the output to a value within the range [0,1] but also maps each output in such a way that the total sum of predictions is equal to 1. Thereupon, it is extensively adopted in multiclass problems. Softmax is a generalization of the binary form of Logistic Regression.

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=0}^{k} e^{x_j}} \tag{5.7}$$

   The softmax function 5.7, gives the probabilities for each class label being easier to user interpretability.

5.2.1.3 Pooling Layers

The pooling layer is another way to reduce the input volume size as the stride of convolutional layers. Beyond the dimension reduction, it will reduce the number of parameters and consequently

help to control overfitting. It can be adopted max or average function in this layer. From a window with a size previously defined, it is retrieved the maximum value or the average value from the input feature map [128].

5.2.1.4 Fully Connected Layers

The fully connected layers are always located at the end of the model. The neurons of these layers are entirely connected to all activations in the previous layer in the case of a feedforward neural network. It is usually implemented one or two of these layers before the softmax or other classifier that computes the final output probabilities [128].

5.2.1.5 Batch Normalization Layer

As the name implies, Batch Normalization is used to normalize the activations of a particular input volume before forwarding it to the following layer in the network. This layer changes the values of the features to have zero mean and unit variance for each mini-batch during training. In addition, avoiding overfitting, high effectiveness is also shown at reducing the number of training epochs, allowing a wider variety of learning rates, even helping to stabilize the loss curve [130].

5.2.2 Regularization

Overfitting refers to a model that performs well on training data but not on unknown input data. Due to the enormous number of free parameters in a conventional CNN, it is prone to overfitting. A sufficiently large CNN may learn to memorize the training data instead of learning appropriate rules that generalize to unseen data. Consequently, several regularization methods have been developed to prevent the CNN from overfitting, including L1 regularization, L2 regularization (commonly called "weight decay"), and Elastic Net, which are used by updating the loss function itself adding a parameter to constrain the capacity of the model.

The dropout technique is considered a regularization technique, which can be explicitly added to the network architecture. Dropout layers randomly eliminate some connections between the preceding layer and the following one for each mini-batch in the training set, considering probability p to dictate the number of disconnections. In this way, the network will be constantly altered during the train, ensuring that no single node is responsible for activating a given pattern [128].

Beyond these, there are other implicit forms of regularization, like the early stopping, which will control the loss evolution. If it is not decreasing for a stipulated number of epochs, the network will end the training.

5.2.3 Learning

The learning capability is a crucial function in a neural network. Learning is achieved through an iterative process, where multiple adjustments are made by different weights applied, being concluded when the network reaches a generalized solution for an optimization class problem. It

can be deduced that the network intelligence is saved in the weights and biases of the units defined through the training process.

Several learning algorithms can be implemented, varying essentially in the way followed to modify the hidden layer weights and the time that modification is made. It is crucial for the network's training process to know how these variables were initialized, the training algorithm employed, and the training time. A good choice of the initial values of the network weights can decrease the time necessary for training. Typically, the initial values of the network weights are random numbers evenly distributed over a defined interval. The wrong choice of these weights can lead to premature saturation [131].

The multiple network options can be divided between two basic types of network, feedback and feedforward networks. On the one hand, the feedback implies a trace of the output values back to the input values. On the other hand, the feedforward intends no feedback having only a direct flow of information. For each input vector established in the network, an output vector is calculated and read from the output neurons.

The feedforward networks are the most used. The backpropagation method is widely adopted in this type of network, based on the error signals propagating across the network from one layer to the next until the output layer. This method will adjust the weights appropriately according to the error function with differentiable activation units to learn the training set. The most common rule used in this process is gradient descent. The main goal is to decrease the value of the loss function, minimizing the quadratic error calculated in the output through the actual output values and the expected ones [132].

The backpropagation algorithm evolves four main steps:
1 - Weights initialization;
2 - Feedforward procedure;
3 - Backpropagation errors;
4 - Weights and bias update.

Even though these methods can be very effective, there is a possibility of premature convergence in local minimums, requiring a more extensive computational effort and a reliance on initial positioning for improved performance. To overcome this disadvantage, some algorithms privilege a local and global search.

Two common problems that could also occur training deep networks with the backpropagation algorithm are the vanishing and exploding gradients. In the first case, the term goes to zero exponentially fast, making it difficult to learn since the units of initial layers do not get helpful feedback on how to adjust their weights, starting to vanishing. In the second case, the term goes to infinity exponentially fast, leading to an unstable update process. This case generally happens in the deeper layers due to the direct influence on the output result [50].

5.2.3.1 Loss Function

A loss function is a way to compute the CNN's performance, or prediction, according to the ground truth and thereby to quantify how well the CNN is doing. The loss is specified in general

for one sample of the dataset, and the CNN's weights are calibrated during the learning process to decrease the mean of the losses.

A usually employed loss function for classification tasks is the cross-entropy loss. A CNN estimates the probabilities of input x_i belonging to each class with, for example, a softmax activation function as the last layer, in which outputs are the probabilities of input x_i belonging to each class. The cross-entropy loss can be defined as

$$CE(p_t) = -log(p_t) \tag{5.8}$$

Where p_t is the estimated probability that the input belongs to the ground truth class Y_i, the cross-entropy loss is named after the fact that it minimizes the cross-entropy between the ground truth label and the label distribution produced by the model.

More recently, in 2017, the Authors of [133] come with a focal loss function to overcome the problem of class imbalance by reshaping the conventional cross-entropy loss such that it down-weights the loss assigned to well-classified examples. The novel focal loss focuses training on a sparse set of complex examples and preventing the detector from being overwhelmed by a large number of simple negatives during this task. It is added a modulating factor $(1-p_t)^{\gamma}$ to the cross-entropy loss with a tunable focusing parameter $\gamma \geq 0$, where p_t is the probability for the ground truth class.

$$FL(p_t) = -(1-p_t)^{\gamma} log(p_t) \tag{5.9}$$

The focusing parameter γ will smoothly adjust the rate at which simpler examples are down-weighted. When $\gamma = 0$, the focal loss function is equivalent to cross-entropy. The Authors found that $\gamma = 2$ was achieved the best results in their experiments.

5.2.3.2 Loss Minimization

As previously stated, the learning stage is accomplished by minimizing a fixed loss function over a given dataset. Local optimization approaches are often used because there is seldom a closed-form solution to the learning problem. The parameters of the CNN are most frequently modified using a version of gradient descent.

An individual loss evaluation and weight update are barely inefficient, i.e., an adoption of stochastic gradient descent. Nevertheless, an evaluation and update just at the end of the train set is also not an optimal option. Thereby, it is used the mini-batch with a size compressed between 1 (stochastic gradient descent) and the samples of the train set. Therefore, the weights will be updated at the end of each mini-batch train, following the equation 5.10.

$$\theta_{i+1} = \theta_i - \eta \sum_{i \in B} \nabla_{\theta} L_i(y_i, f(x_i, \theta)) \tag{5.10}$$

Where B is the batch, a subset of the complete dataset. η defines the learning rate, and ∇, the gradient of the loss function (L) between the desired output, y, and the network output.

There are several variants of the update rule, all of which are intended to reduce noise in the projected gradient and speed up convergence. Gradient descent with momentum, Nesterov momentum, AdaGrad, and Adam are some instances. However, these are first-order methods, which only enable the gradient to be calculated concerning the weights.

Adam, Adaptative Moment Estimation, is an optimization algorithm that arose in 2014 by an improvement of RMSprop (Root Mean Square prop) with momentum added. It is one of the DL optimization methods more applied in computer vision systems, allowing a faster convergence network. Nonetheless, without a good hyperparameters tune to a given optimizer (and associated model), the network could never achieve acceptable accuracy [128].

5.2.4 Hyperparameters

Hyperparameters can be divided into two types, those that determine the structure of the model and those that define the network training process. Within the first group for a CNN model stand out:

- kernel size - defines the size of the filter in a convolutional layer;
- stride - the matrix step within the image;
- padding - add zero layers to make sure that all the image information is collected, including the edges;
- hidden layers;
- activation functions.

Regarding the training process:

- learning rate - regulates the update of the weights at the end of each batch;
- momentum - adjusts the value to allow the preceding update to influence the current weight update;
- number of epochs - the number of iterations that are made to the complete training data;
- batch size - number of patterns presented to the system before update weights.

5.3 Dataset

The experiment data used in the preparation of this work is obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [134], the dataset most reported in the state-of-the-art. It is a global research collection that began in 2004 and continuously assists investigation and development of treatments to stop or slow AD progression with thousands of subjects' data. The ADNI researchers gather, verify and utilize data from MRI and PET imaging, genetics, cognitive

tests, CSF, and blood biomarkers to provide information that can be used to predict the disease. The data provided is verified and segregated into different stages. Depending on the subset, it can be divided into the three main groups, CN, MCI, and AD, or also specifying the middle one into LMCI and EMCI, or joining SMC (Significant Memory Concern) to address the gap between healthy elderly controls and MCI with subjective memory concerns.

The MRI data obtained through 1.5T and 3T scanners from General Electric (GE) Healthcare, Philips Medical Systems, and Siemens Medical Solutions result in T1 images. Considering that machines routinely undergo upgrades, the quality along time is significantly better [134]. Currently, most MRI studies are conducted at 1.5 T, despite studies using 3T scanners are reaching. Nevertheless, using different MRI field strengths did not significantly reduce the discriminative power and can be accepted for further clinical studies [135].

However, all the available images undergo quality control, concluding that T1 MPRAGE files are considered the best choice. The T1 MPRAGES images provided by ADNI are subjected to different preprocessing tools such as gradwarp, a system-specific correction of image geometry distortion caused by gradient non-linearity varying depending on the gradient model. Additionally, B1 correction, to fix non-uniformity intensity consequence of equipment variation, happens when RF transmission is made with a more uniform body coil while the reception is completed with a less consistent head coil. Besides these two corrections, it is also applied the N3 algorithm to decrease the residual intensity non-uniformity due to wave or dielectric effects [134].

In addition to imaging data, a CSV file is provided with demographic data of the subjects like acquisition date, age, sex, and the corresponding disease stage group, beyond the technical information of image and corrections applied, as visible in Table 5.1.

Table 5.1: Example of subject information provided in a CSV file by ADNI dataset.

Image Data ID	Subject	Group	Sex	Age	Visit	Modality	Description	Type	Acq Date	Format
I30968	023_S_0058	CN	M	70	2	MPR	GradWarp, B1 correction, N3, Scaled	Processed	12/12/2005	NiFTI
I31107	023_S_0061	CN	F	77	2	MPR	GradWarp, B1 correction, N3, Scaled	Processed	12/20/2005	NiFTI
I31392	023_S_0376	MCI	M	71	2	MPR	GradWarp, B1 correction, N3, Scaled	Processed	4/28/2006	NiFTI
I31446	023_S_0388	MCI	M	71	2	MPR	GradWarp, B1 correction, N3, Scaled	Processed	05/04/2006	NiFTI
I31468	023_S_0604	MCI	M	87	2	MPR-R	GradWarp, B1 correction, N3, Scaled	Processed	6/20/2006	NiFTI
I31500	023_S_0625	MCI	M	76	2	MPR	GradWarp, B1 correction, N3, Scaled	Processed	07/12/2006	NiFTI
I31540	023_S_0916	AD	M	80	2	MPR	GradWarp, B1 correction, N3, Scaled	Processed	9/29/2006	NiFTI
I31623	023_S_0030	MCI	F	80	2	MPR	GradWarp, B1 correction, N3, Scaled	Processed	10/26/2005	NiFTI
I32755	005_S_0553	CN	M	85	2	MPR	GradWarp, B1 correction, N3, Scaled	Processed	6/28/2006	NiFTI
I32777	005_S_0324	MCI	F	75	2	MPR-R	GradWarp, B1 correction, N3, Scaled	Processed	4/21/2006	NiFTI

In this work, the only images analyzed are T1 MPRAGE due to the high level of quality and the fact that most papers in the state-of-the-art report this type as the selected one with good visibility of grey and white matter. The imaging data is saved in a 3D file acquired through multiple cerebral slices in the axial plane by the MRI equipment according to a specific thickness, allowing a better visualization from different views to evaluate the brain and dementia lesions.

5.3.1 NIFTI file

The file format NIFTI (Neuroimaging Informatics Technology Initiative) is an evolution of the widely used format ANALYZE. Those together with Dicom and Minc are the most commonly

used types of files in medical imaging. These images are 3-dimensional arrays of the entire brain and contain the object information in the metadata. NIFTI overperforms ANALYZE in multiple points, such as the supply of spatial location, including the distinction of right and left sides to avoid ambiguity in brain studies and the affine coordinate definitions relating to the voxel index. Besides that, it has codes to indicate data "meaning" as a dictionary providing the information in a dual file, header(".hdr") and image (".img") or in single file storage (".nii") where all the data is combined.

Therefore, the NIFTI file overcame the use of its predecessor and was adopted by influential software used in neuroimaging processing like AFNI, FSL, and SPM. Those were committed to supporting this format for input and output, fostering the interoperability at the file-exchange level between data analysis [136].

5.4 Computational Infrastructure and Implemented Resources

5.4.1 Computational Infrastructure

In the development of this project, a computer ASUS VivoBook S15 with processor i7-8565 U CPU at 1.80 GHz and 16 GB RAM was used. The 64-bit operating system installed on this PC was Microsoft Windows 10 Home 64-bit, compatible with almost all computational tools utilized for the preprocessing stage. Besides that, the training and evaluation process of the DL network was done in a Linux machine based on a Graphics Processing Unit (GPU) Tesla V100-DGXS with 32GB RAM.

5.4.2 Development language

or the development of this project, the python language was chosen. Python is a high-level, interpreted, and general-purpose dynamic programming language that focuses on code readability [137]. The most significant advantages are the versatility of the development platform and the development speed. Also, it is provided extensive support and many libraries, easy to integrate with functions in other languages like C, C++, or Java. It can be used in small projects associated with hardware or big projects based on just software and user interfaces.

In the medical field, python proves to be a helpful tool that can be inserted in multiple steps like medical imaging segmentation, data extraction, statistical analysis, and a portal to machine learning and artificial intelligence.

5.4.3 Anaconda Environment, Spyder and Jupyter Notebook

Therefore, the preprocessing stage was developed on Spyder, which is a free and open-source scientific environment included with Anaconda, the world's most popular open-source package distribution. To the next step, the jupyter notebook, a web-based interactive development environment, was adopted to facilitate access from the personal computer to the Linux machine.

5.4.4 GPU and CUDA

Considering the use of GPU to run code, it was adopted a library created by NVIDIA, CUDA, a parallel computing platform and programming model for general computing on GPUs. By using the capabilities of GPUs, developers may substantially speed up computational applications with CUDA [138].

5.4.5 TensorFlow and Keras

In the application of DL approaches are used the Application Programming Interface (API) Keras to simplify the algorithm. Keras is a deep learning API running on top of the machine learning platform TensorFlow. It was developed to provide essential abstractions and building blocks for rapid experimentation of solving problems. It offers consistent and most uncomplicated functions minimizing the number of parameters requires from the users, providing error messages and the summary of actions made. Keras can be integrated with TensorFlow to develop new and more complex methods [139].

5.4.6 ITK-SNAP

The ITK-SNAP is a software application used to study 3D biomedical images, including structure segmentation through semi-automatic methods [140]. Besides that, the ITK-SNAP software, adopted in version 3.8.0, provides a modern graphical user interface supporting different 3D image formats and presenting a 3D display with a dynamic control multi-view, as visible in Figure 5.5.

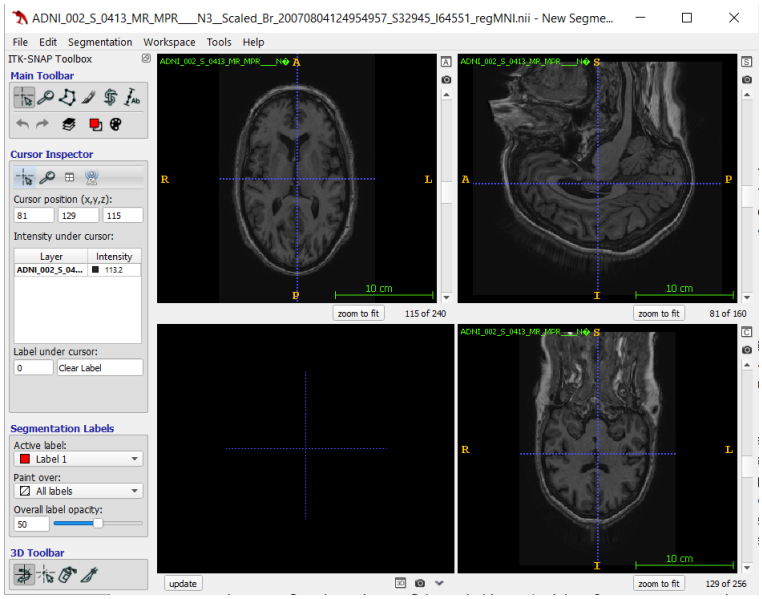


Figure 5.5: ITK-SNAP interface.

5.5 Summary

This chapter presents an overview of the pipeline implemented in the project highlighting three steps: preprocessing, learning process, and the final classification. For the final task of classification, multiple CNN's were implemented managing architectures and hyperparameters.

The dataset adopted also influences the image processing through the network. ADNI presented in advance some essential preprocessing steps such as bias correction and N3 algorithm, which are widely applied in MR images for the Alzheimer's disease study.

The use of Python in the project development allowed the use of libraries specialized to DL assignments extensively applied in this field with updated approaches.

Chapter 6

Experimental Work and Results

This chapter presents the obtained results during the implementation of the adopted methodology. In a first instance, the output from the preprocessing stages is described and displayed the output, followed by the exposition and analysis of the achieved classification results through different approaches.

6.1 Preprocessing

As stated before, preprocessing is an essential step to improve the results obtained in the followed steps, being commonly required a manual or automatic correction depending on a previous study in order to deal with the main characteristics of the studied images.

To process NIFTI files, it is necessary to adopt a specific library, NiBabel, which focuses on neuroimaging files in the read/write operation [141]. In the opening process of the files, this library takes into account the image format to access all possible information, which includes the header and the image data that becomes available through NumPy arrays. In turn, the Numpy library [142] is a powerful and recognized fundamental package for scientific computing in python, which allows performing algorithms with multi-dimensional arrays efficiently of a diversified data type and contains from simpler tools to complex mechanisms of linear algebra.

6.1.1 Skull Stripping

In the reduced volume brain, an automatic skull stripping algorithm is applied. Given the studied algorithms and their characteristics, BET is one of the most present in similar projects[143]. In this method, initially, a mask is created using two thresholds estimated from the image histogram. Then, a spherical deformable model is initialized at the gravity center of the mask and expanded towards the brain's surface by locally adaptive forces. It is a fast and versatile method, but intensity and shape variations related to tumor presence or other anomalies can negatively influence the evolution of the mesh.

Another skull stripping algorithm is presented by the DeepBrain library, which is based on an Unet approach [144], Figure 6.1. This algorithm follows a CNN architecture for fast and precise

segmentation of images. There are two main parts of this network. The first is the contraction or encoder, which extracts global features consisting of convolution block, max pooling and dropout. The second part is the expansion layer or decoder consists of expansion blocks in the same number as the contraction blocks. Each step consists of a transposed convolution, a transformation going in the opposite direction of a regular convolution, followed by concatenation with the respective cropped feature map from the contraction layer. This action would ensure that the features learned while contracting the input will reconstruct the image. It was also applied a dropout technique in each block. After that, the number of feature maps of a convolutional layer gets half to maintain the symmetry resulting in one single output.

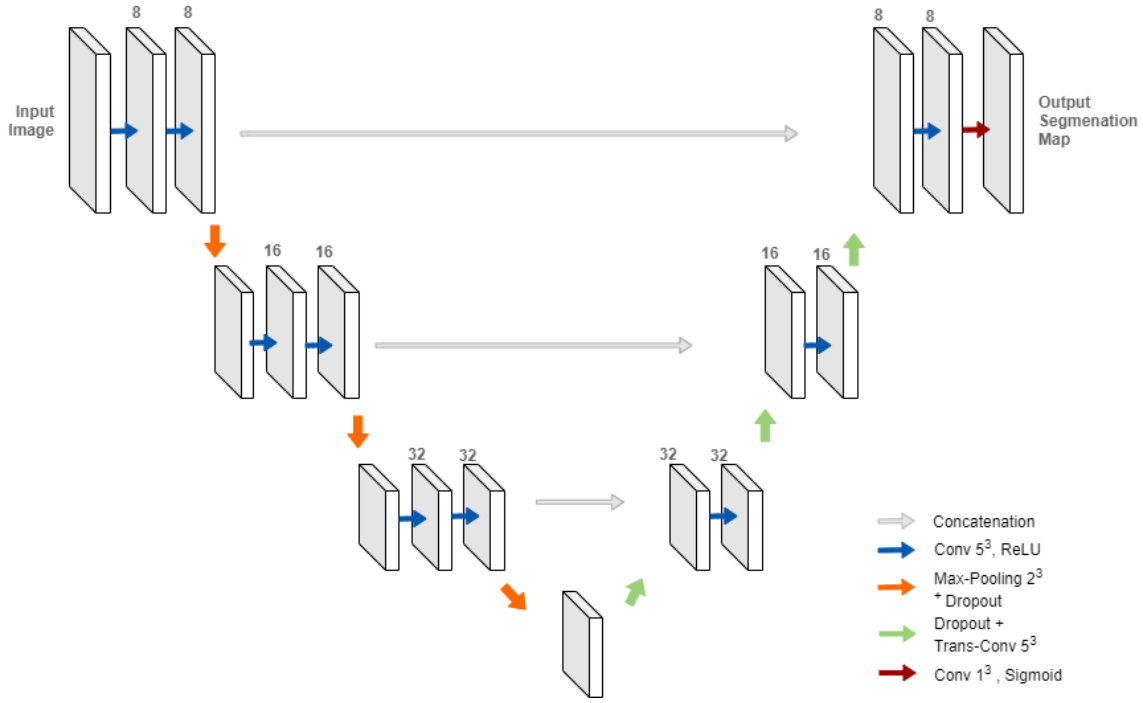


Figure 6.1: DeepBrain library architecture for skull stripping (based on [144]).

The network training process involves a subset of entries from 3 different datasets, CC359, NFBS and ADNI. Considering that network was trained with more than five thousand images from multiple databases to improve the data variety and joining the fact that was also applied data augmentation in numerous rotations and orientations, the learning process was improved and the probability of a successful extraction regardless of the input brain orientation will be more significative.

The skull-stripped output is obtained through a threshold mask. The value can be tuned, being the more acceptable the 0.5. An example of this process is presented in Figure 6.2.

Instances of BET and DeepBrain results are shown in Figure 6.3, where the same input image is used. It is visible that in the upper brain, there is a mass gap produced by the BET algorithm extraction, while DeepBrain presented a satisfactory cut in that region, being distinguishable the white and the gray matter. Therefore, it can be concluded that the last one revealed a better brain

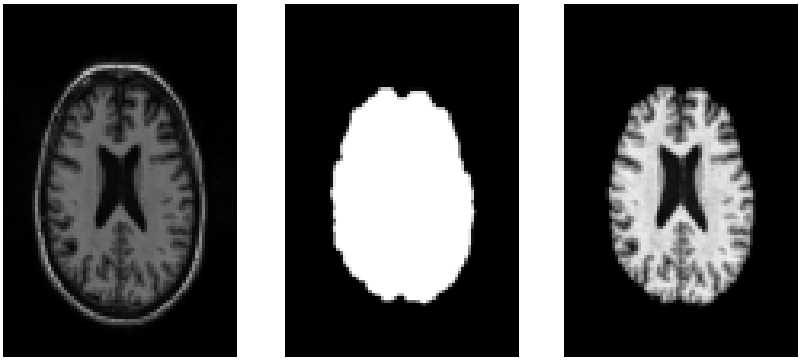
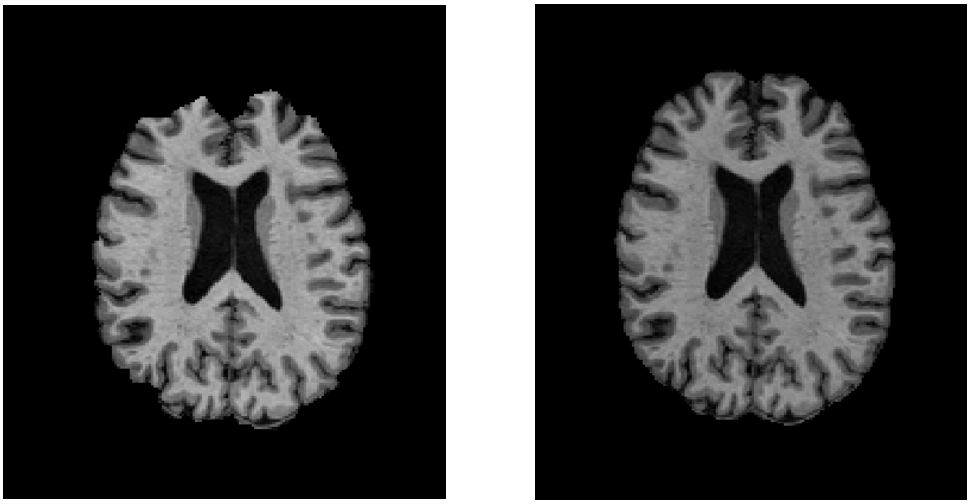


Figure 6.2: Skull stripping process from the original image to the obtained output.

extraction being the one selected to implement in this preprocessing stage.



(a) Bet result. (b) DeepBrain result.

Figure 6.3: Comparison between the results obtained through two different skull stripping algorithms.

6.1.2 Registration

Image registration is the process of converting images into a standard coordinate system in such a matter that analogous pixels represent homologous biological points. To implement this preprocessing stage was used SimpleElastix, an extension of SimpleITK [145] that offers a user-friendly API to the popular image registration algorithms of the Elastix library [146]. This makes state-of-the-art medical image registration easy to do in multiple languages, including python.

The Elastix was developed in the ITK library by the Institute of University Medical Center Utrecht (Netherlands) and consists of software for medical image registration through pixel intensity. It is based on a modular conception that includes different optimization criteria, interpolation methods, transformation models, and similarity criterias. Therefore, it allows the user to compare

different alignment methodologies to select the most suitable configuration for a specific application.

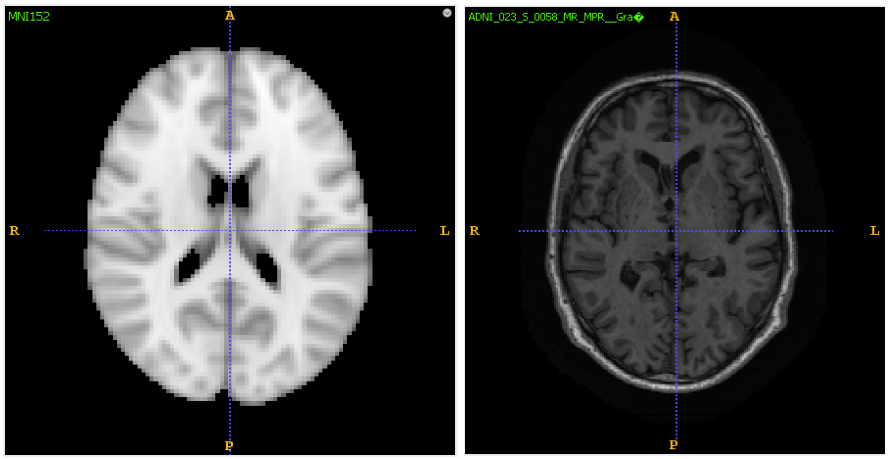
This simplified version allows the user to register the image with a couple of lines of code [147]. The brain MRI images were linearly registered to the target image, undergoing three different transformations based on translation, a 9-parameter affine transformation followed by a non-rigid transformation using the B-Spline to overcome the inter-subject anatomical differences in shape, size and relative orientation.

A 9 parameter affine allows rotation, scaling, and translation in the three different axes. It is typically an excellent choice to initialize a non-rigid transformation like the B-Spline. The last can align images when correspondence is impossible without localized deformations, and it can adequately adjust anatomical, physiological, and pathological diversity between subjects [147].

In this work, for comparison purposes, the images went through 2 different registrations. The first one had the MNI-152 atlas as the fixe object. The MNI-152 template was created in 2001 from 3D brain MRI images of 152 normal subjects. In this brain template, the reference space is based on Talairach and Tournoux coordinate system [148]. Due to the smaller atlas shape, the resulting image dimensions after the registration will be reduced.

In a second way, the registration is done intersubject, i.e., it is chosen by manual criteria utilizing the ITK-SNAP an image with lower visible misalignment to serve as fixe object, in the time that others are aligned using this as reference.

Figure 6.4 presents a central axial slice of each fixed object used. Considering the characteristics of the images, the step registration and skull stripping will be done in a contrariwise order.



(a) MNI-152 atlas. (b) Reference patient scan.

Figure 6.4: Referencial images adopted in the registration step.

Beyond the resulting image, the function used will produce a report based on parameter maps, a collection of key-value pairs that atomically defines the registration components, and any required settings for each process.

Therefore, at the end of this process, it will have two different pathways with images of dimensions (91,109,91) and (256, 256, 180). For the second dimension, the volumes are cut, eliminating

the slices in the extremes of the axes containing useless information to decrease the data to process being the dimension reduced to (112, 176, 120).

6.2 3D Network Implementation

Following the preprocessing steps and aiming to perform the classification of MRI T1-weighted brain images into AD, MCI, and NC classes, a DL algorithm was chosen based on the literature review. As previously referred, the CNN algorithm is the most used in image processing problems. Therefore, it was chosen to accomplish the final task of this project.

For the 3D implementation was used the resulting volume of preprocessing stage considering the respective labels provided by the ADNI dataset having, in the end, a complete set to training, validation, and test phase. A python function called train-test-split is used to perform this division, whereby the data set is divided without a choice subject to human bias. Furthermore, to guarantee that different networks are trained, validated, and tested with the same subsets, a parameter defined as the random number will perform the same division while the number remains the same.

The training set refers to the set of images and labels used in the training stage being the bigger one. At the same time, the validation and the test sets are smaller collections that serve to evaluate the network during the training process and to perform the final test with unseen images, respectively. The split was done with 10% for test data and the rest for the training set, where 15% of the 90% corresponded to the validation set.

Design a CNN architecture from scratch is not the most straightforward way to perform the classification generally for beginners in the field. Thereby, it was done a subset choice of papers within all from the literature review. Considering a 3D approach, the article [90] network was implemented as the baseline architecture due to the good results reported, and also providing its description, being presented in Figure 6.5.

The architecture displayed in Figure 6.5 using the dataset chosen did not obtain acceptable results. Thereby, numerous tests related to the rearrangement of the network were carried out, where was managed:

- Number of convolutional layers in each block;
- Number of blocks;
- Number of filters of each convolutional layer;
- Presence and position of batch normalization and pooling layers;
- Number of FC layers;
- Number of units in each FC layer;
- Presence and probability value of dropout layer.

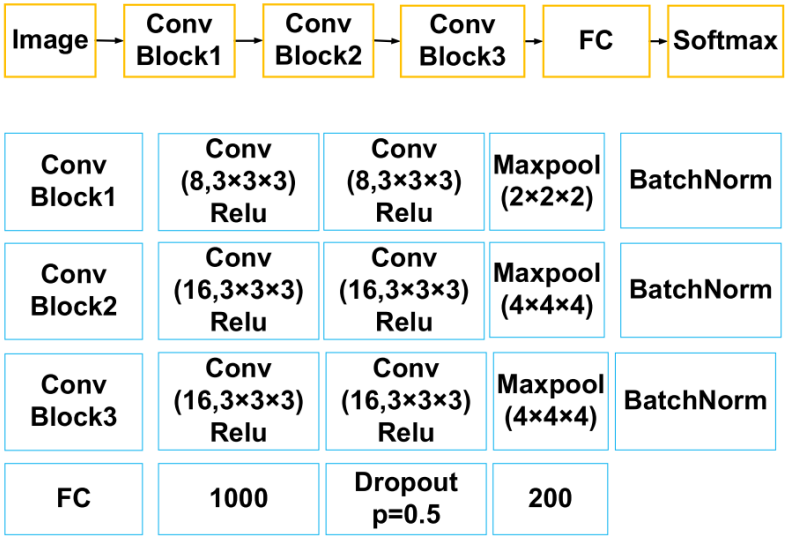


Figure 6.5: Structure of the model implemented by [90]. Where, Conv: convolution layer; Max-pool: Max-pooling layer; FC: Fully-connected layer (from [90]).

Two different experiments were performed to evaluate how the process was influenced considering two different input image dimensions.

Starting with the lower dimension, more than 80 tests were made with different architectures until was reached an acceptable result to operate in further analyses. Therefore, in Figure 6.6 is presented the architecture selected to work.

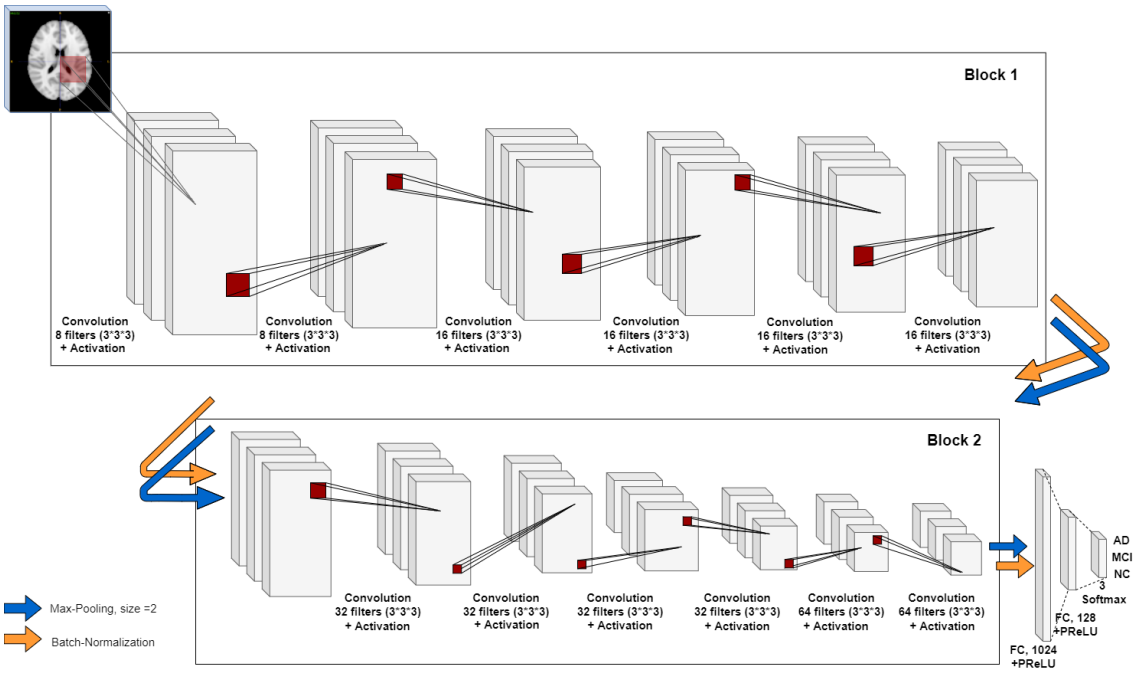


Figure 6.6: Proposed 3D algorithm architecture for ternary classification.

Reckoning the final structure were chosen three different activation functions to assess how each influences the learning process, PReLU, ReLU, and ELU. Besides that, the learning rate was also managed to allow the model to reach the best performance. In the experiments, an upper limit of 200 was set for the epochs. Still, it was also implemented an early stop of 15, i. e., if it elapses 15 epochs without a decrease of validation loss, the network stops training.

The optimizer selected was Adam and remains the same in all the experimental work due to the characteristics previously presented, such as the capability to adapt the learning rate during the training, computing different learning rates for different parameters through gradient moments. That is, the initialization started slowly and pick up speed over time, leading to faster convergence. Moreover, that is the reason to be an almost universal choice in classifiers for this task in state-of-the-art.

Tables 6.1 and 6.2 introduce the obtained results in the multiple experiments regarding [91,109,91] dimensions.

Table 6.1: Obtained results for input images of size [91,109,91] and a learning rate equals to 0.0001.

lr=0,0001	**Accuracy**	**Precision**	**Recall**	**F1-score**
RELU & cross-entropy	71,25%	71,39%	72,07%	71,25%
RELU & focal loss	65,00%	64,83%	65,16%	64,88%
PRELU & cross-entropy	80,00%	80,31%	80,47%	80,27%
PRELU & focal loss	63,75%	65,49%	64,21%	63,90%
ELU & cross-entropy	**85,00%**	**86,28%**	**85,68%**	**84,72%**
ELU & focal loss	36,25%	24,47%	33,70%	28,14%

Table 6.2: Obtained results for input images of size [91,109,91] and a learning rate equals to 0.00001.

lr=0,00001	**Accuracy**	**Precision**	**Recall**	**F1-score**
RELU & cross-entropy	77,50%	77,77%	78,00%	77,38%
RELU & focal loss	72,50%	72,50%	73,00%	72,63%
PRELU & cross-entropy	77,50%	77,66%	78,00%	77,42%
PRELU & focal loss	83,75%	83,59%	84,02%	83,52%
ELU & cross-entropy	**88,75%**	**89,40%**	**89,63%**	**88,75%**
ELU & focal loss	80,00%	80,05%	80,43%	79,72%

Focusing on these tables provided by the atlas registration tests, where the best value was obtained, there are some essential points to discuss. Starting with the hyperparameter learning rate, within the performed tests, an increase of accuracy is visible using a lower learning rate wherein all of the changes that were conducted. There was an improvement between 3 to 20%.

Besides, the choice of the activation function applied in the convolutional layer was a network feature deeply study in this work due to the lack of information and comparison of this point in

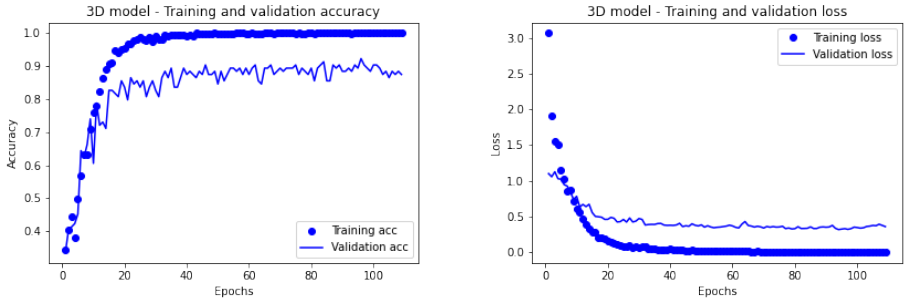
the state-of-the-art being in almost all papers the ReLU activation chosen without given a further reason. Therefore, this work presented an observation between three known functions.

As previously stated, ReLU, the baseline and the widely apprehended activation function has some disadvantages, like the created barrier for negative values where they are annulled. This feature hindered the advance of the gradient descent algorithm, leading in some cases to the death neuron problem. Leaving from there, the Leaky-ReLU allowed a light leakage that could manage some negative values near zero but not precisely, allowing the existence of a bit of gradient avoiding that a neuron never been activated. The PReLU similar to Leaky ReLU but with a new parameter, the learning parameter, that allows different neurons can have different values. Afterward, ELU arose, merging the good features of ReLU and Leaky ReLU.

Therefore, ReLU achieved the lower performance between the three applications followed by PReLU, and the best one was ELU for the two different learning rates using the cross-function. However, with atlas registration using a focal loss, ELU did not accomplish the learning task being reflected in the test results.

The focal loss is applied when facing unbalanced problems, as is the case. This cost function produced a lower loss in all the sets despite that the accuracy did not follow with a positive increase where the cross-entropy function achieved high results.

The best result was obtained by combining the ELU function and a lower learning rate (0.00001). Figures 6.7a and 6.7b demonstrate the training evolution of accuracy and loss, respectively.



(a) Accuracy evolution graph during network training.

(b) Loss evolution graph during network training.

Figure 6.7: Training network report for the best result obtained.

The learning curves are a powerful diagnostic tool in this area and can be used to diagnose the learning and generalization behavior of machine learning models. During the training, in each epoch, the metrics values are saved and display in the graph. On the one hand, with the training line is visible how well the model is learning. On the other hand, the validation learning curve calculated from a hold-out validation dataset gives an idea of how well the model is generalizing. Through the loss plot, it is visible that both curves decrease to a point of stability, presenting a small gap between them, the "generalization gap". From both graphs, it can be concluded that the model performs a good learning behavior without evidence of underfitting or overfitting.

Figure 6.8 shows the confusion matrix for the test set, which reveals a good final classification result.

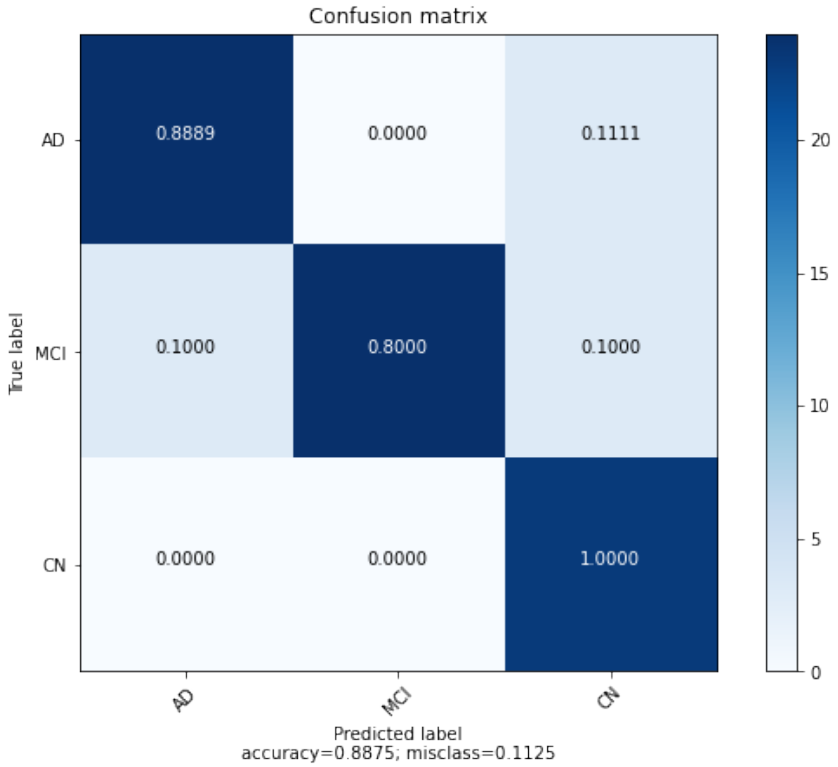


Figure 6.8: Confusion matrix obtained through the best test.

The healthy class has achieved an accuracy of 1, where all the samples were well defined. At the other end, the AD class reveals some misclassed samples, which can be explained through the reduced visible disease pieces of evidence since the only feature that MRI assesses is the structural change exposed by matter atrophy. The middle class achieved a lower accuracy, as expected, due to some subjects present, like the AD case, not solid structural evidence, consequently, it could be inserted in other classes. MCI is also the intermediary class, and thenceforth depending on the disease evolution of each patient, the network can give a different output considering that evolution is not equal for all who have dementia.

Contrarily to ReLU, ELU allowed the passage of information from shallow layers that ReLU would let die since not having a high gradient compared to deeper layers, hence the initial information started to get lost.

For the second dimension, Table 6.3 presents the obtained results.

For internal registered images, using a learning rate of 0.0001 were obtained values significantly lower and therefore are not shown in Table 6.3. Moreover, when applied the focal loss occurred an increase of loss value and a premature stabilization in the learning process, making it impracticable for the network to get enough knowledge to perform the task. This happened in the presence of a higher learning rate, not allowing Adam optimizer to manage the value after the

Table 6.3: Obtained results for input images of size [112, 176, 120] and a learning rate equals to 0.00001.

lr=0,00001	Accuracy	Precision	Recall	F1-score
RELU & cross-entropy	81,25%	81,32%	82,25%	81,36%
RELU & focal loss	73,75%	74,49%	74,02%	74,17%
PRELU & cross-entropy	**86,25%**	**86,22%**	**86,48%**	**86,26%**
PRELU & focal loss	77,50%	77,47%	77,91%	77,55%
ELU & cross-entropy	77,50%	77,73%	78,58%	77,81%
ELU & focal loss	82,50%	82,64%	83,15%	82,83%

initialization. In this way, the gradient will reach outsider values complicating the return to the right path to achieve the minimum due to the focus on the complex and more outside samples by the adopted cost function, occurring vanishing and exploding gradient problems.

To the learning rate presented in Table 6.3, as the tests operated with lower dimensions, the ELU function also reached the best results, emphasizing that ELU should be taken with more consideration in this sort of problems.

Comparing the performance values obtained from 3D experiments, both achieved accuracies within 60% and 90%, out of one exception. Considering the computational cost imposed by the dimensions [172,120,176], this approach did not payoff, even though the network with some specific characteristics outperformed the one with a lower dimension input.

When is completed a ternary classification, the network does not reach an excellent performance due to a considerable data variability to analyze, which can induce error in the learning process and result in misclassified samples. Besides that, the vast number of learning parameters that the network needs to optimize does not facilitate the achievement. However, it showed promising results being within the values present in the literature, as Table 6.4 show, even the application of data augmentation or transfer learning on some reported papers.

Table 6.4: Classification performance comparison with previous researches based on CNNs using sMRI for distinguishing the three stages of the disease (AD vs. MCI vs. CN) with 3D data.

Work	Method	Accuracy
[89]	3D-CNN	61,10%
[90]	3D sVoxCNN	95,05%
[93]	ADNet + logistic regression	52,30%
[96]	3D-DenseNet	97,52%
[98]	3D-DenseNet	97,19%
[101]	3D-FCNN	91,32%
[102]	3D-CNN	64,82%
	Proposed Method	88,75%

Data augmentation and transfer learning techniques were not applied in this network because there are fewer implementations in 3D data when compared with 2D slices, even outside the medical field, making the process more difficult.

6.3 2D Network Implementation

Trying to explore a lighter implementation was decided to study a 2D approach.

As well as the 3D implementation, the 2D was started after the preprocessing stage. The network choice fell again on CNN due to the characteristics and advantages of this approach in computer vision.

Considering the aforementioned problem of designing a network from scratch, in the 2D procedure of this work is selected one widely used and known architecture, the VGG19.

VGG-19 is a 19-layer deep CNN. This network was first designed to classify 1000 different objects of the ImageNet Large-Scale Visual Recognition (ILSVRC-2012) dataset to accomplish the challenge equally named. This dataset includes 1.3 million images to train, 50.000 to the validation phase, and 100.000 to test. The authors expressed that the representation depth and the use of a minimal (3x3) convolution filter are advantageous features for classification accuracy [149]. VGG outperforms baselines on a variety of tasks and datasets outside of ImageNet, being one of the most widely used image-recognition models nowadays [128].

The training process of deep architectures like VGG19 is highly challenging due to its depth. Supposing that these architectures were randomly initialized and trained from scratch, the learning process will often not be accomplished, being difficult to gain any initial knowledge to get on the right path. Thereby, a notion known as pre-training arises to train more profound variations of deep networks. Pre-training is the process of training, firstly smaller versions of the network designed with fewer weight layers and subsequent use these converged network weights as initializations for deeper networks [128].

Besides using the pre-training tool to work with deeper networks, these pre-learned weights can be used within the same network to classify different tasks. Thence, it is possible and widely applied in state-of-the-art deep pre-trained networks for 2D image classification in dementia analysis with weights from ImageNet, i.e., from a classification task to distinguish daily objects such as keyboards, pencils, different animals, and others. As a result, the network has learned a variety of rich feature representations for various images.

Beyond VGG19, other pre-trained networks provided by Keras were implemented, such as InceptionV3, VGG16, and DenseNet. Within these networks, the best results were obtained through VGG19, the architecture presented in Figure 6.9, being the focus of the 2D implementation.

Figure 6.9: VGG19 architecture (from [149]).

Before inserting images in CNN, it was required to prepare the 3D provided data to produce 2D RGB images. Regarding the studied articles, the axial view is the most elected to classify the

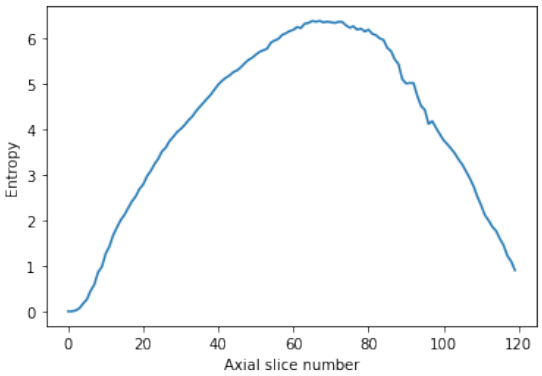
three stages of Alzheimer's disease, consequently also applied in this work due to being the most used and related to recognizing brain diseases.

Having the view selected, the next step was slicing the data volume. The data internally registered was resulted in a total of 120 slices per patient.

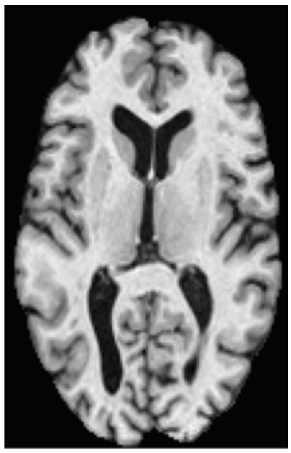
Feeding a network with this amount of parts for each subject could induce errors in the model caused by the variability of slices. Thereupon, a slice selection was performed adapted from [73, 75, 150] where an entropy-based algorithm selection was implemented.

The entropy applied to image analysis is a measure of uncertainty based on the number of possible gray level values in the image and the likelihood of each occurring. Assuming that all intensity values in an image have an equal probability of occurring results in a maximum entropy value because there is complete uncertainty about which intensities could occur. If an intensity value occurs more frequently than the others, the uncertainty decreases since the expectation of that given value is higher. Hence it is considered that an image with a higher entropy value contains more information to be processed [151].

Firstly, a chart of entropy values along the z-axis (axial view) was drawn Figure 6.10a, and a slice example with high entropy value is presented in Figure 6.10b. After, a variable number of slices with the highest values were selected for further steps. Contrarily to [73], where between 8, 16, and 32 slices, the best result was obtained with the higher number, in this work was achieved better results with fewer slices like 7.



(a) Example of an entropy graph obtained.



(b) Axial slice number 62.

Figure 6.10: Entropy in axial view.

The referred architectures to achieve the classification in 2D images were conceived to the same goal, the ILSVR challenge. Thereby the final layer is composed of 1000 units corresponding to the 1000 classes of objects. To adapt the design to this work was required to remove FC layers and to formulate new ones to produce a 3-class output. This process encompassed study and attempts to manage the number of FC layers, the number of units in each one, the activation function adopted, and other techniques applied like the dropout.

Regarding the FC layers, were tested different numbers of layers quantity and units having as a starting point the literature of 2D networks [73, 80, 65, 63], where these values are specified. Concerning the previous tests, the ELU was the selected activation function in the FC layers of this VGG19 before the softmax function to perform the ternary classification.

The results obtained through the network were relative to each slice. To obtain the real subject classification, one must resort to an ensemble technique that combines different predictions to reach a single output. This project adopted majority voting, hard voting, and weighted average probabilities voting or soft voting based on the functions provided by scikit-learn. Regarding the first hypothesis, it is used the predicted class labels for majority rule voting, while in soft voting, the prediction is a result of the maximum value of the predicted probabilities sums for each class. Thereby, the followed Table 6.5 was constructed during the network tests.

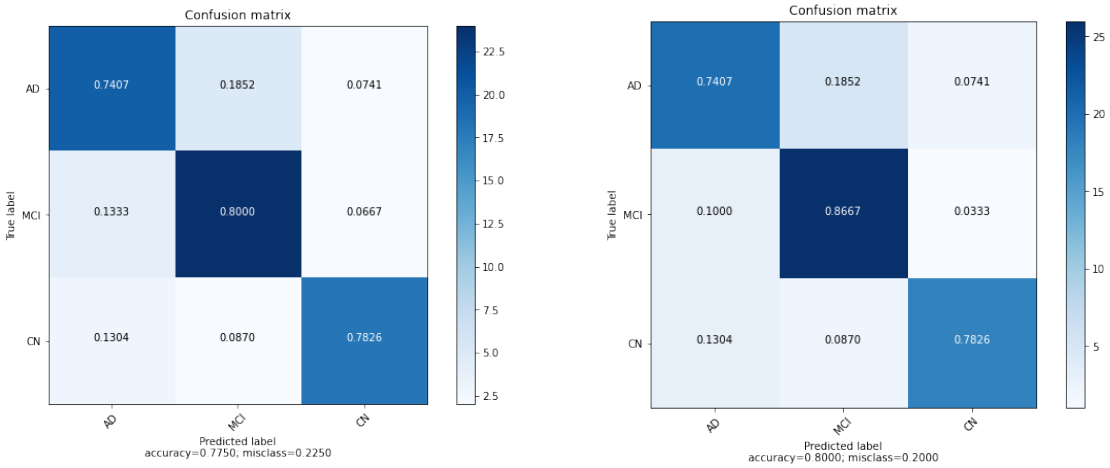
Table 6.5: Results obtained through VGG19 with initialization of ImageNet weights.

		Hard Voting				**Soft Voting**			
	FC units	**Accuracy**	**F1-score**	**Precision**	**Recall**	**Accuracy**	**F1-score**	**Precision**	**Recall**
7 slices	1024-(0,7)-64	71,25%	71,19%	71,53%	71,80%	70%	69,94%	70,47%	70,69%
	1024-(0,7)-128	69,64%	69,84%	70,89%	69,57%	69,64%	69,68%	71,68%	69,24%
	1024-(0,7)-256	70%	69,57%	70,05%	69,55%	70%	69,50%	69,69%	69,55%
	2048-(0,7)-64	71,25%	70,66%	72,15%	70,29%	71,25%	70,84%	72,24%	70,41%
	2048-(0,7)-128	70,00%	68,80%	74,17%	68,29%	71,25%	69,86%	75,52%	69,40%
	2048-(0,7)-256	**77,50%**	**77,59%**	**77,77%**	**77,44%**	**80%**	**79,94%**	**80,48%**	**79,67%**
	4096-(0,7)-64	72,50%	72,67%	73,09%	72,88%	73,75%	73,90%	74,29%	74,33%
	4096-(0,7)-128	75%	75,10%	76,28%	74,76%	72,50%	72,50%	73,56%	72,20%
	4096-(0,7)-256	72,50%	72,12%	75,90%	71,31%	75%	74,84%	78,61%	73,87%
5 slices	1024-(0,7)-64	65%	65,16%	71,32%	64,55%	66,25%	66,28%	72,19%	65,66%
	1024-(0,7)-128	65%	65,25%	66,27%	64,73%	62,50%	62,45%	66,07%	62,02%
	1024-(0,7)-256	75%	75,40%	77,47%	74,55%	73,75%	74,18%	75,74%	73,44%
	2048-(0,7)-64	58,75%	58,72%	61,46%	58,72%	60%	59,24%	62,54%	59,77%
	2048-(0,7)-128	75%	75%	78,97%	73,99%	72,50%	72,45%	75,87%	71,52%
	2048-(0,7)-256	75%	74,64%	77,31%	73,99%	73,75%	73,43%	76,12%	72,76%
	4096-(0,7)-64	70%	69,98%	70,53%	69,86%	70%	70,29%	71,55%	70,07%
	4096-(0,7)-128	67,50%	67,09%	71,10%	66,80%	70%	69,73%	74,57%	69,15%
	4096-(0,7)-256	61,25%	60,39%	61%	60,26%	60%	59,38%	60,91%	59,03%

Preliminary studies evaluated the advantages of the ImageNet weights application and the use of a larger number of slices. Then, it was concluded that the network did not achieve competitive results without weights, i.e., from scratch, presenting lower values in multiple tests and thereby not presented in the Table 6.5, being the displayed results referent to pre-trained networks. Regarding other amounts of slices per subject, the performance results lowered more when using a bigger number of slices and for that reason are also not shown in this table. The main reason was the injection of more data variability in the network, complicating the learning process.

The best combination within the performed tests was then reached with 2 FC layers of 2048 and 256 units and a dropout layer in the middle of 0.7 using seven slices of each subject. The respective confusion matrices for the hard and soft voting for this test are introduced in Figure 6.11.

Comparing the results, the soft voting, which was reached through the sum of probabilities



(a) Confusion matrix for hard voting.

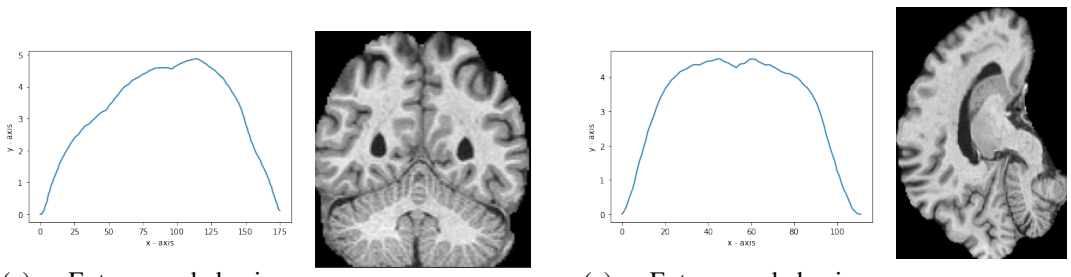
(b) Confusion matrix for soft voting.

Figure 6.11: Obtained matrices for the test set with the 2D implementation.

from the seven slices per class, accomplish the best result. Nevertheless, this is not a general rule since a single slice can result in a high probability and give a clear output of a wrong label leading this ensemble process to be negatively affected.

Even though the axial view is the most applied in these tasks, the coronal and sagittal views were tested to compare results. Wherefore, from the three best test results of 5 and 7 axial slices were implemented the other views assessments.

The slices were selected through the same entropy selection method. An example for each view is presented in Figure 6.12.



(a) Entropy behaviour through coronal view.

(b) Coronal slice number 124.

(c) Entropy behaviour through sagittal view.

(d) Sagittal slice number 62.

Figure 6.12: Entropy in coronal and sagittal views.

Afterward, the most profitable performance for the coronal view was obtained with seven slices and an FC layers set of 2048 and 256 units. On the other side, for the sagittal view, the best outcome was achieved through 2048 and 128 units in FC layers inputting five slices in the VGG19 network. The results for both tests are presented in Table 6.6 with hard and soft voting.

As expected, the coronal and sagittal view experiments were lower than the results obtained

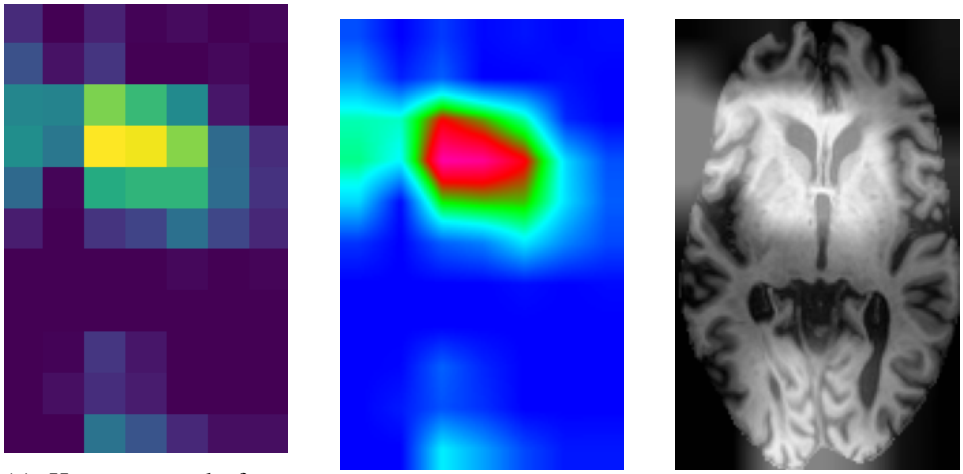
Table 6.6: Best results obtained for coronal and sagittal views.

			Hard Voting				Soft Voting			
	FC units	**n. slices**	**Accuracy**	**F1-score**	**Precision**	**Recall**	**Accuracy**	**F1-score**	**Precision**	**Recall**
Coronal	2048-(0.7)-256	7	71.25%	70.78%	72.51%	71.03%	72.50%	72.05%	73.30%	72.02%
Sagittal	2048-(0.7)-128	5	61.40%	59.99%	68.58%	59.97%	64.91%	63.61%	66.94%	63.68%

with the axial view, although coronal slices achieved performance values that can justify additionally studies.

For further comprehension, a heatmap was created for different axial slices to show the spots that had the most significant attention by the network. To visualize the heatmap was used a technique called Grad-CAM (Gradient Class Activation Map), which is based on taking the gradient concerning the final convolutional layer and then weigh it against the output of this layer to find the importance of a specific class in the model [152].

Intentionally, was chosen a well-classified example, where the prediction matched the ground-truth label, Figure 6.13, while in Figure 6.14, the network prediction was incorrect. In the colormaps presented, the warm color represents the areas that received greater attention from the model. In the brain slice Figure 6.13c and 6.14c, the brighter area corresponds to that attention area.



(a) Heatmap result from the last convolutional layer.

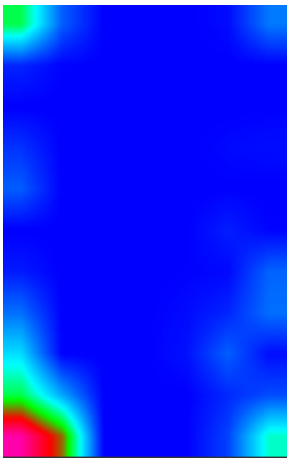
(b) Heatmap in colormap rainbow.

(c) Overlapped image with heatmap.

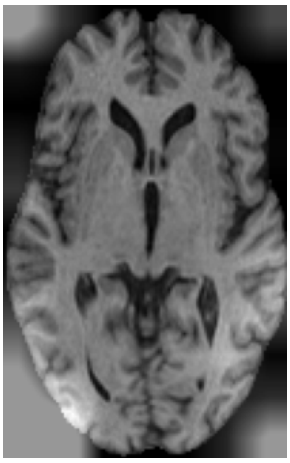
Figure 6.13: Heatmaps and output for a well-classified MCI example.

Therefore, in the misclassified example, the attention is outside the brain area, revealing that the network does not capture relevant features. Contrarily, in the correct example, the attention is around the ventricles area, indicating that this region is essential to classify due to tissue atrophy. This idea is coherent with the existing neuroimaging studies, where it is defended that ventricles and hippocampus areas present significant structural changes verifying and validating the proposed method.

(a) Heatmap result from the last convolutional layer.



(b) Heatmap in colormap rainbow.



(c) Overlapped image with heatmap.

Figure 6.14: Heatmaps and output for a misclassified example. Prediction: Healthy subject; Ground truth: AD.

The 2D implementation was reached lower performance values than the 3D implementation. This aspect was expected as being seen in the literature since information is lost while selecting slices, the relation between contiguous pixels of adjacent layers is unredeemed.

6.4 Summary

Both proposed methodologies were functional. However, the 2D implementation reached lower performance results than the 3D implementation, which can be easily explained by the loss of information between the slices, where the connection is wasted. Regarding the primary approach, section 6.2, was developed and implemented a network that allowed the influence analysis of activation functions. Besides, changing the input dimensions will also produce different metric values. The 3D data with lower dimensions, registered in the MNI-152 atlas, outperformed the data through an internal registration. This can be explained by the reduction of data variability, inducing fewer learning errors.

Chapter 7

Conclusions and Future Work

The final chapter presents the final considerations about the achieved results and the project progress, followed by possible future developments or improvements.

7.1 Final Considerations

The central theme of this dissertation was inserted in the processing and analysis of medical imaging, particularly in the study of preprocessing techniques and automatic classification of magnetic resonance images regarding different stages of Alzheimer's disease. Therefore the main goal was the development of a computational tool able to help in a practical context of the disease diagnosis.

The realization of the state-of-the-art allowed an analysis of multitudinous methods to detect the disease accurately. The different procedures counted on selection data, preprocessing, segmentation (some of them), and the central aim, the classification.

Highlighting the disease complexity, it is not straightforward the distinction between a normal cognition, mild cognitive impairment, and an AD subject being even more difficult in an early stage of the disease where the structural changes are less visible.

The automatic classification systems arise to help in an early diagnosis, having a better detection capacity than human eyes, requiring less effort and time from health professionals. In addition, these systems can detect differences in images that are not noticeable to a human being.

The use of algorithms specialized in processing images like the CNNs, which are widely used in the neuroimaging field, is also adopted in this work by multiple approaches.

In this work, CNN plays a vital role since it can easily extract valuable features at various levels and reduce errors caused by incomplete prior hypotheses. According to previous studies, the MCI stage is considered an intermediary between normal control and a fully developed disease. However, MCI can also be considered a different pathology since some subjects can remain stable or even regress for healthy cognition. Independently of the pathology present, CNN can automatically extract the most discriminative features not based on previous knowledge.

The best project results were obtained through a 3D approach, where were achieved values near the 90% accuracy in the test set. This value was a product of a good selection and combination

of hyperparameters like the activation function and the learning rate. Besides that, the followed steps in the preprocessing phase revealed essential to the proper functioning of the classifier, such as the registration, which directly influenced the decision because this best outcome reported was obtained using input images that undergo an atlas MNI-152 registration.

Although activation functions are not an object of analysis in the literature, they are very relevant for the success of a network, as it is evident by the accomplished results where a function change in convolutional layers produced significant performance differences.

Concerning a 2D implementation, several different architectures based on CNN have been studied, bearing in mind that this type of network is applied to most computer vision problems. Therefore, AD classification is not the exception.

Pre-trained networks have become increasingly popular within multiple applications since they allowed faster training due to early learning regarding low-level features, such as edges, shapes, corners, and intensity, that can be shared across tasks. Hence, this work was compared the application of ImageNet weights followed by a fine-tuning regarding the task, which revealed to improve the model performance.

Due to the almost general use of axial view to a 2D implementation, it was selected in this work. Therefore, the other views were also assessed to prove that choice. Although coronal presented inferior results, the difference was not significant, revealing that it can be helpful for later studies.

Comparing the two different implementations, 3D and 2D, the first one, which has spatial features not included in a slice approach, reported better results.

7.2 Future Work

One of the future goals could be the network test with images provided by the hospital to assess the generalization level and to study the further application in that environment.

The current method only was considered the MRI modality being an advantage and disadvantage. On the one hand, there are more data provided by this imaging modality comparing to others, and it is a medical exam done more frequently and in practically all hospitals, which does not happen with PET, for instance. On the other hand, it is generally more profitable to combine as many modalities as possible to use their richer information. Combining data from structural and functional imaging, genetics, cognition, disease status, and other phenotypic modalities, thence the exploration and implementation of new data could be a network improvement.

Moreover, respectively to the use of MRI only, data augmentation in 3D could also be implemented to increase the amount of data provided to the network, improving the performance. Although it is still not widely used due to the computational cost required, it is still a tool to be explored, and that will gain much prominence as it already happens in 2D networks.

The collection of several images over time highlighting the disease evolution will allow the specialists to get relevant information about the patient's status through quantitative change measurements in structural imaging and other patient variables. Having a computational tool to evaluate and monitor the evolution of the disease, taking into account the application of routines to delay the symptoms' progression, is possible the adjustment of those making the patient tracking more personalized.

References

[1] Christofides S. Maidment A. D.A. McLean I. D. Dance, D. R. and K. H. Ng. *Diagnostic radiology physics: A handbook for teachers and students. Endorsed by: American Association of Physicists in Medicine, Asia-Oceania Federation of Organizations for Medical Physics, European Federation of Organisations for Medical Physics*. Sep 2014.

[2] Cinnamon VanPutte, Jennifer Regan, and Andrew Russo. *Seeley's Essentials of Anatomy and Physiology*, volume 9. 9th edition, 2015.

[3] World Health Organization. Dementia. (Accessed on: 10/09/2020). URL: https://www.who.int/news-room/fact-sheets/detail/dementia.

[4] Alzheimer's Association. "2020 Alzheimer's disease facts and figures". *Alzheimer's and Dementia*, 16(3):391–460, 2020.

[5] Martin Prince, Anders Wimo, Maëlenn Guerchet, Gemma-Claire Ali, Yu-Tzu Wu, and Matthew Prina. "World Alzheimer Report 2015: The global impact of dementia". 2017. (Accessed on: 10/12/2020). URL: https://www.alz.co.uk/research/WorldAlzheimerReport2015.pdf.

[6] Lili Zhagn and Zhiping Li. "Alzheimer and the discovery of Alzheimer's disease". *Zhonghua yi shi za zhi (Beijing, China : 1980)*, 44(5):288–290, 2014.

[7] Marcus E. Raichle. "A brief history of human brain mapping". *Trends in Neurosciences*, 32(2):118–126, 2009.

[8] Adam M Staffaroni, Fanny M Elahi, Dana Mcdermott, D O Kacey Marton, Elissaios Karageorgiou, Simone Sacco, Matteo Paoletti, Eduardo Caverzasi, Christopher P Hess, euroimaging in Dementia Howard J Rosen, and Michael D Geschwind. N. 1(212):510–537, 2017.

[9] The top 10 causes of death. (Accessed on: 10/12/2020). URL: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.

[10] Michael A. Deture and Dennis W. Dickson. "The neuropathological diagnosis of Alzheimer's disease". *Molecular Neurodegeneration*, 14(1):1–18, 2019.

[11] R. Mayeux and Y. Stern. "Epidemiology of Alzheimer Disease". *Cold Spring Harbor Perspectives in Medicine*, 2(8):a006239–a006239, aug 2012.

[12] World Health Organization. "The Epidemiology and Impact of Dementia". *World Health Organization*, pages 3–6, 2018.

[13] Alzheimer's Association. "Medical Tests". (Accessed on 10/10/2020). URL: https://www.alz.org/alzheimers-dementia/diagnosis/medical_tests.

[14] Carol A. Manning and Jamie K. Ducharme. Handbook of Assessment in Clinical Gerontology. pages 155–178, 2010.

[15] Giovanni B. Frisoni, Nick C. Fox, Clifford R. Jack, Philip Scheltens, and Paul M. Thompson. "The clinical use of structural MRI in Alzheimer disease". *Nature Reviews Neurology*, 6(2):67–77, 2010.

[16] Eliana Venturelli, Chiara Villa, and Elio Scarpini. "Cerebrospinal fluid biomarkers for Alzheimer's disease". *Alzheimer's Diagnosis*, 1(April):227–243, 2011.

[17] Ryan A. Cloyd, Shon A. Koren, and Jose F. Abisambra. "Manganese-Enhanced Magnetic Resonance Imaging: Overview and Central Nervous System Applications With a Focus on Neurodegeneration". *Frontiers in Aging Neuroscience*, 10(December):1–22, 2018.

[18] Mark Brown and Richard Semelka. *MRI: Basic Principles and Applications*. Third edition, 01 2005.

[19] Medical Solutions Siemens Healthineers Global. "magnetom sola". Accessed on 10/10/2020. URL: https://www.siemens-healthineers.com/magnetic-resonance-imaging/0-35-to-1-5t-mri-scanner/magnetom-sola.

[20] Vijay P.B. Grover, Joshua M. Tognarelli, Mary M.E. Crossey, I. Jane Cox, Simon D. Taylor-Robinson, and Mark J.W. McPhail. "Magnetic Resonance Imaging: Principles and Techniques: Lessons for Clinicians". *Journal of Clinical and Experimental Hepatology*, 5(3):246–255, 2015.

[21] Tayyabah Yousaf, George Dervenoulas, and Marios Politis. "Advances in MRI Methodology". In *International Review of Neurobiology*, volume 141, pages 31–76. Elsevier Inc., First edition, 2018.

[22] Clifford R. Jack, Val J. Lowe, Stephen D. Weigand, Heather J. Wiste, Matthew L. Senjem, David S. Knopman, Maria M. Shiung, Jeffrey L. Gunter, Bradley F. Boeve, Bradley J. Kemp, Michael Weiner, and Ronald C. Petersen. "Serial PIB and MRI in normal, mild cognitive impairment and Alzheimers disease: Implications for sequence of pathological events in Alzheimers disease". *Brain*, 132(5):1355–1365, 2009.

[23] Clifford R. Jack, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L. Whitwell, Chadwick Ward, Anders M Dale, Joel P Felmlee, Jeffrey L Gunter, Derek L.G. Hill, Ron Killiany, Norbert Schuff, Sabrina Fox-Bosetti, Chen Lin, Colin Studholme, Charles S. DeCarli, Gunnar Krueger, Heidi A Ward, Gregory J Metzger, Katherine T. Scott, Richard Mallozzi, Daniel Blezek, Joshua Levy, Josef P. Debbins, Adam S. Fleisher, Marilyn Albert, Robert Green, George Bartzokis, Gary Glover, John Mugler, and Michael W. Weiner. "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods". *Journal of Magnetic Resonance Imaging*, 27(4):685–691, apr 2008.

[24] Steffen Sammet. "Magnetic Resonance Safety". *Abdominal Radiology (NY).*, 41(3):444–451, 2017.

[25] K. A. Johnson, N. C. Fox, R. A. Sperling, and W. E. Klunk. "Brain Imaging in Alzheimer Disease". *Cold Spring Harbor Perspectives in Medicine*, 2(4):a006213–a006213, apr 2012.

[26] Stefan Zachow, Michael Zilske, and Hans-Christian Hege. "3D Reconstruction of Individual Anatomy From Medical Image Data: Segmentation and Geometry Processing". *Proceedings of the CADFEM Users Meeting*, 01 2007.

[27] Rafael C Gonzalez, Richard E Woods, and Pearson Prentice Hall. *Digital Image Processing*. Third edition, 2008.

[28] Kayvan Najarian and Robert Splinter. *Biomedical Signal and Image Processing*, volume 53. Taylor & Francis, Boca Raton, second edition, apr 2012.

[29] Weili Zheng, Michael W.L. Chee, and Vitali Zagorodnov. "Improvement of brain segmentation accuracy by optimizing non-uniformity correction using N3". *NeuroImage*, 48(1):73–83, 2009.

[30] Mr Amir Ebrahimighahnavieh, Suhuai Luo, and Raymond Chiong. "Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review". *Computer Methods and Programs in Biomedicine*, 187:105242, 2020.

[31] Hamed Taheri Gorji and Naima Kaabouch. "A Deep Learning approach for diagnosis of Mild Cognitive Impairment based on MRI images". 9, 2019.

[32] Jyoti Islam and Yanqing Zhang. "Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks". *Brain Informatics*, 5(2), 2018.

[33] Heung-Il Suk, Seong-Whan Lee, and Dinggang Shen. "Deep ensemble learning of sparse regression models for brain disease diagnosis". *Medical Image Analysis*, 37:101–113, 2017.

[34] Muhammad Febrian Rachmadi, Maria Del C. Valdés-Hernández, Maria Leonora Fatimah Agan, and Taku Komura. "Deep learning vs. conventional machine learning: Pilot study of WMH segmentation in brain MRI with absence or mild vascular pathology". *Journal of Imaging*, 3(4):1–19, 2017.

[35] Aly Al Amyn Valliani, Daniel Ranti, and Eric Karl Oermann. "Deep Learning and Neurology: A Systematic Review". *Neurology and Therapy*, 8(2):351–365, 2019.

[36] Melda Yücel, Sinan Melih Nigdeli, and Gebrail Bekdaş. "Artificial Intelligence and Machine Learning with Reflection for Structural Engineering: A Review". *Advances in Structural Engineering—Optimization: Emerging Trends in Structural Optimization*, pages 23–72, 2021.

[37] Ting Shen, Jiehui Jiang, Yupeng Li, Ping Wu, Chuantao Zuo, and Zhuangzhi Yan. "Decision Supporting Model for One-year Conversion Probability from MCI to AD using CNN and SVM". *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2018-July:738–741, 2018.

[38] 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.

[39] Stephanos Leandrou, Styliani Petroudi, Panayiotis A. Kyriacou, Constantino Carlos Reyes-Aldasoro, and Constantinos S. Pattichis. "Quantitative MRI Brain Studies in Mild Cognitive Impairment and Alzheimer's Disease: A Methodological Review". *IEEE Reviews in Biomedical Engineering*, 11:97–111, 2018.

[40] Peter Bede. "From qualitative radiological cues to machine learning: MRI-based diagnosis in neurodegeneration". *Future Neurology*, 12(1):5–8, 2017.

[41] Stephan Dreiseitl and Lucila Ohno-Machado. "Logistic regression and artificial neural network classification models: A methodology review". *Journal of Biomedical Informatics*, 35(5-6):352–359, 2002.

[42] K. A Nazeer and M. Sebastian. "Improving the Accuracy and Efficiency of the k-means Clustering Algorithm". *Lecture Notes in Engineering and Computer Science*, 2176, 07 2009.

[43] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. A Bradford Book, Cambridge, MA, USA, 2018.

[44] Antonio Coronato, Muddasar Naeem, Giuseppe De Pietro, and Giovanni Paragliola. "Reinforcement learning for intelligent healthcare applications: A survey". *Artificial Intelligence in Medicine*, 109(December 2019):101964, 2020.

[45] Saman Sarraf, Danielle DeSouza, John Anderson, and Ghassem Tofighi. "DeepAD: Alzheimer's Disease Classification via Deep Convolutional Neural Networks using MRI and fMRI". *bioRxiv*, page 070441, 2016.

[46] Yann Lecun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". *Nature*, 521(7553):436–444, 2015.

[47] Marcia Hon and Naimul Mefraz Khan. "Towards Alzheimer's disease classification through transfer learning". *Proceedings - 2017 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2017*, 2017-Janua:1166–1169, 2017.

[48] Taeho Jo, Kwangsik Nho, and Andrew J. Saykin. "Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data". *Frontiers in Aging Neuroscience*, 11(August), 2019.

[49] Heung Il Suk and Dinggang Shen. "Deep learning-based feature representation for AD/MCI classification". *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8150 LNCS(PART 2):583–590, 2013.

[50] Sandra Vieira, Walter H.L. Pinaya, and Andrea Mechelli. "Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications". *Neuroscience and Biobehavioral Reviews*, 74:58–75, 2017.

[51] Adrien Payan and Giovanni Montana. "Predicting Alzheimer's disease a neuroimaging study with 3D convolutional neural networks". *ICPRAM 2015 - 4th International Conference on Pattern Recognition Applications and Methods, Proceedings*, 2:355–362, 2015.

[52] Heung Il Suk, Seong Whan Lee, and Dinggang Shen. "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis". *NeuroImage*, 101:569–582, 2014.

[53] Andrés Ortiz, Jorge Munilla, Juan M. Górriz, and Javier Ramírez. "Ensembles of Deep Learning Architectures for the Early Diagnosis of the Alzheimer's Disease". *International Journal of Neural Systems*, 26(7):1–23, 2016.

[54] Nicola Amoroso, Domenico Diacono, Annarita Fanizzi, Marianna La Rocca, Alfonso Monaco, Angela Lombardi, Cataldo Guaragnella, Roberto Bellotti, and Sabina Tangaro. "Deep learning reveals Alzheimer's disease onset in MCI subjects: Results from an international challenge". *Journal of Neuroscience Methods*, 302:3–9, 2018.

[55] Nagaraj Yamanakkanavar, Jae Young Choi, and Bumshik Lee. "MRI segmentation and classification of human brain using deep learning for diagnosis of alzheimer's disease: A survey". *Sensors (Switzerland)*, 20(11):1–31, 2020.

[56] Silvia Basaia, Federica Agosta, Luca Wagner, Elisa Canu, Giuseppe Magnani, Roberto Santangelo, and Massimo Filippi. "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks". *NeuroImage: Clinical*, 21(October 2018):101645, 2019.

[57] Alexander Khvostikov, Jenny Benois-Pineau, Andrey Krylov, and Gwenaelle Catheline. "Classification methods on different imaging modalities for Alzheimer disease studies". In *GraphiCon 2017 - 27th International Conference on Computer Graphics and Vision*, pages 237–242, 09 2017.

[58] Marina Sokolova, Nathalie Japkowicz, and Stan Szpakowicz. "Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation". *AAAI Workshop - Technical Report*, WS-06-06:24–29, 2006.

[59] Google. "classification: Roc curve and auc; machine learning crash course". (Accessed on 03/07/2021). URL: https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc.

[60] Ammarah Farooq, SyedMuhammad Anwar, Muhammad Awais, and Saad Rehman. "A deep CNN based multi-class classification of Alzheimer's disease using MRI". In *2017 IEEE International Conference on Imaging Systems and Techniques (IST)*, pages 1–6, 2017.

[61] Hamed Taheri Gorji and Naima Kaabouch. "A Deep Learning approach for diagnosis of Mild Cognitive Impairment based on MRI images". *Brain Sciences*, 9(9), 2019.

[62] Swathi S. Kundaram and Ketki C. Pathak. *"Deep Learning-Based Alzheimer Disease Detection"*, volume 673. Springer Singapore, 2021.

[63] Altuğ Yiğit and Zerrin Işik. "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.

[64] Bian, Shuyang. "Diagnosis of Alzheimer´s Disease using Structural MRI and Convolution Neural Network". *E3S Web Conf.*, 185:03037, 2020.

[65] S. Sarraf, D.D. Desouza, J.A.E. Anderson, and C. Saverino. "MCADNNet: Recognizing stages of cognitive impairment through efficient convolutional fMRI and MRI neural network topology models". *IEEE Access*, 7:155584–155600, 2019.

[66] Dan Pan, An Zeng, Longfei Jia, Yin Huang, Tory Frizzell, and Xiaowei Song. "Early Detection of Alzheimer's Disease Using Magnetic Resonance Imaging: A Novel Approach Combining Convolutional Neural Networks and Ensemble Learning". *Frontiers in Neuroscience*, 14(May):1–19, 2020.

[67] Xiuli Bi, Shutong Li, Bin Xiao, Yu Li, Guoyin Wang, and Xu Ma. "Computer aided Alzheimer's disease diagnosis by an unsupervised deep learning technology". *Neurocomputing*, 392:296–304, 2020.

[68] Karim Aderghal, Jenny Benois-Pineau, Karim Afdel, and Catheline Gwenaëlle. "FuseMe: Classification of sMRI images by fusion of deep CNNs in 2D+e projections". *ACM International Conference Proceeding Series*, Part F1301(June), 2017.

[69] Gulshan Mukhtar and Saima Farhan. "Convolutional neural network based prediction of conversion from mild cognitive impairment to alzheimer's disease: A technique using hippocampus extracted from MRI". *Advances in Electrical and Computer Engineering*, 20(2):113–122, 2020.

[70] Mark Claypool, Ragnhild Eg, and Kjetil Raaen. "Classification of sMRI for AD Diagnosis with Convolutional Neuronal Networks: A Pilot 2-D+? Study on ADNI". *MultiMedia Modeling*, 1:226–237, 2017.

[71] Weiming Lin, Tong Tong, Qinquan Gao, Di Guo, Xiaofeng Du, Yonggui Yang, Gang Guo, Min Xiao, Min Du, and Xiaobo Qu. "Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment". *Frontiers in Neuroscience*, 12(NOV):1–13, 2018.

[72] Huanhuan Ji, Zhenbing Liu, Wei Qi Yan, and Reinhard Klette. *"Early Diagnosis of Alzheimer's Disease Based on Selective Kernel Network with Spatial Attention"*, volume 12047 LNCS. Springer International Publishing, 2020.

[73] Naimul Mefraz Khan, Nabila Abraham, and Marcia Hon. "Transfer Learning with Intelligent Training Data Selection for Prediction of Alzheimer's Disease". *IEEE Access*, 7:72726–72735, 2019.

[74] Jyoti Islam and Yanqing Zhang. "Deep convolutional neural networks for automated diagnosis of Alzheimer's disease and mild cognitive impairment using 3D brain MRI". *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11309 LNAI:359–369, 2018.

[75] Bumshik Lee, Waqas Ellahi, and Jae Young Choi. "Using deep CNN with data permutation scheme for classification of Alzheimer's disease in structural magnetic resonance imaging (SMRI)". *IEICE Transactions on Information and Systems*, E102D(7):1384–1395, 2019.

[76] Obioma Pelka, Christoph M. Friedrich, Felix Nensa, Christoph Mönninghoff, Louise Bloch, Karl Heinz Jöckel, Sara Schramm, Sarah Sanchez Hoffmann, Angela Winkler, Christian Weimar, and Martha Jokisch. "Sociodemographic data and APOE-ε4 augmentation for MRI-based detection of amnestic mild cognitive impairment using deep learning systems". *PLoS ONE*, 15(9 September):1–24, 2020.

[77] Congling Wu, Shengwen Guo, Yanjia Hong, Benheng Xiao, Yupeng Wu, Qin Zhang, and The Alzheimer's Disease Neuroimaging Initiative. "Discrimination and conversion prediction of mild cognitive impairment using convolutional neural networks". *Quantitative Imaging in Medicine and Surgery*, 8(10):992–1003, 2018.

[78] Shangran Qiu, Gary H. Chang, Marcello Panagia, Deepa M. Gopal, Rhoda Au, and Vijaya B. Kolachalama. "Fusion of deep learning models of MRI scans, Mini–Mental State Examination, and logical memory test enhances diagnosis of mild cognitive impairment". *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, 10:737–749, 2018.

[79] Huanhuan Ji, Zhenbing Liu, Wei Qi Yan, and Reinhard Klette. "Early diagnosis of Alzheimer's disease using deep learning". *ACM International Conference Proceeding Series*, pages 87–91, 2019.

[80] Linlin Gao, Haiwei Pan, Fujun Liu, Xiaoqin Xie, Zhiqiang Zhang, and Jinming Han. "Brain disease diagnosis using deep learning features from longitudinal MR images". *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10987 LNCS:327–339, 2018.

[81] Zhenyu Cui, Zhiao Gao, Jiaxu Leng, Tianlin Zhang, Pei Quan, and Wei Zhao. "Alzheimer's Disease Diagnosis Using Enhanced Inception Network Based on Brain Magnetic Resonance Image". *Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019*, pages 2324–2330, 2019.

[82] Jingwan Jiang, Li Kang, Jianjun Huang, and Tijiang Zhang. "Deep learning based mild cognitive impairment diagnosis using structure MR images". *Neuroscience Letters*, 730(March):134971, 2020.

[83] Sergey Korolev, Amir Safiullin, Mikhail Belyaev, and Yulia Dodonova. "Residual and plain convolutional neural networks for 3D brain MRI classification". *Proceedings - International Symposium on Biomedical Imaging*, pages 835–838, 2017.

[84] Zongjie Tu and Prabir Bhattacharya. "Fast training of a convolutional neural network for brain MRI classification". *ACMSE 2019 - Proceedings of the 2019 ACM Southeast Conference*, pages 218–221, 2019.

[85] Ruoxuan Cui and Manhua Liu. "RNN-based longitudinal analysis for diagnosis of Alzheimer's disease". *Computerized Medical Imaging and Graphics*, 73:1–10, 2019.

[86] Fan Li and Manhua Liu. "Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks". *Computerized Medical Imaging and Graphics*, 70:101–110, 2018.

[87] Manhua Liu, Fan Li, Hao Yan, Kundong Wang, Yixin Ma, Li Shen, and Mingqing Xu. "A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease". *NeuroImage*, 208(August 2018), 2020.

[88] Ruoxuan Cui and Manhua 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.

[89] Soheil Esmaeilzadeh, Dimitrios Ioannis Belivanis, Kilian M. Pohl, and Ehsan Adeli. *"End-to-end Alzheimer's disease diagnosis and biomarker identification"*, volume 1. Springer International Publishing, 2018.

[90] Xinying Yu, Bo Peng, Jun Shi, Jianbin Zhu, and Yakang Dai. "3D Convolutional Networks Based Automatic Diagnosis of Alzheimer's Disease Using Structural MRI". *Proceedings - 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, CISP-BMEI 2019*, (Mci), 2019.

[91] Kanghan Oh, Young Chul Chung, Ko Woon Kim, Woo Sung Kim, and Il Seok Oh. "Classification and Visualization of Alzheimer's Disease using Volumetric Convolutional Neural Network and Transfer Learning". *Scientific Reports*, 9, 2019.

[92] Fei Gao, Hyunsoo Yoon, Yanzhe Xu, Dhruman Goradia, Ji Luo, Teresa Wu, and Yi Su. "AD-NET: Age-adjust neural network for improved MCI to AD conversion prediction". *NeuroImage: Clinical*, 27(June), 2020.

[93] Guilherme Folego, Marina Weiler, Raphael F. Casseb, Ramon Pires, and Anderson Rocha. "Alzheimer's Disease Detection Through Whole-Brain 3D-CNN MRI". *Frontiers in Bioengineering and Biotechnology*, 8(October), 2020.

[94] Chong Yaw Wee, Chaoqiang Liu, Annie Lee, Joann S. Poh, Hui Ji, and Anqi Qiu. "Cortical graph neural network for AD and MCI diagnosis and transfer learning across populations". *NeuroImage: Clinical*, 23(April):101929, 2019.

[95] Eunho Lee, Jun Sik Choi, Minjeong Kim, and Heung Il Suk. "Toward an interpretable Alzheimer's disease diagnostic model with regional abnormality representation via deep learning". *NeuroImage*, 202(August), 2019.

[96] Hongfei Wang, Yanyan Shen, Shuqiang Wang, Tengfei Xiao, Liming Deng, Xiangyu Wang, and Xinyan Zhao. "Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer's disease". *Neurocomputing*, 333:145–156, 2019.

[97] Juan Ruiz, Mufti Mahmud, Md Modasshir, and M. Shamim Kaiser. "3D DenseNet Ensemble in 4-Way Classification of Alzheimer's Disease". *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12241 LNAI:85–96, 2020.

[98] Shuqiang Wang, Hongfei Wang, Yanyan Shen, and Xiangyu Wang. "Automatic Recognition of Mild Cognitive Impairment and Alzheimers Disease Using Ensemble based 3D Densely Connected Convolutional Networks". *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, pages 517–523, 2019.

[99] Mingxia Liu, Chunfeng Lian, and Dinggang Shen. *"Anatomical-Landmark-Based Deep Learning for Alzheimer's Disease Diagnosis with Structural Magnetic Resonance Imaging"*, volume 171. Springer International Publishing, 2020.

[100] Ping Cao, Jie Gao, and Zuping Zhang. "Multi-view based multi-model learning for MCI diagnosis". *Brain Sciences*, 10(3), 2020.

[101] Hao Tang, Erlin Yao, Guangming Tan, and Xiuhua Guo. *"A fast and accurate 3D fine-tuning convolutional neural network for alzheimer's disease diagnosis"*, volume 888. Springer Singapore, 2018.

[102] Ahsan Bin Tufail, Yongkui Ma, and Qiu Na Zhang. "Multiclass classification of initial stages of Alzheimer's Disease through Neuroimaging modalities and Convolutional Neural Networks". *Proceedings of 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference, ITOEC 2020*, (Itoec):51–56, 2020.

[103] Ahsan Bin Tufail, Yongkui Ma, and Qiu-Na Zhang. "Classification of subjects of Mild Cognitive Impairment and Alzheimer's Disease through Neuroimaging modalities and Convolutional Neural Networks". pages 1–6, 2020.

[104] Wei Feng, Nicholas Van Halm-Lutterodt, Hao Tang, Andrew Mecum, Mohamed Kamal Mesregah, Yuan Ma, Haibin Li, Feng Zhang, Zhiyuan Wu, Erlin Yao, and Xiuhua Guo. "Automated MRI-Based Deep Learning Model for Detection of Alzheimer's Disease Process". *International Journal of Neural Systems*, 30(6):1–14, 2020.

[105] Simeon Spasov, Luca Passamonti, Andrea Duggento, Pietro Liò, and Nicola Toschi. "A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease". *NeuroImage*, 189(January):276–287, 2019.

[106] Y. Shmulev and M. Belyaev. *"Predicting conversion of mild cognitive impairments to alzheimer's disease and exploring impact of neuroimaging"*, volume 11044 LNCS. 2018.

[107] Dan Jin, Bo Zhou, Ying Han, Jiaji Ren, Tong Han, Bing Liu, Jie Lu, Chengyuan Song, Pan Wang, Dawei Wang, Jian Xu, Zhengyi Yang, Hongxiang Yao, Chunshui Yu, Kun Zhao, Max Wintermark, Nianming Zuo, Xinqing Zhang, Yuying Zhou, Xi Zhang, Tianzi Jiang, Qing Wang, and Yong Liu. "Generalizable, Reproducible, and Neuroscientifically Interpretable Imaging Biomarkers for Alzheimer's Disease". *Advanced Science*, 7(14):1–12, 2020.

[108] Emimal Jabason, M. Omair Ahmad, and M. N.S. Swamy. "Deep structural and clinical feature learning for semi-supervised multiclass prediction of Alzheimer's disease". *Midwest Symposium on Circuits and Systems*, 2018-Augus:791–794, 2019.

[109] Heung Il Suk, Seong Whan Lee, and Dinggang Shen. "Latent feature representation with stacked auto-encoder for AD/MCI diagnosis". *Brain Structure and Function*, 220(2):841–859, 2015.

[110] Jun Shi, Xiao Zheng, Yan Li, Qi Zhang, and Shihui Ying. "Multimodal Neuroimaging Feature Learning with Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer's Disease". *IEEE Journal of Biomedical and Health Informatics*, 22(1):173–183, 2018.

[111] Feng Li, Loc Tran, Kim Han Thung, Shuiwang Ji, Dinggang Shen, and Jiang Li. "A Robust Deep Model for Improved Classification of AD/MCI Patients". *IEEE Journal of Biomedical and Health Informatics*, 19(5):1610–1616, 2015.

[112] Tao Zhou, Kim Han Thung, Xiaofeng Zhu, and Dinggang Shen. "Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis". *Human Brain Mapping*, 40(3):1001–1016, 2019.

[113] Yongsheng Pan, Mingxia Liu, Chunfeng Lian, Tao Zhou, Yong Xia, and Dinggang Shen. *"Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis"*, volume 11072 LNCS. Springer International Publishing, 2018.

[114] Chiyu Feng, Ahmed Elazab, Peng Yang, Tianfu Wang, Feng Zhou, Huoyou Hu, Xiaohua Xiao, and Baiying Lei. "Deep Learning Framework for Alzheimer's Disease Diagnosis via 3D-CNN and FSBi-LSTM". *IEEE Access*, 7:63605–63618, 2019.

[115] Karim Aderghal, Alexander Khvostikov, Andrei Krylov, Jenny Benois-Pineau, Karim Afdel, and Gwenaelle Catheline. "Classification of Alzheimer Disease on Imaging Modalities with Deep CNNs Using Cross-Modal Transfer Learning". *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, 2018-June(Dl):345–350, 2018.

[116] Jongin Kim and Boreom Lee. "Identification of Alzheimer's disease and mild cognitive impairment using multimodal sparse hierarchical extreme learning machine". *Human Brain Mapping*, 39(9):3728–3741, 2018.

[117] Parisa Forouzannezhad, Alireza Abbaspour, Chunfei Li, Mercedes Cabrerizo, and Malek Adjouadi. "A Deep Neural Network Approach for Early Diagnosis of Mild Cognitive Impairment Using Multiple Features". *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, pages 1341–1346, 2019.

[118] Yongsheng Pan, Mingxia Liu, Chunfeng Lian, Yong Xia, and Dinggang Shen. *"Disease-Image Specific Generative Adversarial Network for Brain Disease Diagnosis with Incomplete Multi-modal Neuroimages"*, volume 11766 LNCS. Springer International Publishing, 2019.

[119] Yongsheng Pan, Mingxia Liu, Chunfeng Lian, Yong Xia, and Dinggang Shen. "Spatially-Constrained Fisher Representation for Brain Disease Identification with Incomplete Multi-Modal Neuroimages". *IEEE Transactions on Medical Imaging*, 39(9):2965–2975, 2020.

[120] Shaozhuo Li, Na Wang, Xuehui Du, and Aodi Liu. *"Multimodal 3D CNN for classification of Brain disease using sMR and FDG-PET images"*, volume 1058. 2019.

[121] Li Kang, Jingwan Jiang, Jianjun Huang, and Tijiang Zhang. "Identifying Early Mild Cognitive Impairment by Multi-Modality MRI-Based Deep Learning". *Frontiers in Aging Neuroscience*, 12(September):1–10, 2020.

[122] Yechong Huang, Jiahang Xu, Yuncheng Zhou, Tong Tong, and Xiahai Zhuang. "Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network". *Frontiers in Neuroscience*, 13(MAY), 2019.

[123] Donghuan Lu, Karteek Popuri, Gavin Weiguang Ding, Rakesh Balachandar, Mirza Faisal Beg, Michael Weiner, Paul Aisen, Ronald Petersen, Cliford Jack, William Jagust, John Trojanowki, Arthur Toga, Laurel Beckett, Robert Green, Andrew Saykin, and John et all Morris. "Multimodal and Multiscale Deep Neural Networks for the Early Diagnosis of Alzheimer's Disease using structural MR and FDG-PET images". *Scientific Reports*, 8(1):1–13, 2018.

[124] Daniel A. Starr. "A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal MRI". *Physiology & behavior*, 176(1):139–148, 2011.

[125] Mohammad R. Arbabshirani, Sergey Plis, Jing Sui, and Vince D. Calhoun. "Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls". *NeuroImage*, 145:137–165, 2017.

[126] Andrew M. Blamire. "MR approaches in neurodegenerative disorders". *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2018.

[127] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. "Understanding of a convolutional neural network". In *2017 International Conference on Engineering and Technology (ICET)*, pages 1–6, 2017.

[128] Adrian Rosebrock. *"Deep Learning for Computer Vision with Python"*. PyImageSearch, 1 edition, 2017.

[129] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)", 2016.

[130] Sergey Ioffe and Christian Szegedy. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 448–456, Lille, France, 07–09 Jul 2015. PMLR.

[131] Braga Antonio, Carvalho Andre, and Teresa Ludermir. *Redes neuronais artificiais : teoria e aplicacoes*. LTC, 2 edition, 2007.

[132] Kevin Gurney. *An introduction to neural networks*. CRC Press, London, first edition, 1997.

[133] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. "Focal Loss for Dense Object Detection". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):318–327, 2020.

[134] Clifford R. Jack, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L. Whitwell, Chadwick Ward, Anders M Dale, Joel P Felmlee, Jeffrey L Gunter, Derek L.G. Hill, Ron Killiany, Norbert Schuff, Sabrina Fox-Bosetti, Chen Lin, Colin Studholme, Charles S. DeCarli, Gunnar Krueger, Heidi A Ward, Gregory J Metzger, Katherine T. Scott, Richard Mallozzi, Daniel Blezek, Joshua Levy, Josef P. Debbins, Adam S. Fleisher, Marilyn Albert, Robert Green, George Bartzokis, Gary Glover, John Mugler, and Michael W. Weiner. "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods". *Journal of Magnetic Resonance Imaging*, 27(4):685–691, apr 2008.

[135] April J. Ho, Xue Hua, Suh Lee, Alex D. Leow, Igor Yanovsky, Boris Gutman, Ivo D. Dinov, Natasha Leporé, Jason L. Stein, Arthur W. Toga, Clifford R. Jack, Matt A. Bernstein, Eric M. Reiman, Danielle J. Harvey, John Kornak, Norbert Schuff, Gene E. Alexander, Michael W. Weiner, and Paul M. Thompson. "Comparing 3 T and 1.5 T MRI for tracking Alzheimer's disease progression with tensor-based morphometry". *Human Brain Mapping*, 31(4):499–514, April 2010.

[136] U.S. Department of Health and National Institute of Mental Health Human Services. "NIfTI-1 Data Format - Neuroimaging Informatics Technology Initiative". (Accessed on 03/25/2021). URL: https://nifti.nimh.nih.gov/nifti-1.

[137] Medium. "Advantages and Disadvantages of Python Programming Language". (Accessed on 04/27/2021). URL: https://medium.com/@mindfiresolutions.usa/advantages-and-disadvantages-of-python-programming-language-fd0b.

[138] Cuda zone. (Accessed on 04/23/2021). URL: https://developer.nvidia.com/cuda-zone.

[139] Keras Team. "Keras documentation: About Keras". (Accessed on 04/23/2021). URL: https://keras.io/about/.

[140] Paul A. Yushkevich, Yang Gao, and Guido Gerig. "ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images". In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3342–3345, 2016.

[141] Matthew Brett and et all Markiewicz. "Nibabel", April 2020. (Accessed on 04/19/2021). URL: https://doi.org/10.5281/zenodo.3757992.

[142] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, and et al. "Array programming with NumPy". *Nature*, 585(7825):357–362, Sep 2020.

[143] Stephen M. Smith. "Fast robust automated brain extraction". *Human Brain Mapping*, 17(3):143–155, 2002.

[144] Ivan Itzcovich. "Deepbrain · PyPI". https://pypi.org/project/deepbrain/. (Accessed on 05/19/2021).

[145] Ziv Yaniv, Bradley C. Lowekamp, Hans J. Johnson, and Richard Beare. "SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research". *Journal of Digital Imaging*, 31(3):290–303, June 2018.

[146] Stefan Klein, Marius Staring, Keelin Murphy, Max A. Viergever, and Josien P. W. Pluim. "Elastix: A Toolbox for Intensity-Based Medical Image Registration". *IEEE Transactions on Medical Imaging*, 29(1):196–205, 2010.

[147] Bradley Lowekamp, gabehart, Daniel Blezek, Luis Ibanez, Matt McCormick, Dave Chen, Dan Mueller, David Cole, Hans Johnson, Kasper Marstal, Richard Beare, Arnaud Gelas, Kent Williams, David Doria, and Brad King. "SimpleElastix", June 2015. (Accessed on 04/11/2021). URL: https://doi.org/10.5281/zenodo.19049, doi:10.5281/zenodo.19049.

[148] Pravat K. Mandal, Rashima Mahajan, and Ivo D. Dinov. "Structural brain atlases: Design, rationale, and applications in normal and pathological cohorts". *Journal of Alzheimer's Disease*, 31(SUPPL. 3), 2012.

[149] Karen Simonyan and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition". In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[150] Marcia Hon and Naimul Mefraz Khan. "Towards Alzheimer's disease classification through transfer learning". *Proceedings - 2017 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2017*, 2017-Janua:1166–1169, 2017.

[151] J.P.W. Pluim, J.B.A. Maintz, and M.A. Viergever. "Mutual-information-based registration of medical images: a survey". *IEEE Transactions on Medical Imaging*, 22(8):986–1004, 2003.

[152] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization". *International Journal of Computer Vision*, 128(2):336–359, Oct 2019.