Sofia Manuela Gomes de Carvalho

Classification of Dementias Based on Brain Radiomics Features

July 2021

Universidade do Minho
Escola de Engenharia

Sofia Manuela Gomes de Carvalho

Classification of Dementias Based on Brain Radiomics Features

Master Dissertation
Integrated Master's in Informatics Engineering

Dissertation supervised by
Victor Manuel Rodrigues Alves
Tiago Gil Oliveira

July 2021

DECLARATION

Name: Sofia Manuela Gomes de Carvalho

Dissertation Title: Classification of Dementias Based on Brain Radiomics Features

Mentors: Victor Manuel Rodrigues Alves, Tiago Gil Oliveira

Conclusion Year: 2021

Master Designation: Integrated Master's in Informatics Engineering

Universidade do Minho, _____/_____/_____

Signature: _____

ACKNOWLEDGEMENTS

To my supervisor at the University of Minho, Victor Alves, I am truly grateful for all his support and availability throughout all phases of the project. His constant disposition was fundamental to both the developing and writing phases, never denying his assistance even with his busy schedule.

I would like to thank my supervisor at the School of Medicine from the University of Minho, Tiago Gil Oliveira, for sharing his knowledge and guiding me. I would also like to express my appreciation for all the work of the Life and Health Sciences Research Institute team, which has enabled me to develop my dissertation in a high-quality environment with such great minds.

I would like to thank the amazing people I met through my academic journey, for all the moments that have always cheered me up through the course. To all my old friends, thank you for your patience and support during this period. Your presence and friendship are essential to my happiness.

I would also like to thank my parents and my brothers for all the support, advice and guidance in all my decisions. Their wisdom has made me the person I am today.

Finally, to André Martins, for being with me celebrating the good moments and supporting me in bad. For all your love, thank you.

To my parents, brothers and boyfriend, I dedicate this dissertation.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

Universidade do Minho, _____/_____/_____

Signature: _____

This page was intentionally left blank.

ABSTRACT

Neurodegenerative diseases impair the functioning of the brain and are characterized by alterations in the morphology of specific brain regions. Some of the main disorders include Alzheimer's, Parkinson's, and Huntington's diseases, and the number of cases increases exponentially since ageing is one of the main risk factors. Trying to identify the areas in which this type of disease appears is something that can have a very positive impact in this area of Medicine and can guarantee a more appropriate treatment or allow the improvement of the quality of life of patients. With the current technological advances, computer tools are capable of performing a structural or functional analysis of neuroimaging data from Magnetic Resonance Images (MRI). Therefore, Medical Informatics uses these techniques to create and manage medical neuroimaging data to improve the diagnosis and management of these patients. MRI is the image type used in the analysis of the brain area and points to a promising and reliable diagnostic tool since it allows high-quality images in various planes or strategies and MRI methods are fundamental diagnostic tools in clinical practice, allowing the diagnosis of pathologic processes such as stroke or brain tumours. However, structural MRI has limitations for the diagnosis of neurodegenerative disorders since it mainly identifies atrophy of brain regions.

Currently, there is increased interest in informatics applications capable of monitoring and quantifying human brain imaging alterations, with potential for neurodegenerative disorders diagnosis and monitoring. One of these applications is Radiomics, which corresponds to a methodology that allows the extraction of features from images of a given region of the brain. Specific quantitative metrics from MRI are acquired by this tool, and they correspond to a set of features, including texture, shape, among others. To standardize Radiomics application, specific libraries have been proposed to be used by the bioinformatics and biomedical communities, such as PyRadiomics, which corresponds to an open-source Python package for extracting Radiomics of MRIs.

Therefore, this dissertation was developed based on magnetic resonance images and the study of Deep Learning (DL) techniques to assist researchers and neuroradiologists in the diagnosis and prediction of neurodegenerative disease development. Two different main tasks were made: first, a segmentation, using FreeSurfer, of different regions of the brain and then, a model was build from radiomic features extracted from each part of the brain and interpreted for knowledge extraction.

Keywords: Machine Learning, Dementia Classification, Magnetic Resonance Imaging, Radiomics, Brain Morphological Features, Neurodegenerative Diseases, ADNI.

RESUMO

As doenças neurodegenerativas estão associadas ao funcionamento do cérebro e caracterizam-se pelo facto de serem altamente incapacitantes. São exemplos destas, as doenças de Alzheimer, Parkinson e Huntington, e o seu número de casos tem vindo a aumentar exponencialmente, uma vez que o envelhecimento é um dos principais factores de risco. Tentar identificar quais são as regiões cerebrais que permitem predizer o seu aparecimento e desenvolvimento é algo que, sendo possível, terá um impacto muito positivo nesta área da Medicina e poderá garantir um tratamento mais adequado, ou simplesmente melhorar a qualidade de vida dos pacientes. Com os avanços tecnológicos atuais, foram desenvolvidas ferramentas informáticas que são capazes de efetuar uma análise estrutural ou funcional de Ressonâncias Magnéticas (MRI), sendo essas ferramentas usadas para promover a melhoria e o conhecimento clínico. Deste modo, as constantes evoluções científicas têm realçado o papel da Informática Médica na neuroimagem para criar e gerenciar dados médicos, melhorando o diagnóstico destes pacientes.

A MRI é o tipo de imagem utilizada na análise de regiões cerebrais e aponta para uma ferramenta de diagnóstico promissora e fiável, uma vez que permite obter imagens de alta qualidade em vários planos, permitindo assim, o diagnóstico de processos patológicos, tais como acidentes vasculares ou tumores cerebrais.

Atualmente, existem inúmeras aplicações informáticas capazes de efetuar análises estruturais e funcionais do cérebro humano, pois é este o principal órgão afetado pelas doenças neurodegenerativas. Uma dessas aplicações é o *Radiomics*, que permite fazer a extração de *features* de imagens do cérebro. A biblioteca a utilizar será *PyRadiomics*, que corresponde a um *package open-source* em *Python* para a extração de *features Radiomics* de imagens médicas. As *features* correspondem a características da imagem.

Assim sendo, a presente dissertação foi desenvolvida com base em imagens de ressonância magnética e no estudo das técnicas de *Deep Learning* para investigar e auxiliar os médicos neurorradiologistas a diagnosticar e a prever o desenvolvimento de doenças neurodegenerativas. Foram feitas duas principais tarefas: primeiro, uma segmentação, utilizando o *software* FreeSurfer, de diferentes regiões do cérebro e, de seguida, foi construído um modelo a partir das *features* radiómicas extraídas de cada parte do cérebro que foi interpretado.

Palavras-Chave: *Machine Learning*, Classificação de Demências, Imagens Ressonância Magnética, *Radiomics*, *Features* Morfológicas Cerebrais, Doenças Neurodegenerativas, ADNI.

This page was intentionally left blank.

Table of Contents

LIST OF FIGURES

List of Tables

List of Abbreviations and acronyms

A	
AD	Alzheimer's Disease
ADAM	Adaptive Moment Estimation
ADNI	Alzheimer's Disease Neuroimaging Initiative
AI	Artificial Intelligence
ANN	Artificial Neural Network
C	
CE	Cross-Entropy
CN	Cognitively Normal
CNN	Convolutional Neural Networks
CNS	Central Nervous System
CSF	Cerebrospinal Fluid
CSV	Comma Separated Values
CT	Computer Tomography
D	
DICOM	Digital Imaging and Communications in Medicine
DL	Deep Learning
DNN	Deep Neural Network
DT	Diffusion Tensor
DTI	Diffusion Tensor Imaging
F	
FN	False Negative
FNR	False Negative Rate
FP	False Positive
FPR	False Positive Rate
G	
GLCM	Grey Level Co-occurrence Matrix
GLDM	Grey Level Dependence Matrix

GLRLM	Grey Level Run Length Matrix	
GLSZM	Grey Level Size Zone Matrix	
GM	Grey Matter	
I		
ICV	IntraCranial Volume	
ICVS	Life and Health Sciences Research Institute	
L		
LBP	Local Binary Pattern	
LoG	Laplacian of Gaussian	
LR	Learning Rate	
M		
MCI	Mild Cognitive Impairment	
MGZ	MGH compressed version	
MGH	Massachusetts General Hospital file format	
ML	Machine Learning	
MLP	Multi-Layer Perceptron	
MRI	Magnetic Resonance Imaging	
N		
N3	Nonparametric Nonuniform Intensity Normalization	
NGTDM	Neighbouring Grey Tone Difference Matrix	
NIfTI	Neuroimaging Informatics Technology Initiative	
O		
OECD	Organisation for Economic Co-operation and Development	
P		
PET	Positron Emission Tomography	
R		
ReLU	Rectifier Linear Units	
RF	Radiofrequency	
RNN	Recurrent Neural Network	

	ROI	Region Of Interest
S		
	SGD	Stochastic Gradient Descent
T		
	TN	True Negative
	TP	True Positive
W		
	WM	White Matter

GLOSSARY

Activation Function	Used in neural networks, decides whether the neuron is activated or not. The neuron receives an input value from previous neurons. If the value is large enough, the neural unit passes the value to the next neuron, thus acting as a threshold function.
Artificial Intelligence (AI)	The capacity of a computer does some human activities like thinking, making decisions, or learning.
Batch	In Neural Networks, samples are used in one iteration.
Class	Refers to the output category of the data, and a label in a dataset refers to one of the classes.
Dataset	A collection of data. The data contains features and, if using supervised training, a label.
Deep Learning (DL)	A subfield of Machine Learning belongs to representation learning, where the algorithm is trained to learn the feature hierarchy to classify the data provided.
Epoch	One forward pass and one backward pass through all training samples.
Feature	In Machine Learning, a feature is the measurable propriety of an attribute.
Good Fit Model	A model that suitably learns the training dataset and generalizes well to the test dataset.
High Pass Filter	Used to make an image appear sharper. These filters emphasize fine details in the image. High-pass filtering amplifies noise.
Low Pass Filter	A low-pass filter, also called a "blurring" or "smoothing" filter, averages out rapid changes in intensity. Low-pass filtering smooths out the noise.
Machine Learning (ML)	An approach to achieving Artificial Intelligence, where a computer is trained using specific algorithms and a large amount of data to gain the ability to adapt to new situations and predict patterns.
Magnetic Resonance Imaging (MRI)	A non-invasive and painless imaging technique based on the absorption and emission of energy in the radiofrequency spectrum of the electromagnetic spectrum. MRI is widely often used for clinical assistance and also in clinical research.
Medical Imaging Informatics	A discipline that results from the combination of medical images and biomedical informatics. Serves to enhance the improvement and knowledge of clinical care. On the one hand, medical imaging allows the study of a disease state *in vivo*. On the other hand, biomedical informatics is involved in the development of computer science techniques for creating and managing medical data.

Overfitting	Occurs when a model that fits the training data too well fails when applied to new data. The generalizability is negatively impacted.
Pixel	Picture element that defines a unit of information in two-dimensional space.
Soft Tissue	Tissues that support other structures and organs of the body, such as tendons, fibrous tissues, fat, muscles, nerves, and blood vessels.
Test set	The subset of the dataset that was used to test the model after the training phase.
Train set	The subset of the dataset that was used to train a model.
Underfit Model	A model that fails to sufficiently learn the problem and performs poorly on a training dataset and does not perform well on a holdout sample.
Validation set	A subset of the dataset that was used to adjust hyperparameters in the training phase.
Voxel	Volume element that represents a value in the three-dimensional space.

This page was intentionally left blank.

1 INTRODUCTION

Key points:

- Diagnostically, imaging and medical images have moved from a minor exclusionary role to a central position.
- The main objective of this dissertation is the application of Deep Learning techniques to classification of Dementias and understand the features extracted.
- The motivation is to investigate and try to find markers that help neuroradiologists in today's society to diagnose and consequently treat neurodegenerative diseases.

1.1 CONTEXT AND MOTIVATION

MRI is one type of medical image that is widely used to diagnose, plan and guide the treatment and monitoring of the progression of a disease and is essential in the structural and functional study of the human brain, where it is possible to observe all systems present in the brain: the sensory, visual, auditory, olfactory and motor systems, that is, all systems corresponding to the five human senses. To significantly improve the analysis of data obtained from a MRI scan, the use of medical image processing and analysis applications is becoming more necessary [1]. Thus, as a way of increasing what is understood of the complexity of the human brain and its processes, neuroimaging emerged. Also, due to the growing number of images taken in a single scan, it becomes more necessary and useful to use these applications in order to be able to process the entire data volume more precisely. In this way, it is possible to identify patterns in the image more easily in a quantitative manner, as well as to segment regions of interest, to obtain and analyse volumes of different anatomical brain parts, among others. These applications serve as a tool to assist the doctor, allowing a more accurate diagnosis by the doctor/healthcare professional. An example of one of these tools is Radiomics [2]. Image registration techniques are the core for interpretation and analysis of the medical image to connect the spatial information and the pathologies or anomalies. However, image manipulation is possible only with digital images and one type of this kind of image is MRI, which provides variable image contrast and high contrast for soft tissue (i.e., tissues that support other structures and organs of

the body, such as tendons, fibrous tissues, fat, muscles, nerves and blood vessels), and high spatial resolution [3]. Manual segmentation requires experienced neuroradiologists to analyse the MRI information, which takes a long time because they have to analyse several slices of the image. Manual segmentation is also subjective, depending on the neuroradiologist. It has been found that deep learning segmentation and classification techniques are useful for improved diagnosis, growth rate prediction, treatment planning, and treatment costs because their algorithms perform well.

This dissertation's motivation relies on promising medical images (MRI images) and the emerging Deep Learning (DL) techniques to investigate and help doctors in today's society to detect and make the diagnosis of neurodegenerative diseases in an earlier and faster way.

1.2 OBJECTIVES

The main goal of this work is to optimize the use of Radiomics [2], which is an area of investigation where features are extracted from medical images and after that, are analysed for feature extraction in volumetric brain acquisitions to classify dementia and its evolution. The analysis of the application of the method for the classification of dementias will be used to conclude the evolution of neurodegenerative diseases over time, based on the use of brain imaging features. Also, DL methods will be applied, and in order to achieve the main objective of this project, DL models, image pre-processing techniques, and image types that are commonly used in medical imaging have also been studied, as well as segmentation of various brain regions.

1.3 INVESTIGATION METHODOLOGY

Therefore, in the research process, it is necessary to make a study of the methodological principles and methods that are going to be used here. For this purpose, we present here a survey of the state of the art, as well as a reflection on the methodological issues found, to substantiate the chosen research strategies and arrive at reliable solutions to the problems encountered. After preparing the research plan, a review of the relevant literature was performed, starting with the execution of the project.

The research strategy selected for this dissertation was Design Science Research (DSR), which is considered to be effective in computer science and medical informatics projects. The DSR methodology (Figure 1-1), is a rigorous scientific research methodology that includes the techniques, principles, and procedures that must be followed to design and develop successful solutions to the problems [4]. It can be divided into six main steps:

1. Problem identification and motivation - Identification of the research problems to be solved, and appropriate justification of the practical and scientific values of the solution;

2. Definition of the solution's objectives – Identification and presentation of the solution objectives based on the identified research problems;

3. Design and development of the solution – Design the solution, i.e. definition of the features and architecture, and then develop the solution to solve the specified objectives;

4. Demonstration – Experiments and simulations to demonstrate the solutions developed, to evaluate the capacity of the solutions to solve the research problems;

5. Evaluation – Assessment of the solution and comparison of the results achieved in the demonstration step with the objective defined in the second step. This means assessing whether the solution worked out can support the research problems and the objectives set;

6. Communication – Communication of the problem as the developed solution to an audience, i.e., spread the importance of the problem and the solution.
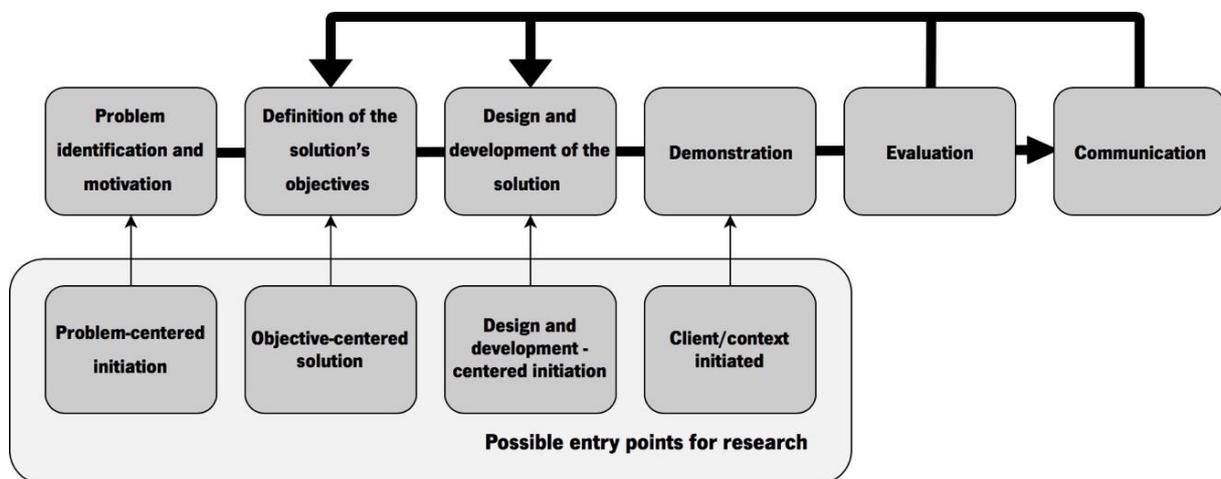
Figure 1-1 - Design Research Process schema (adapted from [4])

1.4 STRUCTURE OF THE DISSERTATION

This dissertation is structured in 5 chapters, including this as a first chapter, which provides a context of the work as well as its motivations and objectives. The next chapter introduces some theoretical information about the concepts addressed in this work, as well as background information on medical concepts. Chapter 3 introduces the literature review, while chapter 4 provides information about the medical images that describe the data and the materials used in this dissertation, as well as technologies that were used in the development of this dissertation. Therefore, this chapter becomes

one of the most important chapters. Finally, chapter 5 ends with the project discussions and conclusions, as well as some open research possibilities that might arise.

2 CONCEPTS AND TECHNOLOGIES

Key points:

- In several neurodegenerative diseases, a progressive and irreversible degeneration process occurs in the nerve cells of Central Nervous System.
- Radiomics is an informatic application capable of doing structural and functional analyses of the human brain.
- The MRI images, 3D volumes, are a widely used medical image type for clinical assistance and research and show promising results when applied in Deep Learning segmentation techniques.
- Machine learning is an approach to achieve Artificial Intelligence, i.e. the capacity of a computer to make some human activities, such as making decisions or learning.

2.1 CENTRAL NERVOUS SYSTEM

The Central Nervous System (CNS) consists of the brain and spinal cord (Figure 2-1). It has been thoroughly studied by anatomists and physiologists, but it still holds many secrets; it controls our thoughts, movements, emotions, and desires. It also controls our breathing, heart rate, the release of hormones, body temperature, and much more. The brain is the most complex organ in the human body; the cerebral cortex (the outermost part of the brain and the largest part by volume) contains billions of neurons (Figure 2-2), each of which is connected to thousands of other neurons. The CNS can be roughly divided into white and grey matter (Figure 2-3). As a very general rule, the brain consists of an outer cortex of grey matter and an inner area that houses the tract of white matter.
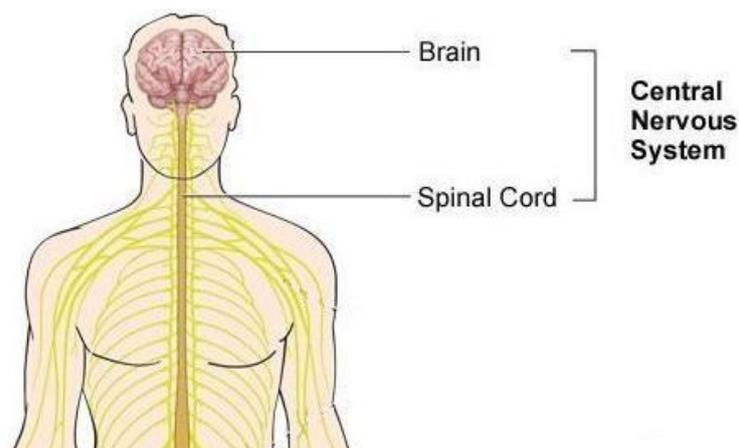
Figure 2-1 - Constitution of Central Nervous System (adapted from [5])

White matter

White Matter (WM) is tissue in the brain composed of nerve fibres. The fibres (called axons) connect nerve cells and are covered by myelin (a type of fat). The myelin is what gives white matter its white colour. Myelin speeds up the signals between the cells, enabling the brain cells to quickly send and receive messages. White matter makes up about half of the brain [6]. Some research has found that abnormalities in white matter were present in imaging studies of the brain before the development of the symptoms of Alzheimer's Disease (AD) [6].
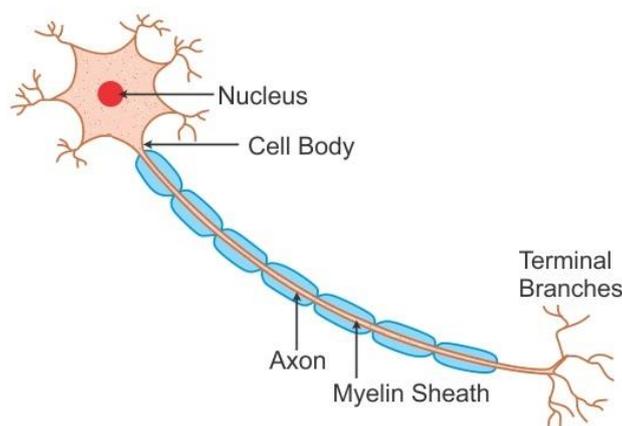
Figure 2-2 - Neuron: diagram of nerve cells (retrieved from [7])

Grey matter

Grey Matter (GM) makes up the other half of the brain [6]. The grey matter is mainly composed of neuronal cell bodies and unmyelinated axons. Axons are the processes that extend from neuronal cell bodies, carrying signals between those bodies. In the grey matter, these axons are mainly unmyelinated, meaning they are not covered by a whitish-coloured, fatty protein called myelin. The grey matter serves to process information in the brain. Structures within the grey matter process signals generated in the sensory organs or other areas of the grey matter [8].
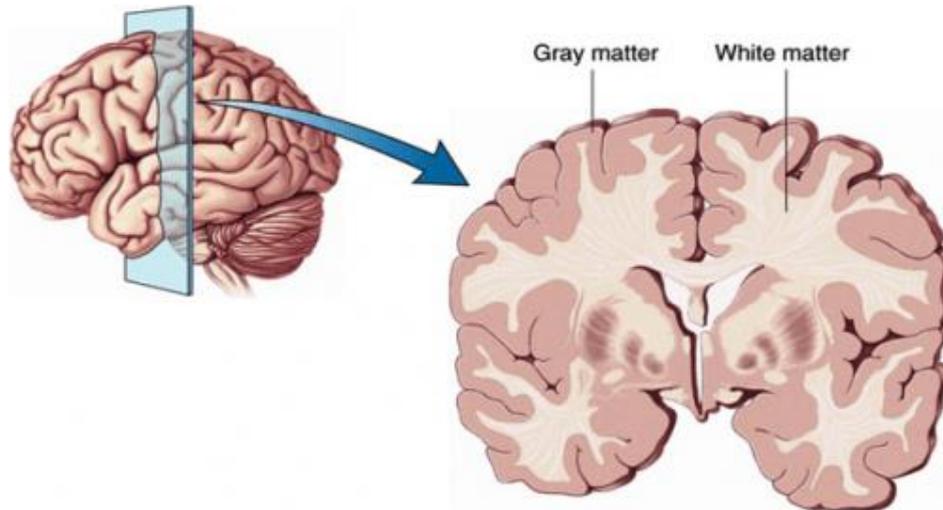
Figure 2-3 - Illustration of white and grey matter in the human brain (retrieved from [9])

2.2 NEURODEGENERATIVE DISEASES

Neurons are building structures of the nervous system that includes the brain and spinal cord. Usually, these small cells do not reproduce or are replaced in the body. In several neurodegenerative diseases, an irreversible process occurs that results in progressive degeneration and/or death of nerve cells, resulting in different manifestations according to the type of neurons affected, such as changes in movement in Parkinson's disease or loss of cognitive functions in dementia. Neurodegenerative diseases are the core of this complex process, which involves several hereditary and sporadic conditions. The term neurodegenerative is generic, but covers, according to Directorate-General Health and Food Safety, an organism of the European Commission, more than 600 conditions that affect the neurons of the human brain. Of the hundreds of disorders, the most common ones are AD, which is estimated to be responsible for 60 to 70 per cent of the cases of dementia diagnosed worldwide, Parkinson's disease, and Huntington's disease. The reasons for the progressive and irreversible destruction of neurons are still unknown. In AD, there are deposits of small plaques of proteins that damage different parts of the brain and lead to progressive loss of memory [10]. Brain cells degenerate and die, causing a progressive decline in memory and other cognitive functions. It is a chronic and progressive brain disease, which affects multiple higher functions, altering the state of consciousness and impacting activities of daily living. In the case of Parkinson's disease, dopamine deficiency affects the basal ganglia of the brain, causing stiffness and tremors in the main muscles of the body, two of the most typical signs of the condition.

AD is a neurodegenerative pathology of uncertain cause, which manifests itself mainly in old age people and it's the most common form of dementia, being a major challenge for healthcare in the twenty-first century. AD is an irreversible, progressive brain disorder marked by a decline in cognitive functioning. The consequences of AD at the brain level are well known, characterized by a progressive loss of neurons (brain cells) and connections between various areas of the brain and a reduction in the number and size of brain cells that irreversibly deteriorate patients' cognitive functions. Language, thinking, attention and concentration are some of the functions affected by the disease that leads to the death of brain cells and the cut in communication within the brain.

AD is an irreversible neurodegenerative disease that results in a loss of mental function caused by the deterioration of brain tissue and it is the most common cause of dementia among people over the age of 65. It can be characterized by synaptic loss and neurodegeneration, leading to cognitive decline, and AD is an irreversible, progressive brain disorder marked by a decline in cognitive functioning with no validated disease-modifying treatment [11]. This results in progressive brain atrophy. There are still no therapies available to cure neurodegeneration, despite the numerous scientific researches that seek to understand the phenomenon [10], as is the case with this dissertation.

It is estimated that in Portugal there are between 67 and 90 thousand people with this disease, in a universe of 5% of the Portuguese population suffering from dementia [12], yet no prevention methods or cures have been developed.

The main risk factor is age, although having a close relative with Alzheimer's disease increases the possibility of having the disease [13].

Some of the Alzheimer's disease symptoms are [12]:

- Memory change: patients have difficulty remembering recent events or information. Memory for more distant events is typically preserved in the early stages of the disease;

- Language disorders: Patients find it difficult to find words in some contexts, there is a reduction in vocabulary, and, in more advanced stages, the speech content becomes poorer;

- Difficulty performing motor tasks: In more advanced stages of the disease, it prevents patients from performing simple tasks such as dressing or eating;

- Changes in personality and behavioural symptoms: People who were calm become hostile and aggressive, and other patients become very passive;

- Difficulty performing complex tasks such as paying bills and dealing with money;

- Psychotic and depressive symptoms;

- Disorientation in known places.

A clinical evaluation, supported by observation and by physical, laboratory, and neurological exams, makes it possible to diagnose the disease. A timely diagnosis can help detect and diagnose other potential rare diseases [12], therefore the possibility of being able to identify which brain areas and which features of the brain predict the development and evolution of the disease is very promising.

Effective and accurate diagnosis of AD is important for the initiation of effective treatments. Particularly, early diagnosis of AD plays a significant role in therapeutic development and ultimately for effective patient care.

2.3 ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a longitudinal multicentre study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of AD. ADNI began in 2004 under the leadership of Michael W. Weiner and, since its launching more than a decade ago, it has led to major contributions in AD research, enabling the sharing of data between researchers around the world. ADNI is a database of images that contains different types of data from study volunteers, such as clinical, genetic, MRI images (structural, diffusion-weighted imaging, perfusion, and resting state sequences), Positron Emission Tomography (PET) image, and biospecimen [14]. Participants are followed and reassessed over time to track the pathology of the disease as it progresses.

The goal of the ADNI study is to track both the progression of AD and changes in the underlying pathology using biomarkers, together with clinical measures to assess the brain's structure and function throughout disease states [15]. In this dissertation, we considered three of these states: Cognitively Normal (CN), Mild Cognitive Impairment (MCI), and AD. During the phases of the ADNI study, to the extent possible, participants were carried forward from previous phases for continued monitoring.

ADNI enrols participants between the ages of 55 and 90 who are recruited at 57 sites in the United States and Canada.

ADNI participants can be divided into three different stages:

1. CN – These participants are the control subjects in the ADNI study are characterized for having normal ageing. They show no signs of depression, MCI or dementia;

2. MCI - These participants have reported a subjective memory concern either autonomously or via an informant or clinician. However, daily living activities are essentially preserved, there are

no significant levels of impairment in other cognitive domains, and no signs of dementia exist. MCI is thought of as a prodromal stage of AD;

3. AD - These participants are subjects with the clinical diagnosis of AD, showing global signs of dementia.

A biomarker, or biological marker, is a substances' measurement, or indicator of a biological state. Biomarkers may exist before the clinical symptoms arise. ADNI uses various biomarkers to help predict the onset of AD [15].

2.4 MAGNETIC RESONANCE IMAGING

A great deal of effort has been made to develop strategies for early detection, especially at pre-symptomatic stages in order to slow or prevent disease progression. In particular, advanced neuroimaging techniques, such as MRI and PET, have been developed and used to identify AD-related structural and molecular biomarkers [11].

A MRI scan is a non-invasive imaging technique that enables the observation of anatomic structures, physiological functions, and molecular composition of tissues and it has become a common procedure around the world.

Since its development, medical doctors and researchers continue to refine MRI techniques to assist in medical procedures and research. MRI has transformed the role of radiology in medicine since its initial applications in structural imaging in the early 1980s and now encompasses wider areas of functional and molecular imaging. The development of MRI revolutionized medicine since it examines the inside of the human body in high detail using a non-invasive tool, since it operates in the RadioFrequency (RF) range, so it does not have any harmful ionizing radiation. The following are examples in which an MRI scanner would be used [16]:

- Anomalies of the brain and spinal cord;
- Tumours and other anomalies in various parts of the body;
- Breast cancer screening for women who face a high risk of breast cancer;
- Injuries or abnormalities of the joints, such as the back and knee;
- Certain types of heart problems;
- Diseases of the liver and other abdominal organs.

This list is by no means exhaustive and the use of MRI technology is always expanding in scope.

MRI can be used to examine macrostructural changes (gross differences in tissue volume that reflect volume variability, parenchymal atrophy, or frank pathology), or microstructural changes (fibre tract integrity and pathologies that can be altered due to subtle changes in myelin) [17].

MRI is a type of scan that uses strong magnetic fields and radio waves to produce detailed images of the organs and tissues within the body. An MRI scanner uses a large powerful magnet, radio waves, and a computer to create a detailed, cross-sectional image of internal organs and structures. The scanner itself typically resembles a large tube with a table in the middle, allowing the patient to slide in (Figure 2-4).

Figure 2-4 - Magnetic Resonance Image scanner machine (retrieved from [18])

This type of scan differs from Computerized Tomography (CT) scans and X-rays, as it does not use potentially harmful ionizing radiation. There is very little preparation required, if any, before an MRI scan. On arrival at the hospital, doctors may ask the patient to change into a gown. As magnets are used, it is critical that no metal objects are present in the scanner. The doctor will ask the patient to remove any metal jewellery or accessories that might interfere with the machine. A patient will rarely experience side effects from an MRI scan.

MRI is based on the absorption and emission of energy in the RF range of the electromagnetic spectrum. It produces images based on spatial variations in the phase and frequency of the radiofrequency energy being absorbed and emitted by the imaged object. The physics behind the MRI is complex and extensive and, therefore, will not be depicted here. However, it will be explained in a high-

level overview. The human body is largely made of water molecules, which are comprised of hydrogen and oxygen atoms. At the centre of each atom lies an even smaller particle called a proton, which serves as a magnet and is sensitive to any magnetic field. Without an external magnetic field, a group of protons assumes a random orientation of magnetic moments. Under the influence of an applied external magnetic field, the protons assume a non-random alignment, resulting in a measurable magnetic moment in the direction of the external magnetic field. By applying RF pulses to the body, the protons absorb this energy and jump into a higher energy state. Then, the energy is released, and the protons reduce the energy state by emitting energy that is absorbed by RF antennas placed around the patient (on the MRI scanner). These signals are processed into MRI scans, which are 3D volumes. Different RF pulses lead to different contrasts between white matter, grey matter, and CerebroSpinal Fluid (CSF), and scans can then be created based on the differences in the signal from hydrogen atoms in different types of tissue.

An obvious strength of MRI is its availability since that European and U.S. guidelines recommend that all subjects with cognitive decline undergo structural imaging (MRI or CT) and that it is part of proposed diagnostic criteria for AD and other dementias. In most centres, MRI is regarded as essential in the investigation of neurodegenerative diseases, which indicates that it is a marker of its utility. Although not as rapid as CT, MRI is safe as it does not involve ionizing radiation making it able for individuals to be imaged serially without concerns about carcinogenicity. MRI offers a range of different sequences that can probe different tissue characteristics providing multiple clinical and research measures in the same session [19].

Table 2-1 - Advantages and Limitations of Neuroimaging Techniques (adapted from [20])

Methods		Measurement Provided	Spatial Resolution	Advantages	Limitations
	CT	Brain structures	0.5 – 1 mm	High spatial resolution	Radiation Low contrast
	PET	Perfusion Metabolism Neurotransmitter dynamics	4 – 10 mm	Fewer motion artefacts High sensitivity	Low spatial and temporal resolution Limited availability Radiation
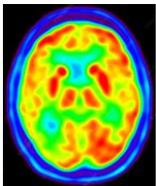
					High cost Not portable
	DTI	Fiber tracks	2.5 mm	High spatial resolution	Limited for grey matter
	fMRI	Hemodynamic activity	< 3 mm	High spatial resolution No radiation Widely available	Not portable Low temporal resolution Sensitive to motion artefacts
	sMRI	Brain structures (e.g. White matter, grey matter, and cerebrospinal fluid)	1 – 2 mm	High spatial resolution No radiation	Relatively low sensitivity High cost Long scanning time
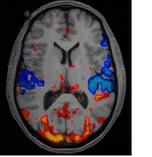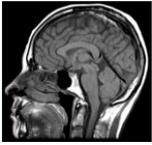
Each of these techniques has advantages and disadvantages related to resolution, safety, availability, and accessibility [20].

Medical imaging data are mainly stored in two types: DICOM and NIfTI. Since the ADNI files were in DICOM format and PyRadiomics does not work with this type of file, there was a need to make some conversion types. The choice was converting DICOM to NIfTI format since this format is compatible with PyRadiomics as well as with FreeSurfer. Therefore, the files could be passed to FreeSurfer to segment MRI from ADNI and finally to PyRadiomics to extract all of the features. This tool outputs in NIfTI format.

2.4.1 DICOM

Digital Imaging and Communications in Medicine (DICOM) is the international standard to transmit, store, retrieve, print, process, and display medical imaging information. It defines the formats for medical images that can be exchanged with the data and quality necessary for clinical use. DICOM is implemented in almost every radiology, cardiology imaging, and radiotherapy device (X-ray, CT, MRI, ultrasound, etc.), and increasingly in devices in other medical domains such as ophthalmology and dentistry. With hundreds of thousands of medical imaging devices in use, DICOM is one of the most

widely deployed healthcare messaging standards in the world [21]. There are billions of DICOM images currently in use for clinical care, as the example of scans provided by ADNI.

Since its first publication in 1993, DICOM has revolutionized the practice of radiology, allowing the replacement of X-ray film with a fully digital workflow, enabling advanced medical imaging applications. From the emergency department to cardiac stress testing, to breast cancer detection, DICOM is the standard that makes medical imaging work [21][22]. DICOM:

- Makes medical imaging information interoperable;
- Integrates image-acquisition devices and printers from different manufacturers;
- Is actively developed and maintained to meet the evolving technologies and needs of medical imaging.

DICOM files usually store 2D images containing the slice of the MRI at a certain time instant. Because of this, the data folders contain many .dcm files, each one is a 2D image [21]. These files were passed to FreeSurfer, in order to make the segmentation of the areas of the brain, so that PyRadiomics could be applied specifically to both hippocampi.

2.4.2 NIfTI

NIfTI is a type of file format in 2D, 3D or 4D [23] for neuroimaging and this type of file is used very commonly in imaging informatics for neuroscience and even neuroradiology research. The standard NIfTI image format was designed for scientific analysis of brain images and is a simple, compact, and versatile format. To keep compatibility with the previous formats used by many different imaging software, data stored in NIfTI format uses a pair of files (*.hdr/.img*) [23]. Single file or pair of files are not the only possible presentations, though. It is very common to have files that occupy large space in the disk and this is the perfect case where compression of NifTI files can achieve excellent results. The brain images can be stored as a pair of files (*.hdr*), or a single file (*.nii*).

NIfTI is an eponymous acronym that comes from the Neuroimaging Informatics Technology Initiative, which was a working group at the National Institutes of Health (NIH) in the USA in the early 2000s. The Neuroimaging Informatics Technology Initiative wanted to create a file type that was similar to the Analyze, another type and common medical imaging file type format used in neuroimaging research, but improved from that.

One of the objectives of NifTI is to enhance the utility of existing informatics tools used widely in neuroimaging research and provide an environment to facilitate convergence on common solutions to widespread problems [23].

Some programs are meant for one file type and will not work well on another. Therefore, file conversion is often necessary. DICOM files are standard in clinical care. However, some imaging informatics tools can convert DICOM to NifTI files automatically [24].

Since the images obtained from ADNI are in DICOM format and PyRadiomics doesn't support this type of data file, there was a need to convert them to NIfTI.

2.4.3 MGZ

The compressed versions of NifTI images have the *.gz* extension appended: *.nii.gz* (in case of a single file) or *.hdr/.img.gz* (pair of files) [23]. The MGZ format corresponds to the compressed version of MGH files [25]. MGH stands for Massachusetts General Hospital and is a binary format used to store high-resolution structural data. Can store 4D (like NIfTI), saving the information of columns, rows, slices, and frames, or more generic information, saving only volumes and surfaces, i.e., "volume-encoded" surface files, with no geometry information.

The output of the extraction of each hippocampus from the ADNI files that were previously processed in FreeSurfer to segment the areas of the brain is in MGZ format.

2.5 RADIOMICS

With technological updates that have been made lately, were developed informatics tools that can carry out a structural or functional analysis of neuroimaging data from magnetic resonances. In this way, the constant scientific evolution has highlighted the important role that Medical Informatics plays in neuroimage, and imaging can noninvasively visualize the radiographic phenotype of a pathology before, during, and after treatment [26]. Currently, there are numerous informatic applications capable of doing structural and functional analyses of the human brain. One of these informatics applications is Radiomics. It aims to quantify phenotypic characteristics on medical imaging through the use of automated algorithms and Radiomic data has the potential to uncover disease characteristics that fail to be appreciated by the naked eye. Radiomics is an emerging concept, which was first proposed by the Dutch scholar Philippe Lambin in 2012 [27], and is a research branch in the field of medical imaging that aims to quantify radiomic features using high-throughput algorithms. Radiomic artificial intelligence

technology, either based on engineered hard-coded algorithms or deep learning methods, can be used to develop non-invasive imaging-based biomarkers [26].

The central hypothesis of Radiomics is that distinctive imaging algorithms quantify the state of diseases, and thereby provide valuable information for personalized medicine. The intuitive understanding for radiomics is to transform the visual image information into deep features for quantitative study [27]. Based on the rapid development of machine learning and image processing techniques, radiomic analyses has gained considerable attention and have been successfully applied in the field of oncology, as a method for supporting clinical decision-making and precision medicine [27][28]. but it can be applied to other medical issues where a disease is imaged. MRI is a routinely used diagnostic tool for AD management and a lot of this pathology information that is not recognized by the human eye remains unmined. In this way, the purpose of Radiomics is to extract high-dimensional radiomic features from medical images to fully exploit the in-depth information of this neurodegenerative disease. Radiomics aims to quantify the diseases' phenotype (the set of observable characteristics of an individual resulting from the interaction of its genotype with the environment) using high-throughput algorithms applied to medical imaging data. The features can be correlated to the clinical outcomes and could, therefore, be combined in most parametric biomarker models for clinical decision support. Radiomics refers to a set of techniques for extracting a large number of quantitative features from medical images and subsequently mining these features to retrieve clinically useful diagnostic and prognostic information. This methodology is based on the hypothesis that a radiologic phenotype may reflect genetic biological alterations and may thus be predictive of the biologic behaviour of a disease [29]. Radiomics is also an effective method for assessing the morphologic and textural changes that can occur in the organs that are associated with various diseases. Unlike visual assessments of clinical images that are usually made by doctors and, therefore, have the risk of being compromised by human error, it may allow for more objective and comprehensive assessments of these changes based on quantitative indices: the features. Radiomic artificial intelligence technology, either based on engineered hard-coded algorithms or deep learning methods, can be used to develop non-invasive imaging-based biomarkers and Radiomics aims to quantify phenotypic characteristics on medical imaging using automated algorithms.
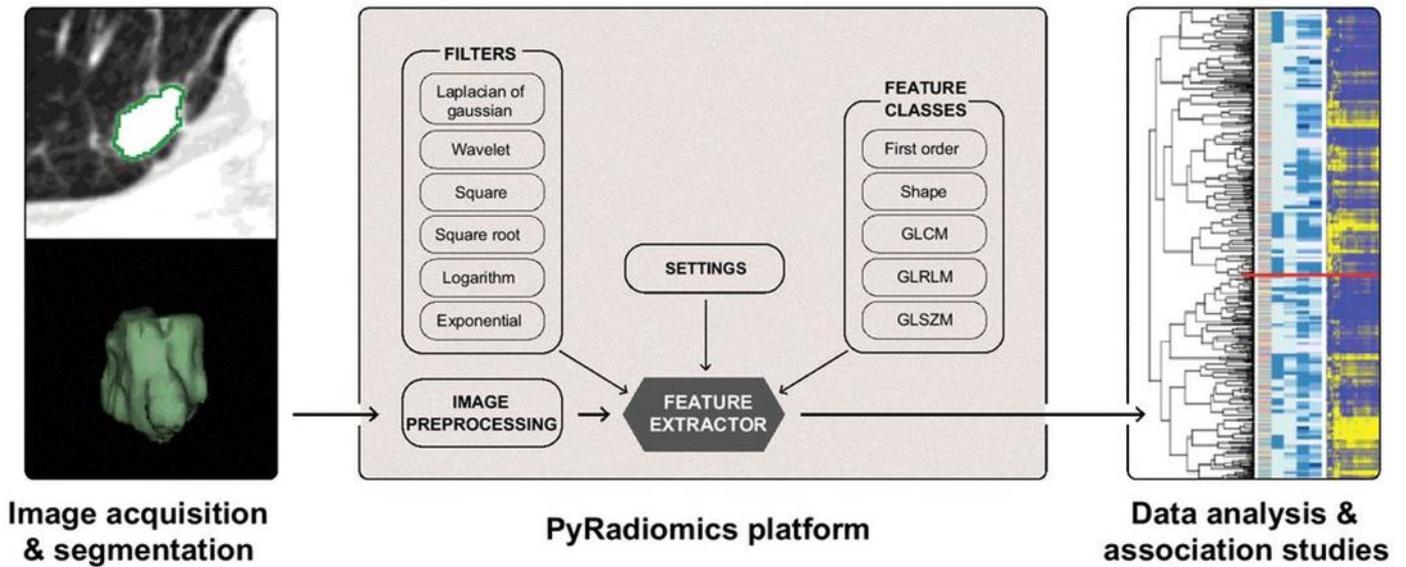
Figure 2-5 - Overview diagram of the process of Radiomics (retrieved from [26])

The process of radiomic analysis usually includes the acquisition of image data, segmentation of Region Of Interest (ROI), feature extraction, feature selection, and classification or prediction (Figure 2-5) [27]. The radiomic analysis of medical scans involves multiple processes, including image pre-processing, segmentation, feature extraction, feature selection, and classification. Image pre-processing is an important step for achieving valid and reproducible radiomic features.

2.5.1 RADIOMIC FILTERS

Features can be calculated on the original image or on images that were pre-processed using a choice of several built-in filters (Figure 2-5). The following are some built-in optional filters [30] [26]:

- Laplacian of Gaussian (LoG);
- Wavelet;
- Square;
- Square Root;
- Logarithm;
- Exponential;
- Gradient;
- Local Binary Pattern 2D (LBP2D);
- Local Binary Pattern 3D (LBP3D).

The filters shown above correspond to the image types (either the original image or derived images using those filters) that can be used to extract features. Moreover, custom settings can be provided for each enabled image type, which will then only be applied during feature extraction for that image type. Currently, available image types are [30]:

- Original - no filter applied;

- Laplacian of Gaussian - edge enhancement filter. This filter emphasizes areas of grey level change, where a sigma value can be personalized. This sigma defines how coarse the emphasized texture should be. A low sigma emphasis on fine textures (change over a short distance), where a high sigma value emphasizes coarse textures (grey level change over a large distance). In this dissertation, Laplacian of Gaussian was applied with 5 different values for sigma: 1, 2, 3, 4, and 5. Applies a Laplacian of Gaussian filter to the input image and yields a derived image for each sigma σ value specified. This filter is sensitive to areas with rapidly changing intensities, enhancing edges. The sigma is a possible setting and is a list of floats or integers that must be greater than 0. The width (mm) of the filter in the Gaussian kernel is determined by σ and can be used to emphasize more fine (low σ values) or coarse (high σ values) textures;

- Wavelet - Applies wavelet filter to the input image and yields 8 decompositions per level (all possible combinations of applying either a high or a low pass filter in each of the three dimensions);

- Square - Takes the square of the image intensities and linearly scales them back to the original range. Negative values in the original image will be made negative again after the application of the filter;

- Square Root - Takes the square root of the absolute image intensities and scales them back to the original range. Negative values in the original image will be made negative again after the application of the filter;

- Logarithm - Takes the logarithm of the absolute intensity + 1. Values are scaled to the original range and negative original values are made negative again after the application of the filter.

- Exponential - Takes the exponential, where filtered intensity is e^(absolute intensity). Values are scaled to the original range and negative original values are made negative again after the application of the filter;

- Gradient - Returns the magnitude of the local gradient;

- Local Binary Pattern 2D - Computes the Local Binary Pattern in a by-slice operation (2D);

- Local Binary Pattern 3D - Computes the Local Binary Pattern in 3D using spherical harmonics.

2.5.2 RADIOMIC FEATURES

The features correspond to visual characteristics of the image that are generated by Radiomics, which include intensity, texture, shape, among others, and have been found useful in several clinical areas, such as oncology and cardiology. Engineered features are hard-coded features that are often based on expert knowledge in the medical area.

Radiomic features can be divided into [29]:

- Morphologic features;

- Histogram features;

- Textural features;

- Higher-order features.

Various features can be extracted using PyRadiomics and currently available feature classes are [26] [30] (Figure 2-5):

- First Order Statistics;

- Shape-based (2D and 3D);

- Grey Level Co-occurrence Matrix (GLCM);

- Grey Level Run Length Matrix (GLRLM);

- Grey Level Size Zone Matrix (GLSZM);

- Neighbouring Grey Tone Difference Matrix (NGTDM);

- Grey Level Dependence Matrix (GLDM).

Feature extraction is supported for both single slice (2D) and whole volume (3D) segmentations [26]. These shape-based features are independent of the grey level intensity distribution in the ROI and are therefore only calculated on the non-derived image and mask.

To summarize, radiomic features are subdivided into the following eight feature classes [30]:

- First Order Statistics (19 features) - Describe the distribution of voxel intensities within the image region defined by the mask, through commonly used and basic metrics;

- Shape-based (3D) (16 features) - In this group of features are included descriptors of the three-dimensional size and shape of the region of interest;

- Shape-based (2D) (10 features) - These features include descriptors of the two-dimensional size and shape of the ROI;

- Grey Level Cooccurrence Matrix (24 features) - Describes the frequency of two neighbouring pixels having certain grey-level pixel values [29];

- Grey Level Run Length Matrix (16 features) - Quantifies grey level runs, which are defined as the length of consecutive pixels that have the same grey level value;

- Grey Level Size Zone Matrix (16 features) - Quantifies grey level zones in an image. A grey-level zone is defined as the number of connected voxels that share the same grey level intensity. A voxel is considered connected if the distance is 1;

- Neighbouring Grey Tone Difference Matrix (5 features) - Quantifies the difference between a grey value and the average grey value of its neighbours within distance δ;

- Grey Level Dependence Matrix (14 features) - Quantifies grey level dependencies in an image. A grey level dependency is defined as the number of connected voxels within distance δ that are dependent on the centre voxel.

The number of extracted features can be variable, largely depending on the number of textural features and the number of filters used for extracting higher-order features [29]. Every feature is identified by a unique name consisting of the applied filter, the feature class, and the feature name.

2.6 MACHINE LEARNING

Machine Learning (ML) is a branch of computer science that studies the design of algorithms that can learn. ML is a common name for several algorithms that identify patterns in data for making predictions.

This dissertation focuses on deep learning, which corresponds to a set of algorithms that are inspired by the structure and function of the brain, and it learns features through a hierarchical learning process. These algorithms are usually called Artificial Neural Networks (ANN). Deep learning is based on representing learning in which the algorithm learns the best features to carry out a given task on its own by navigating the provided data. Deep learning is one of the hottest fields in data science with many case studies that have astonishing results in robotics, image recognition, and Artificial Intelligence (AI) [31].

ML is an approach to achieve AI (Figure 2-6), where a computer is trained using specific algorithms and a large amount of data to gain the ability to adapt to new situations and predict patterns.

AI can be defined, succinctly, as the capacity of a computer to make some human activities such as making decisions or learning.
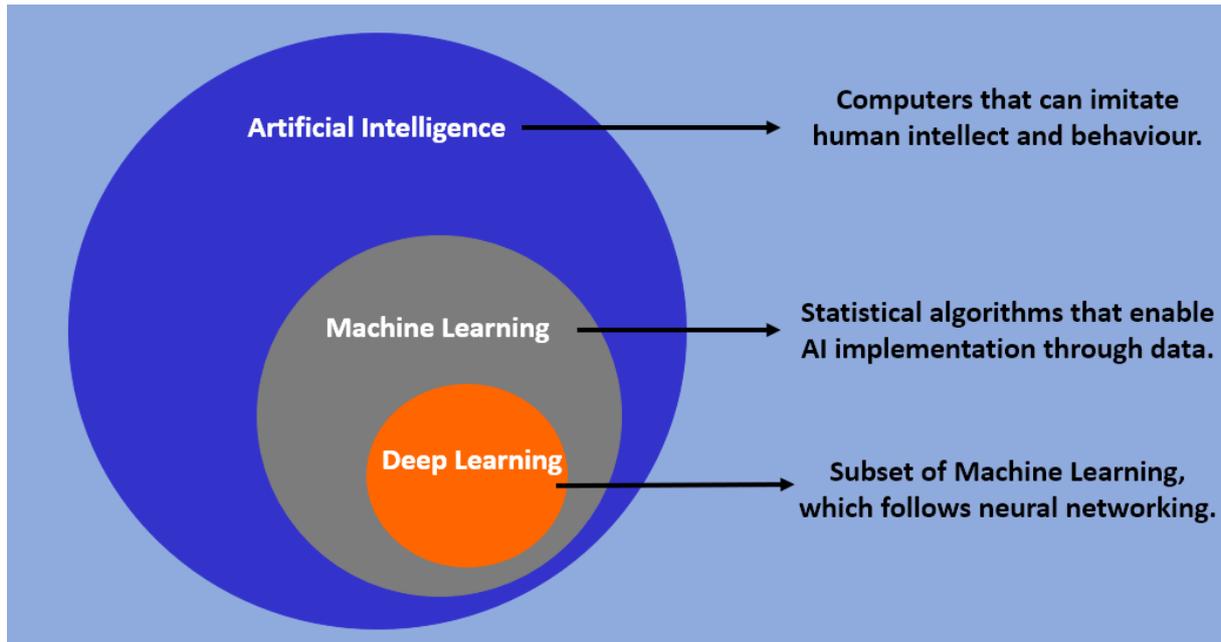
Figure 2-6 - Artificial Intelligence versus Machine Learning versus Deep Learning

ML is generally grouped into two types of learning algorithms: unsupervised learning and supervised learning (Figure 2-7). On one hand, unsupervised learning occurs when ML is used to automatically recognize the patterns in a given dataset. Unsupervised learning algorithms do not know the data (i.e., it is not predefined) for attempting to identify patterns and are most commonly used to identify hidden patterns in data. On the other hand, on supervised learning algorithms, the ground-truth of each dataset case is given and is used known (predefined) input and output data, the model is trained to generate reasonable predictions about the response to new data, where the classification prediction is performed by the computer.
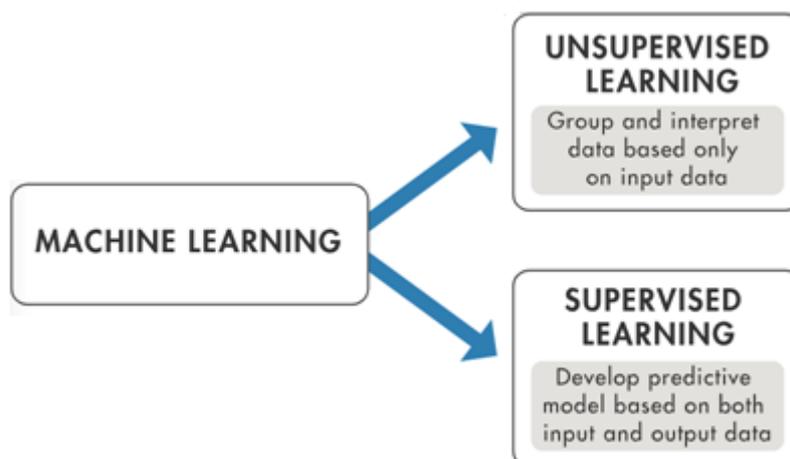
Figure 2-7 - Machine Learning Techniques (adapted from [32])

Deep learning is a subset of machine learning, meaning that it learns features through a hierarchical learning process. Recent trends in the use of deep learning are aimed at faster analysis with better accuracy than human practitioners [11].

Deep learning methods have been reviewed extensively in recent years and it is a field that investigates how simple models of biological brains can be used to solve difficult computational tasks like predictive modelling tasks. The system architecture is based on the biological neural system, but the goal is not to create realistic models of the brain. Instead, the aim is to develop robust algorithms and data structures that can be used to model difficult problems. The brain is capable of performing quite complex computations, and this is where the inspiration for ANN comes from [31]. Neural networks are not models of the brain but are, instead, computational models for solving complex machine learning problems and are comprised of neurons that have weights and activation functions [33].

The predictive capability of neural networks comes from the hierarchical or multi-layered structure of the networks. The data structure can pick out (learn to represent) features at different scales or resolutions and combine them into higher-order features. For example, from lines to collections of lines, to shapes.

The building block for neural networks is artificial neurons. Much like biological neurons, which have dendrites and axons, the single artificial neuron is a simple tree structure, which has input nodes and a single output node, which is connected to each input node. Figure 2-8 shows a visual comparison of the two [31].
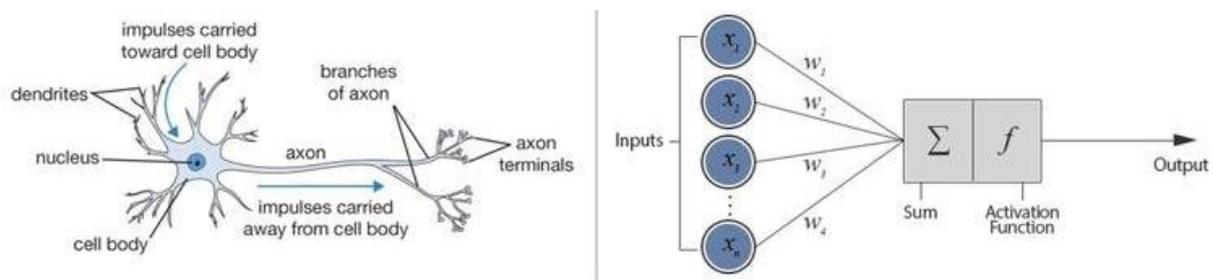
Figure 2-8 - Biological Neuron and Artificial Neural Network (adapted from [34])

Each connection that departs from the input node has a weight associated with it (w_i). Artificial neurons are simple computational units that have weighted input signals and produce an output signal, using an activation function. The weighted inputs are summed and this result will be the input for an activation function, i.e., the weighted sum is passed through an activation function [31]. An activation function is a simple mapping of summed weighted input to the output of the neuron.

Neurons are arranged into networks of neurons [33]. A row of neurons is called a layer and one network can have multiple layers (Figure 2-9).
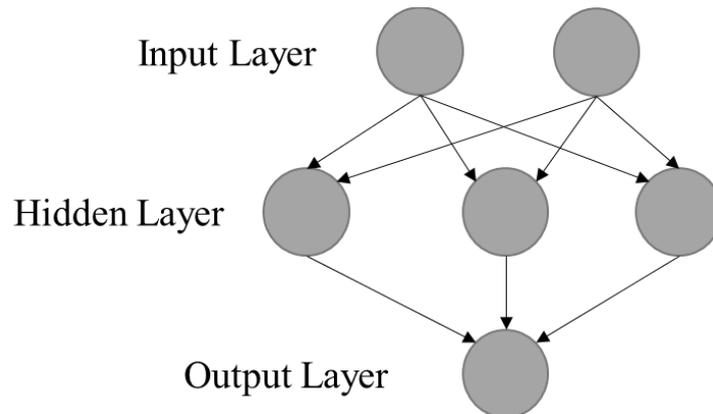
Figure 2-9 - Model of a Simple Artificial Neural Network

The top layer that takes input from the dataset is called the visible layer because it is the part of the network that is directly exposed to the input. Layers that come after the input layer are called hidden layers and the simplest network structure has a single neuron in the hidden layer that directly outputs the value. Given increases in computing power and efficient libraries, very deep neural networks can be constructed, with neural networks can have many hidden layers. The final hidden layer outputs a value or a vector of values and is called the output layer.

A Multi-Layer Perceptron (MLP) is one type of artificial neural network and provides a lot of flexibility and have proven itself over decades to be useful and reliable in a wide range of problems [35]. MLPs are the classical type of a neural network and are suitable for regression, binary, and multi-classification (and typically present good results with classification problems) [36], as well as classification prediction problems where inputs are assigned a class or label [35], which is the case in this dissertation. They are comprised of one or more layers of neurons providing levels of abstraction, where data is fed to the input layer and predictions are made on the output layer.

Once configured, the neural network needs to be trained with the dataset. A consequence of this model is that perceptrons only work with numerical data [31]. This implies converting any nominal data into a numerical format, so a good practice is to prepare the data before training a neural network model [33] and, after that preparation, it can be sent to classification and a score is produced. The information is presented as activation values and is forwarded to the hidden nodes, based on the weights (real numbers that express the importance of the respective inputs to the output) and activation functions, where the activation value is passed from node to node or not. The information flows from node to node

until the output node, which presents the predicted value [33]. Initial feature importance is derived from the delta change in score, from run to run.

Once a neural network has been trained, it can be used to make predictions. It is possible to make predictions on test or validation data to estimate the skill of the model on unseen data [33].

Deep learning identifies optimal features automatically from the data (i.e., the classifier trained by deep learning finds features that have an impact on diagnostic classification without human intervention).

The loss function is used to calculate the difference between the network output and the expected output. The backpropagation calculates the gap repeatedly, changing weights and stopping the calculation when the gap is no longer updated. After the initial loss value is calculated, the weights are updated until the differential value becomes 0 [11].

2.6.1 THE BASIC CLASSIFICATION PROCESS

In order to apply such ML algorithms, appropriate architectural design or pre-processing steps must be predefined. Input images may need to undergo some pre-processing steps, to standardize them. Radiomic features are then extracted, which may include shape, histogram, texture, and high-order features. Feature selection is performed to reduce feature dimension, and the classification model is then developed using selected radiomic features. The final radiomics model is used for the classification of new input images. Classification studies using machine learning generally require five steps [11]:

1. Feature extraction;
2. Feature selection;
3. Dimensionality reduction;
4. Feature-based classification algorithm selection;
5. Classification (training and testing/validation).

Feature extraction involves the transformation of original data into a feature vector that is meaningful to the classifier; this step is mandatory.

Feature selection and dimensionality reduction are steps that prepare for ML and are the transformations capable to run the algorithm most effectively. During the feature selection (reduction) step, more important features are selected for differentiation between classes. Although this step has the potential to improve classification performance, it is optional.

Neural networks are best when the data is one-hot encoded, scaled, and shuffled. Besides, other transformations may be beneficial as well to make relationships easier for the algorithm to identify. One-

Hot Encoding is a process of converting categorical data to sparse data, which has columns of only zeros and ones (also known as dummy variables). All non-numeric data need to be converted to dummy variables. Another transformation that can make the algorithm most effective is to normalize the data, since the neural networks typically perform faster and often with higher accuracy when the features are scaled and/or normalized (also known as standardizing) [36]. The result of standardization is that the features will be rescaled so that they will have the properties of a standard normal distribution and is an important step if measurements that have different units are being compared. Standardizing will process the features so that they are centred around 0 with a standard deviation of 1 [33][37].

After that, the data is split into train and test sets, so it could be used to train and test the multilayer perceptron created, to classify the subjects. The classification consists of 2 sub-steps: training and testing (validation).

Training is used to teach an application to correctly classify data. The ML algorithm is trained through the learning process using training data [29].

Confusion Matrix is another performance measurement for ML and allows the measurement of the effectiveness of the model. A confusion matrix is a breakdown of predictions into a table showing correct predictions and the types of incorrect predictions made. Ideally, the goal is to see only numbers in the diagonal, which means that all the predictions were correct. It is a table (Table 2-2) with 4 different combinations of predicted and actual values [38].

Table 2-2 - Confusion Matrix for Classification Models Evaluation (adapted from [38])

		Actual Values	
		Positive (1)	Negative (0)
Predicted	Positive (1)	TP	FP
	Negative (0)	FN	TN

The values present in Table 2-2 represent:

- True Positive (TP) - predicted positive and it's true;
- True Negative (TN) - predicted negative and it's true;
- False Positive (FP) - (Type 1 Error) predicted positive and it's false;
- False Negative (FN) - (Type 2 Error) predicted negative and it's false.

For the testing sub-step, sensitivity, specificity, and accuracy are the most commonly reported measures of a classifier's performance [11][20]. The metrics used to evaluate the model were the accuracy and other functions that were defined: precision, specificity, False Positive Rate (FPR), sensitivity, and False Negative Rate (FNR).

Precision (equation 2) is a measure of a classifier's exactness: the higher the precision, the more accurate the classifier [31]. Precision is when the model predicts "yes", how often is it actually "yes".

Specificity (equation 3) is defined as the proportion of actual negatives, which got predicted as the negative (or true negative) [39]. It is when the actual value is "no", i.e., is a measure of a classifier's completeness that describes how well the model classifies negative cases as negatives [40]. An example to better understand this is the model used for predicting whether a person is suffering from the disease. Specificity is a measure of the proportion of people not suffering from the disease who got predicted correctly as the ones who are not suffering from the disease. In other words, the person who is healthy actually got predicted as healthy is specificity.

Specificity implies that there will be another proportion of actual negative, which got predicted as positive and could be termed as false positives. This proportion could also be called a false positive rate. The sum of specificity and false positive rate would always be 1 [39].

Sensitivity (equation 5) is a measure of the proportion of actual positive cases that got predicted as positive (or true positive), i.e., sensitivity describes how well the model catches all of the positive cases [40]. Following the same example, sensitivity is a measure of the proportion of people suffering from the disease who got predicted correctly as the ones suffering from the disease. In other words, the person who is unhealthy actually got predicted as unhealthy. This implies that there will be another proportion of actual positive cases, which would get predicted incorrectly as negative (and, thus, could also be termed as the false negative). This can also be represented in the form of a false negative rate. The sum of sensitivity and false negative rate would be 1 [39].

All the metrics used to evaluate the model are defined by the following equations ([39], [40], [38]):

$$1. \quad Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative}$$

$$2. \quad Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$3. \quad Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive}$$

$$4. \quad FPR = \frac{False\ Positive}{True\ Negative + False\ Positive}$$

$$5. \quad Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$6. \quad FNR = \frac{False\ Negative}{False\ Negative + True\ Positive}$$

2.6.2 DEEP NEURAL NETWORKS

Computers can perform numeric computation better than humans. However, it is desirable that they also be able to solve complex perception problems. For this reason, intelligent systems that are inspired by biological neural systems have been extensively studied over the years. Deep Neural Networks (DNN) are based on these systems and have a higher number of hidden layers and are the basic model used in DL. A benefit of very deep neural networks is that their performance continues to improve as they are fed larger and larger datasets. A model with a near-infinite number of examples will eventually plateau in terms of what is the capacity of the network to learn. In DNNs, the DL performs the mapping from a received input to the desired output using the "neurons" of the network, also called a unit, which are nodes that are weighted interconnected (Chartrand, et al., 2017). The outputs of one layer serve as the inputs for the next layer [31].

The objective of a neural network is to have a final model that performs well both on the data that we used to train it (e.g. the training dataset) and the new data on which the model will be used to make predictions. The goal is that the model becomes capable of learning from known examples and then able to generalize from those known examples to new examples in the future. Methods like a train/test split are used only to estimate the ability of the model to generalize to new data.

The capacity of a neural network model and its complexity is defined by both its structure in terms of nodes and layers and the parameters in terms of its weights. If the network has more hidden units, it will be able to learn more complex representations, but it will also be more expensive operations that can be prone to overfitting [31]. Overfitting occurs when the model is too complex: it will describe random error or noise and not the underlying relationship that it needs to describe. Learning, and also generalizing new cases is hard. Too little learning and the model will perform poorly on the training dataset and on new data. The model will underfit the problem. Too much learning and the model will perform well on the training dataset and poorly on new data, the model will overfit the problem (Figure 2-10). An overfitted model performs well on the training data but poorly on other data, reducing the generalizability of the model [29]. In both cases, the model did not generalized. Therefore, to reduce the complexity of a neural network to reduce these problems, two options can be made [41]:

1. Change network complexity by changing the network hyperparameters (values of weights);

2. Change network complexity by changing the network structure (number of weights).

Network Hyperparameters

The neural network parameters influence the results of the network training. These parameters encompass the activation functions, as well as loss functions and optimizers. The optimizer and the loss are two arguments that are required to compile the model [31].

Activation Functions

An activation function decides whether the node is activated or not. The node receives an input value from previous nodes. If the value is large enough, the neural unit passes the value to the next unit. There are three main types of activation functions: Softmax, Sigmoid, and Rectifier Linear Units (ReLU) [11]. Each unit calculates the output differently using the equations in Table 2-3, denote that each unit receives x as input, and f designates the respective output.

Table 2-3 - Activation Functions

Activation Function	Equation	Graphical representation	Proprieties
Softmax	$f(x) = \dfrac{e^{x_n}}{\sum_{c=1}^{n} e^{x_c}}$		Used in multi-classification tasks, normalizing the output of the previous layers to bring the total to one, specifying the probability for each class.
Sigmoid	$f(x) = \dfrac{1}{1 + e^{-x}}$		Usually used in binary classification problems.
ReLU	$f(x) = max\,(0, x)$		Recently, has been used as a hidden unit. Works well with large and consistent gradients.
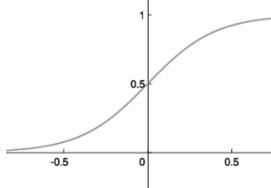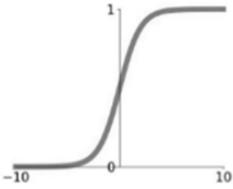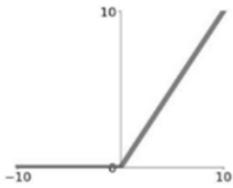
The ReLU function is the most used activation function [11] [31] and was the one chosen in this work, as well as the Sigmoid function.

Loss Functions

The loss function essentially evaluates the predicted value so that the algorithm knows whether the value is correct or not. The predicted value is incorrect if differs from the real value in the training set. Below is an explanation of the binary cross-entropy, the loss function used in this work.

Cross-Entropy (CE) loss for binary classifications can be defined by the equation ([42]):

$$CE = \begin{cases} -log(p), & if \ y = 1 \\ -log(1-p), & otherwise \end{cases}$$

where $y \in \{\pm 1\}$ is the ground-truth class and $p \in [0,1]$ is the estimated probability for the class with the label $y = 1$. This can be easily extended to multi-class problems [42].

Optimizers

To minimize the loss, an optimizer can be used. It searches for the parameters that minimize the loss function. Therefore, in addition to the activation function, advanced gradient descent methods have been developed to solve speed and accuracy issues. A gradient descent method is used to calculate the weights accurately and some of the most popular optimization algorithms used are the Stochastic Gradient Descent (SGD), Adaptive Moment Estimation (ADAM), and RMSprop [31].

Network Structure

A change in the network's structure has an impact on the output produced by the network.

Changing the neural network structure can lead to too little learning (underfitting) or too much learning (overfitting).

Underfitting can be identified from a plot where the accuracy is lower, both on training and testing (Figure 2-10). To address underfitting, which is the case where the model has "not learned enough" from the training data, resulting in low generalization and unreliable predictions, the capacity of the model needs to be increased. Capacity refers to the ability of a model to fit a variety of functions; more capacity means that a model can map more inputs to outputs. Increasing the capacity of a model is easily achieved by changing the structure of the model, such as adding more layers and/or more nodes to layers.

Because an underfit model can be easily addressed, it is more common to have an overfit model. There are two ways to approach an overfit model:

1. Reduce overfitting by training the network on more examples.

2. Reduce overfitting by changing the complexity of the network.

An overfit model is easily diagnosed by monitoring the performance of the model by evaluating it on both a training and a validation dataset. Graphing line plots of the performance of the model during training, called learning curves, will show a familiar pattern: if the model does overfit, it is expected that the value of accuracy on the training set continues to increase and the test set rises and then falls again (Figure 2-10) since the model stops generalizing and just starts memorizing the training data. An overfit model occurs when the validation accuracy diverges [41], as illustrated in Figure 2-10.

Reducing the capacity of the model reduces the likelihood of the model overfitting the training dataset, to a point where it no longer overfits. A Dropout method can be used to reduce overfitting, where inputs are probabilistically removed during training since this type of layer works by probabilistically removing (or "dropping out") inputs to a layer. The main goal of using dropout regularization is to reduce the overfitting of an MLP, as it has the effect of making nodes in the network generally more robust to the inputs [41] since they stop being too specific and can generalize to new examples. The dropout layer is added to a model between existing layers and affects the outputs of the prior layer that are fed to the subsequent layer. Dropout offers a very computationally cheap method to reduce overfitting and improve generalization error in deep neural networks, randomly dropping out nodes during training [43], and this is why it was used a dropout technique to reduce overfitting in the model created in this work.
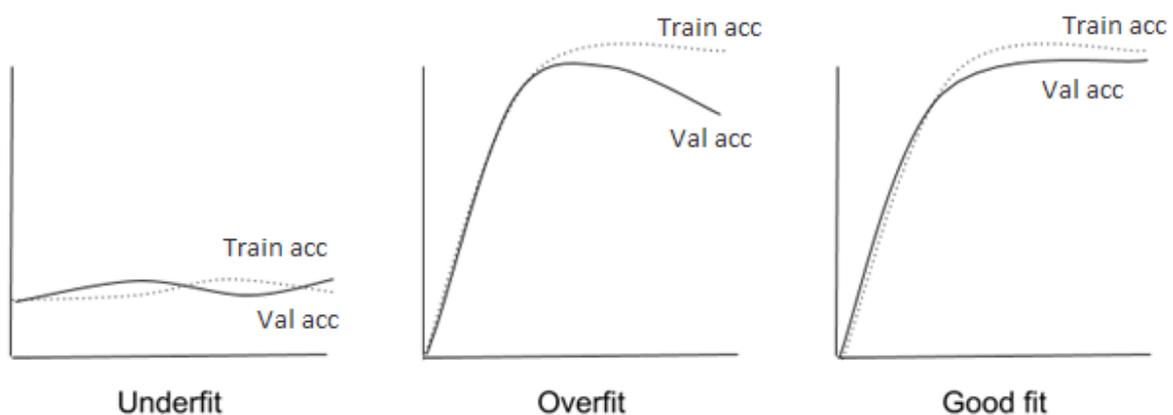
Figure 2-10 –Evaluation metrics during training

2.6.3 MODEL INTERPRETATION

To truly understand how one feature affects the predictions, i.e., what is the feature effect on predictions, there is a need to hold all input values constant and only vary the selected feature to study and understand. To try to investigate which factors influence the final prediction performances is interesting to know which variables are engaged the most, the presence of correlations, the possible causation of relationships, and so on. By measuring the outcome of the prediction, it is possible to draw a relationship between input and prediction. Not all features have the same effect on the target variable: some will have mild inverse relationships and others will have direct positive relationships.

To examine the importance of the features, there is a need to compare the maximum effect that each variable has on the target output.

To address this and discover which features are more significant in terms of classification, several more precise approaches are used to evaluate performance, like Drop-Column Importance, Recursive Feature Elimination, and Permutation Importance. The Permutation Importance was chosen to use in this work.

2.7 WORKBENCH

The tools/framework presented in Figure 2-11 were the supporting technology used in the development environment in order to promote the improvement and knowledge of clinical treatment for AD.

In this dissertation will be implemented a network of perceptrons, i.e., a multilayer perceptron, in Python with the help of Keras (Figure 2-11 – C.) [44]. A MLP consists of multiple neurons that are organized in layers. Keras is a powerful library in Python that provides a clean interface for creating deep learning models and wraps the more technical TensorFlow and Theano backends, and Keras is a package available for deep learning and is one of the most powerful and easy-to-use Python libraries for developing and evaluating deep learning models. It wraps the efficient numerical computation library TensorFlow and Theano [31], which was discontinued, and, therefore, Keras was set to run in the backend, using TensorFlow running on the GPU.

Keras is a high-level neural networks API with a focus on enabling fast experimentation. The default backend is Tensorflow (Figure 2-11 – C.).
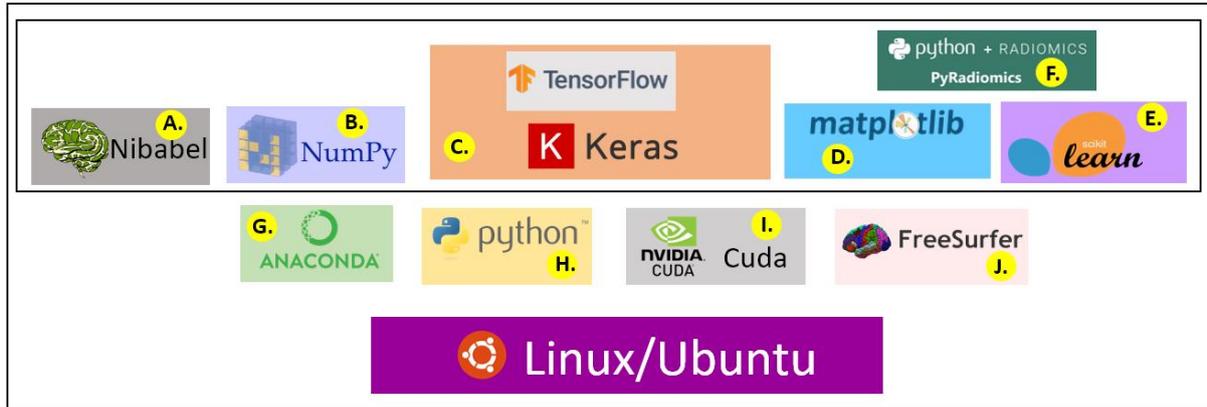
Figure 2-11 – Workbench Support Technology

2.7.1 FREESURFER

FreeSurfer is a software package for the analysis and visualization of structural and functional neuroimaging data from cross-sectional or longitudinal studies. It is developed by the Laboratory for Computational Neuroimaging at the Athinoula A. Martinos Center for Biomedical Imaging [45]. FreeSurfer is used for various modes of analysis and processing of MRI data.

The main goal of using FreeSurfer was to mark the different constituents of the brain with a label.

During the FreeSurfer processing stream, via the recon-all script, some output folders are generated. They are kept in each subjects' directory. An example of one subject is shown in the figure below (Figure 2-12).

Figure 2-12 - Directory structure output of FreeSurfer

The main folder used in this dissertation is the *mri* folder, which contains all the images that were used to apply PyRadiomics. These files correspond to the brain itself (*brain.mgz*) and the image with the

areas of the brain segmented (*aseg.mgz*) that were the output of the application of FreeSurfer to images.

In folder "stats", files are text files with summary information and region of interest summaries. The file *aseg.stats* contains volume summaries and segmentation statistics. The value *SegId* that corresponds to the *StructName* "Left-Hippocampus" was used as a label to extract the left hippocampus from the *brain.mgz* file. The same procedure was followed to extract the right hippocampus. The left hippocampus is identified with SegId = 17 and the right hippocampus with SegId = 53.

2.7.2 PYRADIOMICS

Multiple open-source platforms have been developed for the extraction of Radiomics features from 2D and 3D scans and binary masks and are under continuous development. One of those applications is PyRadiomics [30], which is an open-source radiomics library written in Python that, when applied to medical imaging data, allows the extraction of features from images from an area of the brain.

Since the pre-processing methods can impact significantly the performance of the extracted data, there was a need to standardize feature definitions and image processing, as well as the fact that many studies use in-house developed software, which is often not shared, making reproduction and comparison of the results difficult. These were the main causes that lead to the development of PyRadiomics by Harvard Medical School (Figure 2-13), aiming to develop an easy-to-read and easy-to-use tool. The lack of standardized algorithm definitions and image processing makes reproducibility and comparability of results difficult, and, to address this issue, PyRadiomics was developed. PyRadiomics is a flexible open-source platform capable of extracting a large panel of engineered features from medical images.

Figure 2-13 - Python and Radiomics resulted in PyRadiomics toolbox (adapted from [30])

PyRadiomics is developed as an open-source Python package for the extraction of Radiomics features from medical images [12] and has recently been accepted in Cancer Research [28]. The PyRadiomics platform can extract several radiomics features from medical imaging (such as CT, PET, MRI). It can be used with four main steps (Figure 2-14) [26]:

1. Loading and pre-processing of the image and segmentation masks;

2. Application of enabled filters;

3. Calculation of features using the different feature classes;

4. Returning results.
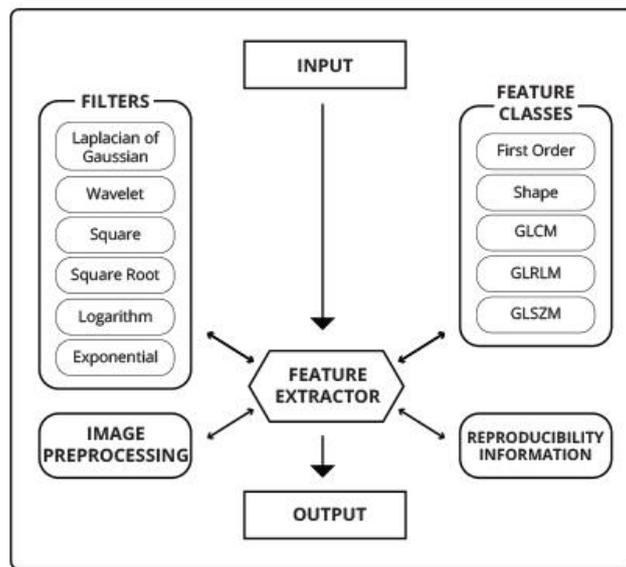
Figure 2-14 - PyRadiomics Toolbox (extracted from [30])

The data is loaded and then converted into NumPy arrays for further calculation using multiple feature classes. Optional filters are also built-in [30], which makes it possible to customize the extraction. For this, PyRadiomics offers 4 different ways of doing this [54], but all 4 categories of customization can be provided in a single YAML or JSON structured text file, which can be used during the initialization of the feature extractor. This removes the need to hard code a customized extraction in a Python script using functions made available in different Python classes instances. Additionally, this also makes it easier to share settings for customized extractions.

2.7.3 TENSORFLOW

TensorFlow is the core open-source library to develop and train ML models [46]. It allows building and training ML models using intuitive high-level APIs, like Keras (Figure 2-11 - C.). TensorFlow allows to

train and deploy the model, no matter what language or platform is used. The relationship between TensorFlow and Keras comes from the fact that TensorFlow's high-level APIs are based on the Keras API standard for defining and training neural networks. TensorFlow allows building and training models using Keras. *tf.keras* is a high-level API to build and train models [46]. Keras is an open-source deep learning library written in Python.

The project was started in 2015 by Francois Chollet and it quickly became a popular framework for developers, becoming one of, if not the most, popular deep learning libraries. Keras was popular because the API was clean and simple, allowing standard deep learning models to be defined, fit, and evaluated in just a few lines of code. A secondary reason Keras took off was that it allowed using one of the many among the range of deep learning mathematical libraries as the backend, such as TensorFlow and Theano. This allowed the power of these libraries to be harnessed (e.g. GPUs) with a very clean and simple interface. In 2019, Google released a new version of their TensorFlow deep learning library that integrated the Keras API directly [47].

TensorFlow offers multiple levels of abstraction and allows building and training models by using the high-level Keras API [46].

TensorFlow is a Python-friendly open-source library for numerical computation that makes ML faster. Implementing ML models is far less daunting and difficult than it used to be, thanks to ML frameworks such as Google's TensorFlow, which ease the process of acquiring data, training models, serving predictions, and refining future results. Created by the Google Brain team, TensorFlow is an open-source library for numerical computation and large-scale ML. The single biggest benefit TensorFlow provides for ML development is abstraction: instead of dealing with the details of implementing algorithms, the developer can focus on the overall logic of the application while TensorFlow takes care of the details behind the scenes [48].

3 LITERATURE REVIEW

Key points:

- The present chapter is not about a detailed description of the technologies, but about relevant aspects that were important in fine-tuning the proposed solutions.

- In this section, the latest ideas having or using the latest techniques or equipment and methods are presented.

- A study was carried out to find out which is the most recent stage in the development of neuroimaging and Deep Learning techniques, relating the most recent stage of technological development incorporating the newest technology, ideas, and features.

- Although some challenges remain, interest in the use of ML algorithms for the prognosis and diagnosis of brain activity and diseases continues to increase.

3.1 NEURODEGENERATIVE DISEASES

The World Health Organisation estimates that there are 47.5 million people with dementia worldwide, a number that could rise to 75.6 million by 2030 and almost triple to 135.5 million by 2050 [1].

AD is a major and most common cause of dementia among people over the age of 65 and accounts for about 60-70% of all dementia cases [49]. Yet, no prevention methods or cures have been discovered. It is an irreversible progressive neurodegenerative disease that results in a loss of mental function caused by the deterioration of brain tissue.

It is estimated that, in Portugal, 153 thousand people are affected by some type of dementia throughout their lives [12]. In the Organisation for Economic Co-operation and Development (OECD) report ''Health at a Glance 2017'' published in November 2017, Portugal is the 4th country with the most cases per thousand inhabitants. The OECD average is 14.8 cases per thousand inhabitants, while for Portugal the estimate is 19.9 cases [49]. By far, AD is the most commonly diagnosed cause of dementia [17].

Progressive cerebral atrophy is a characteristic feature of neurodegeneration that can be visualized in life with MRI. When pathological changes occur in the brain tissue, the texture of the MRI may change correspondingly [27]. This provides an important imaging biomarker for the diagnosis of AD and MCI. The major contributors to atrophy are thought to be dendritic and neuronal losses. Studies of regional

(e.g., hippocampal) MRI volumes have shown these are closely related to neuronal counts at autopsy [50][51].

The fact that pathologically increased cerebral atrophy starts early (even pre-symptomatically), continues relentlessly, at least until individuals are severely affected, and correlates with clinical decline has led to atrophy on MRI being suggested as a marker of disease progression and a potential outcome measure in trials. The amount, distribution, and rate of cerebral atrophy are all closely correlated with cognitive deficit [19].

To date, AD or MCI diagnosis is established after irreversible brain structure alterations. Therefore, the development of new biomarkers is crucial to the early detection and treatment of this disease. At present, there exist some research studies showing that radiomic analysis can be a good diagnosis and classification method in AD and MCI. Radiomic analysis has been applied to some studies about neuropsychiatric diseases, including AD [27].

3.2 NEUROIMAGING OF NEURODEGENERATIVE DISEASES

Imaging has played a variety of roles in the study of AD over the past four decades, as the field of biomedical engineering has infused the clinical neurosciences with powerful neuroimaging instruments equipped to study directly morphological and functional properties of the ageing brain *in vivo*. Advances in the acquisition, visualization, and analysis of neuroimaging data continue to evolve rapidly, with the ongoing development of hardware, software, and conceptual statistical approaches that have already made tremendous scientific contributions [17]. Initially, computed tomography and then MRI were used diagnostically to rule out other causes of dementia [19]. Neuroimaging techniques are widely used in neuroscience to visualize neural activity, improve the understanding of brain mechanisms, and identify biomarkers [20][27]. Especially in the past 20 years, neuroimaging made quite a splash because of the improvements in computing technology and improved the understanding of the mechanisms of the brain as well as the ability to detect the cause of impairment via classification of patients and healthy controls [11][19][26]. There has been a transformation in the part played by neuroimaging in AD research and practice in the last decades. Diagnostically, imaging has moved from a minor exclusionary role to a central position. In research, imaging is helping address many of the scientific questions, providing insights into the effects of AD and its temporal and spatial evolution [19][52]. Furthermore, neuroimaging techniques are invaluable for identifying potential neurobiological markers and generating predictions for preventing the progression of various diseases [20], noninvasively evaluating brain

structure and function in many ways. Historically, imaging (first, computed tomography and then MRI) was used only to exclude potentially surgically treatable causes of cognitive decline. Now its position in diagnosis also includes providing positive support for a clinical diagnosis of AD in symptomatic individuals by identifying characteristic patterns of structural and functional cerebral alterations [19][27]. In particular, advanced neuroimaging techniques, such as magnetic resonance imaging and positron emission tomography, have been developed and used to identify AD-related structural and molecular biomarkers [11]. Some image analysis methods, including functional MRI (fMRI), structural MRI (sMRI) and Diffusion Tensor Imaging (DTI), are currently wildly used in detecting the brain structural and functional changes of AD and MCI [17][27][53].

A great deal of effort has been made to develop strategies for early detection of AD, especially at pre-symptomatic stages in order to slow or prevent disease progression [11][27]. Alongside this increasing specificity for AD, imaging also contributes to differential diagnosis in practice by identifying alternative and/or contributory pathologies. Imaging is central to identifying vascular and non-AD degenerative pathologies and has helped in the recognition of the prevalence of mixed pathology in dementia [19].

Interpretation of findings and identification of biomarkers, especially for neuropsychiatric diseases, is not always an easy process. Therefore, neuroimaging studies that use machine learning as a prognostic/diagnostic tool are becoming more common [20].

Most of the present AD radiomic studies implemented image pre-processing, and the pre-processing methods include skull stripping, denoising, bias field correction, image inhomogeneities correction, and intensity normalization [26][27]. However, there is still no uniform standard for the method selection and performing sequence of image pre-processing.

3.3 BRAIN MORPHOLOGICAL FEATURES

The targeted brain dysfunction that accompanies ageing can have a devastating effect on cognitive and intellectual abilities. A significant proportion of older adults experience precipitous cognitive decline that negatively impacts functional activities. Such individuals meet clinical diagnostic criteria for dementia, which is commonly attributed to AD. Structural neuroimaging, including MRI, has contributed significantly to our understanding of the morphological and pathology-related changes that may underlie the normal and disease-associated cognitive change in ageing. Some studies of texture analysis have found that there are texture differences between AD patients and normal controls in particular structures such as the hippocampus, corpus callosum, and thalamus [27]. Therefore, imaging has

played a variety of roles in the study of AD over the past forty years [11][17][19]. There has been a transformation in the part played by neuroimaging in AD research and practice in the last decades. Simultaneously, the potential of brain imaging has expanded rapidly with new modalities and ways of acquiring images and analysing them. Diagnostically, imaging has moved from a minor exclusionary role to a central position. In research, imaging is helping address many of the scientific questions, providing insights into the effects of AD and its temporal and spatial evolution [19].

MRI became the most used medical imaging technique diagnostically to rule out other causes of dementia. More recently, a variety of imaging modalities, including structural and functional MRI, studies of cerebral metabolism have shown characteristic changes in the brains of patients with AD.

AD is recognized to have a long pre-symptomatic period, during which there is a progressive accumulation of molecular pathology, followed by inexorable neuronal damage. The ability to identify pre-symptomatic individuals with evidence of neurodegenerative change, stage their disease and track progressive changes will be important for early diagnosis and prevention trials. Despite recent advances, particularly in MRI, the ability to identify early neurodegenerative changes reliably is limited. Because of this, the main objective of this dissertation is to try to identify which areas of the brain can predict the development of neurodegenerative, in particular the AD, using, for this, the FreeSurfer [45] tool to segment the areas of the brain and PyRadiomics [30], to extract features from the brain and the areas that were segmented, and compare the value of this features in MRI taken from patients along time. Some studies have already been made, and the results are several. Poor memory ability and small hippocampal volume measurements in magnetic resonance images co-occur in neurological patients. Numerous studies have examined the relationship between memory performance and hippocampal volumes in participants without neurological or psychiatric disorders, with widely varying results [51][54]. Some studies already made used diffusion tensor imaging, instead of MRI, where the main focus was to investigate microstructural changes in cortical and white matter pathways in patients with Alzheimer's disease using DTI files [52][53], for example.

Using MRI, Scola et al. assessed whether brain changes detected by DTI can improve the understanding of structural damage in AD and are associated with different risks of conversion to AD in MCI [52]. The conclusions they reached were that subtle brain diffusivity changes occur from the prodromal stages of AD, mainly in posterior brain regions, and spread throughout the disease to involve the frontal lobe.

Rates of hippocampal and whole-brain atrophy on MRI have to date been the most widely included imaging measures in trials. However, other MRI measures are promising [19][53][55]. Radiomics aims

to quantify phenotypic characteristics of medical imaging through the use of automated algorithms. Radiomic AI technology, either based on engineered hard-coded algorithms or deep learning methods, can be used to develop non-invasive imaging-based biomarkers. At present, radiomic analysis has been applied to a variety of tumour and non-tumour diseases [26][27][29]. The potential of radiomics has been shown across multiple tumour types, including the brain, head and neck, cervix, and lung cancer tumours. Furthermore, these data, extracted from MRI, PET, or CT images, were associated with several clinical outcomes, and hence, potentially provide complementary information for decision support in clinical oncology [19][26].

Radiomics refers to the comprehensive and automated quantification of this radiographic phenotype using data characterization algorithms and can quantify a large panel of phenotypic characteristics, such as shape and texture, potentially reflecting biologic properties like intra- and intertumoral heterogeneities.

3.4 NEUROIMAGING MODEL

ML techniques have recently been introduced for use in identifying diseases and biomarkers of disease [20]. Interpretation of findings and identification of biomarkers, especially for neuropsychiatric diseases, is not always an easy process. Therefore, neuroimaging studies that use machine learning as a prognostic/diagnostic tool are becoming more common. Another motivation to use ML models is the fact that the size of datasets has become too large in which traditional data processing techniques are not able to cope with effectively. However, analyzing complex, high dimensional, and noise-contaminated datasets is a huge challenge, and it is crucial to develop algorithms that can summarize, classify, extract important information and convert them into an understandable form. To undertake these problems, DL models have shown outstanding performances in the recent decade [56]. It is possible to use statistical and ML methods to select the most valuable features for the accurate diagnosis and treatment of clinical diseases.

In recent years, the number of studies based on supervised learning algorithms, especially in a classification framework (e.g., ANN), for the prognosis and diagnosis of diseases has been increasing (neuropsychiatric diseases, depression, uni/bipolar depressive disorder, anxiety disorder, social anxiety disorder, for example) [20][28][29].

Although some challenges (such as sample size) remain, interest in the use of ML algorithms for decoding brain activity continues to increase [19][20][27].

Neuroimaging data have been used to identify structural and molecular/functional biomarkers for AD. It has been shown that volumes or cortical thicknesses in pre-selected AD-specific regions, such as the hippocampus, could be used as features to enhance classification accuracy in ML.

Deep learning, a state-of-the-art ML approach, has shown outstanding performance over traditional ML in identifying intricate structures in complex high-dimensional data, especially in the domain of computer vision. In recent years, various deep learning architectures have been quickly introduced in different domains of application such as medical diagnosis, self-driving cars, natural language and image processing, and predictive forecasting, to develop machines that can perform similar or even better than human learning, with many case studies having great results [31][56].

Deep learning, an emerging area of machine learning research that uses raw neuroimaging data to generate features, is attracting considerable attention in the field of large-scale, high-dimensional medical imaging analysis. Deep learning methods, such as Convolutional Neural Networks (CNN), have been shown to outperform existing machine learning methods and these methods for classification or prediction have been applied in various fields, including computer vision and natural language processing [11][42]. Deep learning techniques have been applied to a wide variety of problems in recent years with the recent advancement in digital technologies, and in many of these applications algorithms based on deep machine learning have surpassed the previous state-of-art performance [57]. The capability of DL models in learning hierarchical features from various types of data, e.g., numerical, image, text and audio, makes them powerful in solving recognition, regression, supervised and unsupervised problems [56].

The application of deep learning to early detection and automated classification of AD has recently gained considerable attention, as rapid progress in neuroimaging techniques has generated large-scale multimodal neuroimaging data. Some model building or machine learning methods have gradually become the focus of AD radiomic studies [27][29][53]. These radiomic features and classification models are considered to have potential as biomarkers for the diagnosis of AD and MCI patients. A systematic review of publications using deep learning approaches and neuroimaging data for diagnostic classification of AD was performed [11][19][27].

Deep learning approaches, such as a convolutional neural network or a Recurrent Neural Network (RNN), that use neuroimaging data without pre-processing for feature selection have been used to select features from neuroimaging data and have yielded accuracies of up to 96.0% for AD classification and 84.2% for MCI conversion prediction [11].

To improve the performance for AD/CN classification and for the prediction of MCI to AD conversion, neuroimaging data such as MRI have commonly been used in deep learning [11][27].

Because of its ease of use and better performance, deep learning has been used increasingly for medical image analysis. The number of studies of AD using deep learning algorithms has increased drastically since 2015 [8][11][26].

4 MATERIALS AND METHODS

Key points:

- In this section, a dataset with MRI images from study volunteers provided by ADNI was used. Participants are followed and reassessed over time to track the pathology of the disease as it progresses. The dataset selected contains 600 patient cases (Figure 4-10).

- In this work, Deep Learning techniques were used. To classify the three classes and distinguish the clinical group of a subject, a multi-perception layer network (Figure 4-21) was designed for classification, using TensorFlow backend.

- The best performance was achieved with an ANN architecture (Figure 4-21), using the dataset distribution showed in Figure 4-20. The best approach achieved a classifier error of 0.83% (Figure 4-31).

4.1 OVERALL WORKFLOW

The main objective of this dissertation is to identify groups of diseases, in specific AD, that can be found in the images that compose the dataset. The goal is to try to distinguish between the different groups (CN, MCI, and AD) using, for this, features obtained through the use of the PyRadiomics tool.

The hippocampus, which corresponds to the region of interest in this work, was segmented from magnetic resonance imaging and radiomic features were extracted and selected to construct and train the MLP.

Combined with appropriate feature selection and classification methods, radiomic features were examined in terms of their performance and stability for predicting prognosis. Decoding mental states aid in predicting the course of diseases using brain signals and, for this purpose, a model is used to examine significant differences, for example between patients and healthy controls.

In this work, was demonstrated an application of PyRadiomics for neurodegenerative diseases, in particular for AD characterization, to discriminate between normal patients (CN group) from those who have mild problems with cognition (MCI group), and from patients that were already diagnosed with dementia (AD group). Was used the data available in the ADNI database [15], which consists of a database that allows the tracking of the progression of AD in the human brain through the process of normal ageing, mild cognitive impairment to dementia or AD.

1217 subjects' information and scans downloaded from ADNI went to FreeSurfer to segment the brain images. This information and scans correspond to 4712 acquisitions. From this 1217 patients, were selected 600 subjects that posteriorly went through a method internally developed in Life and Health Sciences Research Institute (ICVS).

In total, 4312 radiomic features (6 diagnostic features, 186 exponential, 186 gradient, 186 LBP2D, 558 LBP3D, 186 logarithm, 186 features for each of the 5 sigmas of LoG, 214 original features, 186 square, 186 square root, and 1488 wavelet features) were extracted from all MRI images of those 600 subjects. Although the high number of features extracted, not all of them helped to predict clinically significant changes in the hippocampus. Therefore, we used two feature selection methods to determine the best discriminating feature set. First, a cleaning step was performed to eliminate the redundant and irrelevant features. Then, the final MLP model was constructed and *eli5.permutation_importance* analysis was conducted to try to shrink and identify which of these features had a major influence in distinguishing and identifying the three group classes.

The process workflow to study MRI scans using Radiomics is illustrated in the following diagram (Figure 4-1):
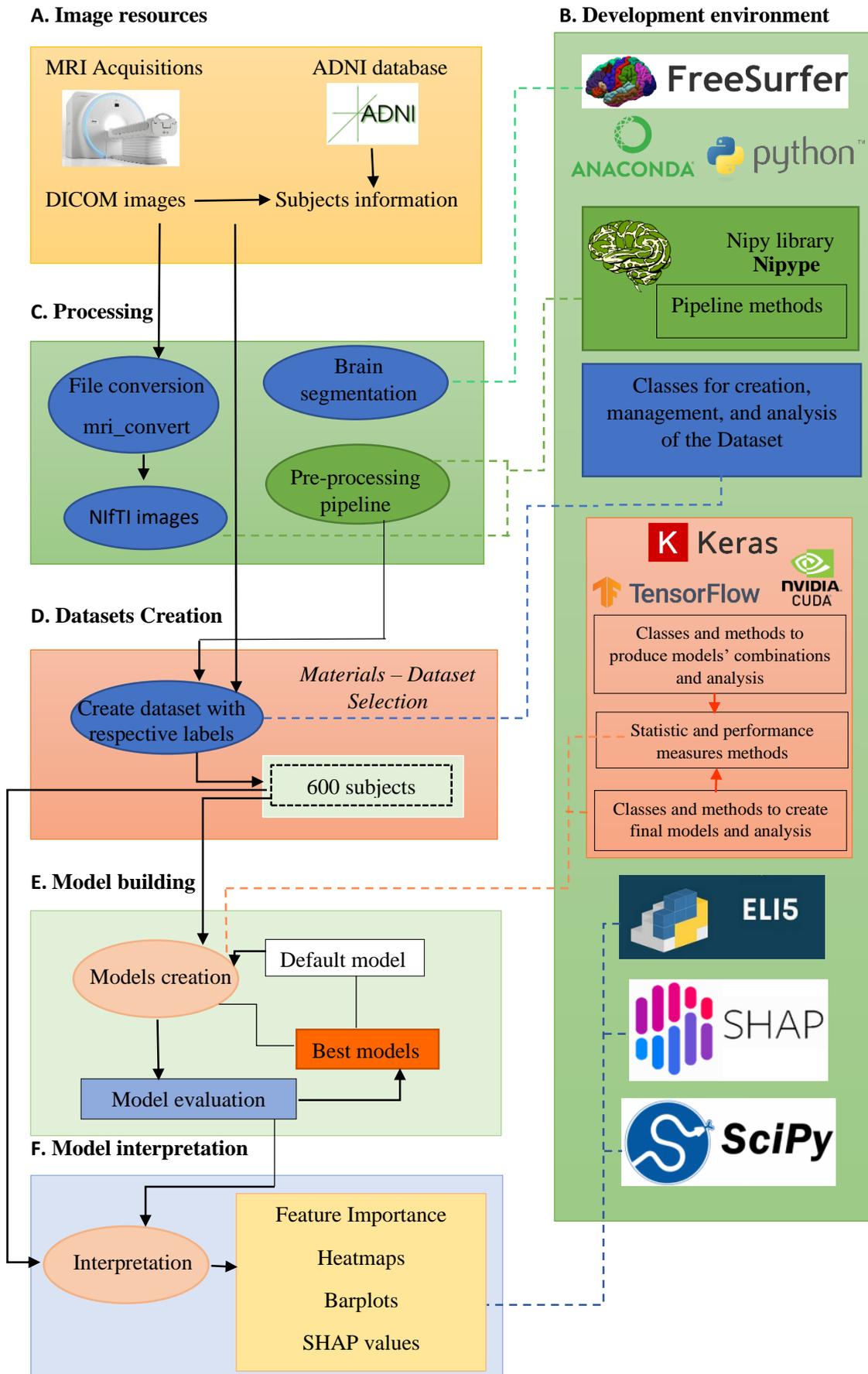
Figure 4-1 - Process workflow to study magnetic resonance scans, using Radiomics

The process pipeline to study magnetic resonance scans, using Radiomics, was developed with 12 steps:

1. Get the MRI scans from the ADNI database;

2. Hippocampus extraction process, using FreeSurfer;

3. Creation of the dataset, with a selection of 600 subjects;

4. Datatypes conversion;

5. Radiomics feature generation;

6. Data cleaning;

7. Encoding of data;

8. Combine outputted data from Radiomics from different groups;

9. Normalization of data;

10. MLP Modelling Process;

11. Bounding Box;

12. Model and results interpretation.

Each step is explained in subsection 4.3.

4.2 DATASET

The dataset creation process is schematised in Figure 4-2.
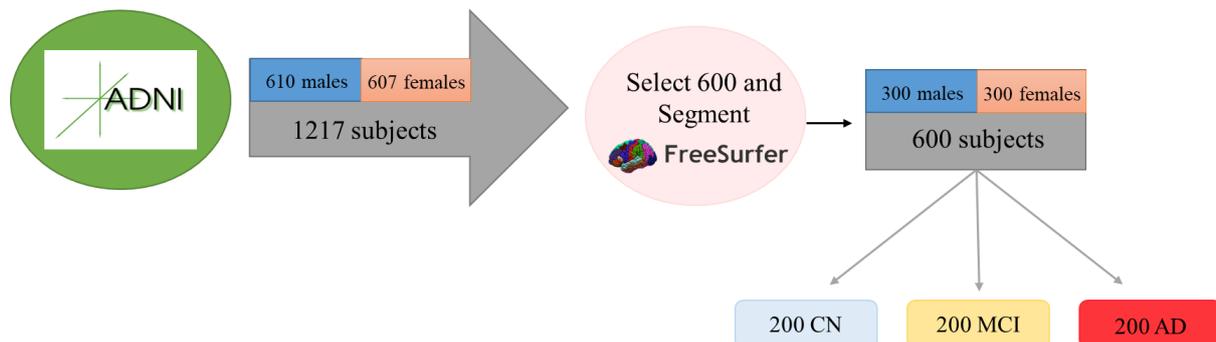
Figure 4-2 - Dataset creation and composition

It is composed of brain medical scans from the ADNI dataset, which were downloaded from the ADNI database. Initially, there were 4712 acquisitions from 1217 subjects. The number of acquisitions differs from the number of subjects since ADNI is a database that unites work developed by researchers to define the progression of Alzheimer's disease, i.e., a patient can have acquisitions made in different stages of the pathology. Of these 1217 subjects, 607 are females and 610 are males.

All the 1217 subjects went to FreeSurfer to be processed and, from these 1217 subjects, a selection of 600 subjects was made. These 600 subjects also went through a method that was internally developed in the ICVS.

The focus was on these 600 subjects, and the selection of them was made so that the composition of the three groups wouldn't have a bias, i.e., would have the same number of subjects and the distribution by age and gender between them would be equal. Thereby, each group is composed by (Figure 4-2):

- 200 subjects of the control group;
- 200 subjects with mild cognitive impairment;
- 200 subjects diagnosed with Alzheimer's disease.

Of the 200 subjects of the control group, half of them are females and the other half are males, i.e., 100 are males and the rest are females. In MCI, the distribution of participants by gender was equally divided and the same happens in the AD groups. This need to equally divide the study data is made to guarantee that these parameters (gender and number of subjects) do not influence the final classification made by the MLP.

In these 600 subjects, the average of ages is, for the CN group, 74.28 years, 74.26 in the MCI group and in the AD group is around 74.35 years.

This distribution and characterization of the dataset are represented in the figures created in the Jupyter Notebook for visualization, present in section 4.3.4.

In this dissertation, the focus is to use MRI scans and the collection of these scans was collected from ADNI participants. MRI data is one component of the comprehensive data set collected in ADNI participants, obtained in studies that have combined public/private collaborations between academia and industry. For this reason, the dataset that will be used in this dissertation is filled with ADNI data, which is in DICOM format. Imaging files available on the ADNI website have every stage of pre and post-processing. The data downloaded from the ADNI contains multiple files and folders and one of them is the folders with MPRAGE in the name that store the DICOM images that were used in this dissertation. These files are considered the best in the quality ratings, and, therefore, this was the reason to choose these files. These files have undergone some image corrections that were already provided by ADNI, including gradwarping, intensity correction, and the images have been scaled for gradient drift using the phantom data. ADNI provided itself intensity normalized and gradient un-warped image volumes for most of all exams in the database, but, as MR manufacturers offered these corrections online as part of the product, ADNI has stopped performing its own pre-processing and instead employed the pre-

processing performed by manufacturers [58]. ADNI provides various MRI techniques of images, but in this dissertation were selected only variations of MPRAGE acquisitions, since this technique is considered to obtain the best images in the quality ratings and were already pre-processed, with corrections of intensity, histogram, corrections made to prevent the errors that can follow when patients move during the exam acquisition, or other variations.

Each image in MPRAGE folders in the database has undergone specific image pre-processing correction steps. These corrections are as follows:

- Intensity correction (or intensity normalization, also known as bias correction) - the outside of the image is much brighter than the inside and this makes grey/white segmentation difficult;

- Gradwarp - gradwarp is a system-specific correction of image geometry distortion due to gradient non-linearity. The degree to which images are distorted due to gradient non-linearity varies with each specific gradient model;

- B1 non-uniformity - this correction procedure employs the B1 calibration scans to correct the image intensity non-uniformity;

- N3 - Nonparametric Nonuniform Intensity Normalization (N3) is a histogram peak sharpening algorithm that is applied to all images. It is applied after gradwarp and after B1 correction for systems on which these two correction steps are performed. This correction allows reducing residual intensity non-uniformity.

4.3 METHODS

4.3.1 GET THE MRI SCANS FROM THE ADNI DATABASE

Deep learning approaches require massive amounts of data to achieve the desired levels of performance accuracy. The MRI scans were obtained from ADNI (Figure 4-3) [15], which provides a large number of images that went already through some types of corrections, to ensure that the images have the best quality possible [58].

In this step, medical images will be loaded into the platform. The large majority of image handling is done using *NiBabel* [59] (Figure 2-11 – A.).
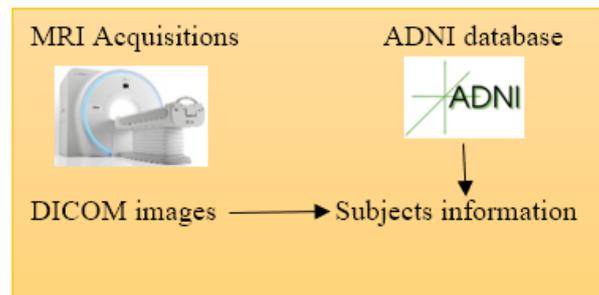
Figure 4-3 - Image resources

4.3.2 DATATYPES CONVERSION

The scans provided by ADNI are in DICOM format and, since PyRadiomics does not support this type of file, there was a need to convert these scans, that previously went through FreeSurfer (Figure 2-11) [45] segmentation, to NIfTI data type.

The outputted files from FreeSurfer are in MGZ type and these files are the input to PyRadiomics. However, PyRadiomics does not work with this type of data. Therefore, it was necessary to convert the MGZ format into NIfTI format before applying PyRadiomics The files returned by FreeSurfer are of type MGZ and these files are the input for PyRadiomics. However, PyRadiomics does not work with this type of data. It was therefore necessary to convert the MGZ format to the NIfTI format before using PyRadiomics (Figure 4-4).
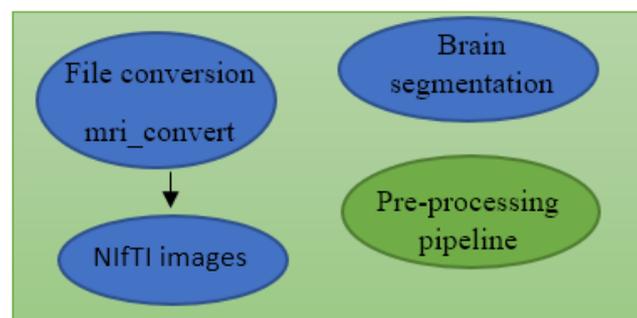
Figure 4-4 – Processing: file conversion step

To make this conversion, the FreeSurfer's command *mri_convert* [60] was used. *mri_convert* attempts to convert images from the proprietary scanner format to the NIfTI format. This command is shown in Figure 4-5 and was used in a bash script:

```
mri_convert aseg.mgz aseg.nii
mri_convert brain.mgz brain.nii
```

Figure 4-5 – Command to convert MRI to NIfTI

4.3.3 HIPPOCAMPUS EXTRACTION PROCESS, USING FREESURFER

The main goal is to segment the MRI images of the brain into the different brain areas. FreeSurfer [45] was used to segment and mark the different constituents of the brain with a label and allow the removal of non-brain tissue from the anatomical brain MRI images. DICOM data from ADNI was processed in FreeSurfer to segment the brain into structures (Figure 4-4). To segment the brain volumes in the different areas, FreeSurfer (Figure 2-11) [45] was used. This step was fundamental to make the extraction of the left and right hippocampus, which are the areas that are known to have more impact on AD.

To extract both left and right hippocampi of every subject, there was a need to merge *aseg.mgz* file with information present in a file located in the folder *stats*, in the file *aseg.stats*. This *aseg.stats* file has information about the thresholds that label and identify all the areas of the brain segmented from FreeSurfer who made it possible to identify and extract both hippocampi, which correspond to the main brain structures affected by AD. This label was then used to extract and segment these two areas. The left hippocampus is referenced with label 17 and the right hippocampus has the label 53. These values were used to extract both hippocampi for every subject and, to make this extraction, the command *fslmaths* [61] was used with each label with the following command line:

```
fslmaths aseg.nii -thr 17 -uthr 17 Left-Hippocampus
fslmaths aseg.nii -thr 53 -uthr 53 Right-Hippocampus
```

Figure 4-6 - Command to extract hippocampi masks from segmented brain

The commands showed in Figure 4-6 yielded *Left-Hippocampus.nii.gz*, a compressed NIfTI file, and the same was obtained for the right hippocampus. To unzip the compressed hippocampus, the following command was made:

```
gunzip -f -k Left-Hippocampus.nii.gz
gunzip -f -k Right-Hippocampus.nii.gz
```

Figure 4-7 - Command to unzip compressed NIfTI hippocampus files

The files outputted with the commands shown in Figure 4-7, which correspond to the left hippocampus and right hippocampus (*Left-Hippocampus.nii and Right-Hippocampus.nii*), were passed to PyRadiomics to be used as a mask.

4.3.4 CREATION OF THE DATASET, WITH A SELECTION OF 600 SUBJECTS

After applying and accessing the data, the subjects' information was downloaded from ADNI. The ADNI participants are staged and classified according to biomarkers and clinical measures. CN participants are the control subjects in this work, showing normal ageing. They show no signs of depression, MCI, or dementia. MCI participants are the ones who have reported a subjective memory concern. Their daily activities are still essentially preserved and there are no significant levels of impairment in other cognitive domains, nor signs of dementia. Participants with AD meet the criteria for AD, having deficits in at least 2 of the 8 areas of cognition, which are memory, language, perceptual skills, attention, constructive abilities, orientation, problem solving and functional abilities, presence of an early and significant episodic memory impairment, presence of medial temporal lobe atrophy and volume loss of hippocampi.

At the time of enrolment, all subjects were between 55 and 92 (inclusive) years old and were in good general health with no diseases precluding enrolment. Work history or 6 grades education was verified. Specific psychoactive medications were excluded. Permitted medications dosage was stable for 4 weeks before evaluation.

The structural scans were converted to NIfTI format, downloaded, and further analysed. This analysis started with selecting 200 CN control participants, 200 MCI individuals, and 200 patients with AD (Figure 4-8). After that, the information about each participant was arranged in a Comma Separated Values (CSV) table (Figure 4-9), appending the label that identifies their group and other relevant information about the subjects. Selected subjects are aged between 55 and 92 years at the moment of MRI acquisition. Men and women were equally distributed in each clinical group (Figure 4-10).

D. Datasets Creation

Figure 4-8 – Creation of the dataset

	Codes	Age_at_img_acquisition	Clinical_group	Sex	intracranial_volume
1	002_S_1261	81.2	CN	F	1481008
2	002_S_1280	80.8	CN	F	1419644
3	002_S_4213	84	CN	F	1470755.31
4	002_S_4225	77.1	CN	M	1909141.22
5	002_S_6009	67.6	CN	M	1593096

Figure 4-9 – Dataset

To visualize and analyze all of the dataset, using graphs and schemes, was created a Jupyter notebook. The data visualization library used was *matplotlib* [62] to plot the distributions quickly. The following images show the output obtained.
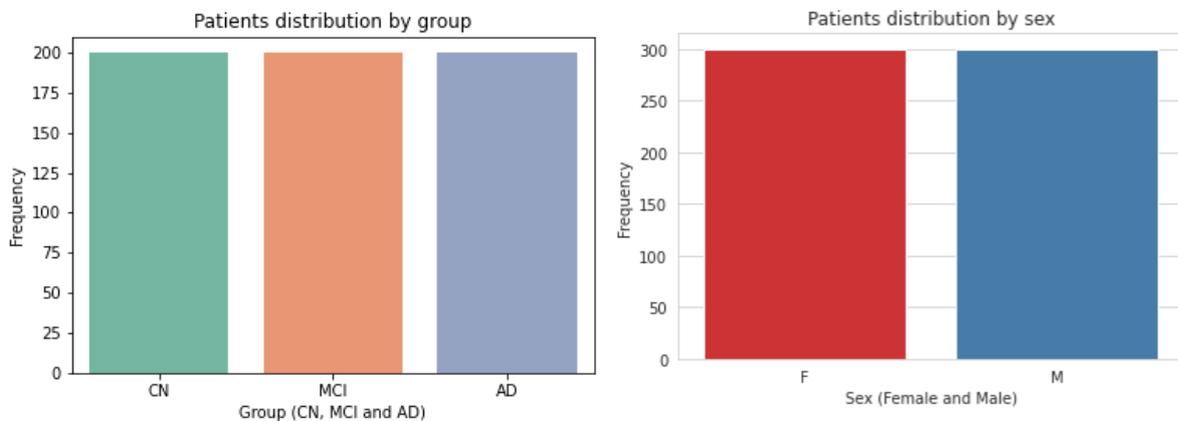
Figure 4-10 - Graph of patients' distribution (by group in the left and by sex in the right)
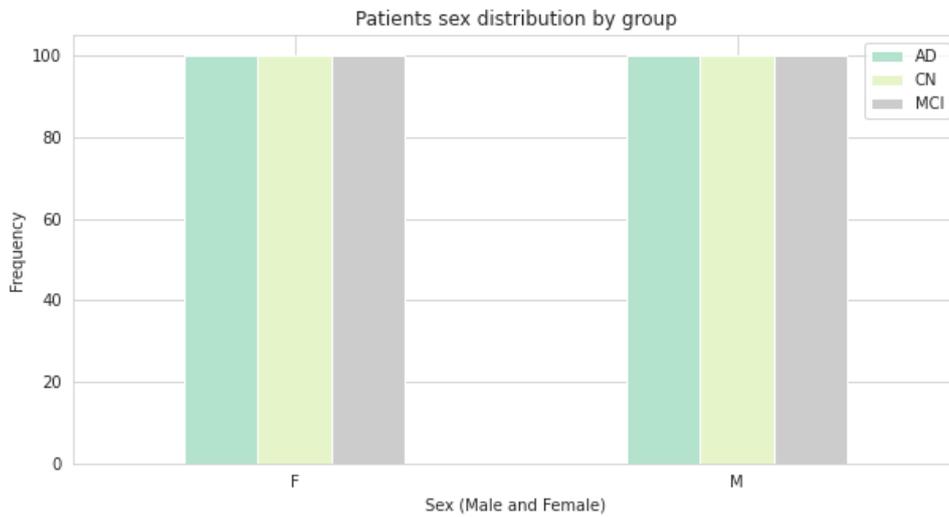
Figure 4-11 - Graph of patients' sex distribution by group

From the left barplot present in Figure 4-10, is possible to verify that the three different classes (CN, MCI, and AD) have 200 subjects each. The barplot on the right makes it able to see that the 600 subjects are equally divided by sex.

The 300 subjects of each gender are also equally distributed by each group: 100 of these 300 subjects belong to the CN group, 100 are MCI and the rest are subjects diagnosed with AD (Figure 4-11).

The age distribution of the patients in each group is shown in Table 4-1:

Table 4-1 - Dataset age distribution

Group	Youngest	Oldest	Average
CN	Male: 56.20 years	Male: 91.40 years	Male: 75.59 years
	Female: 55.10 years	Female: 90.30 years	Female: 72.97 years
MCI	Male: 56.50 years	Male: 91.5 years	Male: 75.57 years
	Female: 55.20 years	Female: 89.00 years	Female: 72.95 years
AD	Male: 55.30 years	Male: 90.60 years	Male: 75.19 years
	Female: 55.20 years	Female: 91.00 years	Female: 73.52 years

Globally, the data of the selected 600 subjects is balanced, i.e, the three groups have the same age distribution, as well as sex and gender, to make sure that these parameters would not influence the results, since accuracy, a measure metrics of the performance characteristics of an algorithm, is sensitive to unbalanced distributions [11].

4.3.5 RADIOMICS FEATURE GENERATION

The PyRadiomics platform can extract radiomic data from medical imaging (such as CT, PET, MRI) using four main steps, which are illustrated in Figure 2-5: loading and pre-processing of the image and segmentation maps, followed by the application of enabled filters. After that, the features are calculated using the different feature classes, and lastly, the results are returned [26].

PyRadiomics input corresponds to two images: one outputted from FreeSurfer (*brain.mgz* with the whole brain converted to NIfTI, i.e., *brain.nii*) and the other one corresponds to the masks (the left and right hippocampus extracted from the brain, i.e., *Left-Hippocampus.nii* and *Right-Hippocampus.nii*).

The pipeline for radiomics feature extraction made in this work consisted of the following 5 steps:

1. Get the dataset, separated by classes;

2. Define the parameters file, to customize the extraction, with the settings and the image types to be used;

3. Create the extractor for each hippocampus, setting the label to the correct value in each hippocampus (17 to extract left hippocampus and 53 for the right hippocampus, since these values correspond to the label that identifies these regions of the brain);

4. Radiomic features extraction for all MRI images present on the dataset (this was made 3 times, one for each group class since the dataset is divided by class);

5. Save the results on a different CSV file, for each hippocampus. This resulted in 6 CSV files.

Medical images and segmentations were loaded into the platform. All the images were arranged in two different pandas DataFrames: one for the left hippocampus and another one for the right hippocampus. To customize the extraction, a global parameter file was developed, specifying the image types to use (Table 4-2) to extract features from. The image types enabled in the parameter file were, besides the original image, derived images using 9 filters (explained in section 2.5.1): LoG with 5 different sigmas (1.0, 2.0, 3.0, 4.0, and 5.0), wavelet, square, squareRoot, logarithm, exponential, gradient, LBP2D, LBP3D.

Table 4-2 - Enabled filters in PyRadiomics feature extraction

Enabled filters in PyRadiomics feature extraction
LoG with 5 different sigmas
Wavelet
Square

SquareRoot
Logarithm
Exponential
Gradient
LBP2D
LBP3D
Original

Since the label of the mask for each hippocampus differs, there was a need to create two different extractors for each one, where the left hippocampus' extractor had the settings label defined to 17 and 53 in the right hippocampus' extractor. These values of the label are identified by the *SegId* value present on file *aseg.stats*, which is a file outputted by FreeSurfer that has segmentation statistics about the image.

Finally, radiomic features were extracted using the open-source Python package PyRadiomics [30]. Calculated features were stored and returned in an ordered dictionary, for each hippocampus. The result was a table with dimensions 200x2177, where lines represent each subject and columns represent PyRadiomics features (Figure 4-12). This table was stored in a different CSV file for each class.

	s_Versions_PyRadiomics	nostics_Versions_Numpy	stics_Versions_SimpleITK	stics_Versions_PyWavelet	nostics_Versions_Python	s_Configuration_Settings	tion_EnabledImage
1	2.2.0	1.18.4	1.2.4	1.1.1	3.7.7	': 17, 'additionalInfo': True}	{}, 'LBP2D': {}, 'LBP
2	2.2.0	1.18.4	1.2.4	1.1.1	3.7.7	': 17, 'additionalInfo': True}	{}, 'LBP2D': {}, 'LBP
3	2.2.0	1.18.4	1.2.4	1.1.1	3.7.7	': 17, 'additionalInfo': True}	{}, 'LBP2D': {}, 'LBP
4	2.2.0	1.18.4	1.2.4	1.1.1	3.7.7	': 17, 'additionalInfo': True}	{}, 'LBP2D': {}, 'LBP
5	2.2.0	1.18.4	1.2.4	1.1.1	3.7.7	': 17, 'additionalInfo': True}	{}, 'LBP2D': {}, 'LBP

Figure 4-12 - Example of a table with values outputted from PyRadiomics

These steps were made 3 times, one for each group class (CN, MCI, and AD), which resulted in 6 different CSV files.

4.3.6 DATA CLEANING

Besides the calculated features, the dictionary with the output of PyRadiomics also contains additional information on the extraction, including the current version, applied filters, settings, and original image spacing, as illustrated in Figure 4-12. Because of this, there was a need to prune the data, removing unnecessary information to classify a subject. Since these columns correspond to irrelevant features for

class prediction, the data went through a cleaning stage, removing the unnecessary features, like the versions of PyRadiomics and Python. Data cleaning, or feature selection, is a process that was performed to reduce the dimensionality of features by selecting informative and reliable features and excluding redundant features among the extracted features.

After this cleaning process, a column with the class of each subject was added, like illustrated in Figure 4-13.

	entage	lbp-3D-k_glszm_ZoneVariance	lbp-3D-k_ngtdm_Busyness	lbp-3D-k_ngtdm_Coarseness	lbp-3D-k_ngtdm_Complexity	lbp-3D-k_ngtdm_Contrast	lbp-3D-k_ngtdm_Strength	classe
1	14346	389532.21	343.0006271915815	0.002571187191140616	0.23247184540316315	0.02793319850602229	0.00264301724551906	AD
2	15464	221258.9333333333	179.17036996592063	0.003974920401288767	0.17290540538745452	0.015095989453786225	0.0040724363819918735	AD
3	17356	313612.0	290.16563083994475	0.002747181540385689	0.2074718922026805	0.021864919083014047	0.002836488754336704	AD
4	87148	277746.854442344	242.8765492138394	0.0032798786298092065	0.20621530648156708	0.02171527087780936	0.0033859192715081413	AD
5	15686	176780.86390532542	284.2200667329022	0.002898843635402767	0.2254674116463751	0.025022129993591703	0.002974432722136946	AD

Figure 4-13 - Resulting CSV file after data cleaning

The result was a table with dimensions 200x2157, where lines represent each subject and columns represent PyRadiomics features and was stored in a different CSV file for each class.

The code lines present in this notebook were executed 3 times, one for each group class (CN, MCI, and AD), which resulted in 6 different CSV files.

4.3.7 ENCODING OF DATA

Since neural networks can only work with numerical data [31], all the data passed to the MLP must be numerical. For this reason, all the categorical values present in the CSV files created with the result of PyRadiomics need to be encoded before passing it to the MLP. The CSV files were encoded (Figure 4-14), after the cleaning step, to ensure that the values were in the correct type and, therefore, usable to train a MLP. Categorical data, such as class and sex attribute, which can take the values "M" and "F", were converted to a real value using, for that, a representation called *One-Hot Encoding*.

	entage	D-k_glszm_ZoneVariance	p-3D-k_ngtdm_Busyness	3D-k_ngtdm_Coarseness	-3D-k_ngtdm_Complexity	bp-3D-k_ngtdm_Contrast	lbp-3D-k_ngtdm_Strength	classe
1	014346	389532.21	343.0006271915815	0.002571187191140616	0.23247184540316315	0.02793319850602229	0.00264301724551906	0
2	515464	221258.9333333333	179.17036996592063	0.003974920401288767	0.17290540538745452	0.015095989453786225	0.0040724363819918735	0
3	017356	313612.0	290.16563083994475	0.002747181540385689	0.2074718922026805	0.021864919083014047	0.002836488754336704	0
4	987148	277746.854442344	242.8765492138394	0.0032798786298092065	0.20621530648156708	0.0217152708778093358	0.0033859192715081341	0
5	215686	176780.86390532542	284.2200667329022	0.002898843635402767	0.2254674116463751	0.025022129993591703	0.002974432722136946	0

Figure 4-14 - Resulting CSV file after data encoding

4.3.8 COMBINE OUTPUTTED DATA FROM RADIOMICS FROM DIFFERENT GROUPS

Since the radiomics extractor is different for each hippocampus, because these areas of the brain have different labels, the output of PyRadiomics resulted in two different CSV files for each class, one for each hippocampus. To combine those files, a Jupyter notebook was created. The combination was made in two different ways: grouping by class and combining by the hippocampus.

The separate resultant files of the left and right hippocampus for each group class were merged into a single CSV file, resulting in one CSV file with dimensions 200x4313 with information about both hippocampi per group (i.e., three CSV files with information of left and right hippocampus). This led to a necessary change in the name of the features: features of the left hippocampus retained the same name, while the ending ".1" was added to features of the right hippocampus, allowing features based on the hippocampus to be distinguished (Figure 4-15).

	_ZoneVariance.1	lbp-3D-k_ngtdm_Busyness.1	lbp-3D-k_ngtdm_Coarseness.1	lbp-3D-k_ngtdm_Complexity.1	lbp-3D-k_ngtdm_Contrast.1	bp-3D-k_ngtdm_Strength.1	AD
1	547.2653061224	380.07184374767183	0.002197842454663286	0.22591443196424355	0.0254422863676168883	0.00226201358117065	1
2	37.77014462813	122.0012223493271	0.005113969845318321	0.11876271083445525	0.007181998708349161	0.005227705267807552	1
3	404269.109375	258.76736634468637	0.0027819851316861037	0.17637662001023904	0.01575735514428531	0.00284768515888676	1
4	266.5179584121	259.06431952732794	0.003073739748295587	0.2017591239152797	0.02119399616734529	0.0031791164097316154	1
5	32.72108843544	401.82007489418805	0.002407523577816691	0.2592787589268001	0.03427501246900039	0.0024606725129794203	1

Figure 4-15 - Rename of PyRadiomics features, to distinguish between left and right hippocampus

Besides this, the results of the left hippocampus for all groups were concatenated into another CSV file (Figure 4-16) with dimensions 600x2157, and the same was made to the PyRadiomics results for the right hippocampus.

	-3D-k_glszm_ZoneVariance	lbp-3D-k_ngtdm_Busyness	lbp-3D-k_ngtdm_Coarseness	lbp-3D-k_ngtdm_Complexity	lbp-3D-k_ngtdm_Contrast	lbp-3D-k_ngtdm_Strength	AD	CN	MCI
1	412349.6388888888	313.35402142197444	0.0029426632593801396	0.2475078033902825	0.030970495830421975	0.0030408195225816925	0	1	0
2	322060.6787109375	180.06135613027826	0.0037208496740920656	0.1505213209672684	0.011307169733924784	0.003834673543329662	0	1	0
3	578179.0612244897	463.1493819910328	0.0021253564836204407	0.2670313870975057	0.03568184654354579	0.0021808990145282306	0	1	0
4	250006.0367036011	182.7052331600622	0.003742881357661865	0.15546921385512358	0.01224754157923695	0.0038761337675161647	0	1	0
5	374718.10204081633	356.17564794187484	0.002187855303388523	0.20073281424856693	0.020565944379857152	0.002246185049185784	0	1	0

Figure 4-16 - Left hippocampus features from the 3 classes

Finally, a global CSV file was created, with all information of 3 groups for both hippocampi, with dimensions 600x4314. The CSV file shown in Figure 4-17 corresponds to this global file, where were appended clinical annotations about the subjects, like sex, IntraCranial Volume (ICV), and age at the time of the image acquisition, to train the MLP. The resulting CSV file has dimensions 600x4318.

	)-k_ngtdm_Coarseness.1	3D-k_ngtdm_Complexity.1	p-3D-k_ngtdm_Contrast.1	p-3D-k_ngtdm_Strength.1	Age_at_img_acquisition	Sex	intracranial_volume	AD	CN	MCI
1	4 0.0029426632593801396	0.2475078033902825	0.030970495830421975	0.0030408195225816925	81.2	F	1481008.0	0	1	0
2	6 0.0037208496740920656	0.1505213209672684	0.011307169733924784	0.003834673543329662	80.8	F	1419644.0	0	1	0
3	8 0.0021253564836204407	0.2670313870975057	0.03568184654354579	0.0021808990145282306	84.0	F	1470755.31	0	1	0
4	2 0.003742881357661865	0.15546921385512358	0.01224754157923695	0.0038761337675161647	77.1	M	1909141.22	0	1	0
5	4 0.002187855303388523	0.20073281424856693	0.020565944379857152	0.002246185049185784	67.6	M	1593096.0	0	1	0

Figure 4-17 - Radiomics features with clinical annotations about the subjects

The main goal of appending this information was to analyze if these values could somehow influence the neural network to identify each class. In volumetric studies using magnetic resonance imaging, values of intracranial volume are often used as variables for premorbid brain volume [55][63]. ICV differs substantially between males and females, males having a 10–12% larger ICV. The effect of age on ICV seems to be minimal. In neurodegeneration research, one approach to compensate for head size variability is to include the ICV as a variate in the regression model. Those clinical annotations did not influence the MLP results, since the dataset was chosen in a way to minimize the bias of the values.

4.3.9 NORMALIZATION OF DATA

Some of the values outputted from PyRadiomics were far apart and neural networks require the input to be scaled consistently [33]. Normalization is a way to deal with these values that lie so far apart. For this, was used the *scikit-learn* package to rescale the data to the range between 0 and 1, by importing the *MinMaxScaler* module from *sklearn.preprocessing* (Figure 2-11 – E.) [64].
The creation of this Jupyter notebook was used in all CSV files to make sure that the values were normalized.

	dm_Busyness.1	lbp-3D-k_ngtdm_Coarseness.1	lbp-3D-k_ngtdm_Complexity.1	lbp-3D-k_ngtdm_Contrast.1	lbp-3D-k_ngtdm_Strength.1	Age_at_img_acquisition	intracranial_volume	F	M	AD	CN	MCI
1	584295290922	0.21369981697781454	0.6838366829445282	0.46298592552152695	0.2210264459940937	0.7170329670329672	0.29358964834363754	1.0	0.0	0.0	1.0	0.0
2	921472506822	0.3186275788480337	0.2538442524574786	0.1192972028652442	0.3272776205257021	0.7060439560439558	0.249069609247492	1.0	0.0	0.0	1.0	0.0
3	7165798663247	0.10349721112330756	0.77039506716833	0.5453340542477113	0.10593278696291408	0.7939560439560438	0.2861512453770665	1.0	0.0	0.0	1.0	0.0
4	7301305453796	0.3215982491186882	0.27578088151635033	0.1357336489044951	0.33282674845228133	0.6043956043956045	0.6042034885880581	0.0	1.0	0.0	1.0	0.0
5	7708751792128	0.11192431953051707	0.4764583889331247	0.2811282387645947	0.11467081395727324	0.3434065934065933	0.37491032727931217	0.0	1.0	0.0	1.0	0.0

Figure 4-18 - Resulting CSV file after data normalization

To train the network, there was a need to shuffle the output of this notebook, shown in Figure 4-18, before loading to the MLP, to make sure that the data used to train the MLP is in the correct form. The shuffled encoded dataset previously normalized was the file used to train the neural network. It corresponds to a table shown in Figure 4-19, with dimensions 600x4318, where 600 are the number of all subjects analyzed and the columns represent PyRadiomics features and other clinically relevant annotations (like sex, age at image acquisition, and ICV).

	ce.1	3D-k_ngtdm_Busyness.1	)-k_ngtdm_Coarseness.1	3D-k_ngtdm_Complexity.1	p-3D-k_ngtdm_Contrast.1	p-3D-k_ngtdm_Strength.1	Age_at_img_acquisition	intracranial_volume	F	M	AD	CN	MCI
1	135	0.011982126976252148	0.33708600982414016	0.23856266712049928	0.11185950372211334	0.34093095068320245	0.6043956043956045	0.18085667339163702	1.0	0.0	1.0	0.0	0.0
2	955	0.02193047895313142	0.2313736832336215	0.48100856311201073	0.2921057466157332	0.239536406653989	0.4423076923076925	0.08895066665428464	1.0	0.0	1.0	0.0	0.0
3	322	0.07025202285801205	0.05732972291649749	0.9049616871762143	0.7236843780565285	0.057924925084393175	0.4587912087912087	0.12290715768236937	1.0	0.0	0.0	1.0	0.0
4	307	0.058152685410852854	0.05291071783990453	0.7985490477007067	0.575723764459474	0.053175830222124465	0.5604395604395607	0.3815813682722242	0.0	1.0	0.0	0.0	1.0
5	375	0.04527165798663247	0.10349721112330756	0.77039506716833	0.5453340542477113	0.10593278696291408	0.7939560439560438	0.2861512453770665	1.0	0.0	0.0	1.0	0.0

Figure 4-19 - Resulting CSV file after data being shuffled

4.3.10 MLP MODELLING PROCESS

To classify the three classes and distinguish the clinical group of a subject, a multi-perception layer network was developed, using the TensorFlow backend (Figure 2-11 – C.). This approach has been developed to classify AD patients from cognitively normal controls or mild cognitive impairment, which is the prodromal stage of AD.

Since the use of ML requires some transformation steps, the data must be pre-processed. To start building the neural network, the data from PyRadiomics was pre-processed before passing it to the neural network, removing unnecessary values.

All the normalization, encoding and concatenation of the PyRadiomics files resulted in 27 files, but only the final CSV file (Figure 4-19) with the shuffled encoded normalized PyRadiomics features concatenated with the three classes and with clinical annotations appended as extra features were used in the MLP.

Once the data has been pre-processed, it is time to construct a neural network model, in particular a multi-layer perceptron. For this, two key architectural decisions need to be made: how many layers to use and how many "hidden units" to choose for each layer.

The development of this network began with the definition of the training info, such as the learning rate and the number of epochs. The learning rate is a configuration parameter that controls the amount that weights are updated. It controls the step or change made to network weight for a given error. Often small weight sizes are used such as 0.1 or 0.01 or smaller [33].

With the model constructed, after defining learning rate, batch size and how many epochs to use to run the fitting, it can be compiled and fit the data to it, passing in the training data.

However, imbalanced data is an aspect to be taken into account in classification problems where the classes are not represented equally. If the size of the training data is imbalanced across different classes, a classification algorithm may have poor classification accuracy for the minority classes [29]. Most classification datasets do not have an exactly equal number of instances in each class, but a small difference often has little influence [31]. Therefore, before training the model, the dataset available for the development of the algorithm may be divided into training and validation datasets. The validation

63

dataset is used for monitoring the performance of the algorithm during the training process. Once the final model is selected and all its parameters are fixed, its performance is evaluated in the test dataset. This resulted in two groups: one for training the network and the other one for testing it. Both groups were divided into x and y cases (Figure 4-20), where the class label of each subject was saved in the y group. The training data was split into 80% for training and 20% for validation. Following this method, the number of examples to train the model was 480 and the other 120 subjects were used to evaluate the fit model's performance. After that, the division of the 480 training subjects and 120 test cases into each class lead to the results present in the following barplot (Figure 4-20), which was created with the use of a function from the NumPy [65] (Figure 2-11 – B.) package to compute the histogram.
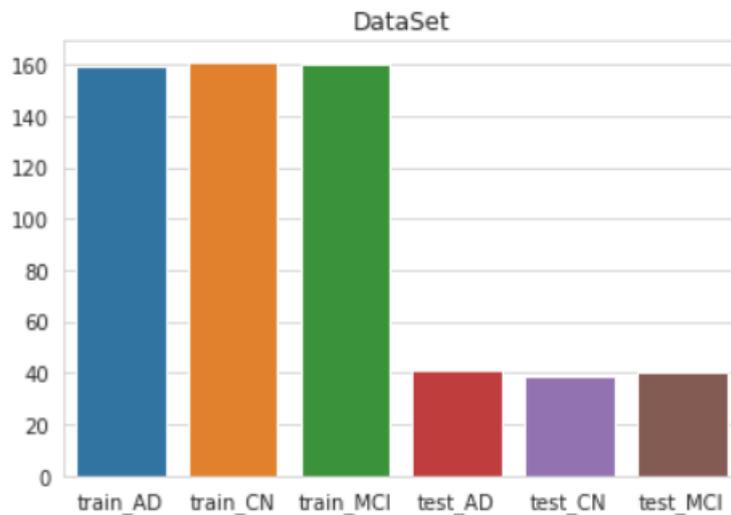
Figure 4-20 - Dataset passed to the MLP

Then, the model was defined (Figure 4-21). The first layer defined is the input layer and there's a need to pass the shape of the input data to it, which is 4316 in this case, since it corresponds to the number of columns of the CSV file used (4319) to train the MLP (Figure 4-19), removing the class columns.

A deep learning algorithm is trained by adjusting network weights. Starting from a random initial configuration, parameters are adjusted to find a set of parameters that perform best on the training dataset. During the training phase, the output of the algorithm is compared with the ground truth by using a loss function that quantitatively measures the error in the prediction in comparison with the ground truth. The error is then backpropagated to optimize network weights [29]. The training phase continues until the loss function reaches a minimum. After some experiences, the best result was obtained by adding 3 hidden Dense layers, which are fully connected layers, and the rectified linear activation function. The first hidden layer uses 1024 "hidden units", the second has 256 and the last

hidden layer has 64 nodes. The final layer will use a *Sigmoid* activation function so that the output is a probability [31]. The *Sigmoid* activation function is used in the output layer to predict the probability of a subject belonging to one of the three classes.

To address the problem of overfitting in the MLP, the dropout regularization technique was used. This was done by inserting a Dropout layer between each layer. These layers normalize the output of a previous layer and also prevents overfitting, probably due to an increase in neural network stability. They are techniques for regularizing the network, i.e. allow the model to generalize over data not previously seen [66]. In this case, a dropout rate, i.e., the probability of setting outputs from the hidden layer to zero, was specified to 40% or 0.4.

```
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 4316)]            0
_____
dense (Dense)                (None, 1024)              4420608
_____
dropout (Dropout)            (None, 1024)              0
_____
dense_1 (Dense)              (None, 256)               262400
_____
dropout_1 (Dropout)          (None, 256)               0
_____
dense_2 (Dense)              (None, 64)                16448
_____
dropout_2 (Dropout)          (None, 64)                0
_____
dense_3 (Dense)              (None, 3)                 195
=================================================================
Total params: 4,699,651
Trainable params: 4,699,651
Non-trainable params: 0
```

Figure 4-21 – Structure of the MultiLayer perceptron

After that, the next step is to compile the model and fit it into the data. In compiling, the model has been configured with the ADAM optimizer (one of the most currently popular gradient descent methods) and the *binary_crossentropy* loss function. Additionally, some evaluation metrics were monitored during the training (like accuracy, precision, false positive rate, false negative rate, specificity, sensitivity, confusion matrix, among others), so that the model can be evaluated, using these metrics.

In this step, the model is ready to be trained. The model has to be trained for a specified number of epochs or exposures to the training dataset. An epoch is a single pass through the entire training set,

followed by testing of the verification set. The batch size specified in the code defines the number of samples that are going to be propagated through the network. Also, by doing this, the efficiency is optimized because it makes sure that are not loaded too many input patterns into memory at the same time [31].

Since deep learning performance is sensitive to the random numbers generated at the start of training and to hyper-parameters (such as learning rates, batch sizes, weight decay, momentum, and dropout probabilities) [31], the defined model was trained for 500 epochs or iterations over all the samples in *x_train* and *y_train*, in batches of 48 samples, with the data being shuffled. Table 4-3 presents the parameters used in the various tests.

Table 4-3 - Parameters used in the tests

Parameter	Value
Batch size	5
Learning rate	6×10^{-4}
Learning rate decay	1.5×10^{-6}
Momentum	0.9
Number of epochs	200

Since batch learning (i.e, saving the errors up across all of the training examples and update the network at the end) is often more stable than online learning (the weights in the network are updated from the errors calculated for each training example) [33], this was the method chosen to be used in the MLP.

When the model is built and trained, it can be used to make predictions on data that it has not seen yet, i.e., it is time to evaluate its performance. It can be made by visually comparing the predictions with the actual test labels (*y_test*), or use other types of evaluation metrics to determine the actual performance. The last option was the one used since some of these metrics were monitored during the training.

The test dataset will be used as a validation dataset and is used only at the final step of this work to report the final model performance, making it possible to evaluate the performance of the model on the test dataset and report the result.

When training the model, two model checkpoints were created to save only the weights of the network when the value of the loss was the best and another one when the accuracy value was the greatest. Since Deep learning models can take hours, days or even weeks to train, and if the run is stopped

unexpectedly, a lot of work can be lost. The checkpoint may be used directly, or used as the starting point for a new run, picking up where it left off. When training deep learning models, the checkpoints are the weights of the model. These two model checkpoints that save the weights make it possible to use the weights to make predictions as is, or used as the basis for ongoing training.

It is important to collect and review metrics during the training of deep learning models and, to visualize the training process and learn about the MLP, it can be made by observing its performance over time during training in Python with Keras. For this, were printed eight different charts:

1. A plot of FNR on the training and validation datasets over training epochs (Figure 4-23);

2. A plot of FPR on the training and validation datasets over training epochs (Figure 4-22);

3. A plot of Precision on the training and validation datasets over training epochs (Figure 4-24);

4. A plot of Accuracy on the training and validation datasets over training epochs (Figure 4-25);

5. A plot of Loss on the training and validation datasets over training epochs (Figure 4-27);

6. A plot of Leaning Rate on the training and validation datasets over training epochs (Figure 4-26);

7. A plot of sensitivity on the training and validation datasets over training epochs (Figure 4-29);

8. A plot of specificity on the training and validation datasets over training epochs (Figure 4-28).

These plots, provided below, were made using the callback features in Keras (Figure 4-1– B.) and show the training metrics. The History callback in Keras updates these training metrics plots automatically after each epoch, making it possible to see a live updated plot from the history data collected during training.

Figure 4-23 - Graph of FNR values during MLP training
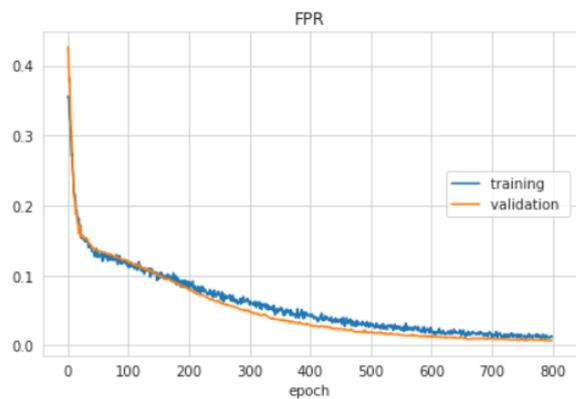
Figure 4-22 - Graph of FPR values during MLP training
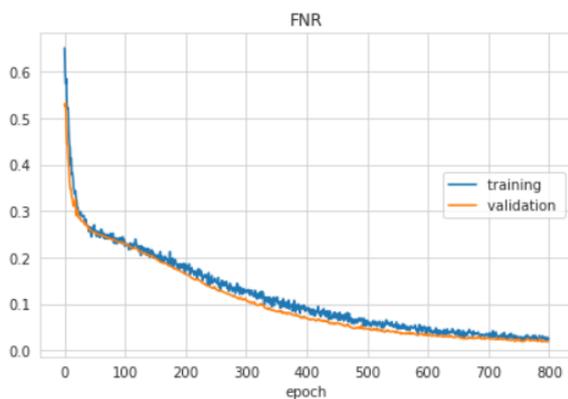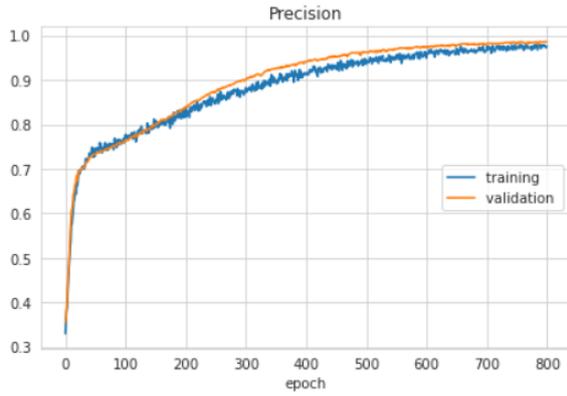
Figure 4-24 - Graph of Precision values during MLP training
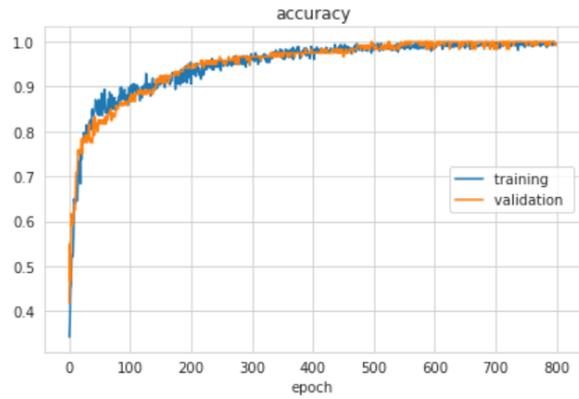
Figure 4-25 - Graph of Accuracy values during MLP training
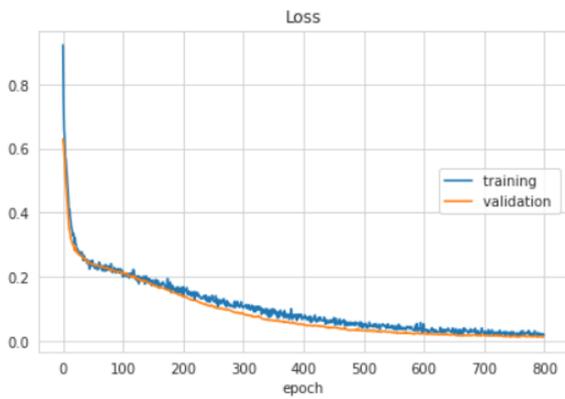
Figure 4-27 - Graph of Loss values during MLP training
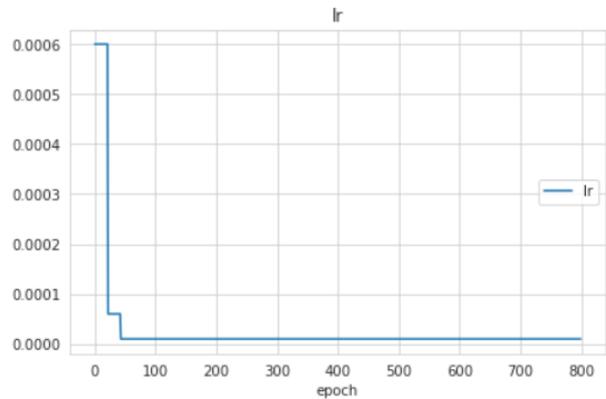
Figure 4-26 – Graph of Learning Rate (LR) values during MLP training
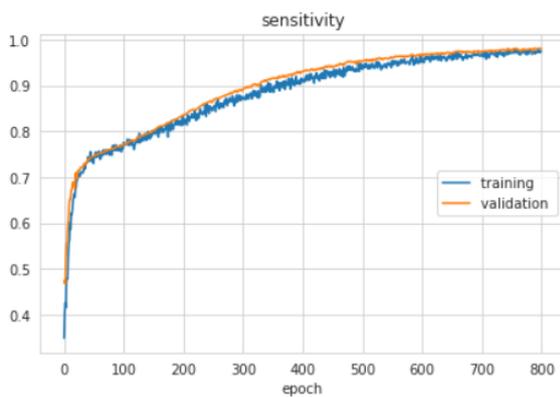
Figure 4-29 - Graph of Sensitivity values during MLP training
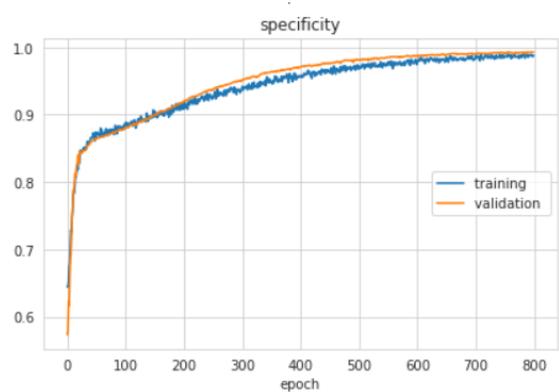
Figure 4-28 - Graph of Specificity values during MLP training

68

4.3.11 MODEL EVALUATION

Since it can be somewhat difficult to interpret graphics of evaluation metrics shown above, a confusion matrix was also plotted. The confusion matrix allows the visualization of the accuracy of the model, so, to measure performance, the confusion matrix was an additional step made in this Jupyter notebook. The confusion matrix obtained is the one below (Table 4-4), which is graphically represented in Figure 4-30.

Table 4-4 – Confusion Matrix of the results obtained with the Multi-Layer Perceptron with the training data

Predicted Actual	CN	MCI	AD
CN	41	0	0
MCI	0	39	0
AD	0	0	40
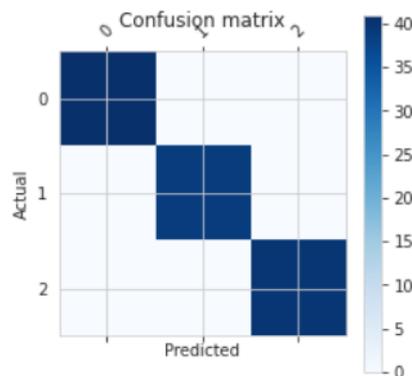
Figure 4-30 – Graphic of obtained Confusion Matrix

From the confusion matrix obtained (Figure 4-30), it is possible to conclude that all the predicted classes correspond to the actual classes.

```
Scores: [0.013652382418513298, 0.9916666746139526, 0.985794723033905, 0.007064592558890581, 0.01931920275092125, 0.9929352402687073, 0.98068046
56982422]
Erro classifier: 0.83%
    FNR
            training                    (min:    0.021, max:    0.651, cur:    0.025)
            validation                  (min:    0.019, max:    0.533, cur:    0.020)
    FPR
            training                    (min:    0.010, max:    0.356, cur:    0.012)
            validation                  (min:    0.007, max:    0.427, cur:    0.007)
    Precision
            training                    (min:    0.329, max:    0.980, cur:    0.976)
            validation                  (min:    0.355, max:    0.986, cur:    0.985)
    accuracy
            training                    (min:    0.342, max:    1.000, cur:    0.996)
            validation                  (min:    0.417, max:    1.000, cur:    0.992)
    Loss
            training                    (min:    0.015, max:    0.923, cur:    0.019)
            validation                  (min:    0.012, max:    0.630, cur:    0.014)
sensitivity
            training                    (min:    0.349, max:    0.979, cur:    0.975)
            validation                  (min:    0.467, max:    0.981, cur:    0.980)
specificity
            training                    (min:    0.644, max:    0.990, cur:    0.988)
            validation                  (min:    0.573, max:    0.993, cur:    0.993)
```

Figure 4-31 – Model evaluation

Since the MLP was trained with a very large number of features, there was a need to create a Jupyter notebook with the purpose to synthesize and analyze all the data obtained from the MLP trained with PyRadiomics features.

To try to identify which of these features had a major influence in distinguishing and identifying the three group classes, the function *get_score_importances* from library *eli5.permutation_importance* was used. This function uses itself another function: the *accuracy_score*. Accuracy is one of the parameters that allow evaluating the efficiency of the MLP (how much the prediction was made correctly and should be as high as possible). *Eli5.permutation_importance* is a module for computing features importance by measuring how the score decreases when a feature is not available [67]. This feature importance could be positive or negative, and, because of this and to make it easier to make interpretations, the first forty features with higher values were selected and saved in a CSV file.

To analyze and visualize data in a better way and to better understand the features enhanced in each hippocampus, a heatmap was created. The choice to use a heatmap was because it is the best way to represent and understand complex data and Data Scientists generally use heatmaps when they want to understand the correlation between various features of a data frame. One of the greatest applications of the heatmap is to analyze the correlation between different features of a data frame, where features are represented in lines, filters in columns and correlation is how many values in these columns are related

70

to each other. Heatmap is a data visualization technique, which represents data using different colours in two dimensions [68].

Two distinct heatmaps were created: one for features of the left-hippocampus (Figure 4-32) and another one for features of the right-hippocampus (Figure 4-33). In the heatmaps generated, each column corresponds to a filter, the lines are the features obtained with PyRadiomics and each cell has the value of that importance.

4316 features were extracted with 23 filters, which correspond to variations in the 9 enabled filters (Table 4-2) in PyRadiomics feature extraction.

Figure 4-32 displays the correlation of the 2139 radiomic features (rows) from the left-hippocampus in 23 applied filters (columns) (Table 4-2). Figure 4-33 shows the same graphic, but for right-hippocampus features instead. Only some of these 2139 features are displayed in these figures, as it is not possible to visualize all of them.

With these heatmaps, is now possible to know which features are most important when assessing the group of a subject in an AD study.
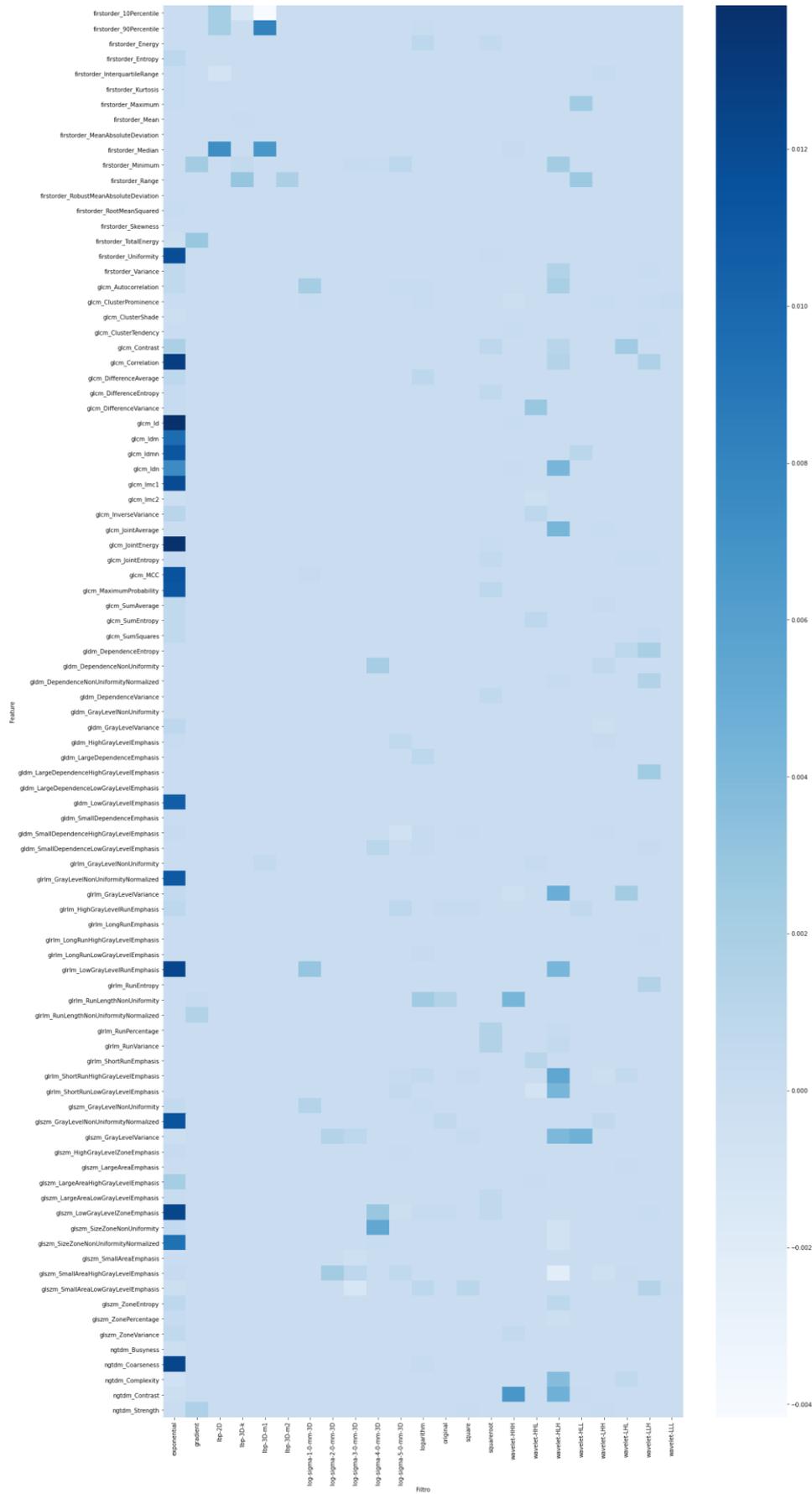
Figure 4-32 – Heatmap of feature importance from Left-hippocampus
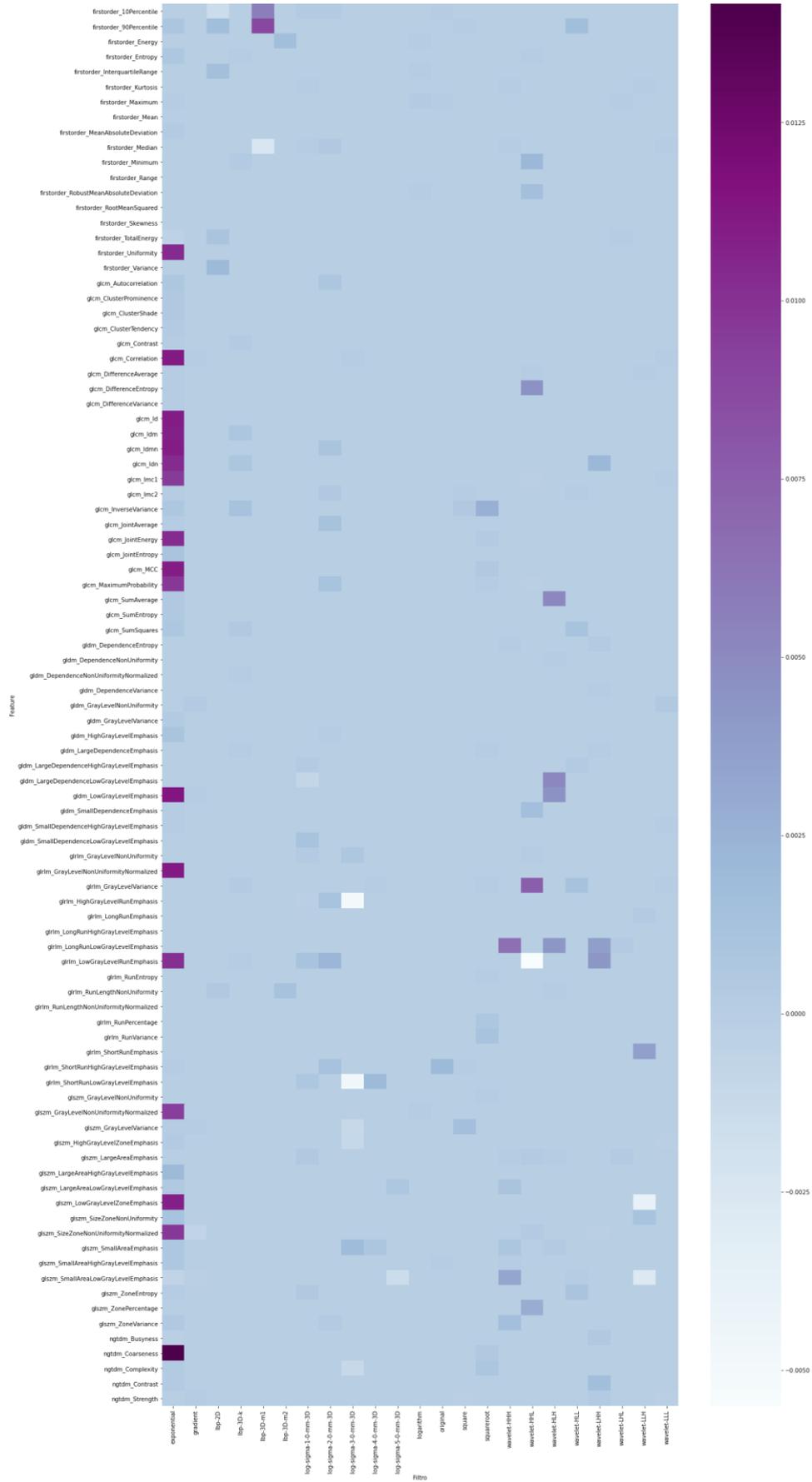
Figure 4-33 – Heatmap of feature importance from Right-hippocampus

4.3.12 BOUNDING BOX

To visualize the effects of each filter in the images, a method was created to "amplify" the area of the hippocampus so that neuroradiologists can analyze the image after applying the filters. This bounding box was made to zoom in on the hippocampus, after applying the filters.

4.3.13 MRI VISUALIZATION WITH ENABLED FILTERS

A Jupyter notebook was created to visualize the effect that every enabled filter has on the MRI image, trying to identify and distinguish visually what parts and zones of the brain each filter highlights. The objective is to try to identify which areas of the hippocampus are emphasised with each filter and where are located the most important features denoted by PyRadiomics. The enabled filters and their influence on the image are shown in Table 4-5.

Table 4-5 – MRI visualization with enabled filters

Filter	MRI Filtered
	MRI original scan
	MRI with Square filter
	MRI with SquareRoot filter
	MRI with Logarithm filter
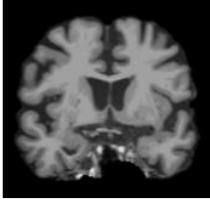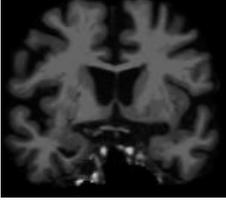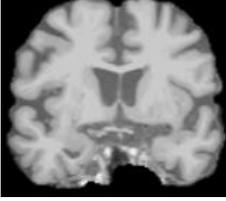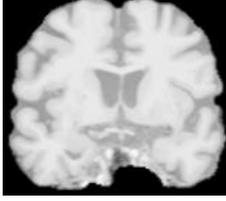
	MRI with LBP2D filter
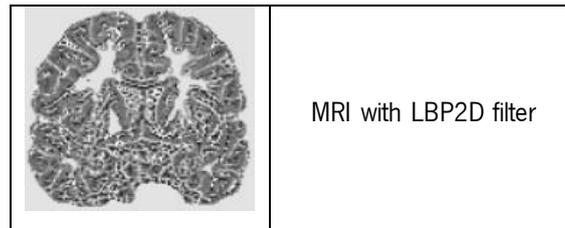
4.4 RESULTS

After being calculated the importance of the features, some analyses can be made:

- Identification of more relevant filters and which relevant features they enable, for both hippocampi;

- Identification of which features are both positive and negative in both hippocampi;

- Identification of which features are repeated in both hippocampi;

- Identification of which features, repeated in both hippocampi, are positive;

- Identification of which features, repeated in both hippocampi, are negative.

The evaluation of the model of the ANN was based on the results of accuracy, which is one measurement of the test set, and the other metrics defined (Figure 4-31). The MLP (Figure 4-21) defined with the parameters present in Table 4-3 is getting about 99% of accuracy (Figure 4-31).

The correlation between values of feature importance's can be visualized in Figure 4-32 for the left hippocampus and Figure 4-33 for the right hippocampus, in the form of heatmaps. These heatmaps help in distinguishing which features are relevant to classify the subjects and attribute a class to each one of them.

From the output of PyRadiomics, it is possible to identify that the first 40 more relevant positive features are the ones shown in Table 4-6.

Table 4-6 – 40 relevant positive features and importance values

Filter	Feature	Importance
Exponential	glcm_Correlation	0.014000
Exponential	glcm_Id	0.013000
Exponential	glcm_JointEnergy	0.012750
Exponential	glcm_MaximumProbability	0.012750
Exponential	gldm_LowGrayLevelEmphasis	0.012667
Exponential	ngtdm_Coarseness.1	0.012667

Exponential	glszm_LowGrayLevelZoneEmphasis.1	0.011917
Exponential	glszm_LowGrayLevelZoneEmphasis	0.011833
Exponential	glszm_GrayLevelNonUniformityNormalized	0.011833
Exponential	glcm_JointEnergy.1	0.011750
Exponential	glcm_Idmn	0.011500
Exponential	glcm_MCC.1	0.011500
Exponential	glcm_Idm	0.011417
Exponential	glcm_Imc1	0.011250
Exponential	glcm_Correlation.1	0.011167
Exponential	glszm_GrayLevelNonUniformityNormalized.1	0.011167
Exponential	ngtdm_Coarseness	0.011000
Exponential	firstorder_Uniformity.1	0.010917
Exponential	glcm_Id.1	0.010667
Exponential	gldm_LowGrayLevelEmphasis.1	0.010417
Exponential	glcm_Imc1.1	0.010417
Exponential	glcm_Idn	0.010333
Exponential	glrlm_LowGrayLevelRunEmphasis	0.010333
Exponential	glrlm_GrayLevelNonUniformityNormalized.1	0.010167
Exponential	glszm_SizeZoneNonUniformityNormalized.1	0.010167
Exponential	glszm_SizeZoneNonUniformityNormalized	0.010167
Exponential	glrlm_GrayLevelNonUniformityNormalized	0.010083
Exponential	glcm_Idmn.1	0.010000
Exponential	glcm_MaximumProbability.1	0.009833
Exponential	glrlm_LowGrayLevelRunEmphasis.1	0.009750
Exponential	glcm_Idn.1	0.009750
Exponential	glcm_Idm.1	0.009667
Exponential	glcm_MCC	0.009417
Exponential	firstorder_Uniformity	0.009000
LBP (3D) m1	firstorder_90Percentile.1	0.008083
LBP (3D) m1	firstorder_90Percentile	0.007333
LBP (2D)	firstorder_Median	0.007083
LBP (3D) m1	firstorder_Median	0.007000
Wavelet HHL	glrlm_GrayLevelVariance.1	0.006583
Wavelet HHH	glrlm_LongRunLowGrayLevelEmphasis.1	0.006500

For the negative features, the 40 more relevant are shown in Table 4-7.

Table 4-7 – 40 relevant negative features and importance values

Filter	Feature	Importance
Wavelet HHL	glrlm_LowGrayLevelRunEmphasis.1	-0.005667
LoG 3.0 (3D)	glrlm_HighGrayLevelRunEmphasis.1	-0.004917
LoG 3.0 (3D)	glrlm_ShortRunLowGrayLevelEmphasis.1	-0.004500
LBP (3D) m1	firstorder_10Percentile	-0.003333
Wavelet LLH	glszm_SmallAreaLowGrayLevelEmphasis.1	-0.003083
Wavelet LLH	glszm_LowGrayLevelZoneEmphasis.1	-0.002917
LBP (2D)	firstorder_10Percentile.1	-0.002417
Wavelet HLH	glszm_SmallAreaHighGrayLevelEmphasis	-0.002250
LoG 3.0 (3D)	glszm_GrayLevelVariance.1	-0.001750
LoG 5.0 (3D)	glszm_SmallAreaLowGrayLevelEmphasis.1	-0.001750
LoG 3.0 (3D)	ngtdm_Complexity.1	-0.001417
LoG 1.0 (3D)	gldm_LargeDependenceLowGrayLevelEmphasis.1	-0.001167
LBP (3D) k	firstorder_10Percentile	-0.001167
Wavelet LHH	glszm_SmallAreaHighGrayLevelEmphasis	-0.001083
LBP (3D) m1	firstorder_Median.1	-0.001083
Wavelet HHL	glrlm_ShortRunLowGrayLevelEmphasis	-0.000833
LoG 3.0 (3D)	glszm_SmallAreaLowGrayLevelEmphasis	-0.000833
LBP (2D)	firstorder_InterquartileRange	-0.000667
LoG 3.0 (3D)	glszm_HighGrayLevelZoneEmphasis.1	-0.000667
LoG 1.0 (3D)	glrlm_HighGrayLevelRunEmphasis.1	-0.000417
Wavelet HLH	glszm_SizeZoneNonUniformity	-0.000417
Wavelet HLH	glszm_SizeZoneNonUniformityNormalized	-0.000417
Wavelet LHH	glrlm_HighGrayLevelRunEmphasis	-0.000333
Wavelet LLL	glcm_ClusterTendency	-0.000333
Exponential	glszm_ZoneEntropy	-0.000250
Wavelet LHH	ngtdm_Strength	-0.000250
Gradient	glszm_SizeZoneNonUniformityNormalized.1	-0.000250
Exponential	firstorder_Entropy	-0.000167
Exponential	glszm_SmallAreaLowGrayLevelEmphasis	-0.000167

Wavelet LLH	glcm_ClusterProminence	-0.000167
Wavelet LHH	glrlm_ShortRunHighGrayLevelEmphasis	-0.000167
Wavelet HLL	glcm_JointEntropy	-0.000167
Gradient	glszm_GrayLevelNonUniformity	-0.000167
Wavelet HHL	glszm_SizeZoneNonUniformity.1	-0.000167
Wavelet LLL	glszm_HighGrayLevelZoneEmphasis.1	-0.000167
Exponential	firstorder_Energy.1	-0.000167
Exponential	firstorder_TotalEnergy.1	-0.000167
Exponential	ngtdm_Complexity	-0.000083
Exponential	ngtdm_Contrast.1	-0.000083
Exponential	glcm_ClusterProminence.1	-0.000083

The column "Filter" of tables shown above represent the filter used, the feature enhanced by PyRadiomics, present in column "Feature". The attributes that end with ".1" correspond to features present in the right-hippocampus, making it possible to distinguish between the hippocampus and, therefore, synthesise the information of features extracted divided by the hippocampus in the following tables (Table 4-8 and Table 4-9).

Table 4-8 – Relevant features in the left-hippocampus and importance values

Filter	Feature	Importance
Exponential	glcm_Correlation	0.014000
Exponential	glcm_Id	0.013000
Exponential	glcm_JointEnergy	0.012750
Exponential	glcm_MaximumProbability	0.012750
Exponential	gldm_LowGrayLevelEmphasis	0.012667
Exponential	glszm_LowGrayLevelZoneEmphasis	0.011833
Exponential	glszm_GrayLevelNonUniformityNormalized	0.011833
Exponential	glcm_Idmn	0.011500
Exponential	glcm_Idm	0.011417
Exponential	glcm_Imc1	0.011250
Exponential	ngtdm_Coarseness	0.011000
Exponential	glcm_Idn	0.010333
Exponential	glrlm_LowGrayLevelRunEmphasis	0.010333

78

Exponential	glszm_SizeZoneNonUniformityNormalized	0.010167
Exponential	glrlm_GrayLevelNonUniformityNormalized	0.010083
Exponential	glcm_MCC	0.009417
Exponential	firstorder_Uniformity	0.009000
LBP (3D) m1	firstorder_90Percentile	0.007333
LBP (2D)	firstorder_Median	0.007083
LBP (3D) m1	firstorder_Median	0.007000
LBP (3D) m1	firstorder_10Percentile	-0.003333
Wavelet HLH	glszm_SmallAreaHighGrayLevelEmphasis	-0.002250
LBP (3D) k	firstorder_10Percentile	-0.001167
Wavelet LHH	glszm_SmallAreaHighGrayLevelEmphasis	-0.001083
Wavelet HHL	glrlm_ShortRunLowGrayLevelEmphasis	-0.000833
LoG 3.0 (3D)	glszm_SmallAreaLowGrayLevelEmphasis	-0.000833
LBP (2D)	firstorder_InterquartileRange	-0.000667
Wavelet HLH	glszm_SizeZoneNonUniformity	-0.000417
Wavelet HLH	glszm_SizeZoneNonUniformityNormalized	-0.000417
Wavelet LHH	glrlm_HighGrayLevelRunEmphasis	-0.000333
Wavelet LLL	glcm_ClusterTendency	-0.000333
Wavelet LHH	ngtdm_Strength	-0.000250
Exponential	glszm_ZoneEntropy	-0.000250
Wavelet LLH	glcm_ClusterProminence	-0.000167
Wavelet LHH	glrlm_ShortRunHighGrayLevelEmphasis	-0.000167
Wavelet HLL	glcm_JointEntropy	-0.000167
Gradient	glszm_GrayLevelNonUniformity	-0.000167
Exponential	firstorder_Entropy	-0.000167
Exponential	glszm_SmallAreaLowGrayLevelEmphasis	-0.000167
Exponential	ngtdm_Complexity	-0.000083

Table 4-9 - Relevant features in the right-hippocampus and importance values

Filter	Feature	Importance
Exponential	ngtdm_Coarseness.1	0.012667
Exponential	glszm_LowGrayLevelZoneEmphasis.1	0.011917

Exponential	glcm_JointEnergy.1	0.011750
Exponential	glcm_MCC.1	0.011500
Exponential	glcm_Correlation.1	0.011167
Exponential	glszm_GrayLevelNonUniformityNormalized.1	0.011167
Exponential	firstorder_Uniformity.1	0.010917
Exponential	glcm_Id.1	0.010667
Exponential	gldm_LowGrayLevelEmphasis.1	0.010417
Exponential	glcm_Imc1.1	0.010417
Exponential	glrlm_GrayLevelNonUniformityNormalized.1	0.010167
Exponential	glszm_SizeZoneNonUniformityNormalized.1	0.010167
Exponential	glcm_Idmn.1	0.010000
Exponential	glcm_MaximumProbability.1	0.009833
Exponential	glrlm_LowGrayLevelRunEmphasis.1	0.009750
Exponential	glcm_Idn.1	0.009750
Exponential	glcm_Idm.1	0.009667
LBP (3D) m1	firstorder_90Percentile.1	0.008083
Wavelet HHL	glrlm_GrayLevelVariance.1	0.006583
Wavelet HHH	glrlm_LongRunLowGrayLevelEmphasis.1	0.006500
Wavelet HHL	glrlm_LowGrayLevelRunEmphasis.1	-0.005667
LoG 3.0 (3D)	glrlm_HighGrayLevelRunEmphasis.1	-0.004917
LoG 3.0 (3D)	glrlm_ShortRunLowGrayLevelEmphasis.1	-0.004500
Wavelet LLH	glszm_SmallAreaLowGrayLevelEmphasis.1	-0.003083
Wavelet LLH	glszm_LowGrayLevelZoneEmphasis.1	-0.002917
LBP (2D)	firstorder_10Percentile.1	-0.002417
LoG 5.0 (3D)	glszm_SmallAreaLowGrayLevelEmphasis.1	-0.001750
LoG 3.0 (3D)	glszm_GrayLevelVariance.1	-0.001750
LoG 3.0 (3D)	ngtdm_Complexity.1	-0.001417
LoG 1.0 (3D)	gldm_LargeDependenceLowGrayLevelEmphasis.1	-0.001167
LBP (3D) m1	firstorder_Median.1	-0.001083
LoG 3.0 (3D)	glszm_HighGrayLevelZoneEmphasis.1	-0.000667
LoG 1.0 (3D)	glrlm_HighGrayLevelRunEmphasis.1	-0.000417
Gradient	glszm_SizeZoneNonUniformityNormalized.1	-0.000250
Wavelet HHL	glszm_SizeZoneNonUniformity.1	-0.000167
Wavelet LLL	glszm_HighGrayLevelZoneEmphasis.1	-0.000167

Exponential	firstorder_Energy.1	-0.000167
Exponential	firstorder_TotalEnergy.1	-0.000167
Exponential	glcm_ClusterProminence.1	-0.000083
Exponential	ngtdm_Contrast.1	-0.000083

From Table 4-8, it is possible to identify that there are 40 relevant features in the left-hippocampus (where 20 of them have positive importance and the other 20 have negative importance). For the right-hippocampus, with information from Table 4-9, there are also 40 relevant features with the same positive and negative distribution (20 with positive importance and 20 with negative importance).

With these last 2 tables, is possible to identify what are the more relevant filters and which significant features they enable, for both hippocampi.

Table 4-10 - Relevant filters and frequency, for left-hippocampus

Filter	Frequency
Exponential	21
LBP (3D) m1	3
LBP (3D) k	1
LBP (2D)	2
Wavelet-HLH	3
Wavelet-HHL	4
LoG 3.0	1
Wavelet-LLL	1
Wavelet-LLH	1
Wavelet-HLL	1
Gradient	1

From Table 4-10, the barplot present in Figure 4-34 was developed.
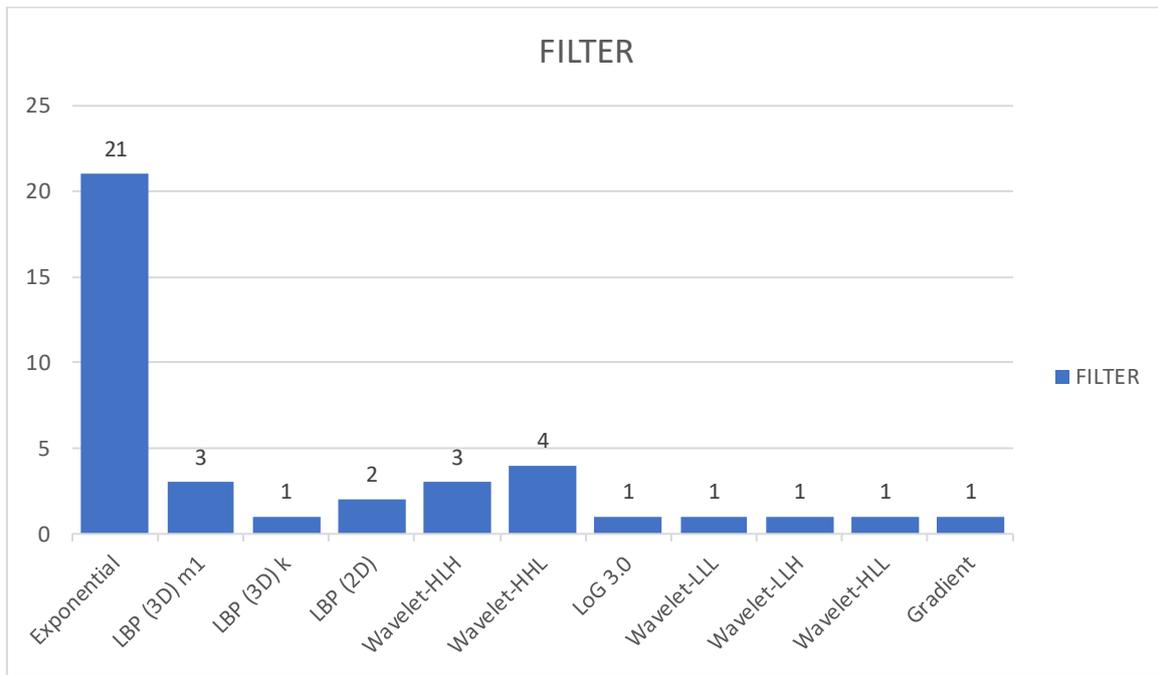
Figure 4-34 - Barplot of relevant filters and frequency, for left-hippocampus

In the left-hippocampus, the more relevant filter is the exponential filter (Figure 4-34) and the feature with more importance is *glcm_Correlation*, with an importance of 0.014000. This feature belongs to the class Grey Level Co-occurrence Matrix and corresponds to the linear dependency of grey level values to their respective voxels in the GLCM. Correlation is a value between 0 (uncorrelated) and 1 (perfectly correlated).

Table 4-11 - Relevant filters and frequency, for right-hippocampus

Filter	Frequency
Exponential	21
LBP (3D) m1	1
Wavelet-HHL	3

Wavelet-HHH	1
LoG 3.0	5
Wavelet-LLH	2
LBP (2D)	1
LoG 5.0	1
LoG 1.0	1
Gradient	1
Wavelet-LLL	1

From Table 4-11, the barplot present in Figure 4-35 was developed.
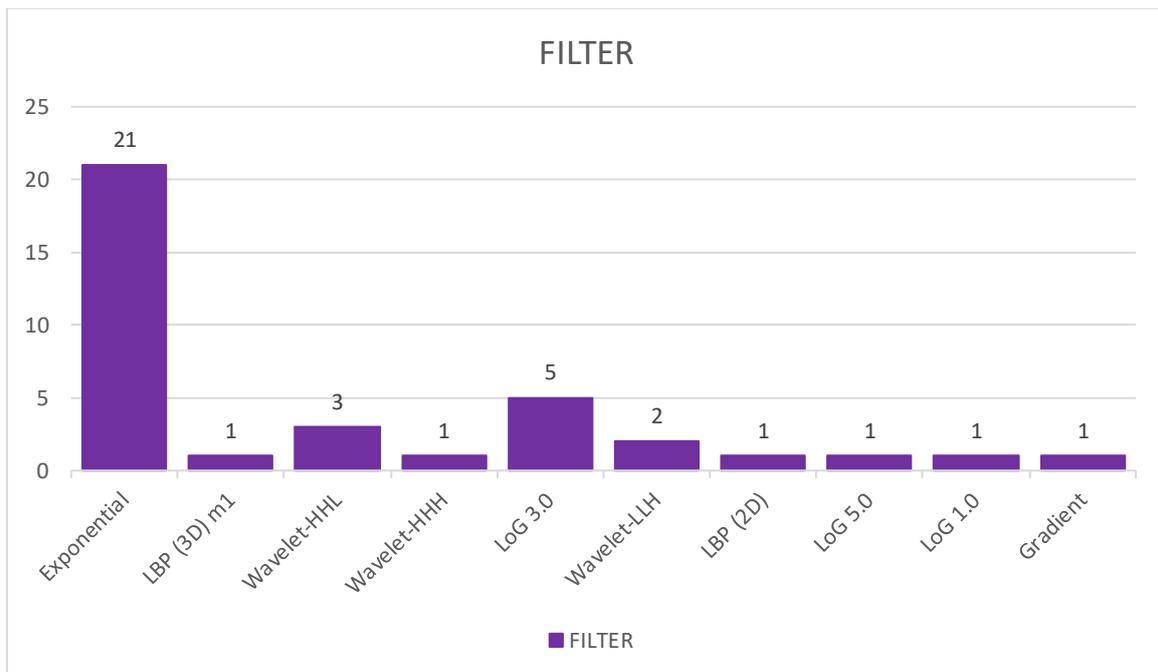

Figure 4-35 - Barplot of relevant filters and frequency, for right-hippocampus

For the right-hippocampus, the more relevant filter corresponds to the exponential filter as well, which allows concluding that the exponential filter is the filter that is capable to enhance more relevant features to classify a subject in the three different stages of AD, in both hippocampi. The feature with more importance is *ngtdm_Coarseness*, with an importance of 0.012667 (Table 4-9). This feature belongs to the class Neighbouring Grey Tone Difference Matrix and calculates and returns the coarseness. Coarseness is a measure of the average difference between the centre voxel and its neighbourhood and is an indication of the spatial rate of change. A higher value indicates a lower spatial change rate and a locally more uniform texture.

83

Another analysis that can be made corresponds to identify which features are repeated in both hippocampi. The features repeated in both hippocampi, with positive importance, are 18 and are the ones present in Table 4-12.

Table 4-12 - Repeated features in both hippocampi, with positive importance

Repeated features in both hippocampi, with positive importance	
Filter	Feature
Exponential	glcm_Idn
Exponential	glcm_Correlation
Exponential	firstorder_Uniformity
Exponential	gldm_LowGrayLevelEmphasis
Exponential	ngtdm_Coarseness
Exponential	glcm_MCC
Exponential	glcm_JointEnergy
LBP (3D) m1	firstorder_90Percentile
Exponential	glcm_MaximumProbability
Exponential	glrlm_GrayLevelNonUniformityNormalized
Exponential	glcm_Id
Exponential	glszm_LowGrayLevelZoneEmphasis
Exponential	glszm_GrayLevelNonUniformityNormalized
Exponential	glcm_Idmn
Exponential	glcm_Imc1
Exponential	glszm_SizeZoneNonUniformityNormalized
Exponential	glcm_Idm

On the other hand, no feature is repeated in both hippocampi with negative importance.

Since there are no repeated features with negative importance in both hippocampi, there are no features that have simultaneously positive and negative importance and were present in both hippocampi.

5 DISCUSSION AND CONCLUSION

5.1 Synopsis

This work aimed to develop techniques to assist researchers and neuroradiologists in the diagnosis and prediction of neurodegenerative disease development based on magnetic resonance images and with the use of Deep Learning. Two overarching goals of this work are to detect AD at the earliest possible stage (pre-dementia) and identify ways to track the disease's progression and, therefore, to support advances in AD intervention, prevention, and treatment through the application of new diagnostic methods at the earliest possible stages (when intervention may be most effective).

This thesis began with a survey and study of the state of the art and an introduction to the application of ML in Neuroimaging, giving greater emphasis and specificity to Deep Learning, as well as a study of the biological part of the brain.

This work was developed in two steps: brain segmentation, using FreeSurfer, and feature extraction, using PyRadiomics. For this, there was a need to define a database with images from ADNI, followed by the implementation of PyRadiomics in these volumetric brain acquisitions. Conducting tests and validating them with case studies allowed to collect the results obtained and writing of the dissertation.

Combining the results of these main steps, the goal was to measure the progression of control patients to MCI and AD.

Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

In this work, a radiomics model was developed to predict the presence of clinically significant features with high diagnostic efficacy and clinical benefit. For this purpose, a MLP was used, which was trained with a dataset of patients diagnosed in one of three group stages: cognitively normal, mild cognitive impairment, and patients with Alzheimer's disease.

The initial dataset downloaded from ADNI was composed of 1217 subjects, with 4712 MRI images. After that, a selection of only 600 subjects was made, to apply this work only on this selection. All the data is balanced and equally divided between the 3 groups. The brain of these 600 subjects was segmented in Freesurfer and, after that, PyRadiomics was applied to the left and right hippocampus, to extract radiomics features. Since the images tend to have different intensities, contrast, and noise, there is a need to make some pre-processing first, before applying Radiomics, so that images appear more similar and to ensure that these variations do not influence the final classification. Images downloaded from ADNI already went through some pre-process steps, so this was not necessary to implement.

When the network was created and then trained with the selected dataset, it was tested. During training, the performance improved significantly when using the model with Batch Normalization layers, which supported the benefits of regularizing the network.

AI has the potential to facilitate or even replace the diagnosis of patients by performing tasks such as early detection and quantification of the probability to develop a neurodegenerative disease. The use of artificial intelligence in medical imaging will allow the professionals to spend more time, e.g. communicating with patients or deliberating with colleagues. They will not be overloaded with the number of medical exams to be analyzed. Deep Learning techniques are promising for analyzing medical exams, as it is possible to identify findings and detect patterns efficiently. Medical professionals should use their time to treat patients and not waste their time treating medical images.

5.2 DISCUSSION

The work began with the investigation of Deep Learning techniques that have already been applied in medical images to predict dementias [11][18][26] and try to classify them, based on brain morphological characteristics [16][59].

The literature shows an improvement in the diagnostic prediction rate (accuracy) and provides a more reliable classification of diseases, with feature-based methods that use brain characteristics of the MRI image becoming the most popular and successful methods for AD diagnosis [27][69].

Recent advances in computer science have enabled the clinical application of computer-assisted analysis in imaging examinations, of which radiomics [26] and Deep Learning are currently the most actively investigated techniques [11]. Both radiomics and Deep Learning use high-dimensional features extracted from images for diagnostic and predictive tasks. Radiomics and Deep Learning may also expand the role of imaging in the assessment of various neurodegenerative diseases beyond the domain of traditional visual image analysis, by obtaining additional diagnostic information from images [26], assessing image features comprehensively and objectively. Radiomics has gained considerable attention in the field of neurodegenerative diseases as a method for supporting clinical decision-making and precision medicine [11][25][69]. This methodology is based on the hypothesis that a radiologic phenotype may reflect genetic alterations in CNS biology and may thus be predictive of the biologic behaviour of neurodegenerative diseases.

Deep Learning techniques are promising for analysing medical exams, as it is possible to identify findings and detect patterns efficiently. Therefore, the results of the present study demonstrate that

radiomic features have potential use in clinical diagnosis for the early detection of AD from CN, with relatively high accuracy (99.2%). We believe that several reasons may account for this high value. First, features were extracted from a dataset with MRI that were previously processed and normalized. Second, the region selected was based on prior knowledge, adopting an automated segmentation of the hippocampus, using FreeSurfer, and only then the features were extracted. At present, morphological analyses of medical imaging scans for traditional diagnosis of AD is usually based on manual extraction of features operated manually by skilled radiologists and imaging technicians, which requires an impractical amount of time and human resources for image pre-processing and can potentially result in reduced efficiency and inaccuracy, as well as overdependence on the analyst's skills [70], so using FreeSurfer prevents this from happening and influencing the final results. Besides this, the use of a selection feature method may have helped improve the robustness of the results, since feature selection represents an active area of development in the field that has been successfully employed in previous radiomic studies [71][72] and thus is of critical importance in this type of studies. We speculate that radiomics features could reflect changes at the microscopic level during the early pathological stages, which occur before changes at the macroscopic level. These results highlight the importance of hippocampal texture abnormalities in AD and support the possibility that textures may serve as a neuroimaging biomarker for the early detection of AD and MCI.

Looking to the future, imaging has helped establish that there is a long preclinical and presymptomatic period where the pathological effects of AD are detectable and, therefore, imaging is starting to provide prognostic information at an early preclinical stage. It has a major role to play in improving the understanding of this disease. In previous studies made, different automatic classification methods to assist in the early diagnosis of Alzheimer's disease using the ADNI database were compared [11][26][73] and most of them classify AD and CN with high accuracy. As stated by Jo et all. (2019), the values of accuracy for AD/CN classification vary between 91.4% and 98.8% and the prediction accuracy of MCI to AD conversion vary between 57.4% and 83.7%. The results with the lowest values were obtained by Li et al. (2015) when was used one of the first models developed to overcome the overfitting problem with a dropout technique and SVM as a classifier. These findings provide evidence showing that radiomic features are beneficial in detecting early cognitive decline. However, at the prodromal stage, their sensitivity was substantially lower [73]. Combinations with other markers and/or more sophisticated prior knowledge seem necessary to be able to detect prodromal AD with high accuracy.

Although radiomics analysis has already been applied to AD and MCI diagnosis and classification [27][52], there still is a long way to go from these computer-aided diagnostic methods to clinical application. The application of hippocampal features as a biomarker still needs to be more studied and confirmed.

5.3 CONCLUSION

The present dissertation aimed to provide an overview of published research based on ML for classifying patients with neurodegenerative diseases, in particular AD, and healthy controls, as well as an approach of ML techniques used for classification.

Overall, PyRadiomics provides a flexible radiomic quantification platform, with a back-end interface within Python allowing automation in data processing and feature definition. Radiomics analysis can extract high-dimensional features of MRI and may identify imaging patterns in the preclinical stages of neurodegenerative diseases that cannot be recognized by human readers. Therefore, radiomics analysis is expected to become a new evaluation method for future cognitive decline in cognitively healthy individuals, which would be of great importance in diagnosing preclinical AD and targeting ultra-early secondary prevention clinical trials.

With the results obtained here (Table 4-4 and Figure 4-31) we propose that it is possible to differentiate between control patients from those with AD. A more difficult task is differentiating MCI subjects from those who will or will not progress to AD soon.

There are some disadvantages to using radiomics approaches. These methods are labour-intensive and time-consuming as they involve segmentation, feature extraction, and ML or modelling processes. Hence, a radiomics study will only produce real clinical value if it generates incremental diagnostic information beyond what can be obtained with classic visual image interpretation. Radiomics features are also highly dependent on the imaging protocol and feature extraction methods. All of these factors may be sources of variation in terms of extracted radiomics features. Nonetheless, deep learning has had an important positive impact on brain disorder research. As the amount of data keeps increasing, good practices become more widespread and deep learning stays a hot research topic and can be a key component to achieve precision medicine and improve disease understanding for brain disorders and medicine in general.

Although some constraints, e.g. dataset dimension, the results of the work developed in this dissertation are encouraging for use in neuroimaging for classifying neurodegenerative diseases.

AI has the potential to facilitate or even replace the diagnosis of patients by performing tasks such as detection, qualification, and quantification. The use of AI in medical imaging will allow the professionals to spend more time, e.g. communicating with patients or deliberating with colleagues. The incorporation of radiomics features and clinical factors into the study of neurodegenerative diseases can provide a quantitative and intuitive approach for neuroradiologists that could more confidently predict the onset and development of AD.

Medical professionals should use their time to treat patients and not waste their time treating medical images. They will not be overloaded with a large number of medical exams to be analysed.

5.4 RESEARCH OPPORTUNITIES

In recent years, the use of ML for early diagnoses of neurodegenerative diseases among researchers gained momentum and is increasing. However, ML still has several challenges that need to be overcome, including population size, since this type of study needs a large number of study populations to have significant and reliable results. Deep learning algorithms typically require large volumes of high-quality ground truth training data. Furthermore, to date there remains a lack of an accurate ML technique for all applications; therefore, it may be useful to test various techniques with each dataset, which could maybe lead to different results. Besides this, Radiomics models or classifiers also have inherent limitations in terms of generalization. The lack of a standardized method for radiomics feature extraction has been an important cause of the poor generalizability of radiomics studies.

Further research could utilize the best statistics and features described herein for estimates of neurodegenerative disease classification to be more robust and, therefore, to overcome the traditional pitfalls of deep neural networks.

Some other options and studies could be made in the future, based on the results and using the MLP developed in this dissertation, like seeing the evolution of the principal features, i.e., see if they change through time or training the MLP only with the features with positive importance, removing from the negative ones. After this, some analysis could be made: see if the MLP metrics get better, compare and see if the ranking of the positive features changes, and if they get a different value of importance. Because only the hippocampus was included, some other important regions, such as the parahippocampus, amygdala and ventricles, should be investigated in future work.

This page was intentionally left blank.

REFERENCES

[1]   "WHO | Medical imaging." [Online]. Available: https://www.who.int/diagnostic_imaging/en/. [Accessed: 30-Jan-2020].

[2]   "Radiomics." [Online]. Available: https://www.radiomics.io/index.html. [Accessed: 30-Jan-2020].

[3]   "Computed Tomography & Magnetic Resonance Imaging Of The Whole Body E-Book - John R. Haaga, Daniel Boll - Google Books." [Online]. Available: https://books.google.pt/books?id=ii3kS2ih8JsC&pg=PA626&lpg=PA626&dq=MRI,+which+provides+variable+image+contrast+and+high+contrast+for+soft+tissue&source=bl&ots=ua5ulWDKhA&sig=ACfU3U176YwDc2kUzq-fazCnhlHMCS58FQ&hl=en&sa=X&ved=2ahUKEwijk86BrLjnAhUO9IUKHQNWB_UQ6AEwCnoECAoQAQ#v=onepage&q=MRI%2C which provides variable image contrast and high contrast for soft tissue&f=false. [Accessed: 04-Feb-2020].

[4]   K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research."

[5]   "Overview of Nervous System Disorders | Spectrum Health Lakeland." [Online]. Available: https://www.spectrumhealthlakeland.org/lakeland-neurosurgery/neurosurgery-health-library/Content/85/P00799. [Accessed: 15-Jun-2021].

[6]   "White Matter in the Brain." [Online]. Available: https://www.verywellhealth.com/what-is-white-matter-in-the-brain-98119. [Accessed: 04-Feb-2020].

[7]   "Access Denied." [Online]. Available: https://www.topperlearning.com/answer/state-the-function-of-nerve-cells-draw-a-neat-diagram-of-a-nerve-cell/z9e46cqq. [Accessed: 16-Jun-2021].

[8]   P. S. J. Weston, I. J. A. Simpson, N. S. Ryan, S. Ourselin, and N. C. Fox, "Diffusion imaging changes in grey matter in Alzheimer's disease: a potential marker of early neurodegeneration," *Alzheimers. Res. Ther.*, vol. 7, no. 1, p. 47, Dec. 2015.

[9]   "Difference between White and Gray Matter | Difference.Guru." [Online]. Available: https://difference.guru/difference-between-white-and-gray-matter/. [Accessed: 16-Jun-2021].

[10]   "Doenças degenerativas explicadas | Lusíadas Saúde." [Online]. Available: https://rotasaude.lusiadas.pt/doencas/sintomas-e-tratamentos/doencas-neurodegenerativas/. [Accessed: 30-Jan-2020].

[11]   J. H. Cole, D. Lu, Z. Wang, T. Jo, K. Nho, and A. J. Saykin, "Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data," *Front. Aging Neurosci. | www.frontiersin.org*, vol. 1, p. 220, 2019.

[12] "O que é a doença de Alzheimer e quais os sintomas? | Lusíadas Saúde." [Online]. Available: https://rotasaude.lusiadas.pt/doencas/sintomas-e-tratamentos/o-que-e-a-doenca-de-alzheimer/.

[13] M. V. F. Silva, C. D. M. G. Loures, L. C. V. Alves, L. C. De Souza, K. B. G. Borges, and M. D. G. Carvalho, "Alzheimer's disease: Risk factors and potentially protective measures," *J. Biomed. Sci.*, vol. 26, no. 1, May 2019.

[14] "ADNI | MRI Analysis." [Online]. Available: http://adni.loni.usc.edu/methods/mri-tool/mri-analysis/. [Accessed: 30-Sep-2020].

[15] "ADNI | Study Design." [Online]. Available: http://adni.loni.usc.edu/study-design/#background-container. [Accessed: 07-Oct-2020].

[16] "Magnetic Resonance Imaging - an overview | ScienceDirect Topics." [Online]. Available: https://www.sciencedirect.com/topics/medicine-and-dentistry/magnetic-resonance-imaging. [Accessed: 26-Aug-2020].

[17] A. M. Brickman, J. Muraskin, and M. E. Zimmerman, "Structural neuroimaging in Alzheimer's disease: Do white matter hyperintensities matter?," *Dialogues in Clinical Neuroscience*, vol. 11, no. 2. pp. 181–190, 2009.

[18] "MRI Systems & Solutions | Philips Healthcare." [Online]. Available: https://www.philips.co.in/healthcare/solutions/magnetic-resonance. [Accessed: 16-Jun-2021].

[19] K. A. Johnson, N. C. Fox, R. A. Sperling, and W. E. Klunk, "Brain Imaging in Alzheimer Disease," *Cold Spring Harb. Perspect. Med.*, vol. 2, no. 4, pp. a006213–a006213, Apr. 2012.

[20] "Multimodal Neuroimaging: Basic Concepts and Classification of Neuropsychiatric Diseases."

[21] M. Larobina and L. Murino, "Medical Image File Formats."

[22] "DICOM." [Online]. Available: https://www.dicomstandard.org/. [Accessed: 22-Dec-2020].

[23] "THE NIFTI-1 DATA FORMAT."

[24] "NIfTI (file format) | Radiology Reference Article | Radiopaedia.org." [Online]. Available: https://radiopaedia.org/articles/nifti-file-format. [Accessed: 22-Dec-2020].

[25] "FsTutorial/MghFormat - Free Surfer Wiki." [Online]. Available: https://surfer.nmr.mgh.harvard.edu/fswiki/FsTutorial/MghFormat. [Accessed: 01-Jul-2021].

[26] J. J. M. Van Griethuysen *et al.*, "Computational radiomics system to decode the radiographic phenotype," *Cancer Res.*, vol. 77, no. 21, pp. e104–e107, Nov. 2017.

[27] Q. Feng and Z. Ding, "MRI Radiomics Classification and Prediction in Alzheimer's Disease and Mild Cognitive Impairment: A Review," *Curr. Alzheimer Res.*, vol. 17, no. 3, pp. 297–309, Mar.

2020.

[28] Y. Wang *et al.*, "Predicting the Type of Tumor-Related Epilepsy in Patients With Low-Grade Gliomas: A Radiomics Study," *Front. Oncol.*, vol. 10, no. March, 2020.

[29] H. J. Park, B. Park, and S. S. Lee, "Radiomics and deep learning: Hepatic applications," *Korean J. Radiol.*, vol. 21, no. 4, pp. 387–401, 2020.

[30] "Welcome to pyradiomics documentation! — pyradiomics 2.2.0.post43+gf719e5e documentation." [Online]. Available: https://pyradiomics.readthedocs.io/en/latest/. [Accessed: 30-Jan-2020].

[31] "(Tutorial) KERAS Tutorial: DEEP LEARNING in PYTHON - DataCamp." [Online]. Available: https://www.datacamp.com/community/tutorials/deep-learning-python. [Accessed: 25-Dec-2020].

[32] "(PDF) Urdu News Headline, Text Classification by Using Different Machine Learning Algorithms." [Online]. Available: https://www.researchgate.net/publication/331889811_Urdu_News_Headline_Text_Classification_by_Using_Different_Machine_Learning_Algorithms. [Accessed: 16-Jun-2021].

[33] "Crash Course On Multi-Layer Perceptron Neural Networks." [Online]. Available: https://machinelearningmastery.com/neural-networks-crash-course/. [Accessed: 16-Jan-2021].

[34] "(PDF) Investigating Keystroke Dynamics as a Two-Factor Biometric Security." [Online]. Available: https://www.researchgate.net/publication/325870973_Investigating_Keystroke_Dynamics_as_a_Two-Factor_Biometric_Security. [Accessed: 16-Jun-2021].

[35] "When to Use MLP, CNN, and RNN Neural Networks." [Online]. Available: https://machinelearningmastery.com/when-to-use-mlp-cnn-and-rnn-neural-networks/. [Accessed: 16-Jan-2021].

[36] "Deep Learning With Keras & LIME in R." [Online]. Available: https://rstudio-pubs-static.s3.amazonaws.com/351198_33fe919a0364428bba3c848f4358fd5e.html. [Accessed: 23-Jan-2021].

[37] "About Feature Scaling and Normalization." [Online]. Available: https://sebastianraschka.com/Articles/2014_about_feature_scaling.html. [Accessed: 23-Jan-2021].

[38] "Understanding Confusion Matrix. When we get the data, after data... | by Sarang Narkhede | Towards Data Science." [Online]. Available: https://towardsdatascience.com/understanding-

confusion-matrix-a9ad42dcfd62. [Accessed: 15-Dec-2020].

[39]   "ML Metrics: Sensitivity vs. Specificity - DZone AI." [Online]. Available: https://dzone.com/articles/ml-metrics-sensitivity-vs-specificity-difference. [Accessed: 26-Jan-2021].

[40]   "Sensitivity, Specificity and Meaningful Classifiers | by Simon Spichak | Towards Data Science." [Online]. Available: https://towardsdatascience.com/sensitivity-specificity-and-meaningful-classifiers-8326738ec5c2. [Accessed: 26-Jan-2021].

[41]   "How to Reduce Overfitting With Dropout Regularization in Keras." [Online]. Available: https://machinelearningmastery.com/how-to-reduce-overfitting-with-dropout-regularization-in-keras/. [Accessed: 29-Oct-2020].

[42]   T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, vol. 2017-October, pp. 2999–3007.

[43]   "A Gentle Introduction to Dropout for Regularizing Deep Neural Networks." [Online]. Available: https://machinelearningmastery.com/dropout-for-regularizing-deep-neural-networks/. [Accessed: 29-Oct-2020].

[44]   "Keras: the Python deep learning API." [Online]. Available: https://keras.io/. [Accessed: 25-Jan-2021].

[45]   "FreeSurfer." [Online]. Available: https://surfer.nmr.mgh.harvard.edu/. [Accessed: 30-Jan-2020].

[46]   "TensorFlow." [Online]. Available: https://www.tensorflow.org/. [Accessed: 24-Feb-2021].

[47]   "TensorFlow 2 Tutorial: Get Started in Deep Learning With tf.keras." [Online]. Available: https://machinelearningmastery.com/tensorflow-tutorial-deep-learning-with-tf-keras/. [Accessed: 10-Mar-2021].

[48]   "What is TensorFlow? The machine learning library explained | InfoWorld." [Online]. Available: https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html. [Accessed: 18-Mar-2021].

[49]   "Prevalência da Demência | Associação Alzheimer Portugal." [Online]. Available: https://alzheimerportugal.org/pt/text-0-9-106-349-prevalencia-da-demencia. [Accessed: 21-Jan-2021].

[50]   C. R. Jack *et al.*, "Antemortem MRI findings correlate with hippocampal neuropathology in typical aging and dementia," *Neurology*, vol. 58, no. 5, pp. 750–757, Mar. 2002.

[51] K. M. Gosche, J. A. Mortimer, C. D. Smith, W. R. Markesbery, and D. A. Snowdon, "Hippocampal volume as an index of Alzheimer neuropathology: Findings from the Nun study," *Neurology*, vol. 58, no. 10, pp. 1476–1482, May 2002.

[52] E. Scola *et al.*, "A diffusion tensor MRI study of patients with MCI and AD with a 2-year clinical follow-up," *J. Neurol. Neurosurg. Psychiatry*, vol. 81, no. 7, pp. 798–805, Jul. 2010.

[53] S. E. Rose, A. L. Janke, and J. B. Chalk, "Gray and white matter changes in Alzheimer's disease: A diffusion tensor imaging study," *J. Magn. Reson. Imaging*, vol. 27, no. 1, pp. 20–26, Jan. 2008.

[54] P. C. Van, "Relationship between hippocampal volume and memory ability in healthy individuals across the lifespan: review and meta-analysis," *Neuropsychologia*, vol. 42, no. 10, pp. 1394–1413, 2004.

[55] O. Voevodskaya *et al.*, "The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and Alzheimer's disease," 2014.

[56] X. Wang, Y. Zhao, and F. Pourpanah, "Recent advances in deep learning," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 3, pp. 747–750, 2020.

[57] R. Anitha, S. Jyothi, and P. S. Krishna, "Medical Image Segmentation to Diagnosis Alzheimer Disease using Neural Networks," *Int. J. Eng. Trends Technol.*, vol. 39, no. 1, pp. 51–56, Sep. 2016.

[58] "ADNI | MRI Pre-Processing." [Online]. Available: http://adni.loni.usc.edu/methods/mri-tool/mri-pre-processing/. [Accessed: 04-Feb-2020].

[59] "Neuroimaging in Python — NiBabel 3.2.0 documentation." [Online]. Available: https://nipy.org/nibabel/. [Accessed: 25-Jan-2021].

[60] "mri_convert - Free Surfer Wiki." [Online]. Available: https://surfer.nmr.mgh.harvard.edu/fswiki/mri_convert. [Accessed: 08-Dec-2020].

[61] "Introduction - FSLUTILS." [Online]. Available: https://fsl.fmrib.ox.ac.uk/fslcourse/lectures/practicals/intro3/index.html. [Accessed: 08-Dec-2020].

[62] "Matplotlib: Python plotting — Matplotlib 3.3.3 documentation." [Online]. Available: https://matplotlib.org/. [Accessed: 25-Jan-2021].

[63] A. Qiu, L. Younes, M. I. Miller, and J. G. Csernansky, "Parallel transport in diffeomorphisms distinguishes the time-dependent pattern of hippocampal surface deformation due to healthy aging and the dementia of the Alzheimer's type," *Neuroimage*, vol. 40, no. 1, pp. 68–76, Mar.

2008.

[64] "sklearn.preprocessing.MinMaxScaler — scikit-learn 0.24.1 documentation." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html. [Accessed: 25-Jan-2021].

[65] "NumPy." [Online]. Available: https://numpy.org/. [Accessed: 25-Jan-2021].

[66] N. Srivastava, G. Hinton, A. Krizhevsky, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," 2014.

[67] "eli5.permutation_importance — ELI5 0.9.0 documentation." [Online]. Available: https://eli5.readthedocs.io/en/latest/autodocs/permutation_importance.html. [Accessed: 15-Dec-2020].

[68] "Matplotlib Heatmap: Data Visualization Made Easy - Python Pool." [Online]. Available: https://www.pythonpool.com/matplotlib-heatmap/. [Accessed: 15-Dec-2020].

[69] R. Perneczky *et al.*, "Radiomic Features of Hippocampal Subregions in Alzheimer's Disease and Amnestic Mild Cognitive Impairment," 2018.

[70] delEtoile J and A. H, "Graph Theory and Brain Connectivity in Alzheimer's Disease," *Neuroscientist*, vol. 23, no. 6, pp. 616–626, Dec. 2017.

[71] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, and J. Tohka, "Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects," *Neuroimage*, vol. 104, pp. 398–412, Jan. 2015.

[72] P. P. Foster *et al.*, "Hippocampus Radiomic Biomarkers for the Diagnosis of Amnestic Mild Cognitive Impairment: A Machine Learning Method."

[73] R. Cuingnet *et al.*, "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database," *Neuroimage*, vol. 56, no. 2, pp. 766–781, May 2011.