CATÓLICA
FACULTY OF BIOTECHNOLOGY
PORTO

Artificial Intelligence System for the Automatic Detection of Alzheimer Disease through Electroencephalographic Signals

by
Teresa Guerra Mendonça de Sousa de Araújo

November 2021

CATÓLICA
FACULTY OF BIOTECHNOLOGY

PORTO

Artificial Intelligence System for the Automatic Detection of Alzheimer Disease through Electroencephalographic Signals

Thesis presented to Escola Superior de Biotecnologia of the
Universidade Católica Portuguesa to fulfill the requirements of Master of Science Degree in
Biomedical Engineering

by
Teresa Guerra Mendonça de Sousa de Araújo

Supervisor: Prof. Doutor Pedro Miguel de Luís Rodrigues
Escola Superior de Biotecnologia, Universidade Católica Portuguesa

Co-Supervisor: Prof. Doutor João Paulo Ramos Teixeira
Instituto Politécnico de Bragança

November 2021

Abstract

Alzheimer's Disease (AD) stands out as one of the main causes of dementia. This neurodegenerative disease is characterised by the deterioration of human cognitive functions – the accumulation of toxic substances in the brain causes the progressive death of neuronal cells. Worldwide, AD represents around 65% of all dementia cases, affecting mainly elderly people. This disease is composed by four evolutionary stages and the asymptomatic period can last up until 20 years. With respect to the researcher's community, this topic remains a huge challenge since it is crucial to create a tool to assist the diagnosis in the early stages with the aim of halting the disease progression. In this way, the main purpose of this dissertation is to develop a system that would be able to differentiate each disease stage. Thereby, a nonlinear multiband analysis of the Electroencephalographic Signals (EEG) was performed enabling to study its behaviour and to extract several features from each study group. After a feature selection per electrode, it was executed, by means of Classic Machine Learning (ML) and Deep Learning (DL) methods, the data classification through a process of leave-one-out cross validation. The maximum accuracies obtained were 78.9% (C vs MCI), 81.0% (C vs ADM), 84.2% (C vs ADA), 88.9% (MCI vs ADM), 93.8% (MCI vs ADA), 77.8% (ADM vs ADA) and 56.8% (All vs All). Considering the topographic maps, it can be concluded that the central and parietal brain regions are the ones that present the most significant differences when the study groups were discriminated. In conclusion, it can be stated that entropy features are the most relevant and that DL did not over performed Classic ML results. Regarding the state of the art with the same EEG database, the proposed method outperforms by 2% in the binary comparison MCI vs ADM. This improvement reflects the performance of this powerful tool in detecting AD.

Keywords: Alzheimer Disease, Nonlinear Multiband Analysis, Electroencephalographic Signals, Classic Machine Learning, Deep Learning.

Resumo

A Doença de Alzheimer (DA) destaca-se como uma das principais causas de demência. Esta doença neurodegenerativa é caracterizada pela deterioração das funções cognitivas humanas – a acumulação de substâncias tóxicas no cérebro causa a morte progressiva das células neuronais. A nível mundial, a DA representa cerca de 65% de todos os casos de demência, afetando principalmente as pessoas idosas. Esta doença é composta por quatro fases evolutivas e o período assintomático pode durar até 20 anos. No que diz respeito à comunidade de investigadores, este tópico continua a ser um enorme desafio, uma vez que é crucial criar uma ferramenta para auxiliar o diagnóstico nas fases iniciais, com o objetivo de travar a progressão da doença. Desta forma, o principal propósito desta dissertação é desenvolver um sistema que seja capaz de diferenciar cada fase da doença. Assim, foi realizada uma análise não linear multibanda dos Sinais Eletroencefalográficos (EEG), permitindo o estudo do seu comportamento e a extração de várias características de cada grupo de estudo. Após uma seleção de características por elétrodo, foi realizada, através de métodos de *Machine Learning* (ML) Clássico e *Deep Learning* (DL), a classificação dos dados através de um processo de validação cruzada de *leave-one-out*. As precisões máximas obtidas foram 78,9% (C vs MCI), 81,0% (C vs ADM), 84,2% (C vs ADA), 88,9% (MCI vs ADM), 93,8% (MCI vs ADA), 77,8% (ADM vs ADA) e 56,8% (All vs All). Atendendo aos mapas topográficos, pode concluir-se que as regiões do cérebro central e parietal são as que apresentam diferenças mais significativas quando se discriminam os grupos de estudo. Em conclusão, pode-se afirmar que as características de entropia são as mais relevantes e que DL não apresentou resultados superiores ao ML Clássico. Relativamente ao estado da arte com a mesma base de dados EEG, o método proposto supera em 2% na comparação binária MCI vs ADM. Esta melhoria reflete o desempenho desta poderosa ferramenta na detecção da AD.

Palavras-chave: Doença de Alzheimer, Análise Não Linear Multibanda, Sinais Eletroencefalográficos, Machine Learning Clássico, Deep Learning.

Acknowledgements

Com a conclusão da tese, dou por terminado o meu percurso académico. Foram 5 anos de muito trabalho e dedicação e nada disto teria sido igual sem o contributo de pessoas especiais que estiveram constantemente presentes.

Gostaria de expressar a minha gratidão ao Professor Doutor Pedro pela sua prontidão durante o desenvolvimento da tese. O seu apoio e encorajamento como orientador foram fundamentais. Adicionalmente, gostaria de reconhecer as sugestões construtivas do Professor Doutor João Paulo.

Gostaria de agradecer o apoio e amor que o Pai, a Mãe e o Francisco me deram durante o curso. A força e motivação que me transmitiram foi essencial para o meu crescimento pessoal e profissional. Aos Pais, obrigada por me terem proporcionado tudo isto. Ao Francisco, obrigada por ter caminhado sempre ao meu lado.

À Joana, pela amizade e cumplicidade. Nesta fase da vida tão importante, desenvolvemos uma ligação única e um sentimento de interajuda peculiar.

À Carolina, pelo orgulho que demonstrou ter em mim.

À Beatriz, pela compreensão em momentos importantes.

À Eunice, ao André, à Maria, ao Bernardo e à Rita, por Roma nos ter unido para sempre.

À Rebeca, pelo carinho e preocupação permanente.

Ao João e ao António, pela amizade verdadeira e interminável.

A todos, um enorme obrigada.

Teresa

"All progress takes place outside the comfort zone"

Michael John Bobak

Outline

List of Figures

List of Tables

Acronyms and Abbreviations

AD	Alzheimer Disease
APP	Amyloid Precursor Protein
APOE	Apolipoprotein E
MCI	Mild Cognitive Impairment
IT	Imaging Techniques
EEG	Electroencephalogram/Electroencephalographic/Electroencephalography
CFS	Cerebrospinal Fluid
MRI	Magnetic Resonance Imaging
PET	Positron Emission Tomography
sMRI	Structural MRI
fMRI	Functional MRI
BOLD	Blood Oxygen Level-Dependent
FDG	Fluorodeoxyglucose
NMDA	N-Methyl-D-Aspartic
ANN	Artificial Neural Networks
PSD	Power Spectral Density
RP	Relative Power
SVM	Support Vector Machine
KNN	K-Nearest Neighbor
DT	Decision Tree
SDT	Surrogate Decision Tree
CNN	Convolutional Neural Networks
MMSE	Mini Mental State Examination
FCBF	Fast Correlation Based-Filter
MLP	Multi-Layer Perceptron Artificial Neural Network
DWT	Discrete Wavelet Transform
RNN	Recurrent Neural Networks
GRU	Gated Recurrent Unit
MUSIC-EWT	Multiple Signal Classification and Empirical Wavelet Transform

BD	Box Dimension
HFD	Higuchi Fractal Dimension
KFD	Katz Fractal Dimension
HE	Hurst Exponent
EPNN	Enhanced Probabilistic Neural Network
AD1	Mild AD
AD2	Moderate AD
cMCI	Converters MCI
sMCI	Stable MCI
A/D	Analogue-Digital
ERS	Event-Related Synchronization
C	Controls
ADM	Alzheimer Disease Mild and Moderate
ADA	Alzheimer Disease Advanced
CWT	Continuous Wavelet Transform
DWPT	Discrete Wavelet Packet Transform
Haar	Haar
Db	Daubechies
Sym	Symlets
Coif	Coiflets
Bior	Biorthogonal
Rbio	Reverse Biorthogonal
Dmey	Discrete Approximation of Meyer
Fk	Fejer-Korovkin
AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
m	Mean
v	Variance
sd	Standard Deviation
Asy	Asymmetry
Kur	Kurtosis
ShEn	Shannon Entropy
PEn	Permutation Entropy
SEn	Sample Entropy
LEn	Log Entropy
AEn	Approximate Entropy

LEx	Lyapunov Exponent
HEx	Hurst Exponent
HFD	Higuchi Fractal Dimension
ReLU	Rectified Linear Unit
FFT	Fast Fourier Transform
FDA	Food and Drug Administration

Chapter 1

Introduction

The first chapter intends to present the work performed, as well as the intrinsic motivation. Furthermore, the considered contributions are described. Lastly, the thesis structure is exposed.

1.1 Motivation

Actually, there are many factors that have repercussions on ageing. During life, social and physical environment, as well as behavioural attitudes have a major impact on human ageing process. Nowadays, as a result of medical improvements, people are living longer – it has been observed both an increase in the average life expectancy and a drop in the fertility rate. The World Health Organization (WHO) states that there are 125 million of people living with 80 years or older, expecting to increase up to 434 million until 2050 [1]. Considering this shift in the population distribution, many countries are facing new challenges in their societies – public health stands out as one of the most important. In this way, each country should organise the government budget in order to ensure a quality life for its citizens.

Over the years, it is predictable that each more ageing health problems arise. Thus, it is crucial to consider not only the physical deterioration (the visible one), but also the nervous system deterioration. Concerning elderly people, it is proved that they have a greater propensity to develop neurodegenerative diseases such as dementia. Neurodegenerative disease is a serious clinical condition since it truly affects the nervous system functioning and progresses irreversibly. Epidemiological data states that 1 in 3 seniors dies with dementia [2, 3].

Alzheimer Disease is known as the most common form of dementia. It is characterised by the deterioration of human cognitive functions, affecting behaviour, language, memory and reasoning, among others. In the disease advanced stages, patients become highly dependent and incapacitated. So far today, there is no cure for this disease [4].

It is important to note that the diagnosis in the early stages (asymptomatic periods) is extremely tough, as far as only a highly robust diagnostic method can postpone the damaging effects of this complex brain disease. Given the large number of people affected and the inherent severity, it is urgent to find a method capable of assisting in the AD early stages diagnosis. For this reason, the main goal of this dissertation is to create an intelligent system that can be useful for that purpose. In other words, a tool that could detect anomalies in the EEG signal of an Alzheimer carrier during the asymptomatic period so that the disease evolution can be delayed.

1.2 Work Contributions

The realization of this dissertation was achieved due to the MATLAB software. A nonlinear multiband analysis of the EEG signals was carried out. As a result of its behaviour, the extraction of this type of features characterises it well. Initially, the Wavelet Packet Transform tool was the one that enabled the signal processing. Subsequently, the data classification allowed to conclude about the accuracy that both Classic Machine Learning and Deep Learning present. The final results proved to be interesting, fitting the state of the art.

1.3 Thesis Structure

In terms of structure, five chapters were idealized. This first chapter presents the thesis work in terms of motivation and objectives. Thereafter, both chapter 2 and 3 cover detailed bibliographic reviews – Alzheimer Disease and Electroencephalography, respectively. Chapter 4 focus on the methodology and tools used and, hereinafter, chapter 5 spotlights the presentation and discussion of the obtained results. Finally, chapter 6 makes remarks about conclusions and expresses possible future investigations.

CHAPTER 2

Alzheimer Disease

The content of this chapter aims to provide complete information regarding AD. Initially, the focus is a clinical introduction about this neurodegenerative disease which severely affects cognitive functions. Subsequently, biological changes that occur in the ill brain are described, as well as symptoms and risk factors. In addition, a review about the evolutionary phases is crucial. Lastly, both complementary diagnostic exams and treatments that may slow down the symptoms are described.

2.1 Clinical Introduction

Due to medical advances, the average life expectancy is increasing more and more in several countries. Consequently, as the population gets older, health problems related to ageing are evident. Among those, dementia stands out being characteristic from a set of pathologies marked by a gradual decline in a person's overall functioning – loss of memory, damaged reasoning, changes in emotional reactions and loss of intellectual capacity [5].

Every 3 seconds, a new case of dementia appears. The WHO estimates that there are around 50 million people suffering from dementia worldwide. It is estimated that this value may reach about 82 million and 152 million in 2030 and 2050, respectively. In terms of geographic distribution, there is an expected higher incidence in low and middle income countries comparing with high income countries (figure 2.1). More specifically, it is verified that China, India, and their South Asian and Western Pacific neighbours are the countries where there is a fastest growth of the elderly population [6].

In particular, AD is the most common way of dementia, representing roughly 65% of all dementia cases. Epidemiological data states that it is the seventh main cause of death and there is still no cure. Alzheimer is a neurodegenerative disease in which the accumulation of certain toxic substances in the brain leads to the progressive death of neuronal cells. As neurodegenerative it means that there is a continuous destruction of the brain cells, thus affecting Nervous System's connections and functions.

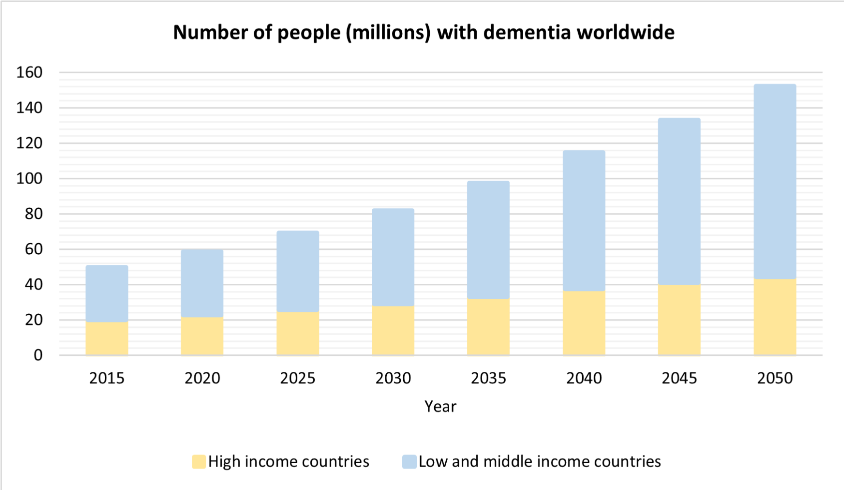

Figure 2.1: Dementia worldwide [6].

Marked by an insidious installation, AD is a neurological disorder that causes progressive memory loss and cognitive decline [4, 7–9].

2.1.1 Pathology

The major brain changes that AD causes are the accumulation of an atypical form of the tau protein within neurons, as well as the accumulation of the β-amyloid fragments outside neurons. Several studies indicate that tau and β-amyloid are interconnected since β-amyloid increases tau phosphorylation [10, 11].

Neurons have internal structures called microtubules which provides the passage of crucial molecules and nutrients along the neuron. The main function of tau protein is to attach to those microtubules in order to stabilise them [12]. This protein suffers phosphorylation and dephosphorylation processes and in its phosphorylated form it is not able to stabilise microtubules. In this way, tau protein disconnects from microtubules and polymerizes with other tau proteins. Therefore, there is an abnormal accumulation of several molecules of this protein, forming straight filaments and, consequently, neurofibrillary tangles. These neurofibrillary tangles are typically present in specific parts of the brain – entorhinal cortex, hippocampus, amygdala, association cortices of the frontal, temporal and parietal lobes and certain subcortical nuclei that project into these regions. This results in an obstruction in the neurons transport system, compromising the synapse process and, eventually, leading to cell death [11].

In biochemical terms, the sequential cleavage of Amyloid Precursor Protein (APP) results in the formation of a peptide named β-amyloid. Present in the brain, this peptide can be found in one of three isoforms – APP695 (shortest isoform), APP751 and APP770 (longest isoform) [10]. In a person with AD, compact deposits of β-amyloid are formed, notably of β-amyloid 42 (particularly toxic). Thus, outside the neuron, extracellular plaques of this peptide are formed. This accumulation disturbs the communication between neurons and compromising their cellular functions. This phenomenon is clearly associated with cognitive decline, as well as neuronal losses in the entorhinal cortex, hippocampus and association cortices [12].

As exemplified below in figure 2.2, the diseased neuron reveals the microtubules disintegration and a disorganised structure of tau proteins.

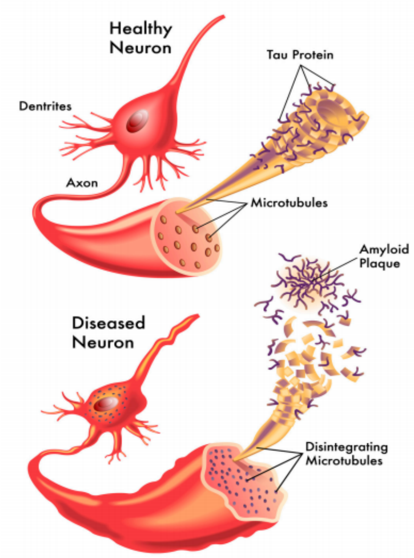

Figure 2.2: Accumulation of tau protein and β-amyloid fragments [13].

2.1.2 Symptoms

Dementia symptoms are usually confused with aging signs. One of the main symptoms, memory loss, refers specifically in forgetting the most recent information. Patients often needs memory aids to remember important events. Regarding the loss of spatial and temporal perceptions, these signs frequently occur when patients do not recognize the physical space concerned, as well as when having difficulty in understanding the passage of time [10].

Alzheimer patients tend to leave objects in inappropriate places, being unable to reconstruct information, mainly visual one, which would lead them to know what they did in the previous moment. Over time, this symptom occurs increasingly, combined with language and expression problems [10].

As the disease progresses, it is frequent to notice changes in the personality and behaviour, as the temperament becomes different when patients feel out of their comfort zone. People with AD cease to have an active social life and they also reveal difficulties in speaking and writing [10].

2.2 Risk Factors

There are several factors that increase the risk of developing AD. Genetics, age, sex, lifestyle and other diseases are considered the main ones.

Concerning family history, having one and two first-degree relatives with AD increases the risk of developing the disease by four and eight times, respectively. In genetic terms, the Apolipoprotein E (APOE) role is to transport cholesterol through bloodstream. This gene has three alleles – e2, e3 and e4. Everyone inherits one allele from each parent. Research have shown that having the e4 form of APOE (APOE4) increases the risk of developing Alzheimer. More specifically, e4 homozygous and e4 heterozygous individuals have a risk of developing AD that is 50% and 25% higher than those who do not have any e4 in their genetic code, respectively. In contrast, having e2 form reduce the risk of developing AD, contrasting with e3 [10, 14].

As it can be expected, age is a preponderant factor when it comes to the risk of developing this disease. Although it can happen to anyone at any age, dementia is scarce before the age of 65. The prevalence of this disease tends to increase with ageing, reaching around 32% of the population with 85 years or older. It is important to enhance that ageing does not necessarily lead to any type of dementia [5, 10].

Concerning the gender, statistical data shows that AD is more common in women [15].

As far as lifestyle habits are concerned, smokers and inactive people develop a propensity for this disease. The level of education also has impact since more years of education means a constant brain stimulation. There is also a strong evidence that having a healthy diet may reduce the probability of developing dementia. These issues are modifiable risk factors thus each person can take behaviours to be less likely to face this disease [10, 15].

People diagnosed with depression are considered equally susceptible. Besides that, cardiovascular diseases are likewise closely linked to dementia. Having a healthy heart and blood vessels will allow the brain to be pumped with blood with sufficient amounts of oxygen and nutrients (a well-functioning brain depends on a well-functioning cardiovascular system). Problems such as hypertension, diabetes, obesity and high cholesterol increase the risk to develop Alzheimer or other kind of dementias [10, 15, 16].

2.3 Evolutionary Stages

The AD progress can be described over four development stages: Pre-clinical, Mild Cognitive Impairment (MCI), Moderate AD and Severe AD. Generally, patients who survive until the most advanced stage die because of disease's natural effects, increased vulnerability to falls, pressure wounds and infections. It is estimated an average time of 8-10 years from diagnosis until death [15].

In the first stage, there are no associated symptoms (asymptomatic period). Pre-clinical phase is simply characterised by imperceptible lesions in the blood, brain and cerebrospinal fluid. Investigators believe that AD arises at least 20 years before symptoms appear hence the patient could stay in this initial phase for a long period of time. If AD could be detected at this early stage, therapeutic interventions would be tested in an effort to combat the disease progression [10, 17].

During the second phase, MCI, symptoms like memory lapses begins. This is the moment when friends and family notice some differences in the person concerned, although the patient still lives independently and with normal functional capabilities. In the research field, the concept of MCI was created with the aim of including individuals with mild symptoms who might eventually progress to AD. It is stated that 6% to 25% of MCI patients later develop AD [17, 18].

In both Moderate and Severe stages the patient is officially AD diagnosed. The person's behaviour and autonomy has been affected and the common symptoms are clear. The Moderate stage is typically the longest one. At this point, the person themselves start to notice the signs, becoming frustrated and upset. Frequent memory lapses and difficulty in performing daily routine tasks are faced, as well as other psychological symptoms [17].

Lastly, the Severe stage is the moment that patient loses the ability to interact with those who are around, having cognitive and motor functions extremely altered. There is requirement of full-time care since the person is no longer independent [17]. Figure 2.3 illustrates the AD evolution until this stage.

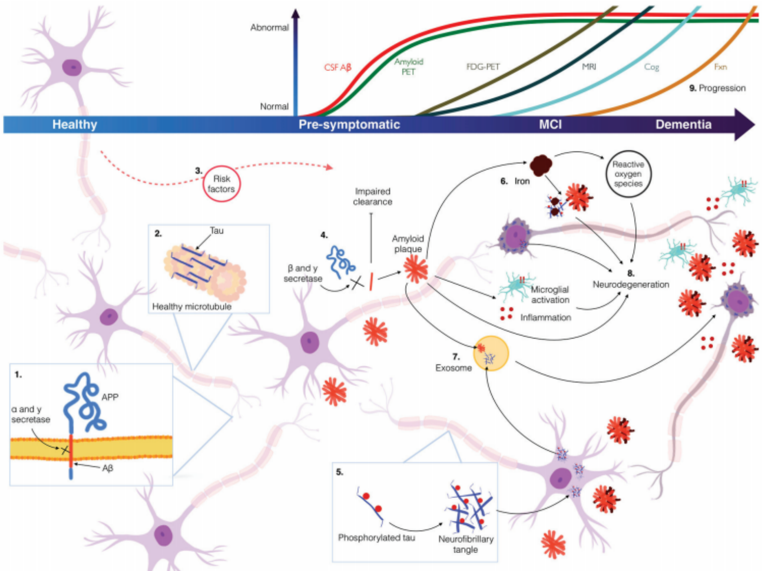

Figure 2.3: Disease progression [15].

2.4 Medical Diagnosis

In accordance with Pietrzak *et al.* [13], the true diagnosis of Alzheimer is only achieved after patient's death [19]. However, there are several tools that support the diagnostic when someone reveals signs of dementia during life. Neurologists consider the medical history and perform several neurological and cognitive tests. Besides that, complementary diagnostic exams like Imaging Techniques (IT), Electroencephalogram (EEG) and Cerebrospinal Fluid (CFS) Analysis are helpful.

In the field of IT, Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) play a crucial role in the study of this disease.

Regarding MRI, it is a non-invasive and common technique and its principle of operation is based on nuclear magnetic resonance of the atomic nuclei [20]. There are two specific techniques – structural MRI (sMRI) and functional MRI (fMRI).

One the one hand, sMRI is used to examine brain anatomy, more specifically to visualize neurodegeneration in the form of brain atrophy – a macroscopic manifestation which corresponds to neurons loss. Studies reveal that brain volume evaluated in sMRI is correlated with neuronal counts after death – reduced brain volume means reduced number of active neurons [20].

On the other hand, fMRI is useful for studying brain activity, in particular to quantify neuronal activity through changes in Blood Oxygen Level-Dependent (BOLD) MRI signal. When changes are detected in the blood flow, volume and oxygenation, modifications in the BOLD MRI signal occur [21].

PET is a modern technique that takes benefit from a biologically active molecule combined with a radioactive tracer. After the injection into the human body, the gamma rays produced by the biochemical reactions of the tracer are detected. Particularly, Fluorodeoxyglucose (FDG) PET is a proper indicator for evaluating brain metabolism since glucose is the most relevant energy source [20, 21].

EEG is a painless medical exam that aims to record and evaluate brain electrical activity. By placing electrodes on the scalp, EEG enables to identify potential abnormalities in brain wave patterns. Those abnormalities may be linked with brain illnesses like dementia [22]. This topic will be more deeply discussed in the following chapter.

Concerning CFS Analysis, it is still considered to be a quite invasive exam as it may cause headaches for the patient. Since CFS is a liquid in constant direct contact with the brain's extracellular space, it is possible to infer about biochemical modifications. Currently, CFS have been increasingly used on diagnostic biomarkers for AD – phospho-tau protein, total tau protein, β-amyloid 42 and total β-amyloid and are the most frequent [19, 23].

2.5 Treatment

Up till now, it has not been possible to find a cure for AD. Although the progression of this disease is irreversible, there are methods that try to slow it down.

2.5.1 Pharmacological Treatment

It is proven that cognitive functions can be improved by administering cholinesterase inhibitors or N-Methyl-D-Aspartic (NMDA) acid receptor antagonists [14].

While cholinesterase inhibitors aim to increase the quantity of neurotransmitters in the brain, memantine is used to prevent neurons damage by blocking certain receptors. During the disease milder phase, cholinesterase inhibitors show positive impacts on the stabilization of cognitive performance during the first year of medication. Over treatment, cholinesterase inhibitors combined with memantine may have certain benefits in both moderate and severe phases [10, 14]. It is important to note that the effectiveness of pharmacological treatments depends on the patient clinical situation.

The most common drugs prescribed for dementia treatment are donepezil, rivastigmine, galantamine and memantine. Donepezil, rivastigmine and galantamine are examples of cholinesterase inhibitors and memantine is an example of NMDA acid receptor antagonist [14]. The proper dose and posology vary between patients according to the stage of their disease.

2.5.2 Non-pharmacological Treatment

In contrast, there are other interventions focused on providing a better quality of life and improving patient's behaviour and independence in order to constantly stimulate the brain. This kind of treatment do not use drugs and is less costly [24].

Over the years, specialists have been proving that certain foods and supplements have a positive impact on dementia since they slow down the cognitive deficit. Thus, it is important that patients take care with their diet, not only to maintain a healthy lifestyle, but also to try to combat the disease damaging effects. Fish and Ginkgo Biloba are recommended since they help in brain stimulation. Walnuts stands out for being rich in vitamins, proteins and omega-3 fatty acids, having benefits on brain activity. Besides that, there are also foods that are considered preventive – apples, coffee, chocolate and cinnamon are examples [24].

Reality Orientation Therapy deal with lack of orientations concerning place and time. In the disease early stages, it is crucial to stimulate cognition, especially when it comes to the first memory failures (usually of short-term). Interventions such as having a notebook with relevant information and a calendar help the patient to remember important knowledge and events, feeling less lost and confused [24].

Reminiscence Therapy focus on remember previously acquired knowledge. In other words, the main purpose is to promote certain stimuli to recall certain memories. During a conversation, the strategy may include the use of tangible aids such as meaningful objects (for examples old pictures and home stuff). It is proved that this revitalizing therapy has positive impacts on the patient's mood and social interactions [24].

Snoezelen Therapy aims to provide a good environment for patients, making them more peaceful and less aggressive. This therapy takes place in comfortable spaces with specific characteristics of light, sound and smell, combined with taste and touch activities – external senses stimulation. By feeling in a pleasant space, patient tends to create good memories [24].

Validation Therapy preserves respect and acceptance of the other's opinion. This moral support therapy aims to deal with and end up with conflicts in order to stimulate good moods and to accept emotions. Consequently, patient feels self-confidence and willingness to communicate with the caregiver, passing on good energy [24].

Afterward, tables 2.1 and 2.2 summarise the state of the art concerning studies about AD diagnosis.

Table 2.1: State of the art.

Study	Accuracy	Methodology
[25]	C vs AD - 95% C vs MCI - 77% MCI vs AD - 83% C vs MCI vs AD - 90%	This paper aims to classify different stages of AD by taking advantage of Artificial Neural Networks (ANN). Firstly, several features were extracted from each conventional band (δ, θ, α, β and γ) of those electroencephalographic EEG) signals. After selecting the best features combination, the classification tool ANN with LCV and with 10FCV allowed to conclude about the accuracy of the several binary comparisons.
[26]	C vs AD - 94%	In order to improve the AD diagnosis, researchers extracted both spectral and wavelet features. More specifically, through Power Spectral Density (PSD) it was feasible to calculate the Relative Power (RP) value for each signal sub band. Through wavelet decomposition, it was calculated the mean and variance of each coefficient. Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) are used to classify healthy subjects against diseased ones. The best result was achieved with spectral features combined with SVM.
[27]	C vs AD - 83% C vs MCI - 92% MCI vs AD - 79%	The authors applied both a Fourier and Wavelet analysis in order to extract as many features as possible in a time-frequency context. The classifier Decision Tree (DT) was used to predict which study group the signal belongs to, based on the extracted features. It was concluded that Wavelet Transform shows the best results.
[28]	All vs All - 95.55% C vs MCI - 98.06% C vs ADM - 95.99% MCI vs ADM-ADA - 93.85%	Rodrigues *et al.* applied a novel tool named Lacsogram for signal processing with the purpose of supporting the AD diagnosis in the evolutionary stages. The methodology of this work is distinguished by a cepstral and a lacstral analysis prior to feature extraction. Surrogate Decision Tree (SDT), SVM and ANN were the used classifiers, having ANN the best performance for the classification.
[29]	C vs AD - 76.33%	This paper presents a study concerning an automatic method in which the purpose is to detect the presence of AD. Two-dimensional Convolutional Neural Networks (CNN) are used to extract several features (one feature value for each data set). Subsequently, SVM classifies the EEG images in the wave bands δ, θ, α, and β. It was performed 3 classifications and the average value successfully concluded the study.
[30]	C vs AD - 99.2% C vs MCI - 78.5% MCI vs AD - 91.3%	The study carried out by Thapa *et al.* aimed to understand the impact of combining neurophysiological scores such as Mini Mental State Examination (MMSE) with MRI features for the classification of AD. Measurements of right and left hippocampus volumes were extracted from brain images. Using the SVM classifier, it was concluded that the best accuracy values are found in the combination of MRI features along with MMSE.

Table 2.2: State of the art (follow-up).

Study	Accuracy	Methodology
[31]	C vs All - 78.43% AD vs All - 76.47%	This study intended to detect MCI and early AD patients, thus discriminating the different stages. It was extracted spectral and nonlinear features from the signals. Fast Correlation-based Filter (FCBF) was useful to select the best features. Multi-Layer Perceptron Artificial Neural Network (MLP) reveal the best classification.
[32]	C vs MCI vs AD - 96.5%	The authors developed a novel technique based on EEG. In one approach, EEG signals are decomposed into the conventional bands using the Discrete Wavelet Transform (DWT) and the PSD of each band is obtained applying Burg's method. In another approach, coherence values are extracted. After feature extraction, the classifier Bagged Trees is trained in order to elaborate an AD three-way classification.
[33]	C vs AD - 91.2% C vs MCI - 78.9%	Liu *et al.*, through the acquiring of FDG-PET brain images, suggest an improved method based on the combination of CNN with Recurrent Neural Networks (RNN) with the intention of classify AD. On the one hand, the 2D CNN is applied to extract features from the slices. On the other hand, the Gated Recurrent Unit (GRU) of the RNN is useful to incorporate the inter-slice features for image classification.
[34]	AD vs MCI - 90.3%	The researchers developed a methodology using Multiple Signal Classification and Empirical Wavelet Transform (MUSIC-EWT) and extracting Box Dimension (BD), Higuchi Fractal Dimension (HFD), Katz Fractal Dimension (KFD) and Hurst Exponent (HE). The most discriminative features were trained by Enhanced Probabilistic Neural Network (EPNN) algorithm to discriminate EEG.
[35]	C vs AD1 vs AD2 - 97.64%	With the goal of detect AD in an automatic framework, this paper presents Hjorth parameters as a good option to characterize the signals. The performance before and after the addition of these parameters with the other common features was evaluated. DWT decomposition technique and KNN classifier demonstrate the best results for this three-class classification.
[36]	C vs AD - 99.2% C vs cMCI - 87.7% C vs sMCI - 76.4% AD vs cMCI - 75.8% AD vs sMCI - 86.3% cMCI vs sMCI - 75.1%	According to Basaia *et al.*, it is possible to support the AD diagnose taking benefit from a single cross-sectional brain structural MRI scan (3D T1-weighted images). The developed powerful deep-learning algorithm does not use feature extraction techniques. This work was performed using two datasets, both providing high results from CNN tests.

AD1:Mild AD; AD2:Moderate AD; cMCI:Converters MCI; sMCI:Stable MCI.

Chapter 3

Electroencephalography

The aim of this chapter is to deepen topics concerning Electroencephalography (EGG). Firstly, the EEG concept is described since this medical exam provides the recording of brain's electrical activity. Subsequently, it is carried out an important review about signal recording system (more specifically the role of electrodes) and conventional brain waves. Finally, signal artefacts are explored, as well as the effects that AD might have on EEG signal.

3.1 Human Brain

Brain, an essential organ of the human body, is composed by millions of neurons. These nerve cells communicate with each other through nervous impulses and are responsible for the entire human body functioning. Neurons are formed by three important regions – cell body, axon and dendrites [37, 38].

The action potential is the event that causes the transmission of information between neurons. At the membrane level, when a modification of electrical charge occurs, a nervous impulse is produced – it is generated a current flow along the axon. At the terminal zone, neurons release neurotransmitters into the synaptic cleft, also known as chemical messengers. These molecules are transmitted from one neuron to another, carrying out important information [38].

Anatomically, the brain is divided into two hemispheres – right and left. In turn, each hemisphere is divided into lobes – frontal, occipital, parietal and temporal. Each lobe is responsible for specific functions. Frontal lobe, located behind the forehead, is the largest one. Organizing, problem solving, emotions, motor functions are examples of what frontal lobe controls. Regarding occipital lobe, it covers visual information and it is situated at the lower back. Parietal lobe, located behind the frontal lobe, recognizes sensorial stimuli such the pain feeling and is responsible for recognition and orientation. Concerning temporal lobe, it is situated on both two sides of the brain – behind and under frontal and parietal lobe, respectively. It involves memory and both speech and sound processing. Indeed, and as stated in Siuly et al. review [38], each brain region is responsible for specific functions, meaning that each lobe is concerned with processing different types of electrical activities [37, 38].

3.2 Electroencephalographic Signal

A signal corresponds to the potential difference measured between two points. More specifically, a biomedical signal corresponds to the signal produced through a biological system. The occurrence of a certain event has impact on signal modification. In this way, it is essential to extract information about the system state in order to extract conclusions [39].

By extracting characteristics about brain waves, it is possible to study brain electrical activity. An electroencephalogram is considered a non-invasive and painless exam. EEG signal is acquired by placing electrodes on the scalp and its frequency and amplitude values are contained in the range of 1 to 100 Hz and 10 to 100 μV, respectively. Researchers conclude that tiny electrical changes in EEG signals allows a diagnosis to be reached [37, 40].

3.2.1 Electrodes

The international 10-20 electrode system was created in order to establish standards. This system considers 21 electrodes – 19 electrodes and 2 reference ones (placed on the ears) as shown in figure 3.1. The goal is to measure the distance between nasion (corresponds to the point between the forehead and the nose, at eyes level) and inion (corresponds to the bony prominence located on the midline at the back of the head). Electrodes are placed at 10 or 20% of the distance from these skull reference points [38, 41].

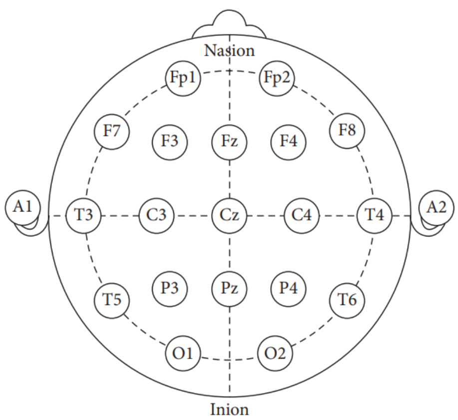

Figure 3.1: Placement of EEG electrodes [41].

Electrodes are identified using letters and numbers. On the one hand, the nomenclature used to distinguish the region of the scalp are the capital letters C (central), F (frontal), O (occipital), P (parietal) and T (temporal). On the other hand, the hemisphere location is distinguished through odd and even numbers while the electrode is in the left or right hemisphere, respectively. In the specific case of being placed in the midline (between hemispheres), the letter z is assigned [38].

There are different types of electrodes when it comes to the material they are made of. The most commonly used are Ag-AgCl electrodes – composed by silver and with a silver chloride coating. These electrodes show good mechanical and electrical properties, as well as a good accuracy in recognising minimal changes in electrical potential. It is necessary to use a conductive gel, in order to decrease the impedance on the electrode-scalp surface [42].

Apart from electrodes, the amplifiers, the Analogue-Digital (A/D) converter and the recording device play an important role in the EEG signal measurement system. Briefly, the purpose of each electrode is to capture the electrical charges which reflect the brain activity at a given point. Subsequently, all these potentials are amplified to an appropriate voltage range and converted to the digital format. The recording device registers and stores the collected information. At the end of the exam, the medical specialist examines visible differences in patient's brain waves, comparing with a healthy subject [38, 42].

3.2.2 Brain Waves

In the clinical branch, it is possible to evaluate brain abnormalities in patient's brain activity by measuring the signal frequency. In this way, It is considered five conventional frequency bands – delta, theta, alpha, beta and gamma (from the lowest to the highest frequency). It is important to note that there is no standard regarding the frequency range of each brain wave. In the literature, it is possible to find insignificant differences in these same values, which are not substantial. Figure 3.2 shows the behaviour of each rhythm, exhibiting electrical voltage oscillation [38, 42].

Delta waves (δ) are the slowest and highest amplitude waves, belonging to the frequency range of 0.5 to 4 Hz [38]. These waves are related to events such as deep sleep and wakefulness. Usually, artefacts caused by the activity of the mandible and neck muscles are found. Since they are near the surface, they end up producing large signals, due to the lack of attenuation [43].

Theta waves (θ) have amplitude greater than 20 μV and are contained within the frequency range of 4 to 8 Hz. The presence of these waves is associated with moments of emotional stress, deep meditation and creative inspiration. Theta waves play an important role in childhood. In this way, when is detected a higher occurrence in a waking adult, neurological pathologies are suspected[38, 43].

Alpha waves (α) belong to the frequency and amplitude range of 8 to 13 Hz and 30 to 50 μV, respectively. When a person is with the eyes closed or in a relaxed state, these waves appear mostly in the occipital lobe. Alpha wave activity is reduced when the subject is submitted in intense mental activity. This brain rhythm is the most prominent [38, 43].

Beta waves (β) refer to rhythms in the frequency range of 13 to 30 Hz, having amplitude values lesser than 30 μV. This wave type is commonly observed in the brain of a healthy adult and it is associated with active brain activity, active thinking and problem solving. Beta waves are usually present in the central and frontal regions. Panic moments are characterized by high-level beta waves [38, 43].

Gamma waves (γ) are the fastest and lowest amplitude waves, lying above the frequency of 30 Hz (normally the upper limit is 100 Hz). The gamma band has been recognized as a great indicator for Event-Related Synchronization (ERS). The activity of these waves are of rare occurrence – a brain disease can be confirmed by an abnormal presence of this rhythm [38, 43].

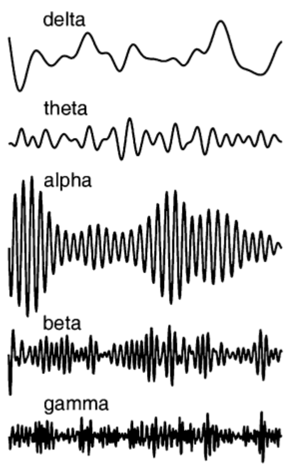

Figure 3.2: Brain waves [42].

3.2.3 Artefacts

An artefact (noise) is defined as an electrical potential which is not originated from patient's brain, therefore considered undesirable for analysis. Compared to the EEG signal, noise is a signal with a different morphology and a higher amplitude. Fortunately, it can be removed through signal processing techniques allowing the medical specialists to perform an accurate EEG interpretation. Artefacts may be of physiological or non-physiological type since they come from the patient or not, respectively [40, 42, 44].

Concerning physiological artefacts, it means the existence of biological signals related to another type of human activity than EEG. Events such as breathing, heartbeat (electrocardiography), perspiration, speech, eye movements and muscle activities (electromyography) may consist physiological artefacts. These noises may be difficult to avoid during the exam since they are often involuntary [42, 44].

In contrast, non-physiological artefacts are easier to control, as they concern the electrodes and recording device operation. When electrodes are misplaced or placed in an incorrect position, the EEG signal recording is compromised. It is important to note that gel must not be applied in excessive quantities. Additionally, problems such as impedance fluctuation, broken electrical wires and cable movements are also artefacts to be avoided [42, 44].

3.3 Alzheimer Disease Effects on Electroencephalographic Signal

An AD patient reveals changes in the EEG signal compared to the normal patterns observed in a healthy individual. Firstly, in the disease early stages, there are an increase in theta band activity and a decrease in alpha band activity. Later on, in more advanced stages, there are both an increase in delta and theta bands activity and both a decrease in beta and alpha bands activity [18]. According to Sanei and Chambers [43], in a serious clinical condition, triphasic waves and epileptiform discharges may appear.

The main phenomenon is named slowing effect. As the disease progresses, there is an increase in power in the lower frequency bands (δ and θ), as well as a decrease in power in the higher frequency bands (α, β and γ). Additionally, shift-to-the-left phenomenon is also observed. A shift of the power spectrum peak at the alpha band to the lower power bands occurs. The distribution of these rhythms becomes emphatic in anterior regions, instead of in the occipital region (healthy subject) [18, 25, 28].

Briefly, these signal power modifications are essentially due to the lack of acetylcholine in the AD brain. Insufficient amounts of acetylcholine result in failures in the synchronisation of synaptic potentials [28].

CHAPTER 4

Methodology

The methodology reflects the developed work regarding the creating of a novel system that detects abnormalities in EEG signals, being capable to discriminate each disease stage. Firstly, the database used is described. Then, the Wavelet Packet tool is characterised and it was performed a detailed review about the signals nonlinear analysis. Before the exposure of the developed method, the feature selection technique and the classification algorithms are explained.

4.1 Database

The EEG signals have been collected at Hospital de São João in Porto, Portugal. With the approval of the local ethics committee and the hospital's administration board within the project CES198-14. Following the international 10-20 electrode system, the recording was achieved by placing 19 electrodes on the scalp of the subjects. During the acquisition, it was ensured that all the study subjects were relaxed and with the eyes closed.

All signals were acquired at a sampling frequency of 256 Hz and the DC component was removed. It is important to note that only signals without any kind of artefacts are selected. This database includes 37 subjects split into 4 distinct groups – 11 healthy subjects called controls (C), 8 MCI, 10 AD Mild and Moderate (ADM) patients and 8 AD Advanced (ADA) patients. The MMSE average and age average of each group are presented in Table 4.1.

Table 4.1: EEG Database.

Subjects	**C**	**MCI**	**ADM**	**ADA**
Number	11	8	10	8
Age Average	74	80	79	79
MMSE Average	28.68	26.29	18.89	11.50

4.2 Wavelet Packet Decomposition

The EEG signals properties change over time, thus they are classified as non-stationary signals. A time-frequency analysis is very helpful in order to locate events of interest. Thereby, Wavelet analysis emerges as an alternative to Fourier analysis. Wavelet analysis is a method that benefits from the power of a variable-sized regions window, being able to analyse a specified area of a large signal. On the one hand, the use of a large window size (long time intervals) provides the capture of low frequency information. On the other hand, the capture of high frequency information is achieved by applying a reduced size window (short time intervals) [45, 46].

This technique aims to decompose the original signal (mother wavelet) into several versions of it (single wavelet). The signal is decomposed into low (approximation coefficients) and high frequencies (detail coefficients). Generally, the low frequency components are the most important since they characterize the signal identity. In turn, the high frequency components analysis reflects the more specific details of a signal [46].

Based on Wavelet analysis principles and depending on whether the signal is continuous or discrete, it is distinguished the DWT from the Continuous Wavelet Transform (CWT), respectively [45]. Besides that, Discrete Wavelet Packet Transform (DWPT) emerges as an extension of DWT. DWPT offers a good temporal resolution and a powerful evaluation of EEG signal.

More specifically, DWPT provides a simultaneous decomposition of low and high frequencies. Unlike the DWT, it also considers high frequencies after the 1st level of decomposition. This procedure is iterative, as the signal is successively filtered both by a low-pass and a high-pass filter (obtaining approximation and detail coefficients, respectively) [47, 48]. The following figure (figure 4.1) represents the decomposition of the signal S, through low-pass filters A and high-pass filters D.

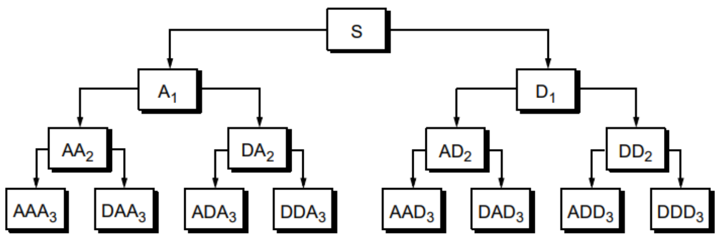

Figure 4.1: Wavelet Packet Tree [46].

The DWPT 1-D coefficients equations are present below (formula 4.1) [47].

$$\begin{cases} d_i^m(l) = \sum_m h_0(m-2l) d_{i-1}^{\frac{m}{2}}(m), & if\ rem(m,2) = 0 \\ \\ d_i^m(l) = \sum_m h_1(m-2l) d_{i-1}^{\frac{j-1}{2}}(m), & if\ rem(m,2) \neq 0 \end{cases} \tag{4.1}$$

where m is the frequency factor meaning the position in the Wavelet Packet Tree ($m = 0, 1, 2, \ldots, 2^m - 1$), i represents the decomposition level $i \in \mathbb{Z}$, l is the sample number, h_0 and h_1 refer to the high-pass and low-pass filters, respectively. The filters applied vary according to the wavelet family and subfamily.

4.2.1 Wavelet Families

For the analysis of discontinuous signals, eight wavelet families stand out – Haar (Haar), Daubechies (Db), Symlets (Sym), Coiflets (Coif), Biorthogonal (Bior), Reverse Biorthogonal (Rbio), Discrete Approximation of Meyer (Dmey) and Fejer-Korovkin (Fk). With the exception of Haar, each wavelet family has, in turn, several subfamilies. The order N indicate which subfamily it refers to. The number of subfamilies varies among them – for instance, Daubechies have 45 subfamilies (N = 1, 2, 3, ..., 45) and Fejer-Korovkin wavelets only have 6 s (N = 4, 6, 8, 14, 18, 22) [47].

There is no best wavelet, in the sense that the selection varies depending on the analysis purpose. Each wavelet has distinct characteristics, so the ideal is to focus on the properties of each one. Thus, the chosen wavelet can be optimized. Table 4.2 summarises the specifications of each wavelet family.

Table 4.2: Wavelet Families [49–51].

Properties	**Wavelet**							
	Haar	Db	Sym	Coif	Bior	Rbio	Dmey	Fk
Symmetry	X				X	X	X	
Asymmetry		X						X
Near symmetry			X	X				
Infinitely Regularity								
Arbitrary Regularity		X	X	X	X	X		X
Orthogonal Analysis	X	X	X	X	X			X
Biorthogonal Analysis	X	X	X	X	X	X		
FIR Filters	X	X	X	X	X	X	X	X
Continuous Transform	X	X	X	X	X	X		X
Discrete Transform	X	X	X	X	X	X	X	X
Exact Reconstruction	X	X	X	X	X	X		X
Fast Algorithm		X	X	X	X	X	X	X
Arbitrary Number of Vanishing Moments		X	X	X	X	X		
Existence of φ	X	X	X	X	X	X		X

4.3 Nonlinear Analysis

According to the nonlinear dynamic theory, a complex system (such as the human brain) is characterised by nonlinear dynamic properties. Due to physiological events, the brain environment is constantly changing. Consequently, brain waves exhibit a nonlinear and chaotic behaviour. Additionally, the degree of complexity of the brain represents the time series randomness. Depending on the intensity of the brain activity, the EEG signal can be more or less complex, containing more or less information per signal fragment. Briefly, many methods of extracting nonlinear features have been increasingly explored, making them a powerful approach for EEG signal characterization. Indeed, through the detection of patterns in the time series, it is possible to infer about EEG signal behaviour and predict the same kind of occurrences in the future [42].

4.3.1 Entropy

Entropy concept arises with the intention of describing the molecules distribution in a given system. This thermodynamic definition explains how molecules are organised, considering the size and atomic configuration of each one. Thus, the assessment of entropy allows the quantitative evaluation of the degree of randomness and uncertainty of a given sequence of data. Entropy is a measure that considers the amount of energy present in the complex system. There are several types of entropies which are properly used to analyse the EEG signal behaviour [42, 43, 48].

4.3.1.1 Shannon Entropy

Shannon Entropy offers the capacity to obtain relevant information resulting from unusual events. In specific time intervals, this measure captures information of rare character to be analysed. Based on the probability distribution of the signal's amplitude values, this metric calculates that probability density. Shannon Entropy is defined through the following equation (formula 4.2) [42].

$$ShEn = \sum_{i=1}^{n} p(si) \log_a \frac{1}{p(si)} \tag{4.2}$$

where p(s_i) is the probability of acceptance by the variable S.

4.3.1.2 Approximate Entropy

Approximate Entropy aims to quantify the predictability of a certain occurrence in order to foresee the reappearance of certain amplitude values based on previous values. Thus, this parameter evaluates the regularity of a time series data. Approximate Entropy is calculated through the following equation (formula 4.3) [42, 52].

$$ApEn(m,r,N) = H^m(r) - H^{m+1}(r) \tag{4.3}$$

where N is the data length (suggested to be 1000 of the signal standard deviation), r is the similar tolerance (between 0.1 and 0.25) and m represents the embedding dimension (between 2 and 3). H is the Heaviside function that results from intermediate calculations.

4.3.1.3 Sample Entropy

Sample Entropy emerged based on the Approximate Entropy in order to create a modified version of it. This measure, which evaluates the signals complexity, removes the bias caused by self-matching to enhance the performance. Note that Sample Entropy is a parameter independent of the signal length. This parameter is defined through the following equation (formula 4.4) [42, 47].

$$SampEn(m,r,N) \quad = \quad \ln \frac{A^m(r)}{B^m(r)} \tag{4.4}$$

as the factors N, r, m are the same used in Approximate Entropy, the recommended values are similar.

4.3.1.4 Permutation Entropy

Permutation Entropy, a parameter of low computational complexity, aims to convert a temporal data series into an ordered series. Thus, the neighbouring points of a given series are compared, in order to explore the order relationship established between them. The following equation (formula 4.5) enables the calculation of Permutation Entropy [42, 47].

$$PeEn(m) \quad = \quad -\sum_{j=1}^{m!} p_j \ln p_j \tag{4.5}$$

where m is the embedding dimension and p_j represents the probability of the j^{th} permutation occurring.

4.3.2 Chaos Theory

Chaos theory is an approach closely related to dynamic systems. A dynamic system does not share the properties of a system in equilibrium, wherefore certain unpredictable disturbances may influence its behaviour. In this way, those perturbations cause the system transition from one state to another. The concept of phase space represents the set of all possible states through which a dynamic system can pass over time. There are three main exponents which provide a comprehensive framework of chaos [43, 53].

4.3.2.1 Hurst Exponent

Hurst Exponent is a parameter that is acquired from the signal spectrum (spectral analysis). The calculation of this exponent provides the measurement of important properties such as self-similarity and correlation of EEG signals. This measure evaluates the time series smoothness. Hurst Exponent is defined through the following equation (formula 4.6) – it is assumed that the brain activity follows the fractional Brownian motion random field model [42].

$$HE = \frac{\beta - 1}{2} \tag{4.6}$$

where β derives from power-law.

4.3.2.2 Kolmogorov Exponent

The Kolmogorov Exponent is a parameter that evaluates the chaos level of a system. The phase space is clearly associated to this metric, meaning that each state signifies a single point. The measurement of Kolmogorov Exponent is achieved by dividing the phase space into multidimensional hypercubes – every parameter of the system concerns an axis of the multidimensional space. This measure is calculated through the following equation (formula 4.7) [43].

$$KE = \lim_{N \to \infty} \frac{1}{NT} \sum_{n=0}^{N-1} (K_{n+1} - K_n) \tag{4.7}$$

where NT represents the trajectories and $K_{n+1} - K_n$ is the information needed to predict which hypercube is in a given trajectory.

4.3.2.3 Lyapunov Exponent

The Lyapunov Exponent provides the ability to distinguish a system with chaotic, unstable and unpredictable behaviour (LE $>$ 0), from one with ordered, stable and predictable properties (LE $<$ 0). Thus, Lyapunov Exponent is a metric that calculates the average growth rate of infinitesimally small errors – perturbations which occur along the time series. The following equation (formula 4.8) enables the calculation of this exponent in the initial point x_0 [43].

$$LE(x_0) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \ln |f'(x_k - 1)| \tag{4.8}$$

where f' is the derivate of the iterator function f.

4.3.3 Fractal Dimension

Fractal structures present self-similarity properties that enable detailing both irregular processes and structures. Regarding brain activity study, fractal geometry has been showing to be a useful approach to identify neurological pathologies. In fact, monitoring the self-similarity of the brain rhythms allows the characterization of the patient clinical condition. Fractal features extraction such as the Higuchi and Katz fractal dimensions are considered to be helpful in EEG signal classification [53].

4.3.3.1 Higuchi Exponent

The Higuchi Exponent is a measure commonly used in biological signal investigations in order to quantify its fractal dimension, being considered efficient and precise . This method consists in an iterative process that aims to reconstruct the given temporal sequence, creating a new one. Higuchi Exponent is derived from the following equation (formula 4.9) [54, 55].

$$L(k)\alpha k^{-D} \tag{4.9}$$

where k indicates the time interval, $L(k)$ is the length of the curve in the k time interval and D is the Higuchi Exponent.

4.3.3.2 Katz Exponent

The Katz Exponent is a fractal dimension easy to calculate since it is directly obtained from the time series data. The following equation (formula 4.10) enables the calculation of this exponent [54, 55].

$$KE = \frac{\log N}{\log N + \log \frac{d}{L}} \tag{4.10}$$

where L is the total length of the waveform, N represents the signal length and d means the maximum value of distance between the initial point to the others points in space.

4.4 Feature Selection Technique

Regarding feature selection techniques, f-score stands out since it constitutes a simple approach and provides good accuracy values in the classification process. Through statistic components, it evaluates the features individually and selects the subgroup containing the best features (best features combination). In this way, the average value of the f-score of all features (threshold) is calculated. A characteristic is considered interesting and relevant when the f-score value is higher than the threshold value, being the irrelevant characteristics discarded. According to Fisher discriminate analysis, it is feasible to expand the f-score two class analysis to multiclass assessment. The equation presented in formula 4.11 refers to the calculation of the f-score value [56, 57].

$$f - score(f_i) = \frac{\sum_j \frac{n_j}{c-1}(\mu_j - \mu)^2}{\frac{1}{n-c}\sum_j (n_j - 1)\sigma_j^2} \tag{4.11}$$

where f_i is the i^{th} feature from the dataset, n_j represents the total instances of the class j and n refers to the total instances of all the class, μ is the mean values of all the class features, μ_j represents the mean value of all the j^{th} class features, σ means the standard deviation and c refers to the number of the output class.

4.5 Machine Learning

Artificial Intelligence (AI) describes the capacity of a technological system to reproduce human reasoning, thinking and decision making. The main purpose is to develop software systems and hardware devices that simulate the human cognitive functions through logic processes. In this way, machines would be able to learn new knowledge based on human intellectual behaviour [58, 59].

In turn, Machine Learning (ML) is a subcategory of AI that represents the ability of a system to make predictions without requiring direct and explicit instructions. Mathematical models are built based on a set of data, therefore the computer can learn its behaviour and make decisions according to the acquired experience [58, 60].

ML techniques has had a very positive impact, specifically regarding the analysis of the EEG signal, since they are able to classify data sets based on previous knowledge achieved through training. A ML classifier refers to a mathematical function that use significant information from a given data set (for example, values of features from brain activity). Those values are utilized as independent variables or predictors from a condition. The role of the classifier is to predict to which class that the condition belongs to [42].

The evaluation of the classifier remains an important step that aims to verify its pertinence. In cross-validation approach, the data is randomly divided into k mutually exclusive subsets of equal size – one is applied for testing and the others are used for training. This method is iterative over all the k folds [61].

In theoretical terms, the classifier accuracy is calculated according to the formula 4.12 [26].

$$Accuracy = \frac{TP + TN}{TP + FN + FP + FN} \tag{4.12}$$

where TP, TN, FN and FP mean true positive, true negative, false negative and false positive.

4.5.1 Support Vector Machine

SVM is commonly applied in the detection of subtle patterns in complex datasets, given the power it offers comparing with others. This supervised classification method takes benefits from a decision boundary, also known as hyperplane, in order to split two distinct classes of data (as shown in figure 4.2). The closest points of each group (support vectors) are selected and then, the hyperplane is traced to separate the data in a way to maximize the distance between the classes. Kernel function is useful to include additional space dimensions, making a nonlinear system become linear [62, 63].

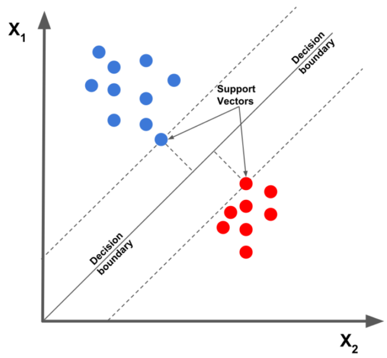

Figure 4.2: SVM classifier [63].

4.5.2 Naive Bayes

Naive Bayes algorithm is considered very efficient and simple in computational terms, namely regarding its good performance in training input attributes that are truly independent from each other. Derived from Bayes theorem, this classifier makes predictions based on probabilities learned from previous data. The goal is to maximize the probability P $(X_i|E)$ – the probability of the event X_i given the occurrence of event E (formula 4.13). In this way, the predicted class will be the one correspondent to the maximum value of that probability [64].

$$P(X_i|E) = \frac{P(X_i)}{P(E)} \prod_{k=1}^{n} P(v_k|X_i) \tag{4.13}$$

where X_i signifies the i^{th} class, E is the test sample, n represents the number of independent attributes given the class and v_k is the value of the k^{th} attribute in the sample E.

4.5.3 Decision Tree

The algorithm intends to create a hierarchical data organization formed by several nodes (tree).

The purpose of this algorithm is to create a branching map of attributes that best forecast class labels, through training set of cases labelled with classes. The first node of the tree is named root node and the ones derived from it are the internal ones (child nodes). Each internal node corresponds to a test and its answer represents an attribute (decision point). For a discrete and continuous attributes there are several and only 2 possible results, respectively. At the end of the tree, the leaf nodes identify the class assigned (as illustrated in figure 4.3) [64, 65].

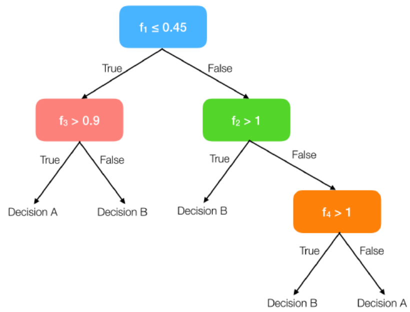

Figure 4.3: DT classifier [65].

4.5.4 K-Nearest Neighbor

The KNN classifier consists of an algorithm based on the nearest neighbour rule. This reasoning assumes that, when evaluating the provided sample, it corresponds to the category of the sample which is closest (the one who is more similar). More specifically, the nearest instances are analysed and the most common class is returned as prediction – the sample is grouped to its nearest neighbour (as shown in figure 4.4). When there are k number of nearest neighbours, the classification process depends on the distance, having greater ponderation the ones with a smallest distance [61, 66].

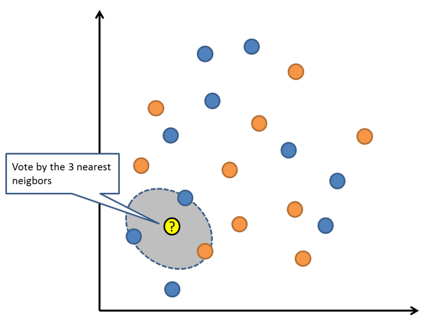

Figure 4.4: KNN classifier [61].

4.5.5 Ensemble

Ensemble classifier intends to aggregate several ML classifiers with the purpose of obtaining better results in the prediction process, thus improving the classification performance (contrasted with each individual classifier). The number of classifiers that will be combined (assemble size) depends on the accuracy and speed of each classifier. It is understandable that larger ensembles present longer training time for prediction and with too large classifiers over-trained classification may occur. Ensemble learning has three different approaches – random subspace, bagging and boosting. The figure above (figure 4.5) shows the diagram of this type of classification [67–69].

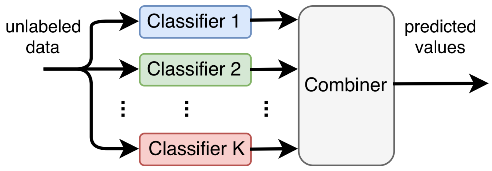

Figure 4.5: Ensemble classifier [69].

4.6 Deep Learning

Deep Learning (DL) is a subgroup of ML. The principle of DL concerns the development of algorithms using ANN to make prediction on data. ANN is valuable since these networks are created based on the structure of neurons situated in the human brain. Thereby, it is believed that DL optimises learning and intelligent decision making. Recently, and propelled by the increased computational power, it is possible to develop machines that understand and manipulate different kinds of large datasets – DL techniques have shown good capabilities namely when it comes to time series prediction. CNN and RRN are the two different types of DL algorithms [42, 70].

Regarding the network design, the system is fed with raw data (for example the EEG raw signal), thus developing the representations required for pattern recognition. This network is composed of numerous layers of representations – these layers are organized sequentially and composed of many primitive (non-linear) functions. The principle states that the representation of a given layer serves as feed to the next layer. As data passes across the network layers, the input space gets distorted until the data points become distinct [70].

4.6.1 Convolutional Neural Networks

CNN is a deep learning approach that learns relevant information directly from data without the need of manual feature extraction. These networks have a good architecture since they are able to learn crucial features. Due to their very accurate results, CNN have been applied in the medical field not only to detect patterns in images (medical imaging) but also in time series (EEG signal). This technique is based on convolution calculations and can perform both supervised and unsupervised learning. Figure 4.6 shows an example of a CNN network [29, 71].

The basic architecture of this network contains an input layer, multiple hidden layers and an output layer. The number and type of layers may change depending on the final purpose. Data is fed into the network through the input layer, allowing the insertion of different kinds and dimensions of datasets. The learning process is carried out until reaching the output layer. The output layer makes classification labels, taking advantage of a logistic function or a normalized exponential function. Typically, for classification problems, a softmax layer and then a classification layer are used as outputs [29, 71, 72].

The convolutional layer is the one that extract features from the input data by applying several convolutional filters (convolve the image). In turn, the activation layer selects only the activated features into the next layer. The Rectified Linear Unit (ReLU layer) is an example of a nonlinear activation function. Besides that, the role of the pooling layer is to filter the information by selecting certain characteristics, thus reducing the number of features that the network requires to learn (down sampling). The fully connected layer multiplies the input by a weight matrix and then adds a bias vector (combines all the features learned by the previous layers) [29, 71, 72].

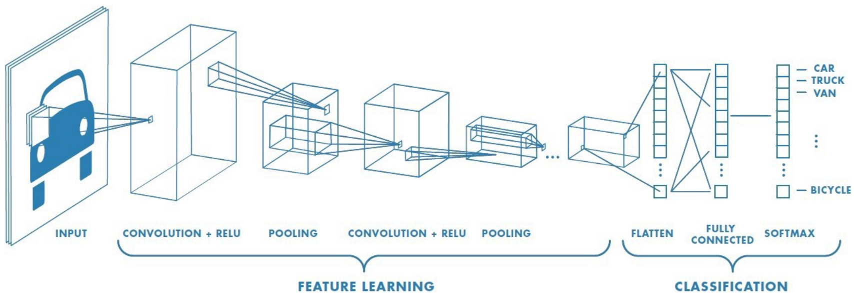

Figure 4.6: CNN network [71].

The specifics of the developed CNN network are presented in chapter 4.7.

4.7 Method Implementation

This work was developed taking benefit from MATLAB R2019a software. The methodology is divided into five steps – (1) Pre-processing, (2) Feature Extraction, (3) Feature Selection, (4) Classic ML and (5) DL. The following figure, figure 4.7, shows the block diagram that summarizes the methodology. At the end of this chapter, another more detailed diagram is presented (figure 4.9).

(1) The Pre-processing is a crucial step in terms of data planning and organisation. In this first stage, the EEG signals belonging to the database were loaded into the Matlab and then split in 5 second segments – 1280 samples per signal. Each signal has 19 channels and the segments number vary between subjects. Thereafter, each segment was normalized per channel according to the formula

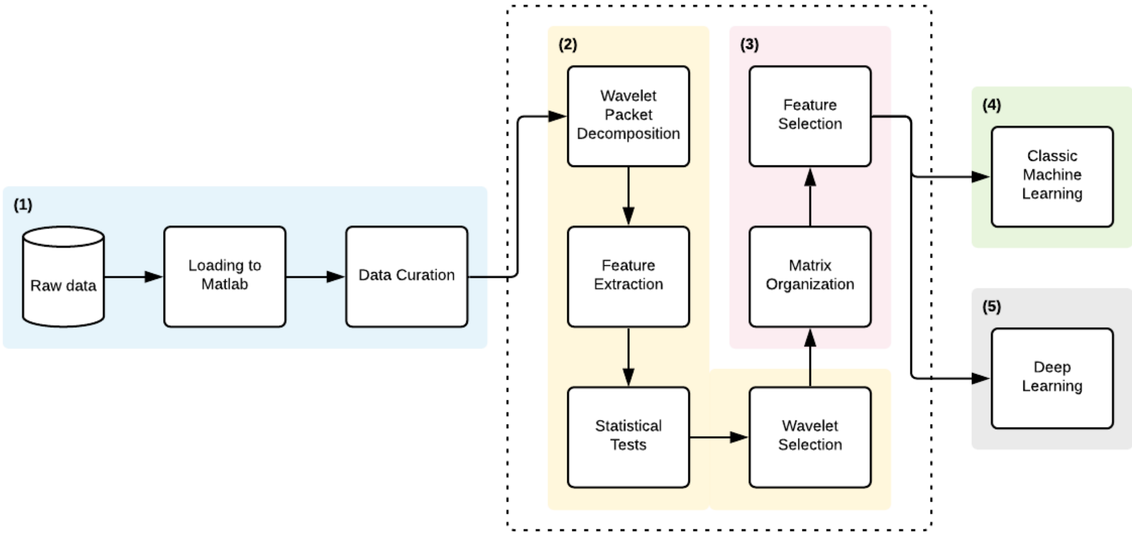

Figure 4.7: Block diagram.

4.14. It is important to emphasize that these signals had already been submitted for noise removal and they were digitally filtered by a 1-40 Hz band-pass filter. This filter was implemented precisely because it restricts the frequency range of interest (activity of the conventional EEG bands).

$$x(n) \quad = \quad \frac{x(n)}{\sum_{n=0}^{N-1} x^2(n)} \tag{4.14}$$

(2) Subsequently, the main goal of this step is to extract several features with discriminative power, meaning characteristics that have the ability to differentiate the groups in terms of their clinical situation. For this purpose, all signals were decomposed by the Wavelet Packet Transform until level 6. Signals were decomposed until this level with the purpose of selecting the frequency range of the desired conventional bands. Nodes were selected from the Wavelet Packet Tree thus obtaining 18 frequency bands of interest. Subsequently, 24 nonlinear and statistic features were extracted from each node and from each one of the 19 channels of all participants signals segments.

During this process, all the wavelets mentioned before (chapter 4.2.1) were analysed, meaning a total of 132 wavelets. This exhaustive approach enabled to make the best choice for this specific data. From taking benefit from Kruskal-Wallis statistical test, it was chosen the wavelet which provided the greatest results for the selected features. More prominently, a binary statistical analysis (one for each binary comparison C vs MCI, C vs ADM C vs ADA, MCI vs ADM, MCI vs ADA and ADM vs ADA) and another statistical analysis involving all study participants (All vs All) were performed. To make the results analysis more visually appealing and focused, the following figure (figure 4.8) displays the

10 wavelets that presented the best results – the 10 wavelets with the highest number of acceptable *p*-values. In this way, it is possible to concluded that the wavelet db34 revealed a higher number of accepted *p*-values ($p \leq 0.05$).

According to the state of the art, it is found that db is not the most commonly used. In fact, the most recent work carried out by Rodrigues *et al.* [28] points out that bior, in particular, bior3.5 proven to fit well with EEG signals from AD patients. However, Fiscon *et al.* [27] mentions that db4 was used since db ensures a precise feature extraction in the brain waves frequencies. This conclusion allows to state that the best wavelet depends from study to study, meaning that the database and the signal processing techniques applied have a strong impact on its choice. In the present work, db is undoubtedly responsible for the best results. Eventually, it would be interesting to extract other parameters to check if this decision would remain the same.

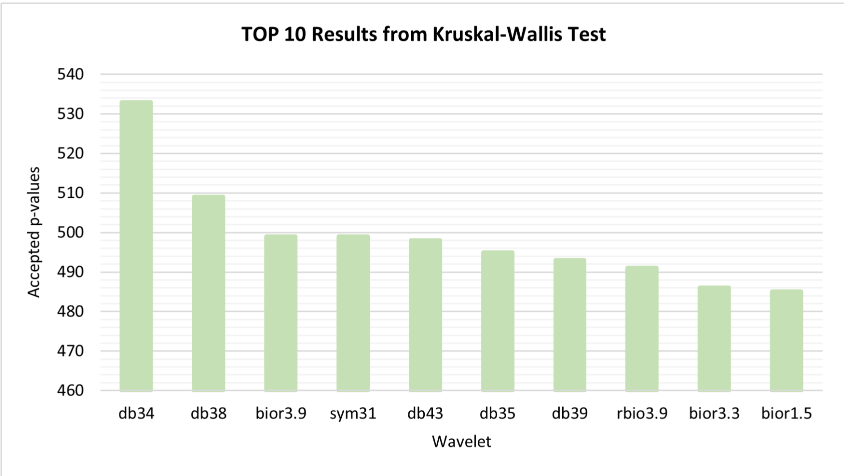

Figure 4.8: Kruskal-Wallis Test.

(3) Afterwards, it was applied the f-score feature selection technique in order to determine which features are considered the relevant ones (not redudant). Before starting this Feature Selection stage, the data was organized in 19 matrices since the evaluation is performed per channel. Instead of using all extracted features from all signals segments, 3 statistics were calculated from the time series (mean, standard deviation and variance). It is important to note that 4 (Threshold Wavelet Entropy, Sure Wavelet Entropy, Kolmogorov Exponent and Katz Fractal Dimension) of the 24 features initially extracted were removed since the f-score values did not vary between subjects. In fact, those values were not useful to the analysis due to the fact that they did not discriminate any differences between groups. Briefly, each one of the 19 matrices contains a matrix of 37 rows (37 subjects) by 1080 columns (1080 features – 20 metrics x 18 nodes x 3 statistics). Each input pair of each of these 19 matrices is representative of 1 patient.

Thereby, the main purpose is to select which is the best feature combination to classify the binary comparations (C vs MCI, C vs ADM C vs ADA, MCI vs ADM, MCI vs ADA and ADM vs ADA) and the multiclass (All vs All). Thus, 11 combinations of features were tested - 2 features, 3 features, 4 features, 5 features, 10 features, 15 features, 20 features, 5% of features, 10% of features, 20% of features and 100% features. Table 4.3 presents the maximum accuracy values obtained for each binary comparison. It was concluded that the range up to 5 features is the one containing the ones with the highest discriminative power.

According to the literature, there are some researchers who are in favour of feature selection, while others are not. In this specific case, the feature selection proves to be a crucial step since using the total number of features does not reflect into better results. According to Ruiz-Gómez *et al.* [31], this study applied a selection technique which concluded that 3 features was the optimal set. This paper, instead of the f-score method, uses another technique called Fast Correlation Based-Filter (FCBF). Although the technique used is distinct, the use of a reduced set of characteristics is corroborated.

Table 4.3: Features Combination.

Features	**Maximum Accuracy**					
	C vs MCI	C vs ADM	C vs ADA	MCI vs ADM	MCI vs ADA	ADM vs ADA
1080	73.7%	76.2%	68.4%	83.3%	87.5%	72.2%
20%	73.7%	76.2%	78.9%	88.9%	93.8%	72.2%
10%	78.9%	76.2%	78.9%	88.9%	93.8%	72.2%
5%	78.9%	76.2%	78.9%	83.3%	93.8%	77.8%
20	73.7%	81.0%	78.9%	88.9%	93.8%	77.8%
15	78.9%	81.0%	78.9%	88.9%	93.8%	77.8%
10	78.9%	76.2%	78.9%	88.9%	87.5%	77.8%
5	78.9%	76.2%	84.2%	88.9%	93.8%	77.8%
4	78.9%	76.2%	84.2%	83.3%	93.8%	77.8%
3	78.9%	81.0%	84.2%	83.3%	93.8%	77.8%
2	73.7%	81.0%	84.2%	83.3%	93.8%	77.8%

(4) Regarding the Classic ML section, and through Classification Learner app from Matlab, it was feasible to evaluate the performance of 25 classifiers. Statistics and Machine Learning Toolbox provides detaild information about the structure of these predictive models [73]. Consequently, the user only needs to select the input data since the algorithms structure is already developed. As output, the accuracy (in %) that each algorithm achieves in the classification procedure will be generated.

As concluded in the previous step, the set of 2, 3, 4 and 5 features reflect the best results. Thereby, these combinations of features fed those Classic ML classifiers, more specifically, each row of these 19 matrices represents an entrance (information from each patient). The results obtained reflect the accuracy of each classifier, based on leave-one-out cross-validation model. It is important to note that the classification is achieved per channel in order to understand the distinctive behaviour of each electrode. Besides that, topographic maps were also developed in order to visualize the scalp differences between groups. In this way, it is feasible to conclude which brain lobes show the most significant differences (best accuracy values).

(5) Finally, the ultimate goal was to train CNN in order to conclude about the potential of this DL technique to differentiate each study group, predicting the class. To feed this neural network, the input is the same as with Classic ML algorithms. As in the Classic ML section, accuracy values were obtained, based on leave-one-out cross-validation model. From those results, the work accomplishment allows a direct comparison between the efficiency of Classic ML contrasting with DL.

Regarding the CNN specificities, the architecture is composed by different layers - imageInputLayer, convolution2dLayer, reluLayer, fullyConnectedLayer, softmaxLayer and classificationLayer. Initially, the imageInputLayer feeds the network with the signals (the information from each input pair), applying data normalization. Thereafter, the convolution2dLayer implements sliding convolutional filters to the input. Subsequently, the reluLayer executes a threshold operation to each input element and set any value less than zero to zero. Afterwards, the fullyConnectedLayer multiplies the input by a weight matrix and then adds a bias vector. It is important to note that it were used 2 and 4 fullyConnectedLayers when refering to binary comparions and multi-class, respectively. Then, the softmaxLayer employes a softmax function. Finally, the classificationLayer calculates the cross-entropy loss for classification and weighted classification tasks with mutually exclusive classes [74].

In relation to the training algorithms (adam, sgdm and rmsprop), the 3 options were tested in order to use the one that achieves the best results. As they all achieved the same accuracy values, the adam training algorithm was chosen as it is the default option.

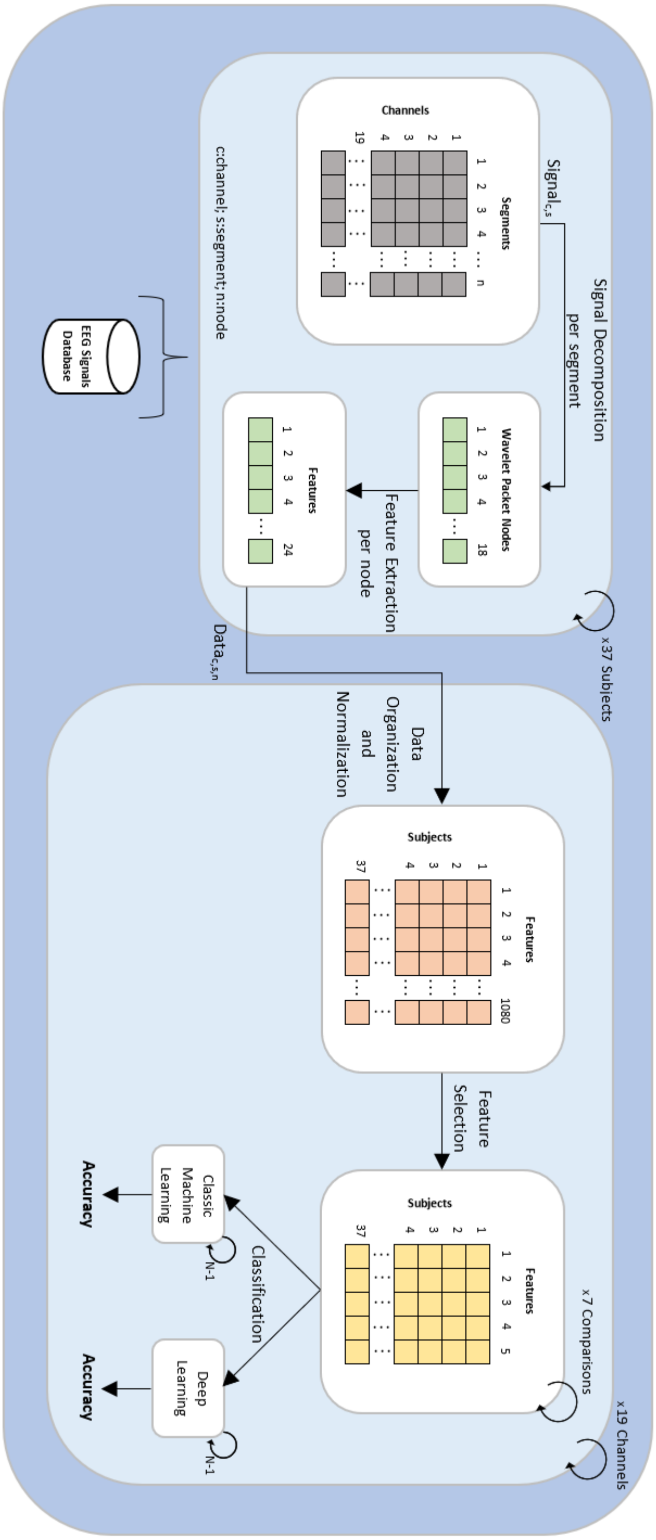

Figure 4.9: Methodology Summary.

CHAPTER 5

Results and Discussion

This chapter aims to present the results achived in the present work and their discussion. The main purpose is to infer about the performance of both Classic ML and DL classification per channel. The features and scalp regions which correspond to the best accuracy values for each comparison are mentioned. The results obtained are corroborated with the literature in order to discuss assertively.

Concerning, Classic ML the binary classification of C vs MCI exhibits a maximum accuracy of 78.9% visible in the P7 and Pz electrodes, being noticeable that those are the ones with the greatest discriminative power. Regarding this comparison, the classifier which presented this best performance was Decision Tree (Fine, Medium and Coarse Tree). In turn, DL selected the P8 electrode as the scalp region which present more significant differences between groups, reaching 78.9% as maximum accuracy.

Regarding Classic ML, and when comparing C vs ADM, it is observed that the C4 and P7 electrodes present a maximum accuracy of 81.0%, being the ones that reveal more significant differences. The classifiers Cubic SVM and Fine Gaussian SVM were the ones who correspond to the highest accuracy result. In contrast, through DL, a maximum accuracy of 76.2% was achieved in the Pz electrode, being this the one that correspond to the scalp region with the highest discriminative power between subjects.

Concerning Classic ML, the binary classification of C vs ADA exhibits a maximum accuracy of 84.2% visible in the F7, C4 and T8 electrodes, being noticeable that those are the ones with the greatest discriminative power. Regarding this comparison, the classifiers which presented this best performance were Linear SVM and Gaussian Naive Bayes. In turn, DL selected the F7 and F8 electrodes as the scalp regions which present more significant differences between groups, reaching 78.9% as maximum accuracy.

Regarding Classic ML, and when comparing MCI vs ADM, it is observed that the P7 electrode present a maximum accuracy of 88.9%, being the one that reveal more significant differences. The classifier Cosine KNN was the one who correspond to the highest accuracy result. In contrast, through DL, a maximum accuracy of 83.3% was achieved in the Pz electrode, being this the one that correspond to the scalp region with the highest discriminative power between subjects.

Concerning Classic ML, the binary classification of MCI vs ADA exhibits a maximum accuracy of 93.8% visible in the O1 electrode, being noticeable that this is the one with the greatest discriminative power. Regarding this comparison, the classifier which presented this best performance was Decision Tree (Fine, Medium and Coarse Tree). In turn, DL selected the P4 electrode as the scalp region which present more significant differences between groups, reaching 93.8% as maximum accuracy.

Regarding Classic ML, and when comparing ADM vs ADA, it is observed that the F3, F8, C3, C4, O1 electrodes present a maximum accuracy of 77.8%, being the ones that reveal more significant differences. The classifiers Ensemble Subspace Discriminant and Fine KNN were the ones who correspond to the highest accuracy result. In contrast, through DL, a maximum accuracy of 72.2% was achieved in the Fz, F4 and C4 electrodes, being these the ones that correspond to the scalp regions with the highest discriminative power between subjects.

Concerning Classic ML, the multiclass classification of All vs All exhibits a maximum accuracy of 56.8% visible in the Pz electrode, being noticeable that this is the one with the greatest discriminative power. Regarding this comparison, the classifier which presented this best performance was Medium Gaussian SVM. In turn, DL selected the Pz electrode as the scalp region which present more significant differences between groups, reaching 51.4% as maximum accuracy.

In fact, Classic ML shows better results than DL with the exception of 2 out of 7 comparisons. Being this the best technique, it is relevant to deepen the results obtained. In this way, the topographic maps were elaborated (figure 5.1) with the Classic ML results in order to visualise the scalp regions that present the best values of accuracy. Besides that, table 5.1 provides the features name that corresponds to the channels that exhibited the best accuracy values for each comparison through the Classic ML algorithms. The feature identification involves its name, the Wavelet Packet Tree node from which it was extracted and, lastly, the operation performed to the signal segments – $Feature\_node\_windows$. In this last parameter, it was calculated the mean (m), variance (v) or standard deviation (sd) of the segments of each signal.

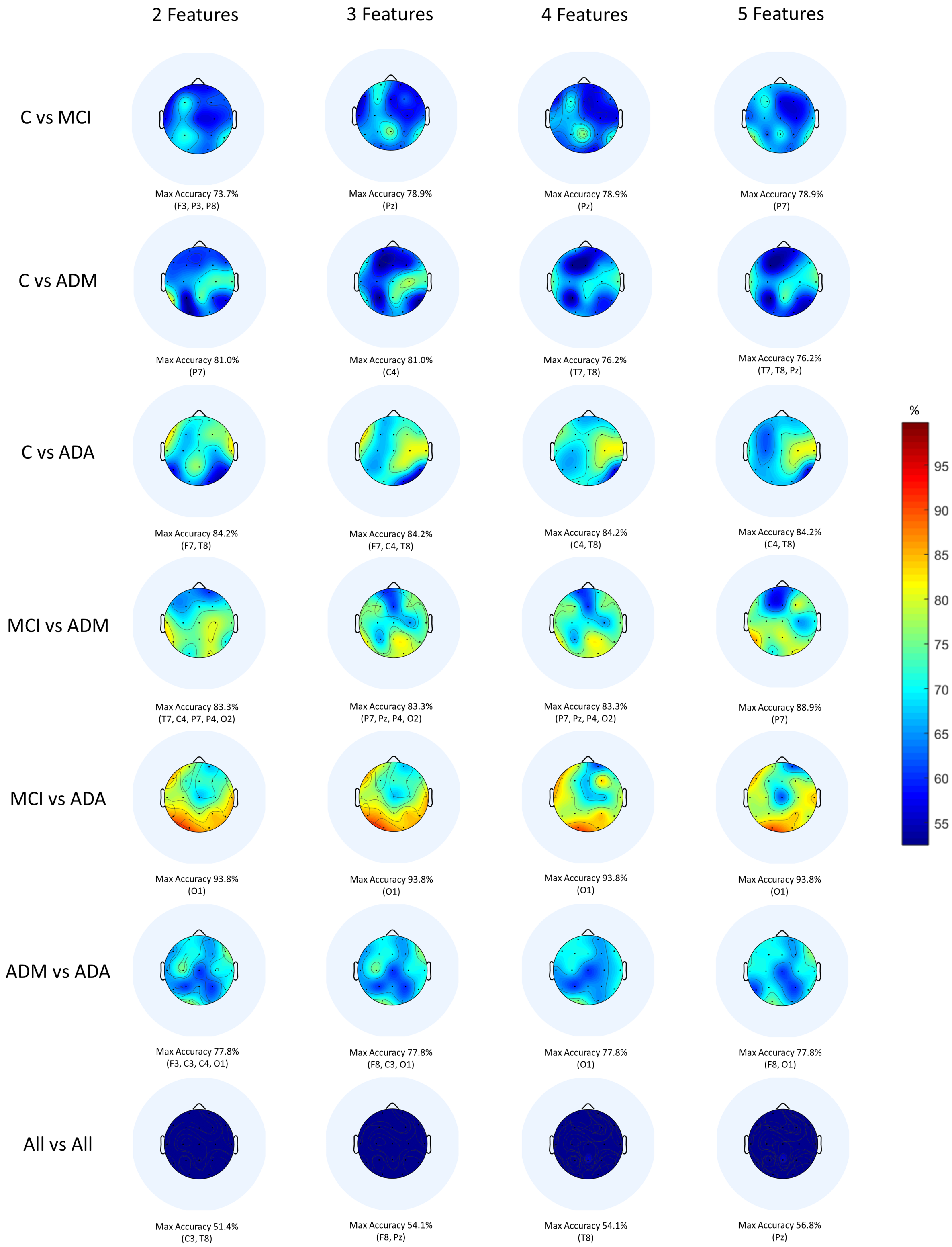

Figure 5.1: Topographic maps from Classic ML classification.

Table 5.1: Most important features in the classification procedure.

C vs MCI	C vs ADM	C vs ADA	MCI vs ADM	MCI vs ADA	ADM vs ADA	All vs All
Asy_12_v Asy_13_sd AEn_9_v AEn_9_sd Kur_10_v Kur_10_sd LEx_4_v LEx_4_sd LEx_6_v	Kur_10_v Kur_10_sd LEx_10_sd LEx_10_v HEx_8_m	PEn_16_v PEn_2_m PEn_12_m LEx_10_sd LEx_10_v HEx_8_m LEx_13_m HEx_4_m Kur_6_sd Asy_6_v Kur_6_v Asy_6_sd AEn_3_m	Kur_10_v Kur_10_sd LEx_4_v LEx_4_sd LEx_6_v	LEn_9_sd LEx_16_m LEn_9_v HFD_9_m Kur_9_m	HFD_4_sd HFD_4_v HEx_4_m PEn_13_m PEn_13_v HEx_2_m PEn_13_sd SEn_6_m HEx_17_v HEx_7_sd LEx_10_sd LEx_10_v LEn_9_sd LEx_16_m LEn_9_v HFD_9_m Kur_9_m	Asy_12_v Asy_13_sd AEn_9_v AEn_9_sd ShEn_7_sd

Asy:Asymmetry; Kur:Kurtosis; ShEn:Shannon Entropy; PEn:Permutation Entropy; SEn:Sample Entropy; LEn:Log Entropy; AEn:Approximate Entropy; LEx:Lyapunov Exponent; HEx:Hurst Exponent; HFD:Higuchi Fractal Dimension.

It is possible to assume that the results achieved corroborate the state of the art. Firstly, it is supposed that, when comparing the control group with the other ones, the highest accuracy would be obtained in the binary comparison C vs ADA. Obviously, a healthy patient will present a greater difference than a patient in AD severe stage does. The same interpretation is applied when contrasting MCI with the remaining subjects. Besides that, when comparing the ADM with the other ones, the lowest accuracy would be achieved in ADM vs ADA. These patients have fewer disparities between them, since they were both AD diagnosed. Regarding All vs All, this multiclass comparison represents the worst accuracy value. A possible justification is based on the fact of the extracted features, when combined, are not sufficiently discriminative when detecting all stages of Alzheimer's. Considering the variety of subjects (different disease stages), it is important to extract very characteristic and unique parameters of each clinical situation to achieve better results.

Besides that, C4 and P7 are the electrodes that which best distinguish the differences between subjects. Those electrodes correspond to the brain central region and to the parietal lobe, respectively. Depending on the disease stage and which stages are being compared, the area which is considered to be the most affected may vary. Although the temporal lobe is typically responsible for memory functions, it does not mean that the other scalp regions are not equally affected by the disease. According to Siuly and Zhang [38], the parietal area is very important in recognition and orientation. Actually, AD patients tend to lose these abilities, for example, at the moments when they do not recognise their family or when they feel lost in a physical space they previously knew. This conclusion sustains the results obtained.

Aditionally, it is crucial to reflect on the work considered closest to this one in order to draw some further conclusions. In this way, tables 5.2, 5.3 and 5.4 present a direct comparison between the developed work and others – same EEG database (table 5.2), different EEG database (table 5.3) and MRI and PET databases (table 5.4). It is important to note that the detailed methodology of all the mentioned studies was previously presented in the state of the art (chapter 2) and that each accuracy result is rounded to the nearest integer number.

Compared to other methods of diagnosing AD through EEG signals from the same database (table 5.2), the proposed method outperformed the study developed by Rodrigues *et al.* [28] by 2% in the binary comparison MCI vs ADM. In contrast, the remaining comparisons performed (C vs MCI, C vs ADM, C vs ADA, MCI vs ADA, ADM vs ADA and All vs All) did not show better results. Besides that, it can be seen that CNNs have never been applied to this dataset, so this work is the first and the only one that follows this approach. Indeed, this works presents added value to the scientific community, as it has the potential to be improved and become a powerful tool for AD diagnosis in all its stages.

Compared to other techniques of diagnosing AD through EEG signals from different databases (table 5.3),it is observed that the present study outperformed the work carried out by Fiscon *et al.* [27] by 13% in the pair MCI vs AD. Conversely, the remaining comparisons analysed (C vs MCI, C vs ADM, C vs ADA, ADM vs ADA and All vs All) did not present better results. It is noteworthy that the present study has the peculiarity of being the only one that applied the F-score technique, so it may have highly contributed to good classification results.

Finally, compared to other methods of diagnosing AD through images (table 5.4), the proposed method outperformed the study developed by Thapa *et al.* [30] concerning the binary comparison MCI vs AD by 1%. In contrast, the remaining comparisons (C vs MCI, C vs ADM, C vs ADA, ADM vs ADA and All vs All) did not achieved better results. It should be noted that the observation of the present study compared with different methods (imaging methods) is the one that refers to a less direct comparison.

Briefly, it is feasible to conclude that the features extracted from the EEG signals (particularly the entropy parameters which characterise so well the signals nonlinearity) show potential insofar as the overall results validate the state of the art. In general, Classic ML prove to be better than DL regarding the classification procedure since DL is commonly used when analysing large amounts of data. According to Esteva *et al.* [70], and particulary regarding the medical field, DL methods are benefic because of the capacity to generate sheer volume of data. In fact, this database only contains 37 participants so the simpler methods (such as Classic ML) were sufficient to have better results than CNN. Although the results obtained derive from the database used, it is believed that this tool is also promising with other patients. Assuming this, it is possible to state that this tool has potential to assist doctors in the AD diagnosis in different stages.

Table 5.2: Comparison with previous work with the same EEG database.

Study	Signal Processing	Features	Feature Selection	Best Classifier	Classification Accuracy
[25]	Multiband Spectral Analysis via DWT	RP, Spectral Ratios, Maxima, Minima and Zero Crossing	KW Test	ANN	C vs MCI - 77%
					C vs AD - 95%
					MCI vs AD - 83%
					All vs All - 90%
[28]	Multiband Cepstral and Lacstral Analysis via DWT	Cepstral and Lacstral Distances	KW Test	ANN	C vs MCI - 98%
					C vs ADM - 96%
					C vs ADA - 96%
					C vs ADM-ADA - 96%
					MCI vs ADM - 87%
					MCI vs ADA - 99%
					MCI vs ADM-ADA - 94%
					All vs All - 96%
Present Study	Nonlinear and Multiband Analysis via DWPT	Nonlinear and Statistic Parameters	F-score	SVM	C vs MCI - 79%
					C vs ADM - 81%
					C vs ADA - 84%
					MCI vs ADM - 89%
					MCI vs ADA - 94%
					ADM vs ADA - 78%
					All vs All - 57%

Table 5.3: Comparison with previous work with different EEG databases.

Study	Signal Processing	Features	Feature Selection	Best Classifier	Classification Accuracy
[27]	Fourier and Wavelet Analysis via FFT and DWT	Fourier and Wavelet Coefficients	Not applied	DT	C vs AD - 83%
					C vs MCI - 92%
					MCI vs AD - 79%
[31]	Not applied	Spectral and Nonlinear Features	FCBF	MLP	C vs All - 78%
					AD vs All - 76%
[26]	Spectral and Wavelet Analysis via PSD and DWT	RP, Mean and Variance	Not applied	SVM	C vs AD - 94%
[32]	Multiband Analysis via DWT and Burg's Method	Amplitude Summation, Variance and Amplitude Summation of the Coherence Values	Not applied	Bagged Trees	C vs MCI vs AD - 97%
[35]	Multiband Analysis via DWT and EMD	Variance, Kurtosis, Skewness, Shannon Entropy, Sure Entropy and Hjorth Parameters	Not applied	KNN	C vs AD1 vs AD2 - 98%
Present Study	Nonlinear and Multiband Analysis via DWPT	Nonlinear and Statistic Parameters	F-score	SVM	C vs MCI - 79%
					C vs ADM - 81%
					C vs ADA - 84%
					MCI vs ADM - 89%
					MCI vs ADA - 94%
					ADM vs ADA - 78%
					All vs All - 57%

Table 5.4: Comparison with previous work with MRI and PET databases.

Study	Image Processing	Features	Feature Selection	Best Classifier	Classification Accuracy
[30]	Not applied	MMSE Score and Right and Left Hippocampus Volumes from MRI	Filter-based	SVM	C vs AD - 99%
					C vs MCI - 79%
					MCI vs AD - 91%
[36]	Normalization and Segmentation	No Feature Extraction	Not applied	CNN	C vs AD - 99%
					C vs cMCI - 88%
					C vs sMCI - 76%
					AD vs cMCI - 76%
					AD vs sMCI - 86%
					cMCI vs sMCI - 75%
[33]	Normalization	Intra-slice and Inter-slice Features	Not applied	RNN	C vs AD - 91%
					C vs MCI - 79%
Present Study	Nonlinear and Multiband Analysis via DWPT	Nonlinear and Statistic Parameters	F-score	SVM	C vs MCI - 79%
					C vs ADM - 81%
					C vs ADA - 84%
					MCI vs ADM - 89%
					MCI vs ADA - 94%
					ADM vs ADA - 78%
					All vs All - 57%

Chapter 6

Conclusions and Future Prespectives

Alzheimer is a neurodegenerative disease marked by a rapid progression, leading the human to totally lose the cognitive functions. As with other dementias, the prevalence is increasing and forecasts show no improvements. During the early stages, patients do not have symptoms – the disease is silent. As soon as the symptomatic phase begins, there is a worsening of the clinical condition.

Nowadays, there are some medical exams that help the AD diagnosis, as well as medication to reduce the symptoms. Out of curiosity, on June 7^{th} 2021, the Food and Drug Administration (FDA) approved a new drug named Aducanumab (brand name Aduhelm). This monoclonal antibody aims to halt the disease progression by attaching the amyloid plaques and removing them. Thus, by stopping the formation of these plaques, tau tangles will not be form – there will be no death of neuronal cells and the cognitive functions will not be deteriorated [75]. This accomplishment confirms that if the AD detection could be achieved in the asymptomatic phase, there will be drugs capable of acting and fighting the disease progression.

To assist these diseases diagnosis, the branch of Biomedical Engineering emerged since it is a rapidly growing field that spotlight on creating valuable solutions to improve healthcare quality. Due to the advances in medicine, several researchers have been focussed on developing novels and powerful tools. Since AD is considered one of the most severe diseases and there is still no cure, the main goal of this work was to develop a strong system proficient of assisting in the AD diagnosis.

Hence, and taking benefit from Wavelet Packet Transform, several statistic and nonlinear features were extracted from various frequency bands of the EEG signals. The most interesting features were used to evaluate the significant differences between the study groups through Classic ML and DL classification algorithms. The main results indicate that Classic ML show best results than DL, except in the binary comparisons C vs MCI and MCI vs ADA (exactly the same accuracy).

In conclusion, the aim of this dissertation was accomplished accordingly to what was expected since it corroborate the state of the art, having exceeded some accuracies comparing to others studies.

Regarding the state of the art with the same EEG database, the proposed method outperforms by 2%, in the binary comparison MCI vs ADM. It is relevant to highlight that the remaining comparisons performed (C vs MCI, C vs ADM, C vs ADA, MCI vs ADA, ADM vs ADA and All vs All) did not show better results. Indeed, this improvement reflects the impact that this tool can have particularly in distinguishing these two consecutive stages of AD.

It is important to enhance that this work was presented in the Affect, Personality and the Embodied Brain (APE2021) conference from Nottingham Trent Univeristy on September 20^{th}, 2021. The preliminary results of this research were exposed in a poster format [76].

After the work main conclusions, it is also relevant to describe future lines of research. There are some aspects that can be improved in the near future, as far as the author capacities are successful. The future perspectives are exposed:

- **Database** – With the aim of generalise the results to any database, it would be a good idea to test this tool with other databases. Additionally, also evaluate the performance of the same database with a larger number of participants equally balanced;

- **Methodology** – In terms of methodology, it would be useful to extract more features from the frequency bands to understand how these describe the EEG signal, being or not relevant in its analysis;

- **Deep Learning** – As far as DL techniques are concerned, there are other algorithms for classification. For example, it would be interesting to feed the LSTM neural network to understand whether it performs better than CNNs. Besides that, CNNs should be optimised to assess whether the results would be better;

- **Market** – Carry out a robust economic analysis of this medical solution as possible market integration. The main purpose would be to analyse the market opportunity as well as understand the solution viability, among other economic parameters. In other words, it is interesting to realize if it represents a good investment for the medicine field;

- **Asthetic** – It would also be curious to design the graphical interface of this solution which would be implemented in the medical devices. It is relevant to consider aesthetic aspects, as well as being user-friendly for health professionals;

- **Other diseases** – Extend the developed AI system to assist the diagnosis of other brain diseases or other kinds of dementia. Thereby, it would be possible to understand the features that best describe each disease, making a distinction between the various neurologic pathologies;

Bibliography

[1] *Ageing and health*. Visited on 16/04/2021. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health.

[2] Y. Hou, X. Dan, M. Babbar, Y. Wei, S. G. Hasselbalch, D. L. Croteau, and V. A. Bohr. "Ageing as a risk factor for neurodegenerative disease". In: *Nature Reviews Neurology* 15.10 (2019), pp. 565–581. DOI: 10.1038/s41582-019-0244-7.

[3] A. Association. "2021 Alzheimer's disease facts and figures". In: *Alzheimer's & Dementia*. Vol. 17. 3. Wiley, 2021, pp. 327–406. DOI: 10.1002/alz.12328.

[4] *Dementia*. Visited on 16/04/2021. URL: https://www.who.int/news-room/fact-sheets/detail/dementia.

[5] *O que é a Demência?* Visited on 17/04/2021. URL: https://alzheimerportugal.org/pt/text-0-9-32-18-o-que-e-a-demencia.

[6] *Dementia statistics*. Visited on 17/04/2021. URL: https://www.alzint.org/about/dementia-facts-figures/dementia-statistics/.

[7] *Global Health Estimates: Life expectancy and leading causes of death and disability*. Visited on 20/04/2021. URL: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates.

[8] *Alzheimer Diagnóstico Precoce*. Visited on 20/04/2021. URL: https://www.hospitaldaluz.pt/pt/guia-de-saude/dicionario-de-saude/D/97/alzheimer-diagnostico-precoce.

[9] *Doença de Alzheimer*. Visited on 20/04/2021. URL: https://www.cuf.pt/saude-a-z/alzheimer.

[10] A. Association. "2019 Alzheimer's disease facts and figures". In: *Alzheimer's & Dementia* 15.3 (2019), pp. 321–387. DOI: 10.1016/j.jalz.2019.01.010.

[11] T. Revett, G. Baker, J. Jhamandas, and S. Kar. "Glutamate system, amyloid β peptides and tau protein: functional interrelationships and relevance to Alzheimer disease pathology". In: *Journal of Psychiatry & Neuroscience* 38.1 (2013), pp. 6–23. DOI: 10.1503/jpn.110190.

[12] *What Happens to the Brain in Alzheimer's Disease?* Visited on 10/05/2021. URL: https://www.nia.nih.gov/health/what-happens-brain-alzheimers-disease.

[13] J. Aaseth, A. Buha, D. R. Wallace, and G. Bjørklund. "Xenobiotics, Trace Metals and Genetics in the Pathogenesis of Tauopathies". In: *International Journal of Environmental Research and Public Health* 17.4 (2020), pp. 1–11. DOI: 10.3390/ijerph17041269.

[14] P. Scheltens, K. Blennow, M. M. B. Breteler, B. de Strooper, G. B. Frisoni, S. Salloway, and W. M. V. der Flier. "Alzheimer's disease". In: *The Lancet* 388.10043 (2016), pp. 505–517. DOI: 10.1016/s0140-6736(15)01124-1.

[15] D. Eratne, S. M. Loi, S. Farrand, W. Kelso, D. Velakoulis, and J. C. Looi. "Alzheimer's disease: clinical update on epidemiology, pathophysiology and diagnosis". In: *Australasian Psychiatry* 26.4 (2018), pp. 347–357. DOI: 10.1177/1039856218762308.

[16] M. I Kester and P. Scheltens. "Dementia: THE BARE ESSENTIALS". In: *Practical Neurology* 9.4 (2009), pp. 241–251. DOI: 10.1136/jnnp.2009.182477.

[17] R. Cassani, M. Estarellas, R. San-Martin, F. J. Fraga, and T. H. Falk. "Systematic Review on Resting-State EEG for Alzheimer's Disease Diagnosis and Progression Assessment". In: *Disease Markers* 2018 (2018), pp. 1–26. DOI: 10.1155/2018/5174815.

[18] N. Houmani, F. Vialatte, E. Gallego-Jutglà, G. Dreyfus, V.-H. Nguyen-Michel, J. Mariani, and K. Kinugawa. "Diagnosis of Alzheimer's disease with Electroencephalography in a differential framework". In: *PLOS ONE* 13.3 (2018), pp. 1–19. DOI: 10.1371/journal.pone.0193607.

[19] K. Pietrzak, K. Czarnecka, E. Mikiciuk-Olasik, and P. Szymanski. "New Perspectives of Alzheimer Disease Diagnosis – the Most Popular and Future Methods". In: *Medicinal Chemistry* 14.1 (2018), pp. 34–43. DOI: 10.2174/1573406413666171002120847.

[20] P. Vemuri, M. E. Murray, and C. R. Jack. "Neuroimaging in Dementias". In: *Rosenberg's Molecular and Genetic Basis of Neurological and Psychiatric Disease*. Elsevier, 2015, pp. 107–118. DOI: 10.1016/b978-0-12-410529-4.00010-3.

[21] K. A. Johnson, N. C. Fox, R. A. Sperling, and W. E. Klunk. "Brain Imaging in Alzheimer Disease". In: *Cold Spring Harbor Perspectives in Medicine* 2.4 (2012), a006213–a006213. DOI: 10.1101/cshperspect.a006213.

[22] K. Blocka. *EEG (Electroencephalogram): Purpose, Procedure, and Risks*. Visited on 28/05/2021. URL: https://www.healthline.com/health/eeg.

[23] K. Blennow. "Cerebrospinal fluid protein biomarkers for Alzheimer's disease". In: *NeuroRX* 1.2 (2004), pp. 213–225. DOI: 10.1602/neurorx.1.2.213.

[24] T. Fertalova and I. Ondriova. "Non-pharmacological Treatment of Alzheimer's". In: *Redirecting Alzheimer Strategy - Tracing Memory Loss to Self Pathology*. IntechOpen, 2019. DOI: 10.5772/intechopen.84893.

[25] P. M. Rodrigues, D. R. Freitas, J. P. Teixeira, D. Alves, and C. Garrett. “Electroencephalogram Signal Analysis in Alzheimer's Disease Early Detection”. In: *International Journal of Reliable and Quality E-Healthcare* 7.1 (2018), pp. 40–59. DOI: 10.4018/ijrqeh.2018010104.

[26] V. Bairagi. “EEG signal analysis for early diagnosis of Alzheimer disease using spectral and wavelet based features”. In: *International Journal of Information Technology* 10.3 (2018), pp. 403–412. DOI: 10.1007/s41870-018-0165-5.

[27] G. Fiscon, E. Weitschek, A. Cialini, G. Felici, P. Bertolazzi, S. D. Salvo, A. Bramanti, P. Bramanti, and M. C. D. Cola. “Combining EEG signal processing with supervised methods for Alzheimer's patients classification”. In: *BMC Medical Informatics and Decision Making* 18.1 (2018), pp. 1–10. DOI: 10.1186/s12911-018-0613-y.

[28] P. M. Rodrigues, B. C. Bispo, C. Garrett, D. Alves, J. P. Teixeira, and D. Freitas. “Lacsogram: A New EEG Tool to Diagnose Alzheimer's Disease”. In: *IEEE Journal of Biomedical and Health Informatics* 25.9 (2021), pp. 3384–3395. DOI: 10.1109/jbhi.2021.3069789.

[29] Z. Jiang, P. Liu, Y. Xia, and J. Zhang. “Application of CNN in EEG Image Classification of AD Patients”. In: *The 2nd International Conference on Computing and Data Science*. ACM, 2021, pp. 1–5. DOI: 10.1145/3448734.3450473.

[30] S. Thapa, P. Singh, D. K. Jain, N. Bharill, A. Gupta, and M. Prasad. “Data-Driven Approach based on Feature Selection Technique for Early Diagnosis of Alzheimer's Disease”. In: *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–8. DOI: 10.1109/ijcnn48605.2020.9207359.

[31] S. Ruiz-Gómez, C. Gómez, J. Poza, G. Gutiérrez-Tobal, M. Tola-Arribas, M. Cano, and R. Hornero. “Automated Multiclass Classification of Spontaneous EEG Activity in Alzheimer's Disease and Mild Cognitive Impairment”. In: *Entropy* 20.1 (2018), pp. 1–15. DOI: 10.3390/e20010035.

[32] B. Oltu, M. F. Akşahin, and S. Kibaroğlu. “A novel electroencephalography based approach for Alzheimer's disease and mild cognitive impairment detection”. In: *Biomedical Signal Processing and Control* 63 (2021), p. 102223. DOI: 10.1016/j.bspc.2020.102223.

[33] M. Liu, D. Cheng, and W. Y. and. “Classification of Alzheimer's Disease by Combination of Convolutional and Recurrent Neural Networks Using FDG-PET Images”. In: *Frontiers in Neuroinformatics* 12 (2018), pp. 1–12. DOI: 10.3389/fninf.2018.00035.

[34] J. P. Amezquita-Sanchez, N. Mammone, F. C. Morabito, S. Marino, and H. Adeli. “A novel methodology for automated differential diagnosis of mild cognitive impairment and the Alzheimer's disease using EEG signals”. In: *Journal of Neuroscience Methods* 322 (2019), pp. 88–95. DOI: 10.1016/j.jneumeth.2019.04.013.

[35] M. S. Safi and S. M. M. Safi. “Early detection of Alzheimer's disease from EEG signals using Hjorth parameters”. In: *Biomedical Signal Processing and Control* 65 (2021), p. 102338. DOI: 10.1016/j.bspc.2020.102338.

[36] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, R. Santangelo, and M. Filippi. “Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks”. In: *NeuroImage: Clinical* 21 (2019), p. 101645. DOI: 10.1016/j.nicl.2018.101645.

[37] J. S. Kumar and P. Bhuvaneswari. “Analysis of Electroencephalography (EEG) Signals and Its Categorization–A Study”. In: *Procedia Engineering* 38 (2012), pp. 2525–2536. DOI: 10.1016/j.proeng.2012.06.298.

[38] S. Siuly, Y. Li, and Y. Zhang. “Electroencephalogram (EEG) and Its Background”. In: *Health Information Science*. Springer International Publishing, 2016, pp. 3–21. DOI: 10.1007/978-3-319-47653-7_1.

[39] N. Dey, A. S. Ashour, W. S. Mohamed, and N. G. Nguyen. “Biomedical Signals”. In: *SpringerBriefs in Speech Technology*. Springer International Publishing, 2018, pp. 7–20. DOI: 10.1007/978-3-319-92225-6_2.

[40] D. P. Subha, P. K. Joseph, R. A. U, and C. M. Lim. “EEG Signal Analysis: A Survey”. In: *Journal of Medical Systems* 34.2 (2008), pp. 195–212. DOI: 10.1007/s10916-008-9231-z.

[41] H. Cai, J. Han, Y. Chen, X. Sha, et al. “A Pervasive Approach to EEG-Based Depression Detection”. In: *Complexity* 2018 (2018), pp. 1–13. DOI: 10.1155/2018/5238028.

[42] L. Hu and Z. Zhang. *EEG Signal Processing and Feature Extraction*. Springer Singapore, 2019. ISBN: 978-981-13-9112-5. DOI: 10.1007/978-981-13-9113-2.

[43] S. Sanei and J. Chambers. *EEG Signal Processing*. John Wiley & Sons Ltd, 2007. ISBN: 978-0-470-02581-9. DOI: 10.1002/9780470511923.

[44] M. Sazgar and M. G. Young. “EEG Artifacts”. In: *Absolute Epilepsy and EEG Rotation Review*. Springer International Publishing, 2019, pp. 149–162. DOI: 10.1007/978-3-030-03511-2_8.

[45] N. Bajaj. “Wavelets for EEG Analysis”. In: *Wavelet Theory*. IntechOpen, 2020, pp. 1–16. DOI: 10.5772/intechopen.94398.

[46] M. Misiti, Y. Misiti, G. Oppenheim, and J.-M. Poggi. *Wavelet Toolbox User's Guide Version 3*. Ed. by MathWorks. 2004.

[47] A. Y. Al-Qammaz, Y. Yusof, and F. K. Ahamd. “An enhanced Discrete Wavelet Packet Transform for Feature Extraction in Electroencephalogram Signals”. In: *Proceedings of the International Conference on Imaging, Signal Processing and Communication - ICISPC 2017*. ACM Press, 2017, pp. 88–93. DOI: 10.1145/3132300.3132303.

[48] N. Sairamya, M. Subathra, E. S. Suviseshamuthu, and S. T. George. “A new approach for automatic detection of focal EEG signals using wavelet packet decomposition and quad binary pattern method”. In: *Biomedical Signal Processing and Control* 63 (2021), p. 102096. DOI: 10.1016/j.bspc.2020.102096.

[49] M. Misiti, Y. Misiti, G. Oppenheim, and J.-M. Poggi. *Wavelet Toolbox User's Guide*. Ed. by MathWorks. 2021.

[50] O. Rashid, A. Amin, and M. R. Lone. “Performance Analysis of DWT Families”. In: *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*. 2020, pp. 1457–1463.

[51] M. Nielsen. “On the Construction and Frequency Localization of Finite Orthogonal Quadrature Filters”. In: *Journal of Approximation Theory* 108.1 (2001), pp. 36–52. DOI: 10.1006/jath.2000.3514.

[52] Y. Zhang, Y. Zhang, J. Wang, and X. Zheng. “Comparison of classification methods on EEG signals based on wavelet packet decomposition”. In: *Neural Computing and Applications* 26.5 (2014), pp. 1217–1225. DOI: 10.1007/s00521-014-1786-7.

[53] T. D. Albertovich and R. I. Aleksandrovna. “The Fractal Analysis of the Images and Signals in Medical Diagnostics”. In: *Fractal Analysis - Applications in Health Sciences and Social Sciences*. InTech, 2017, pp. 57–86. DOI: 10.5772/intechopen.68167.

[54] I. Wijayanto, R. Hartanto, and H. A. Nugroho. “Higuchi and Katz Fractal Dimension for Detecting Interictal and Ictal State in Electroencephalogram Signal”. In: *2019 11th International Conference on Information Technology and Electrical Engineering (ICITEE)*. IEEE, 2019, pp. 1–6. DOI: 10.1109/iciteed.2019.8929940.

[55] C.-T. Shi. “Signal Pattern Recognition Based on Fractal Features and Machine Learning”. In: *Applied Sciences* 8.8 (2018), pp. 1–15. DOI: 10.3390/app8081327.

[56] N. Sevani, I. Hermawan, and W. Jatmiko. “Feature Selection based on F-score for Enhancing CTG Data Classification”. In: *2019 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)*. IEEE, 2019, pp. 18–22. DOI: 10.1109/cyberneticscom.2019.8875656.

[57] Q. Song, H. Jiang, and J. Liu. “Feature selection based on FDA and F-score for multi-class classification”. In: *Expert Systems with Applications* 81 (2017), pp. 22–27. DOI: 10.1016/j.eswa.2017.02.049.

[58] *Diferença entre IA e machine learning*. Visited on 13/08/2021. URL: https://azure.microsoft.com/pt-pt/overview/artificial-intelligence-ai-vs-machine-learning/#introduction.

[59] R. S. T. Lee. *Artificial Intelligence in Daily Life*. Springer Singapore, 2020. ISBN: 978-981-15-7694-2.

[60] X.-D. Zhang. "Machine Learning". In: *A Matrix Algebra Approach to Artificial Intelligence*. Springer Singapore, 2020, pp. 223–440. DOI: 10.1007/978-981-15-2770-8_6.

[61] *Machine Learning Classifiers*. Visited on 20/08/2021. URL: https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623.

[62] "Applications of Support Vector Machine (SVM) Learning in Cancer Genomics". In: *Cancer Genomics & Proteomics* 15.1 (2018), pp. 41–51. DOI: 10.21873/cgp.20063.

[63] S. Mallick. *Support Vector Machines (SVM): LearnOpenCV*. Visited on 20/08/2021. URL: https://learnopencv.com/support-vector-machines-svm/.

[64] A. Awaysheh, J. Wilcke, F. Elvinger, L. Rees, W. Fan, and K. L. Zimmerman. "Review of Medical Decision Support and Machine-Learning Methods". In: *Veterinary Pathology* 56.4 (2019), pp. 512–525. DOI: 10.1177/0300985819829524.

[65] I. Mollas, G. Tsoumakas, and N. Bassiliades. "LionForests: Local Interpretation of Random Forests through Path Selection". In: *ArXiv* abs/1911.08780 (2019), pp. 1–8.

[66] W. Xing and Y. Bei. "Medical Health Big Data Classification Based on KNN Classification Algorithm". In: *IEEE Access* 8 (2020), pp. 28808–28819. DOI: 10.1109/access.2019.2955754.

[67] S. Ekiz and P. Erdogmus. "Comparative study of heart disease classification". In: *2017 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*. IEEE, 2017, pp. 1–4. DOI: 10.1109/ebbt.2017.7956761.

[68] A. S. Ashour, Y. Guo, A. R. Hawas, and G. Xu. "Ensemble of subspace discriminant classifiers for schistosomal liver fibrosis staging in mice microscopic images". In: *Health Information Science and Systems* 6.1 (2018), pp. 1–10. DOI: 10.1007/s13755-018-0059-8.

[69] *Ensemble Learning Toolbox*. Visited on 22/08/2021. URL: https://ww2.mathworks.cn/matlabcentral/fileexchange/68383-ensemble-learning-toolbox?s_tid=FX_rc3_behav.

[70] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean. "A guide to deep learning in healthcare". In: *Nature Medicine* 25.1 (2019), pp. 24–29. DOI: 10.1038/s41591-018-0316-z.

[71] *Convolutional Neural Network*. Visited on 02/09/2021. URL: https://www.mathworks.com/discovery/convolutional-neural-network-matlab.html.

[72] *Specify Layers of Convolutional Neural Network*. Visited on 02/09/2021. URL: https://www.mathworks.com/help/deeplearning/ug/layers-of-a-convolutional-neural-network.html.

[73] *Statistics and Machine Learning Toolbox*. Visited on 02/09/2021. URL: https://www.mathworks.com/products/statistics.html.

[74] *List of Deep Learning Layers*. Visited on 05/09/2021. URL: https://www.mathworks.com/help/deeplearning/ug/list-of-deep-learning-layers.html.

[75] *A new Alzheimer's drug has been approved. But should you take it?* Visited on 10/09/2021. URL: https://www.health.harvard.edu/blog/a-new-alzheimers-drug-has-been-approved-but-should-you-take-it-202106082483.

[76] T. Araújo, J. P. Teixeira, and P. M. Rodrigues. "Artificial Intelligence System for the Automatic Detection of Alzheimer Disease through Electroencephalographic Signals". In: *Proceedings of Affect, Personality and the Embodied Brain Conference "APE2020"* (2021), p. 33.