CATÓLICA
ESCOLA SUPERIOR DE BIOTECNOLOGIA

PORTO

Gray Level Co-Occurrence Matrix Structural MRI Texture Analysis for Alzheimer's Disease Prediction

by
Maria João Bezerra Ferreira de Castro Oliveira

May 2024

CATÓLICA
ESCOLA SUPERIOR DE BIOTECNOLOGIA
PORTO

Gray Level Co-Occurrence Matrix Structural MRI Texture Analysis for Alzheimer's Disease Prediction

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
Maria João Bezerra Ferreira de Castro Oliveira

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

May 2024

Abstract

Alzheimer's disease (AD) is a progressive and irreversible neurodegenerative condition that increasingly impairs cognitive functions and daily activities. Affecting millions worldwide, AD's symptoms gradually worsen over time, starting years before they become evident with mild cognitive impairment (MCI) states. However, once symptoms manifest, the effectiveness of most therapeutic options diminishes, underscoring the urgent need for early detection methods. Given the incurable nature of AD and its profound impact on the elderly, early diagnosis and intervention are crucial, focusing on delaying disease progression and improving patients' quality of life. Thus, this work aims to develop an automatic method to detect AD in 3 different stages, namely, control (CN), MCI, and AD itself, utilizing structural magnetic resonance imaging (sMRI) images, one of the adjunct techniques to diagnosis. For such purpose, brain sMRI images from the ADNI database were pre-processed and a set of 22 texture statistical features from the gray level co-occurrence matrix (GLCM) were extracted from various slices corresponding to the different anatomical planes. Different combinations of features have been used to feed classic machine learning (cML) algorithms to analyze their discrimination power between groups. The cML algorithms achieved the following classification accuracies: 85.2% for *AD vs. CN*, 98.5% for *AD vs. MCI*, 95.1% for *CN vs. MCI*, and 87.1% for *All vs. All*. These results are particularly significant in the field of AD classification, providing a comprehensive set of metrics beyond mere accuracy. Moreover, when compared with a similar state-of-the-art sMRI study, this approach was able to outperform the classification accuracies by 10.8%, 6.9%, and 11.8% in the *AD vs. MCI*, *CN vs. MCI*, and *All vs. All* study groups, respectively.

Keywords: Alzheimer's Disease, Mild Cognitive Impairment, Magnetic Resonance Imaging, Gray Level Co-occurrence Matrix, Classic Machine Learning.

Resumo

A doença de Alzheimer (DA) é uma doença neurodegenerativa progressiva e irreversível que compromete as funções cognitivas e atividades diárias de milhões de pessoas em todo o mundo. Os sintomas da DA pioram gradualmente ao longo do tempo, começando anos antes de se tornarem evidentes em estados de défice cognitivo ligeiro (DCL). No entanto, uma vez manifestados os sintomas, a eficácia da maioria das opções terapêuticas reduz, destacando a necessidade urgente de métodos de deteção precoce. Dada a natureza incurável da DA e o seu impacto significativo nos idosos, torna-se crucial o diagnóstico e a intervenção precoces, com foco em atrasar a progressão da doença e melhorar a qualidade de vida dos doentes. Este trabalho propõe o desenvolvimento de um método automático para detetar a DA em três diferentes estágios: controlo saudável (CN), DCL e a própria DA, utilizando imagens de ressonância magnética estrutural (RMs), uma das técnicas adjuvantes ao diagnósico. Para o efeito, RMs cerebrais da base de dados ADNI foram pré-processadas e um conjunto de 22 estatísticas de textura da matriz de co-ocorrência de níveis de cinza (MCNC) foram extraídas de vários cortes correspondentes aos diferentes planos anatómicos. Diferentes combinações de características foram utilizadas para alimentar algoritmos de *Machine Learning* clássico (MLc), visando analisar a capacidade de discriminação entre grupos. Os algoritmos MLc alcançaram as seguintes precisões de classificação: 85,2% para *DA vs. CN*, 98,5% para *DA vs. DCL*, 95,1% para *CN vs. DCL* e 87,1% para *Todos vs. Todos*. Estes resultados são particularmente significativos no campo da classificação da DA, fornecendo um conjunto abrangente de métricas para além da precisão. Adicionalmente, quando comparado com um estudo semelhante, ao nível do estado da arte, esta abordagem superou as precisões de classificação em 10,8%, 6,9% e 11,8% nos grupos de estudo *DA vs. DCL*, *CN vs. DCL* e *Todos vs. Todos*, respetivamente.

Palavras-chave: Doença de Alzheimer, Défice Cognitivo Ligeiro, Ressonância Magnética, Matriz de Co-ocorrência de Níveis de Cinza, *Machine Learning* Clássico.

Agradecimentos

Primeiramente, gostaria de agradecer ao meu orientador, Professor Doutor Pedro Rodrigues, pela sua disponibilidade imediata e encorajamento, convencendo-me da minha capacidade para concluir com êxito este último desafio do meu mestrado, num tema tão interessante quanto exigente. A sua orientação foi indispensável.

Do mesmo modo, agradeço ao Engenheiro Pedro Ribeiro pelo suporte inestimável que me proporcionou durante este trabalho. A partilha contínua de conhecimento foi fundamental para superar as dificuldades encontradas.

Paralelamente, agradecer o contributo da *Alzheimer's Disease Neuroimaging Initiative* (ADNI) e afiliados, que permitiu a recolha e partilha de dados para este trabalho.

É impossível não expressar a minha profunda gratidão aos meus pais, que foram pilares fundamentais ao longo de toda a minha trajetória académica (e não só), amparando-me incondicionalmente em todos os momentos, fossem eles de alegria ou de adversidade. Apesar das vossas dúvidas quanto ao meu eterno 'estou quase a terminar a tese', é a vocês que devo estar prestes a tornar-me a terceira engenheira da família (mas a primeira mestre!).

Ao João Pedro, um desmedido obrigada por seres a pessoa que me ouve, e, mesmo sem muitas vezes entender, me apoia e encoraja incondicionalmente. És verdadeiramente o meu companheiro para todas as ocasiões. *Ride or die*.

Às amigas Bia, Sara, Catarina e Sofia, por sobrevivermos ao caos da vida e arranjarmos tempo no calendário para fofocas e desabafos.

Às amigas da ESB, feliz por todas as aventuras e desaventuras que nos uniram nesta casa.

À minha querida família - Avô Ferreira, Avó Lina, Titinha, Tio Rui, Camila e Gugu - agradeço de coração por todo o carinho e pelos almoços de domingo que tanto nos unem.

Por último, dedico este trabalho às minhas estrelinhas, à Avó Maria e, em especial, ao Avô Augusto. Embora o Alzheimer tenha levado a melhor, este é o meu contributo na esperança de que, no futuro, possamos vencer esta doença mais frequentemente.

Maria João

"As long as you are learning, you are not failing"

Bob Ross

Outline

List of Figures

List of Tables

Abbreviations

Aβ	Amyloid-beta
AD	Alzheimer's Disease
ADA	Advanced Alzheimer's Disease
ADM	Moderate Alzheimer's Disease
ADNI	Alzheimer's Disease Neuroimaging Initiative
AIBL	Australian Imaging, Biomarkers and Lifestyle
AUC	Area Under the Curve
APP	Amyloid Precursor Protein
BagT	Bagging Trees
CN	Cognitively Normal (Control)
CNN	Convolutional Neural Network
CSF	Cerebrospinal Fluid
CSI	Critical Success Index
DA	Doença de Alzheimer
DCL	Défice Cognitivo Ligeiro
DNN	Deep Neural Network
DWT2D	2D Discrete Wavelet Transform
EEG	Electroencephalogram
FDA	Food and Drug Administration
FGSVM	Feature Generating Support Vector Machine
FHS	Framingham Heart Study
FID	Free Induction Decay
fMRI	Functional Magnetic Resonance Imaging
FS	Feature SElection
GLCM	Gray Level Co-occurrence Matrix
GM	Gray Matter
H	Hydrogen
IRCCS	Centro Neurolesi "Bonino-Pulejo" of Messina, Italy

IWG	International Working Group
KNN	K-Nearest Neighbors
KW	Kruskal-Wallis
LMCI	Late Mild Cognitive Impairment
LR_SVM	Logistic Regression Support Vector Machine
MCC	Matthews Correlation Coefficient
MCI	Mild Cognitive Impairment
MRI	Magnetic Resonance Imaging
MRMR	Minimum Redundancy Maximum Relevance
NACC	National Alzheimer's Coordinating Center
NIA-AA	National Institute on Aging - Alzheimer's Association
NMDA	N-methyl-D-aspartate
OASIS	Open Access Series of Imaging Studies
PCA	Principal Component Analysis
PET	Positron Emission Tomography
QSVM	Quantum Support Vector Machine
RF	Radiofrequency
RMs	Ressonância Magnética Estrutural
ROIs	Regions of Interest
SL-RELM	Semi-supervised Learning Regularized Extreme Learning Machine
sMRI	Structural Magnetic Resonance Imaging
SNP	Single Nucleotide Polymorphism
SVM	Support Vector Machine
SubKNN	Subspace K-Nearest Neighbors

CHAPTER 1

Introduction

1.1 Motivation

Alzheimer's disease (AD) stands as the primary cause of dementia, contributing to 60–80% of dementia cases [1]. Its direct impact on cognitive abilities and neurocognitive function presents significant challenges, particularly in populations with increasing life expectancies [2].

Despite clinical measures being available for diagnosing AD progression, the process is notably time-intensive. Furthermore, expert evaluation is essential, yet even for professionals, accurately assessing the disease's development can be challenging until symptoms become overt. Consequently, early detection becomes a focal point for the scientific community. Early diagnosis not only alerts to the necessity of precautionary measures but ultimately potentially mitigates the disease's non-curable adverse effects on the patient's daily life in the foreseeable future [2].

All of the facts stated above prove that it is valuable to consider new ways of supporting early and accurate diagnosis. In that sense, this work aims to contribute to an automated diagnostic tool for detecting AD in MCI and dementia (AD) stages using sMRI texture features, thus aiming to support medical doctors and clinicians in their decisions.

1.2 Purpose and Contributions

This dissertation aims to devise an artificial intelligence model for the automatic detection of AD and MCI stages, leveraging structural MRI imaging. To achieve this, various structural images from different diagnostic groups - AD (Alzheimer's Disease), MCI (Mild Cognitive Impairment) and CN (Control) - were processed and 22 texture statistical features were computed from the Gray Level Co-occurrence Matrix (GLCM) of each image. These features were then fed into Machine Learning

algorithms for group discrimination. The studied groups comprised *CN vs. AD*, *CN vs. MCI*, *AD vs. MCI* and *All vs. All*. The classification metrics obtained from the proposed model were found to be satisfactory compared to state-of-the-art results, achieving accuracies from 85.2% to 98.5%, thus offering new insights in the field.

1.3 Thesis Structure

This work is organized into seven chapters: Chapter One introduces the dissertation's motivation, purpose, and key contributions; Chapter Two delves into Alzheimer's disease, offering comprehensive insights into its nature and diagnostic methods; the Third Chapter explores the impact of Alzheimer's disease on structural MRI, elucidating the basics of this imaging modality; Chapter Four provides an overview of Machine Learning tools; outlining the proposed methodology and how it was employed comes in Chapter Five; Chapter Six presents the results and their analysis; Chapter Seven concludes the dissertation and outlines potential future directions.

CHAPTER 2

Alzheimer's Disease

Representing one of the most expensive, lethal, and burdening diseases of this century, AD is a progressive neurodegenerative disease characterized by memory loss and multiple cognitive impairments (memory, attention, concentration, speech, and thinking, among others). This deterioration leads to alterations to an individual's behaviour, personality, and functional capacity, impeding their ability to perform activities of daily living [3].

In this chapter, an exploration of Alzheimer's Disease will be conducted, characterizing the condition and the resultant alterations. Furthermore, the risk factors, current diagnostic approaches and treatment modalities will be scrutinized.

2.1 Epidemiology

The World Health Organization (WHO) estimates that more than 55 million people are living with dementia, a number that could reach 78 million by 2030 and nearly triple to 139 million by 2050 [4]. Within this scenario, Alzheimer's disease assumes a prominent position, accounting for approximately 60% to 80% of all instances of dementia [1].

Given that the predominant risk factor for dementia is advancing age, the ongoing rise in life expectancy and demographic aging further amplifies the probability of individuals developing this condition [5]. When looking at Europe's numbers, more than one-fifth of the EU population was aged 65 or over in 2020, which represents an increase of 3% compared to 2010. Life expectancy is also increasing in Europe, rising from an average of 77.7 years for someone born in 2002, to 81.3 years for someone born in 2019 [5]. Concurrently, studies in Europe confirm that the overall prevalence of dementia increases with age, as shown in Table 2.1. Furthermore, it is more prevalent among women, in part because of their increased longevity, persisting as a condition that disproportionately affects

the female population. In Europe, it was reported in 2019 that 6,650,228 women live with dementia, in contrast to 3,130,449 men [5].

On a global scale, dementia (all causes) ranks as the fifth leading cause of mortality, accounting for 4.4% of all deaths in 2016. Furthermore, it is estimated that the number of people with dementia in Europe will almost double by 2050 [5]. AD dementia-related deaths have concomitantly risen, having more than doubled from 1994 (41,255 deaths) to 2013 (86,822 deaths), among patients aged ≥ 50 years. Adding to that information, the median survival for individuals with AD dementia is roughly estimated to be around 7 years from the onset of cognitive decline [6].

These are clear indicators that this dementia disorder will pose huge challenges to public health and elderly care systems, not only in Europe but across the world [7].

Table 2.1: Dementia prevalence rates overview for 2018, calculated by age range and presented as percentages for the overall population and segmented by gender [5].

Age range	**Overall prevalence**	**Prevalence in Females**	**Prevalence in Males**
60-64	0.6%	0.9%	0.2%
65-69	1.3%	1.5%	1.1%
70-74	3.3%	3.4%	3.1%
75-79	8%	8.9%	7%
80-84	12.1%	13.1%	10.7%
85-89	21.9%	24.9%	16.3%
90+	40.8%	44.8%	29.7%

2.2 Pathophysiology

The progressive cognitive decline in AD is characterized by two underlying pathological hallmarks: the abnormal accumulation of amyloid-beta (Aβ) and tau proteins (Figure 2.1) [1, 8]. Aβ is derived from the sequential cleavage of amyloid precursor protein (APP) by beta-secretase and gamma-secretase. In the Alzheimer's brain, the extracellular aggregation of Aβ forms senile plaques, which are toxic to the neurons [8]. These plaques and smaller accumulations of Aβ may damage neurons by interfering with neuron-to-neuron communication at synapses [9]. Plaque deposition typically occurs $\sim$20 years before the onset of cognitive impairment [1]. Tau, on the other hand, is derived from alternative splicing from the microtubule-associated protein tau (MAPT) gene to form soluble protein isoforms [8]. However, in AD, aberrant chemical alterations induce the detachment of tau protein from microtubules, leading it to adhere to other tau molecules. This process results in the formation of threads, which subsequently aggregate to create neurofibrillary tangles (NFTs), blocking the neuron's transport system and harming the synaptic communication between neurons [10]. NFTs can be detected 10–15 years before the onset of AD symptoms [1].

While the precise sequence of events remains uncertain, it is suggested that the accumulation of Aβ may precede the manifestation of abnormal tau [9]. Moreover, heightened Aβ accumulation in the most affected area of the brain, the medial temporal lobe and neocortical structures, is correlated with subsequent increases in tau, as when Aβ reaches a tipping point, there is a rapid spread of tau throughout the brain [9–11].

Furthermore, additional cerebral alterations linked to AD are inflammatory responses, neuron atrophy and reduction in brain volume, which comes as a consequence of cellular loss [9]. The onset of chronic inflammation may ensue when microglia, immune system cells within the brain, encounter challenges in effectively clearing the accumulating substances, such as Aβ and tau, as microglia undertake the task of eliminating these toxic proteins along with the widespread debris resulting from deceased and deteriorating cells [9]. Moreover, the normal metabolic processes of glucose, the primary fuel for the brain, are compromised, further impairing regular brain function. The decline in glucose metabolism may initiate up to 18 years before the anticipated onset of expected symptoms [9].

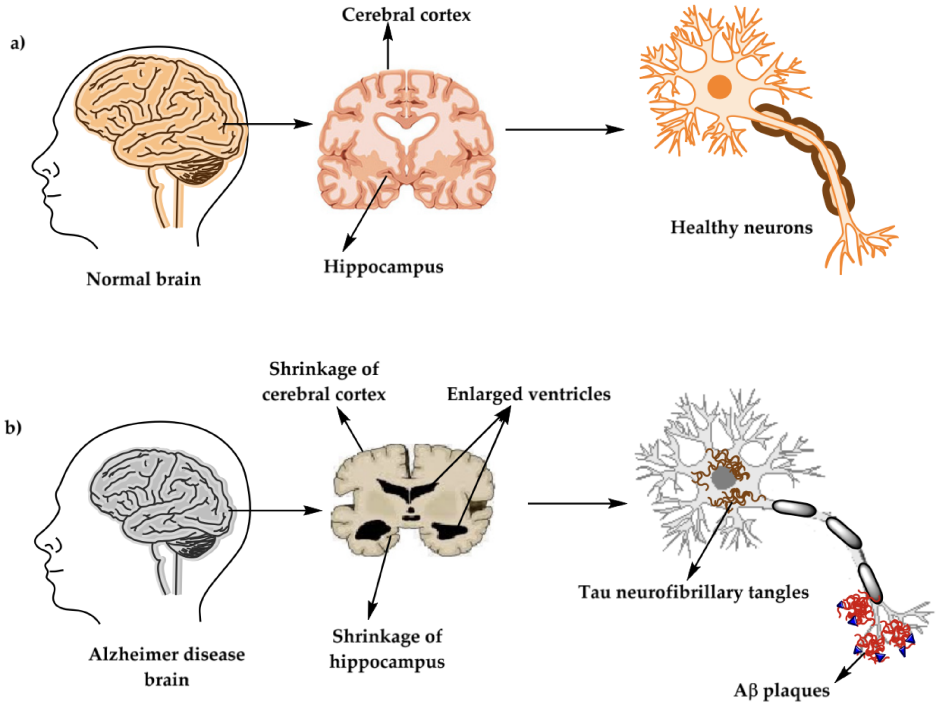

Figure 2.1: Brain and neuron's physiological structure in (a) healthy brain and (b) AD brain [11].

2.3 Alzheimer's Disease Continuum

AD follows a progressive disease sequence that extends from an asymptomatic phase with biomarker evidence of AD, through minor cognitive and/or neurobehavioral changes to, ultimately, AD dementia [1]. This delineates the AD continuum, characterized by three broad phases: preclinical

Alzheimer's disease, mild cognitive impairment (MCI) due to AD and dementia due to AD, also called Alzheimer's dementia [9]. Nonetheless, as evidenced in Figure 2.2, diverse perspectives exist in the examination of this continuum, and while the delineation of each stage may differ, all encompass the presence or absence of pathological Aβ and NFTs, as well as deficits in cognition, function, and behavioral changes leading to AD dementia [1].

During the preclinical phase, individuals exhibit signs of AD pathology without apparent cognitive or functional deterioration, comprising a long asymptomatic phase where day-to-day routines remain unperturbed [1]. When the early changes of AD occur, the brain can employ compensatory mechanisms, allowing individuals to maintain normal functioning and even not progressing to actual MCI or AD dementia states [9].

In turn, MCI identifies individuals who do not have dementia, but who do have biomarker evidence of Alzheimer's brain changes plus new but subtle deficits in cognition. Difficulties in memory, language, and thought processes arise when the brain reaches a point where it can no longer offset the detrimental impact and loss of neurons [6, 9]. Due to divergent coping mechanisms and differing levels of cognitive reserve among individuals, a wide range of variations in patients' experiences and symptomatology exist. Nonetheless, it is generally observed that, during this stage, patients maintain a relatively independent status, notwithstanding potential minor deficits in function [1]. Although additional cognitive decline may slow down or even revert, MCI due to AD is an obligatory phase for those who end up developing AD dementia. Within the subset of individuals experiencing MCI, around 15% undergo progression to dementia within two years. Additionally, approximately one-third of those with MCI develop dementia associated with AD within five years [9]. Some studies also refer to Mild Behavioral Impairment (* in Figure 2.2), a construct that describes the emergence of sustained and impactful neuropsychiatric symptoms that may surface in patients $\geq$ 50 years occurring in advance of or in concert with mild cognitive impairment. These, viewed as "non-cognitive" symptoms, can include impairments of mood, anxiety, drive, perception, sleep, and appetite, as well as behavioral disturbances such as agitation or aggression [1, 12].

The progression to AD dementia means patients will encounter significant cognitive impairments that hinder social functioning, necessitating support with daily activities. These symptoms indicate the extent of neuronal damage across various regions of the brain. AD dementia stage can be phased into mild, moderate and severe, depending on the magnitude of the disease's manifestation and the resulting decline in the patient's level of independence. In the mild phase of AD dementia, patients exhibit mild symptoms, including challenges in daily functioning, such as difficulty concentrating and remembering, disorientation in both time and place, alterations in mood, and the onset of depressive signs [11]. The moderate stage of AD dementia marks a progression of the disease to cerebral cortex regions, leading to heightened memory loss, difficulty in recognizing family and friends, impaired impulse control, and challenges in reading, writing, and speaking [11]. The severe stage of AD dementia entails the widespread dissemination of the disease throughout the entire cortex, accompanied by a significant accumulation of senile plaques and NFTs. This stage results in

progressive functional and cognitive deterioration, rendering patients unable to recognize their own families. Individuals may become bed-bound, experiencing difficulties in swallowing and urination. Ultimately, these complications may lead to the patient's demise [11].

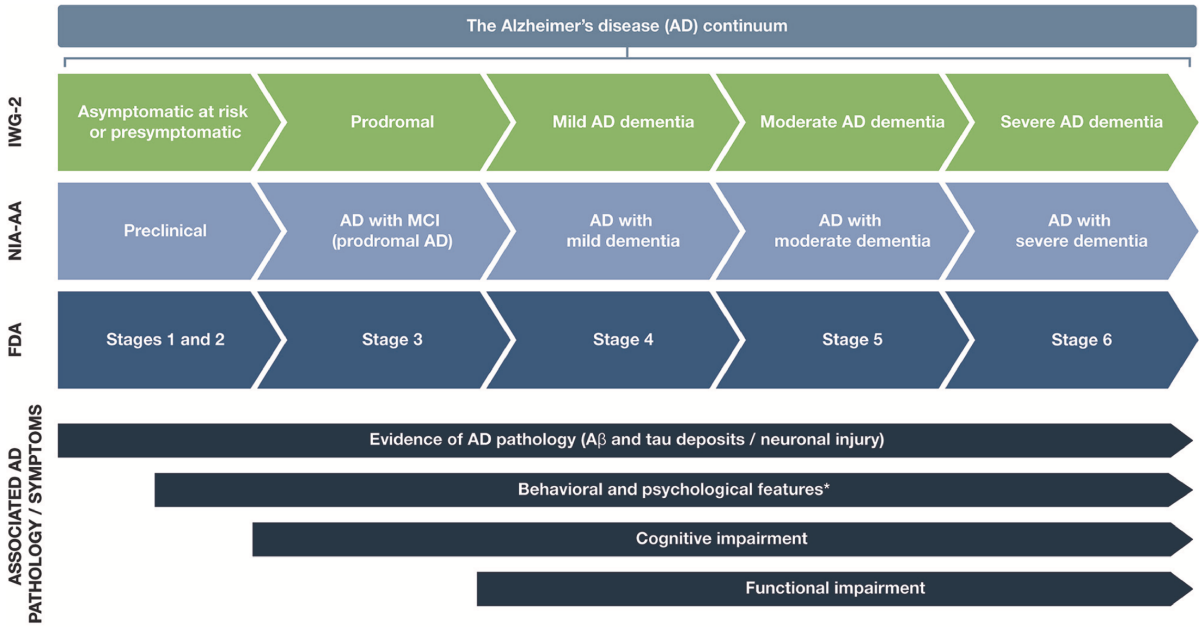

Figure 2.2: Stages within the AD continuum, providing a summary of the different naming conventions that are used within the AD community and the symptoms associated with each stage of the continuum [1, 12].

2.4 Risk Factors

The predominant demographic for individuals developing AD comprises those aged 65 or older, indicating that the ageing-related biological processes may be implicated in the pathogenesis of the disease [7, 9]. Given that age encapsulates the cumulative influence of various risk and mitigating factors throughout an individual's life, including the complex interactions between genetic predisposition, psychological components, biological factors, and environmental exposures, experts suggest that Alzheimer's, like other common chronic illnesses, is the result of multiple factors coming together rather than a single cause. Being elderly does not automatically result in AD dementia [7, 9].

The most relevant conditions associated with senescence and simultaneously with the occurrence of AD dementia are reductions in brain volume and weight, synaptic loss, ventricular enlargement, deposition of senile plaques, and formation of NFTs [11]. All of these are consequences of the complex and irreversible process that happens alongside ageing, which inevitably affects multiple organs and cell systems. Nevertheless, these changes also manifest in typical aging, posing challenges in distinguishing early AD instances [11].

Characterizing the genetic risk factors in AD is a current major objective of research as 70% of the AD cases were related to genetic factors [11]. This strong genetic component opens the door to new prognostic/diagnostic markers and new therapeutic targets. For instance, in 2022, researchers identified 31 novel genes that seem to influence biological processes known to be involved in Alzheimer's disease [13].

One of the genes showing a strong impact on AD dementia propensity is apolipoprotein E (APOE) [9]. The APOE gene plays a crucial role in synthesising the glycoprotein apolipoprotein E, responsible for facilitating the transport of cholesterol and other lipids within the bloodstream [11, 14]. Each individual inherits one of three forms (alleles) of the APOE gene — e2, e3, or e4 — from each parent, leading to six potential APOE pairs: e2/e2, e2/e3, e2/e4, e3/e3, e3/e4, and e4/e4 [9]. The presence of APOE-e4 elevates the susceptibility to AD and correlates with an earlier manifestation of the condition [14]. Those inheriting one copy of the e4 form face approximately three times the risk of developing AD dementia compared to those with two copies of the e3 form. Moreover, those possessing the e4 form are more prone to Aβ accumulation and the onset at a younger age than individuals with the e2 or e3 forms of the APOE gene [9].

The role of familial background is also noteworthy in influencing the likelihood of developing AD, as an individual with a parent affected by dementia faces an increased risk, irrespective of established genetic risk factors like APOE-e4 [9].

While age, genetic predisposition, and familial background remain immutable factors, certain risk elements influenced by general lifestyle, wellness choices, and effective management of other health conditions, can be altered or attenuated to mitigate the likelihood of cognitive decline and dementia [14]. Maintaining brain stimulation, a nutritious balanced diet, healthy habits, active social life and physical activity can effectively prevent or delay the onset of AD dementia [9]. All of these cautions contribute to an overall healthy ageing process, avoiding medical conditions such as cardiovascular disease, obesity, diabetes, and others, which are all associated with increased risk of AD [11].

2.5 Diagnosis

The first steps into diagnosis take action when a patient presents symptoms consistent with early stages of AD. In clinical practice, a two-stage process is often employed. This involves an initial "triage", normally conducted through primary care, where it is important to review clinical background, family history and exclude possible reversible causes of cognitive impairment, such as depression, or vitamin, hormone, and electrolyte deficiencies by routine-like examinations, namely blood analyses [1]. Furthermore, escorts/caregivers have a decisive role in outlining concerns and recent changes in behaviours. One of the frequently reported manifestations involves memory-related symptoms. Therefore, clinical evaluations are performed to gauge the existence of cognitive and functional impairments. A cognitive evaluation that can be efficiently conducted (within 10 minutes)

is the Mini-Mental State Examination (MMSE), perhaps the most common for AD screening [1]. Questions focus on current location and date, subtracting numbers, spelling backwards, recalling unrelated words and reproducing diagrams, all aimed at assessing different cognitive domains affected by AD [1, 15]. MMSE scores range in a scale from 0 to 30, with lower scores meaning more severe impairments.

As for the second stage, once an initial assessment is undergone and the presence of cognitive impairment is confirmed, specialists employ supplementary clinical assessments to ascertain the underlying causes of impairment and validate the diagnosis, namely brain imaging tests [1].

Presently, the predominant imaging modalities of significance revolve around the identification of either neurodegeneration or amyloid deposition.

- Structural MRI (sMRI), which is considered safe and non-invasive, furnishes valuable clinical insights in the examination of factors contributing to cognitive impairment: 1) assess brain volume reductions and other structural changes; 2) identify vascular compromised brain areas; 3) exclude other pathologies non AD-related [1, 16].
- Functional MRI (fMRI), provides understanding regarding the functional integrity of cerebral networks by using signals to measure the synaptic activity of neurons [16].
- FDG-PET, given that the brain predominantly relies on glucose as its primary energy substrate. Fluorodeoxyglucose, a glucose analogue, is employed in conjunction with Positron Emission Tomography (PET) to discern regional patterns of diminished brain metabolism and neurodegeneration, thus serving as potential indicators of AD [1].
- Amyloid-PET, as it utilizes tracers that selectively bind to identify Aβ and highlight regions of senile plaque deposition, one of the first manifestations in AD [1].

Other approaches may include the analysis of Cerebrospinal Fluid (CSF) or even the use of Electroencephalogram (EEG).

- CSF biomarkers present a viable complement or substitute for amyloid PET [1]. Despite being a more invasive procedure, the extraction of CSF through lumbar puncture is a cost-effective and widely accessible method when compared to expensive imaging technology equipments [16]. CSF serves as a reflective medium for molecular processes within the brain, rendering it a suitable avenue for identifying potential biomarkers linked to AD. The principal CSF biomarkers associated with AD pathophysiology include CSF Abeta42, total tau (t-tau), and phosphorylated tau (p-tau). These biomarkers are widely acknowledged to possess superior diagnostic accuracy, particularly in the early stages of AD diagnosis, even in situations where noticeable clinical symptoms are not apparent [16].
- EEG measures the electrical signal generated by brain neurons, thus having the capacity to evaluate brain activity. Three key consequences of AD on EEG signals can be identified: signal

deceleration, a decrease in EEG complexity, and a shift in the usual state of EEG synchrony [17]. Furthermore, this analysis contributes to a less sophisticated and invasive screening [18].

2.6 Treatment Options

AD treatment should be hybrid, conciliating a non-pharmacological approach with pharmacological intervention [19].

Adjusting to the diagnosis can be challenging for both the patient and the family. Therefore, the adaptation process should be supported primarily. In cases of an early diagnosis, preserving or enhancing cognitive function and sustaining the individual's capacity to engage in daily activities is imperative, being this the fundamental of non-pharmacologic therapy. Dietary modifications, engaging in physical exercise, and cognitive stimulation have shown modest yet noteworthy enhancements in cognitive function, contributing to life quality maintenance [1].

When stepping into mild or severe symptomatology, pharmacologic treatments become necessary to ameliorate cognitive and functional decline extensively affecting the patient's routines. Despite the widespread prevalence of this condition and its exponential emergence as a public health concern, not many categories of medications have received approval for the treatment of AD. Frequent current prescriptions include:

- Cholinesterase inhibitors (donepezil, rivastigmine, and galantamine). As AD is associated with a decline in acetylcholine levels, this pharmaceutical enhances cholinergic pathways through the inhibition of acetylcholinesterase, providing moderate and temporary alleviation of symptomatic decline in cognitive function [20].

- N-methyl-D-aspartate (NMDA) receptor antagonist (memantine). Glutamate neurotransmission, particularly associated with the functionality of NMDA in cortical and hippocampal brain regions, contributes significantly to the pathogenesis of AD, fostering cell death and synaptic dysfunction. NMDAR antagonist prevents overactivation of NMDAR glutamate receptor, consequently restoring its normal activity. Once again, this shows relative and transient symptomatic benefit [11, 20].

More recently, in the United States, the first monoclonal antibody anti-Aβ was approved by the FDA, although some concerns arise around its benefits and price-point [20].

2.7 Relevant State-of-the-art Review

Understanding the research conducted on Alzheimer's Disease classification is crucial for the focus of this work. To facilitate this understanding, two comprehensive overview tables have been compiled. These tables contain critical information about relevant studies in this field, highlighting the accuracy

Table 2.2: sMRI state-of-the-art studies overview.

Study	Accuracy	Summary
[2]	MCI vs. AD, CN and LMCI - 99.5% LMCI vs. AD, CN and MCI - 98.2% CN vs. AD, MCI and LMCI - 98.3% AD vs. CN, MCI and LMCI - 99.7%	This research explores the segmentation and classification of AD using MRI through transfer learning and customization of a CNN. The focus is on utilizing images segmented by GM of the brain. Instead of training a model from scratch, the pre-trained deep learning model is used as a foundation, followed by transfer learning. The accuracy of the model is tested across varying numbers of epochs (10, 25, and 50), resulting in an overall accuracy of 97.84%.
[21]	CN vs. AD - 93.3% CN vs. MCI - 87.7% MCI vs. AD - 88.2% All vs. All - 75.3%	This study suggested to automate AD detection across three stages (control, MCI and AD) using sMRI. Texture statistical measures were extracted from sMRI images via a two-level discrete wavelet transform. The most discriminative measures were chosen as features for classical machine learning algorithms and a CNN. Results showed that cML outperformed CNN in discrimination, with the proposed method achieving a 4% increase in discrimination accuracy compared to current state-of-the-art methods for All vs. All comparison.
[22]	CN vs. AD - 95.0% CN vs. MCI - 85.6% AD vs. MCI - 87.5%	This study introduces a lightweight neural network for AD detection using structural MRI data. The network, based on ShuffleNet V1 architecture, efficiently handles limited MRI data and device resources. It incorporates Efficient Channel Attention for improved feature representation in disease-related brain regions. Model optimization employs cross entropy and triplet loss functions to align predicted probabilities with ground-truth labels. The model offers competitive performance compared to other lightweight methods and its computational efficiency enables quick and accurate processing of large datasets, making it suitable for devices with limited computing capabilities and advancing AD detection.
[23]	CN vs. AD - 94.3% CN vs. MCI - 92.9% AD vs. MCI - 92.1%	This study introduces "PKG-Net," a new method for accurately predicting AD from complex input features. It uses a hierarchical pseudo-3D CNN network with a kernel attention mechanism and a global context block. The network extracts multi-scale features from pre-processed images and transforms MRI data into compact high-level features hierarchically. A joint loss function enhances generalization performance by generating more distinguishable features. The method is tested with various architectures and hyper-parameters, showing superior performance on the ADNI dataset compared to existing methods.
[24]	CN vs. AD - 87.5% CN vs. MCI - 79.2% AD vs. MCI - 83.3%	This study introduces a dense neural network tailored for AD binary classification. It examines the influence of different activation functions on prediction accuracy, employing 5-fold cross-validation to compare results. Data from the ADNI database, processed with FreeSurfer software, trains the model to classify AD and CN groups. Results show the proposed DNN surpasses traditional methods and previous studies in accuracy for various classifications.

attained and providing summaries of each study. Table 2.2 focuses on studies that utilized sMRI, while Table 2.3 covers research employing other diagnostic signals, such as EEG or fMRI.

Table 2.3: Non sMRI state-of-the-art studies overview.

Study	Accuracy	Summary
[25]	CN vs. AD - 97.0% CN vs. MCI - 95.0% AD vs. MCI - 83.0% All vs. All - 75.0%	The goal of this study was to differentiate between AD, MCI, and healthy controls. This method involves computing power intensity for high and low-frequency bands and using their absolute differences for classification via supervised machine learning. EEG data from 105 subjects (48 AD, 37 MCI, and 20 CN) were analyzed. The findings outperform previous studies, validating the method's effectiveness. Additionally, its efficiency suggests potential for low-cost real-time diagnosis on embedded devices.
[26]	CN vs. EMCI - 85.6% CN vs. LMCI - 92.90% CN vs. AD - 96.8% EMCI vs. LMCI - 83.3% LMCI vs. AD - 84.9% EMCI vs.˙ AD - 89.5%	This study assesses a classification framework merging graph theory and machine learning in fMRI to distinguish AD, early and late MCI and healthy controls (96 subjects). A new multi-feature selection method, incorporating a dual graph theoretical approach, is introduced. Using the joint human connectome project multimodal parcellation of 180 areas per hemisphere, brain areas are selected through graph-theory analyses. Optimal features are identified via the minimal redundancy maximal relevance method based on support vector machine linear from graph measures for 36 out of 360 areas. The findings suggest that combining graph measures and machine learning could aid AD diagnosis, especially utilizing local measures for better capturing cognitive impairment-related functional changes in local brain regions.
[27]	CN vs. MCI/AD - 95.0%	This study explores dementia detection using blood gene expression, single nucleotide polymorphisms, and clinical data from ADNI. Analyzing 623 participants, a Support Vector Machine classifier with Mutual Information (MI) feature selection achieved high accuracy when trained on all three data types. Gene expression and SNP data alone also showed good performance, but significantly lower. SHapley Additive exPlanations (SHAP) identified potential biomarkers, including those related to axon myelination and synaptic vesicle formation, offering promise for early AD detection. This approach integrates machine learning and SHAP, providing insights for disease understanding and non-invasive diagnosis.
[28]	CN vs. AD - 92.9% CN vs. MCI - 78.0% AD vs. MCI - 87.2% All vs. All- 90.5%	This study explored the use of Amyloid PET scans as a biomarker for AD classification using a 3-Dimensional Ensemble Net. It compared three feature extraction methods (3D Subject, 3D Slice, and 3D Patch Extraction) and three slicing algorithms (Subset Slice, Uniform Slice, and Interpolation Zoom) to analyze PET scans. It found that the 3D Patch extraction approach was most accurate, followed by the Subject-based approach and the 3D Slice approach. Accuracy varied based on patch size, with larger patches achieving the highest accuracy. Uniform slice extraction and interpolation zoom techniques outperformed subset slice selection in the 3D-Slice-based approach. This study highlights the effectiveness of the Ensemble Net model and the 3D patch-based feature extraction approach in Alzheimer's disease classification.
[29]	CN vs. AD - 85.2%	The objective of this study was to identify AD by analyzing non-linear characteristics of speech signals. Signal sub-bands, derived from wavelet transform, were utilized for feature extraction, with their descriptive statistics serving as input for various classifiers. Logistic regression classifiers achieve detection accuracies of 100%, 77.8%, and 85.2% for women, men, and the overall population, respectively.

CHAPTER 3

Magnetic Resonance Imaging

Over recent decades, imaging has played various roles in AD diagnosis and prognosis, with MRI being among them as a detailed, non-invasive technique. Although initially used mostly to rule out other causes of cognitive impairment, MRI has demonstrated a positive predictive value for AD now new imaging analysis modalities surface [30].

This chapter will cover the fundamentals of MRI and an exploration of how AD is perceived in structural MRI clinical images.

3.1 MRI physics

MRI operates on the principle of the interaction between an externally applied magnetic field and a nucleus possessing a nuclear magnetic moment or spin [31]. Therefore, nuclei deemed suitable for MRI are characterized by the presence of either an unpaired proton or neutron. This suitability arises from the association of spin with electrical charge, resulting in the generation of a magnetic field within nuclei containing impaired nucleons. Consequently, these nuclei exhibit magnetic properties akin to a magnetic dipole [32]. The human body's natural magnetic properties are harnessed to produce the images, since its most abundant components, water and fat, have unpaired protons as one of their constituents (1H). Thus, the hydrogen proton has the best MRI sensitivity, also supported by the advantage of a high gyromagnetic ratio (γ), given as the ratio of its magnetic dipole moment (μ) to its angular momentum (L), as reflected in equation (3.1). For hydrogen, (γ) = 42.58 MHz/T [31, 33].

$$\gamma = \frac{\mu}{L} \tag{3.1}$$

In the absence of a magnetic field, hydrogen atom nuclei exhibit random orientations, resulting in a net magnetization of zero (Figure 3.1) [33]. Nevertheless, when an entity is positioned within a robust external magnetic field (B_0), the nuclei display a propensity to align themselves in the direction of that same field, generating a net magnetisation (M), parallel to B_0, which is subsequently manipulated to produce images [31, 33]. When oriented away from the magnetic field, the spins undergo precession around the (B_0) axis (typically designated as the z-axis) at a frequency directly proportional to the strength of the external magnetic field [33]. The frequency in question is referred to as the Larmor frequency, and its calculation is expressed by the Larmor equation, Equation (3.2):

$$w_0 = \gamma B_0 \tag{3.2}$$

where w_0 is the Larmor frequency (MHz), B_0 the magnetic field strength (T) and (γ) is the gyromagnetic ratio (MHz/T) [31].

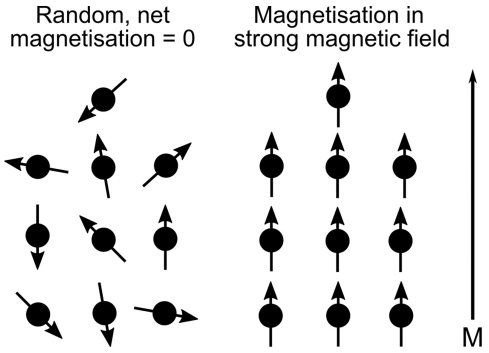

Figure 3.1: As illustrated on the left, each proton exhibits rotational motion around a randomly oriented axis in the absence of a magnetic force. Conversely, in the presence of a strong magnetic field, as depicted on the right, protons are inclined to align with the direction of the magnetic field. The optimal alignment state is the configuration demanding minimal energy that is parallel to the magnetic field B_0. Consequently, the majority of associated spins also tend to align in that direction, along the z-axis. This alignment gives rise to the measurable yet minute magnetization (M), utilized in MRI [33].

3.2 MRI Image Generation

When subjecting the system to a time-dependent magnetic field perpendicular to the static magnetic field B_0, achieved through the application of a radiofrequency (RF) pulse at the Larmor frequency, the supplied energy induces excitation. In the tissue, protons with a Larmor frequency absorb energy, causing them to deviate or reorient from the direction induced by the imaging magnet. When ceasing the RF pulse application, the relaxation process takes place with the return of nuclei to their original spin state at a lower energy (equilibrium). The energy loss is detected as a signal,

referred to as the free induction decay (FID). To convert the FID signal into a meaningful image, the mathematical Fourier Transform is applied [32, 33].

Different tissues have distinct relaxation times, leading to variations in the signal intensity and contrast in the final MRI images. Two-time constants are involved in relaxation, T1 and T2. These relaxation times are employed to characterize the duration required for the majority of hydrogen protons to revert to their equilibrium states, contributing to the image contrast in MRI [33]. The manipulation of image contrast among different tissues within the body is achieved by adjusting the transmission rate of the RF pulse [32]. Images displaying T1 contrast are referred to as T1-weighted images (T1w), useful for highlighting anatomy. They exhibit brightly tissues having short T1, such as fat, due to a brief recovery and consequently increased signal emission, while fluid-filled structures, brain ventricles and spinal canal containing CSF, will appear dark. Those depicting T2 contrast are denoted as T2-weighted images (T2w). Abnormal tissues appear bright in T2w due to exhibiting an extended T2 when compared to healthy tissues. Thus, these images are commonly used used to target visualization of inflammation, edema, and certain lesions [32, 34].

3.2.1 Spatial Information

One distinguishing feature of the magnetic resonance technique is its ability to generate three-dimensional images. This is achieved through the exclusive excitation of hydrogen atoms oscillating at equal frequencies, induced by RF pulses [35].

In addition to quantifying it, identifying the source location of the signal within the patient undergoing the procedure is crucial for imaging. To obtain this spatial information, the magnetic field is linearly varied using magnetic field gradients in three dimensions, activated only during specific, short intervals at various stages of the scanning procedure. These gradients partition the examined body into voxels, which are volume elements approximately one cubic millimetre in size. Each voxel produces a distinct signal with a unique amplitude, ultimately contributing to the construction of an image with a spatial resolution equal to the volume of each voxel. The presentation of the image is in a three-dimensional format, offering the capability for rotation and manipulation. However, the conventional preference leans towards examining a series of two-dimensional images within a selected anatomical plane (Figure 3.2) [35].

When conducting imaging and reaching the midpoint of the magnetic gradient running along the body in a longitudinal direction (by convention), a brief RF pulse is directed into the patient, only encompassing a limited range of closely positioned frequencies centered around the target frequency. The RF pulse stimulates protons within a narrow cross-section of the body, in this case, an axial slice, as the slice orientation is determined by the gradient direction. Adjusting to a shallower gradient or a broader frequency band leads to the generation of a thicker slice, consequently yielding lower resolution [35]. The derivation of information about the two additional dimensions relies on the implementation of newer perpendicular gradients. This application renders signals distinctive in both

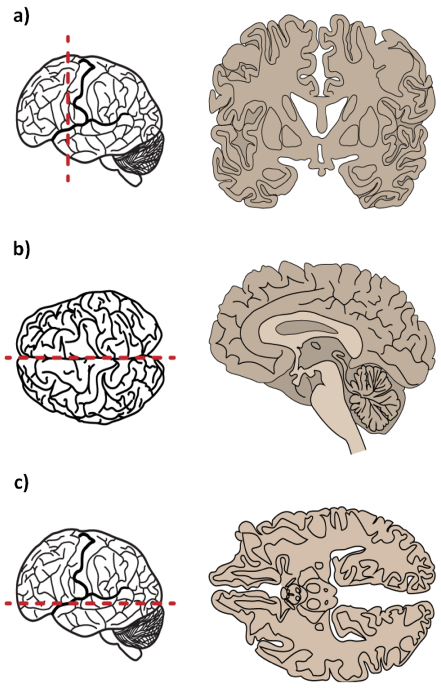

Figure 3.2: Anatomical views of the human brain: a) Coronal plane b) Sagittal plane c) Axial plane [36].

phase and frequency, leading to the manifestation of a unique frequency-phase combination for each voxel, contributing to the creation of detailed and spatially resolved images [35].

3.3 Alzheimer's Disease in Structural MRI

It is possible to look at the pathological advancement of AD from two viewpoints: the microscopic neurobiological level, where the gradual accumulation of Aβ and/or tau results in ongoing damage to neurons, and the macroscopic level, in which these pathological alterations lead to the atrophy of specific brain structures [37]. Structural MRI serves as a pivotal tool for mitigating the expenses and high patient burden associated with obtaining information about "micro" AD biomarkers, despite their high sensitivity. sMRI directs its focus towards the physical modifications occurring in brain regions affected by the neuropathology of AD, such as atrophy and changes in tissue characteristics [30, 37].

The primary utility of sMRI lies in its ability to visualize cerebral atrophy, a well-established characteristic feature of neurodegeneration - dendritic and neuronal losses - induced by pathological changes related to AD [30, 37]. First manifestations occur in the medial temporal lobe, specifically in the entorhinal cortex (part of the parahippocampal cortex), and spread from there to structures like hippocampus, amygdala, and parahippocampus [30, 37]. This is consistent with initial memory impairments and progressive language, praxis, visuospatial and behavioral deficits [38]. Hippocampal

degeneration is correlated with increased amyloid and tau, leading to a smaller hippocampal volume (HV) in AD patients showing declined cognitive function. Thus, HV is a recommended biomarker, indicating neuronal damage, which supports the diagnosis of AD in the presence of clinical symptoms. Furthermore, periodically evaluating HV and hippocampal atrophy improves the ability to predict the progression of AD in preclinical individuals [37]. Elevated rates of hippocampal loss are discernible in AD and MCI patients as opposed to healthy controls (reduced by 15–30% when comparing MCI to controls) [38]. The accelerated atrophy of the hippocampus also serves as a reliable predictor for the increased likelihood of conversion from MCI to AD [37].

AD and MCI states are also associated with increasing gray matter (GM) loss, as depicted in Figure 3.3. This GM loss is of interest due to its correlations with cognitive dysfunction and neuropsychiatric symptomatology [39]. Additionally, sMRI scans can reveal white matter hyperintensities (WMHs), indicative of demyelination and axonal loss. Patients with AD typically exhibit more pronounced WMHs, primarily localized in the frontal lobe, compared to control subjects [39].

Despite the absence of molecular specificity, sMRI is regarded as secure and effective, furnishing valuable measures of brain atrophy. These metrics encapsulate the accrued neuronal damage, which, in turn, is a direct determinant of the clinical state [37].

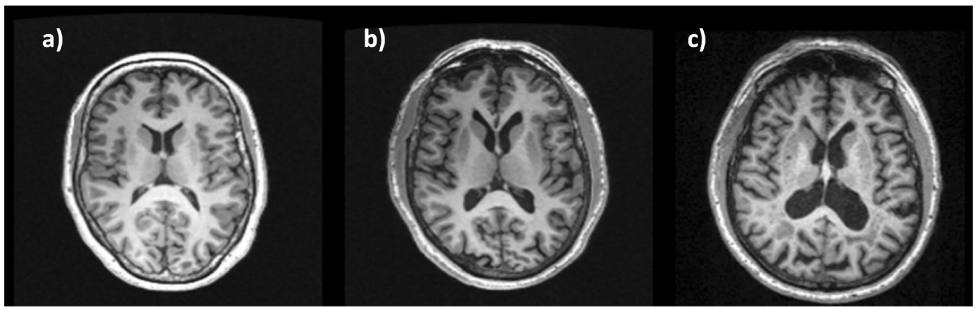

Figure 3.3: T1-weighted sMRI imaging showing decreased GM volume in an c) AD patient compared to a) healthy control and intermediate GM decline in b) patient with MCI [39].

CHAPTER 4

Machine Learning

Machine learning (ML), a branch of artificial intelligence (AI) and computer science, concentrates on leveraging data and algorithms to simulate the learning process observed in humans. Through iterative improvements, these algorithms enhance their accuracy over time. As a result, they become adept at intelligently predicting or detecting patterns within datasets. Alongside the current exponential proliferation of data, ML stands as a fundamental catalyst for technological progress [40, 41].

This chapter's main goal is to elucidate ML methodologies, providing a more in-depth exploration of classifier algorithms and examining some of the commonly employed conventional models.

4.1 Machine Learning Methods

Growing attention is being directed towards the integration of artificial intelligence into the realm of medical imaging, and ML is playing a fulcrum role.

ML applications can be broadly classified into three distinct approaches based on the nature of the data: supervised, unsupervised, and semi-supervised [42, 43].

Supervised Learning uses labelled datasets to train algorithms enabling accuracy in data classification or outcome prediction. Thus, the model acquires knowledge through exposure to a set of input data, each already paired with its correct output. Supervised learning is extensively utilized in image classification tasks due to its focus on establishing associations between features about the learning target and the desired outcome measures in the dataset. The primary objective is to make informed classification decisions based on these identified connections [40–42].

Unsupervised Learning involves machine learning algorithms analyzing and grouping unlabeled datasets into subsets, commonly referred to as clusters. In this scenario, the learning system only has access to input data, and the model autonomously trains itself to uncover latent patterns or groupings within the data [40, 41, 43].

Semi-supervised is a conjunction of both methods referred above, where a portion of the data contains partial labelling, and the labelled subset is employed to extrapolate information for the unlabeled segment. With this, in the training phase, a reduced labelled dataset is employed to facilitate the classification and feature extraction process from a more extensive, unlabeled dataset. This approach is particularly beneficial when facing challenges related to limited labelled data, a common issue in supervised learning algorithms [40, 43].

Furthermore, Reinforcement Learning represents a unique approach within machine learning, distinctly different from both Supervised and Snsupervised Learning. Unlike those methods, it does not rely on data-predefined labels or pattern discovery. Instead, Reinforcement Learning involves an autonomous agent that learns to perform tasks through trial and error, operating within a system of rewards and penalties. This goal-directed learning process allows the agent to make decisions based on interacting with its environment and observing the consequences of its actions, thereby adapting to achieve specific objectives. This approach has been given attention due to its effective applications in personalized medicine [44, 45].

4.2 Classifiers

Classifier algorithms, frequently applied in the realm of medical imaging, are typically trained using a Supervised Learning approach. In this approach, the model attempts to predict the correct discrete label associated with a given input dataset [41].

The supervisory role is assigned to a dataset in which each data point x_i, for instance, an image or signal, is paired with an associated outcome y_i, indicating the specific classification among K possibilities to which x_i belongs. The data is fractionated into training data, utilized to determine the parameters $\hat{\theta}$ that yield model results $\hat{y}_i$ closest to the actual outcomes y_i, and test data, mainly utilized for assessing the efficacy of the model [41].

In the case of a simple yet recurrent example of clinical machine learning which can perform binary classification where $K = 1$, such as determining whether a patient has a disease or not, the standard output of a classifier model is a single probability value, denoted as $\hat{y}_i$. This value represents the probability that the input x_i belongs to the positive class (class 1, the patient has the disease), and $1 - \hat{y}$ represents the probability that the input belongs to the negative class (class 0, the patient does not have the disease). In the training data, the actual outcome y_i is a binary value [41, 46]:

- $y_i = 1$ if x_i belongs to class 1 (patient has the disease).
- $y_i = 0$ if x_i belongs to class 0 (patient does not have the disease).

While binary classification is designed to distinguish between two outcomes (class 0 or class 1), multi-class classification extends this concept to more than two classes. In a multi-class setting, K can represent several categories, such as 2, 3, ..., depending on the specific class correctly predicted.

4.2.1 Multi-Class Classification Algorithms

There are different types of classification tasks in ML, depending on the training data labeling. Binary classification refers to the classification tasks having two class labels such as "true and false" or "yes and no". For classification tasks having more than two class labels it is applied a multi-class classification where the goal is to predict to which class a given input belongs [46].

In this work's context, the class labels are CN, MCI and AD, differentiating the clinical diagnosis within the brain sMRI images used.

Many binary classification algorithms can extend their usage to multi-class classification tasks, often requiring binary transformations to accommodate this transition. Strategies like the *One vs. One* approach are employed to adapt inherent binary classification algorithms for multi-class classification. In the *One vs. One* strategy, the training involves creating a classifier for each unique pair of labels. In the context of a 3-class classification, this results in three pairs of labels, and consequently, three classifiers are trained. Whereas in a *One vs. Rest* (also known as *One vs. All*) approach, each label is considered as an independent label and the rest combined as a single label. [47].

In this work, multi-class classification is conducted using a *One vs. One* strategy, which involves creating three distinct labels: *CN vs. AD*, *AD vs. MCI*, and *CN vs. MCI*. Additionally, instead of adopting a *One vs. All* approach, this study employs an *All vs. All* strategy, where all three conditions are combined into a single label allowing for a comprehensive comparison across all groups.

4.3 Conventional Machine Learning Algorithms

Conventional (or classic) machine learning (cML) algorithms rely on learning from a training set to develop a trained model. Some of the popular supervised cML algorithms, which can contribute to classifying processes, are reviewed here:

4.3.1 Support Vector Machines

Support Vector Machines (SVMs) represent a collection of supervised learning techniques used for tasks such as classification, regression, and the identification of outliers. SVMs operate by discerning target classes and its primary aim is to ascertain the optimal decision boundary, characterized by the widest margin, to classify new data entries. The simplest classifier is the favored choice, even if there may be more than one decision boundary in an n-dimensional space. This optimal boundary is represented as a hyperplane in SVM, and its size depends on the properties of the dataset. The objective is to establish hyperplanes that maximize margins, serving as buffers to define the proximity between data points. An essential element of SVMs is the kernel, a mathematical function used to map the original input data points into high-dimensional feature spaces. It is responsible for calculating distances between data points, assigning higher values when data points are in closer proximity [48, 49].

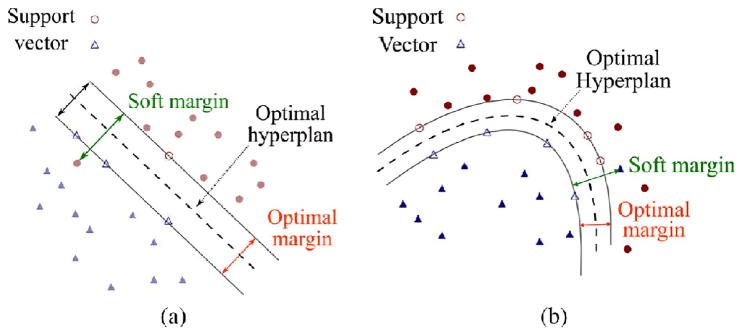

Figure 4.1: Visual representations of a) Linear and (b) Nonlinear SVM [49].

SVMs can be classified into two distinct types: Linear and Nonlinear (Figure 4.1). A Linear SVM, which utilizes a linear kernel, is used primarily for data that can be divided into different classes or categories by a single straight line, meaning that the dataset is structured in such a way that a linear decision boundary is sufficient to classify the data points into their respective categories. Conversely, when data cannot be distinctly classified into unique categories using a simple linear boundary it is

Table 4.1: Different SVM kernel functions [50, 51].

Kernel	Formula	Description
Linear	$K(x,x_i) = x \cdot x_i$	The linear kernel function is appropriate for cases where samples within a low-dimensional space exhibit clear separability. When employed in an SVM, the linear kernel establishes a decision boundary represented by a linear hyperplane, effectively distinguishing between various classes within the feature space.
Polynomial	$K(x,x_i) = [\gamma * (x \cdot x_i) + coef]^d$	In SVMs, Polynomial kernels assess the similarity of training samples in a feature space using polynomials of the original variables. These kernels consider both the individual features of input samples and their combinations, known as interaction features. The parameter d denotes the dimensionality of the kernel function, with the dimension of the mapping function increasing alongside it.
RBF	$K(x,x_i) = exp(-\gamma * \|x - x_i\|^2)$	The Radial Basis Function kernel is essentially a similarity function that computes the Euclidean distance - used to find the distance between two points on a plane - between two feature vectors, employing a parameter σ that sets the "spread" of the kernel, influencing how data similarity is measured. In the context of SVMs, this kernel is utilized to form the Gaussian RBF, which is the most commonly used kernel for classifying training data.
Sigmoid	$K(x,x_i) = tanh(\gamma(x \cdot x_i) + coef)$	The Sigmoid kernel function can be viewed as analogous to a two-layer neural network within the framework of kernel functions. It is frequently utilized as an activation function in artificial neural networks. However, its application as a kernel in SVMs is less favoured, and its effectiveness needs empirical validation.

described as nonlinear. This type of data requires a Nonlinear SVM, where classification is achieved by mapping the data to a higher-dimensional space rather than relying solely on two-dimensional representations [52]. In these scenarios, where the data points are not linearly separable in the original input space, kernel functions such as the polynomial, sigmoid, and Gaussian kernels, facilitate the placement of the hyperplane by transforming the data into a format where separation is feasible [53]. Table 4.1 illustrates four common types of kernel functions.

4.3.2 Logistic Regression

The Logistic Regression Algorithm, as a supervised learning technique, is utilized to predict specific variables based on a set of independent variables. As a discriminative classifier, it operates by deriving a set of weighted features from the input and combining them linearly. This linear combination involves multiplying each feature by a respective weight and summing them up, ultimately predicting the probability of an event occurrence through the fitting of data to a logistic function. Logistic Regression applies to numerous predictor variables, both numerical and categorical types, although most useful in cases where the outcome is predicted in a binary way (true or false) [54, 55]. The logistic function equation is expressed as follows (Equation 4.1):

$$p = \frac{1}{1 + e^{-(b_0 + b_1 x + b_2 x^2)}} \tag{4.1}$$

Here, e represents Euler's number, approximately 2.71828, b_0 is the intercept, and b_1 and b_2 are the coefficients corresponding to the linear and quadratic components of the predictor variable x. These coefficients are determined using the statistical technique of maximum likelihood estimation, applied to training data [55].

4.3.3 Decision Trees

The Decision Tree algorithm is commonly applied in scenarios related to both classification and regression tasks. Two primary types of nodes - leaves and decision nodes - are connected by branches, forming a hierarchical tree-like structure. Each decision node strategically splits the instance space into multiple sub-spaces based on specific discrete functions of the input values. As the algorithm progresses, the tree is systematically partitioned into sub-trees, visually representing solutions to the problem. The objective is to construct a predictive model for the target variable, utilizing straightforward decision rules inferred from the features present in the dataset. The Decision Tree Classifier is a classification model designed for multi-class classification tasks. Similar to other classifiers, it requires two input arrays: array X, containing the training samples, and array Y, comprising integer values representing the class labels for the corresponding training samples. The input array X can be of varying density [56, 57].

The creation of numerous individual decision trees from the complete training dataset happens in the Extra Trees classifier. At the onset, the algorithm selects a split rule for the root node by considering a random subset of features and employing a partially random cut point. Then, a random split is chosen to partition the parent node into two random child nodes. This iterative process continues within each child node until a leaf node is reached, defined as a node without any child nodes. The consolidation of predictions from all generated trees results in the determination of the final prediction through a majority vote [58].

4.3.4 K-Nearest Neighbour

The K-Nearest Neighbours (KNN) algorithm is a supervised machine learning method designed for classifying time series data using labels of its nearest neighbors. The parameter K determines the number of neighbors considered when assigning a label to a newly introduced time sequence. Recognized for its simplicity, the KNN classifier proves effective across diverse data types, even in the presence of complex datasets or noise. Classification of the training dataset involves utilizing the distance matrix among the sequences in the training set. The model does not necessitate reconstruction when introducing new data and upon encountering an unfamiliar sequence, the label is determined based on the count of the K nearest neighbors for each class [59].

4.3.5 Ensemble Learning

Ensemble learning is a machine learning framework that uses various cML algorithms to train a collection of models which demonstrated to be weak classifiers. These models are developed using the recovered features, and combining their outcomes using multiple voting techniques produces higher performance than utilizing each strategy independently. The creation of prediction models in a variety of clinical areas, including the identification of diseases like AD, has made substantial use of ensemble approaches [60].

Common ensemble classification approaches include Bagging, AdaBoost, Random Forest, and Gradient Boosting [61].

- The Bagging method creates subsets of samples by randomly sampling from the training dataset, which are then used to individually train the basic models to incorporate into the model. The training of basic models in the Bagging framework is conducted in parallel [61].

- AdaBoost directs its attention towards misclassified samples by iteratively adjusting their weights, thereby enhancing the classification performance of basic models for the ultimate integration. In this case, the training of basic models occurs sequentially rather than concurrently [61].

- The Random Forest algorithm trains several decision trees on diverse subsets of a dataset. This technique accounts for both the dataset's and the feature dimensions, solving decision

trees' sensitivity to overfitting. This problem is alleviated by aggregating voting outcomes from many decision trees. Much like the parallelization employed in Bagging, the training process of individual base models within the Random Forest is also conducted in parallel [61].

- Gradient Boosting employs random sampling to obtain subsets of data, constructing and training each learner sequentially to minimize residuals generated by the preceding learner. As a result, the technique effectively diminishes the cumulative impact of residuals from integrated models, compelling predictions to closely align with the actual values. Similarly to AdaBoost, the foundational models in Gradient Boosting are trained sequentially [61].

CHAPTER 5

Methodology

5.1 Proposed Methodology

This proposed methodology, illustrated in Figure 5.1, is divided into 3 parts: data collection/pre-processing, feature extraction, and machine learning classification.

5.2 Dataset Description

The Alzheimer's Disease Neuroimaging Initiative (ADNI) database provided the brain sMRI datasets used in this work. Specifically, the first phase of ADNI1, which is available for download at http://adni.loni.usc.edu [62]. The ADNI1 MRI procedure used dual echo T2-weighted and T1-weighted sequences to provide consistent longitudinal structural imaging on 3T and 1.5T scanners. The average total scan time for each patient was about forty-five minutes per session. Many imaging vendors, such as Siemens, GE Healthcare, and Philips Healthcare, adopted this uniform procedure. Strict quality control methods were applied to every examination to find and eliminate images that were judged unsuitable due to subject mobility or inadequate anatomical coverage. 504 images from individual participants in the ADN1 acquisitions, representing a range of diagnostic groupings, are

Table 5.1: Database demographics overview.

Group	# of Subjects	Age Average ± SD	Gender F	Gender M	MMSE Average ± SD
CN	167	77.9 ± 5.33	82	85	29.1 ± 1.16
MCI	235	77.0 ± 7.04	77	158	25.0 ± 4.22
AD	102	77.0 ± 7.33	50	52	18.9 ± 6.12

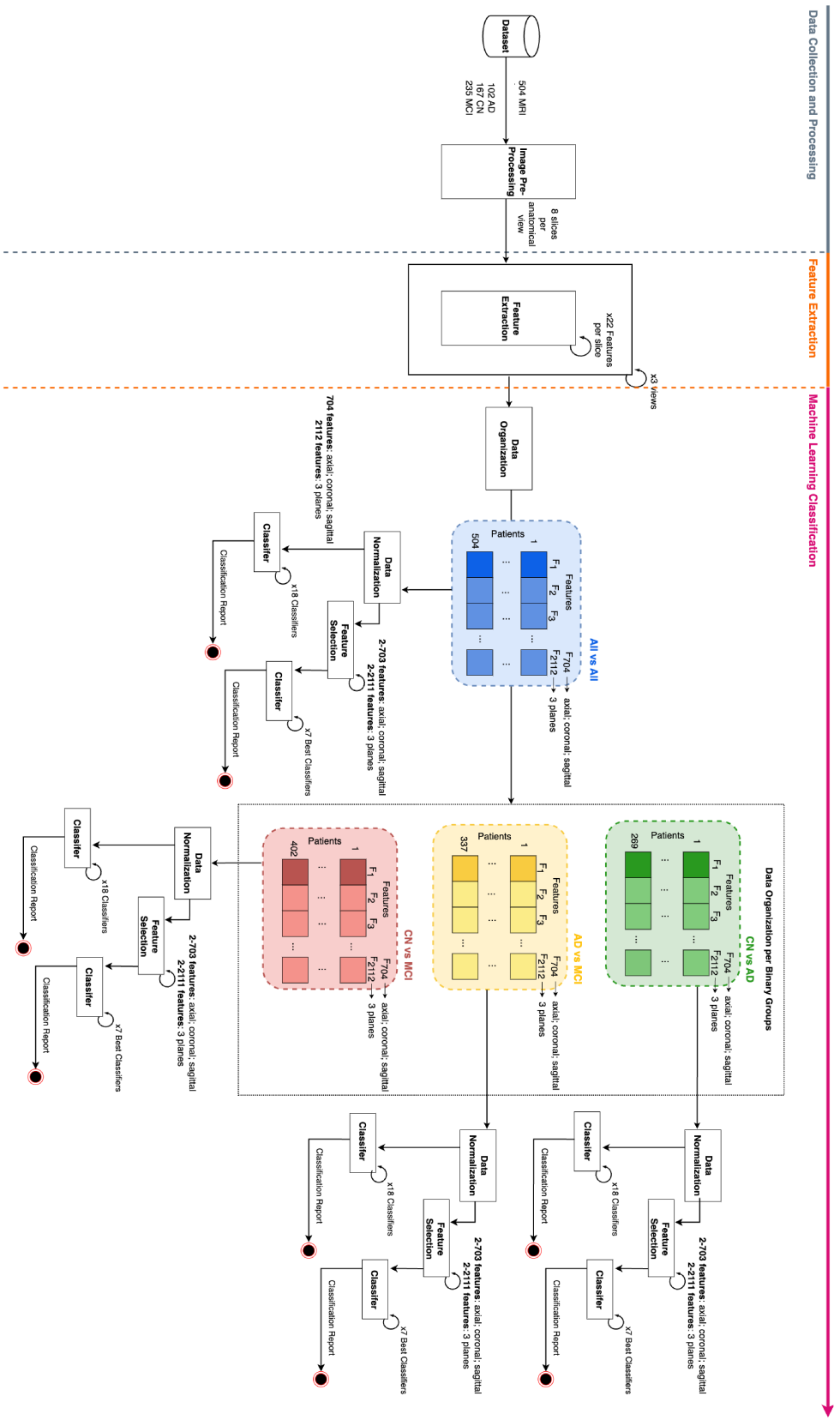

Figure 5.1: Methodology workflow diagram.

included in the dataset: 167 represent healthy controls, 235 patients with MCI and 102 patients diagnosed with AD. Details regarding the demographic characteristics of each group are presented in Table 5.1.

5.3 Image Pre-Processing

5.3.1 Data Cleaning

From the different ADN1 collections, the images were selected in order to have single images from specific subjects, overall going for the one associated with the most recent date.

One of the objectives of ADNI1 included evaluating structural MRI metrics at both 1.5T and 3T to assess whether the strength of the magnetic field significantly influenced quantitative measures critical for AD. The researchers concluded that the performance of both field strengths was similar, with neither presenting significant disadvantages distinct to one or the other [63]. Thus, 3T and 1.5T scans were both used to maximize the number of single subjects' images, as the differences in the image quality between 1.5T and 3T MRI are minimal. This approach enabled the expansion of the study groups' sizes for the present investigation.

5.3.2 Skull-Stripping

Although some image corrections were already provided by ADNI, such as correction of image geometry distortion or intensity non-uniformity, the selected images were downloaded to Matlab® 2023b software where a set of structural pre-processing steps were performed with the help of SPM12 (Statistical Parametric Mapping) package (freely available online at https://www.fil.ion.ucl.ac.uk/spm/), which is designed for the analysis of brain imaging data sequences. These structural steps were applied according to [64, 65] and included segmentation, skull-stripping and normalization. Segmentation is important for the differentiation of tissue types and the calculation of the deformation matrix, which dictates the location of the result on a certain brain coordinates template. Extracting the skull region is a crucial step in brain segmentation assignments for clinical analysis, and it is known as skull stripping. Effective skull stripping becomes essential because of the complex anatomical structure of the brain and the existence of intensity variations in MRIs. The precision and efficiency of this process are of paramount importance for accurate diagnostics [66]. The normalization step at this point was important to normalize the image into a standard space, defining the boundaries around the brain, from a set origin. An example of the resulting image from the structural steps is portrayed in Figure 5.2.

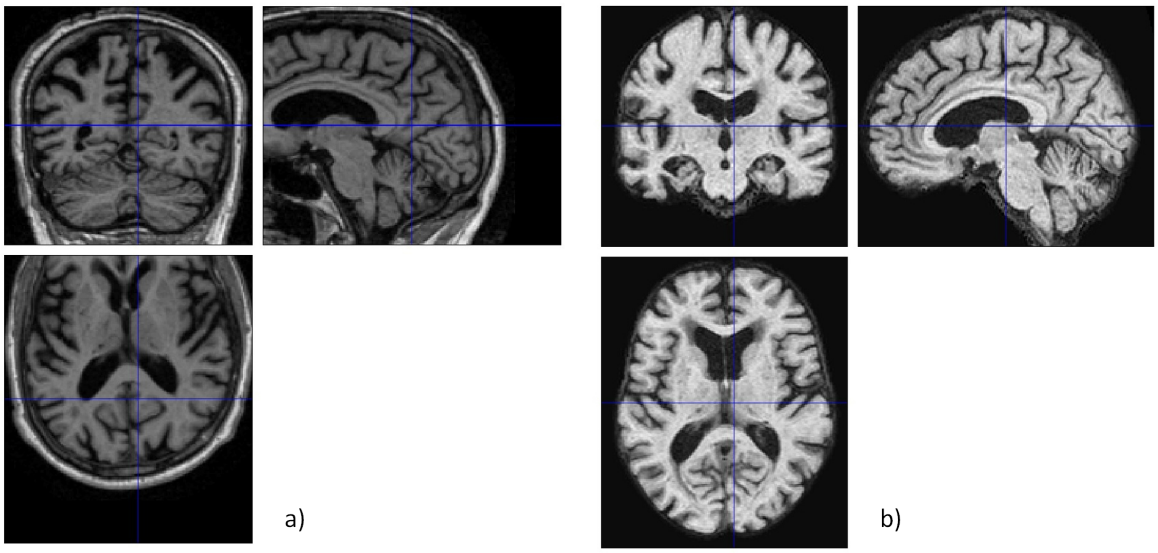

Figure 5.2: Skull stripping process in SPM: a) Original Image b) Processed Image.

5.3.3 Slicing and Final Pre-Processing Steps

The resulting images underwent decomposition into 2D slices representing three distinct anatomical planes: coronal, sagittal, and axial. Employing a Matlab® 2023b routine, eight slices were generated from each image for every anatomical plane. This specific number of slices was chosen due to computational constraints, as it balances detailed anatomical representation with manageable data volume and processing time. Additionally, it is important to note that these slices were consistently extracted from identical positions across all images to ensure uniformity in the data analysis.

Subsequently, all acquired 2D slices were converted from the initial Neuroimaging Informatics Technology Initiative (NIfTI) format to Portable Network Graphics (PNG) and subjected to transformation. The transformation involved resizing the input images to dimensions of 156×156. This resizing is particularly significant in reducing the computational time during the feature extraction phase. Moreover, the use of square images streamlines the subsequent analysis. Following the image resizing, a normalization process was implemented for each slice. Normalization scales the pixel values in the image from a predefined range (0 to 255) to a standardized range between 0 and 1 [67].

5.4 Feature Extraction

Feature extraction refers to the process of converting input data into a set of features. Its primary objective is to capture the essential information from the original dataset and represent it in a reduced dimensionality space [68].

For each subject, eight images per anatomical view, resulting in a total of 24 images per subject, and a cumulative total of 12096 images for the entire dataset consisting of 504 study participants, were employed for feature extraction. From each image, 22 texture features were computed. For that, the Gray Level Co-Occurrence Matrix (GLCM) method was used.

5.4.1 Texture Analysis via GLCM

Texture analysis (TA) facilitates the mathematical identification of alterations in MRI signals that may not be perceptible at the level of individual image pixels. Consequently, it furnishes a quantitative and replicable approach for extracting image characteristics. The term "image texture" typically pertains to the spatial fluctuations observed in pixel intensity levels within a tissue. In AD, variations in image intensity attributed to the accumulation of Aβ and NFTs, along with their consequences, may manifest as specific textural patterns [42]. Predominant techniques in TA revolve around examining the spatial correlation or concurrent presence of pixel intensity values. One example is the GLCM method, widely recognized for its effectiveness in extracting texture features from medical images [42, 69]. Thus, the GLCM, a statistical method employed for texture analysis, considers the spatial arrangement of pixels within an image. It produces a square matrix sized according to the number of gray levels within the image. Each element in this matrix signifies the frequency of co-occurring gray levels, reflecting the distribution of pixel colors or grayscale values at specified offsets [70]. So, each element (i,j) represents the frequency by which the pixel with gray level i is spatially related to the pixel with gray level j. Statistical metrics are then derived from the GLCM to characterize the texture of the image.

By using Matlab® 2023b and the function *graycomatrix* (available in Image Processing Toolbox), it was possible to generate this matrix by determining the frequency of occurrences of a pixel. In this work's context, the function output consisted of multiple GLCMs, specified to the *graycomatrix* function by an array of offsets, which define pixel relationships of varying direction and distance. Specifically, the array utilized defines four offsets [0 1; -1 1; -1 0; -1 -1], representing the distance between a pixel of interest and its neighboring pixels in four distinct directions: horizontal, vertical, and two diagonals (Figure 5.3). Employing multiple directions facilitates a more sensitive characterization of texture within the image, encompassing texture patterns across various orientations [71].

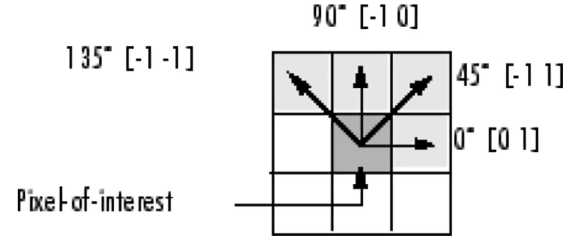

Figure 5.3: Offset array used: [0 1; -1 1; -1 0; -1 -1] [71].

Additionally, the number of gray levels used was 8, the default value for numeric images, which is the case for MRI images. This determines the discrete intensity levels considered in GLCM computation and ultimately the size of the GLCM (8x8). Moreover, this selection strikes a balance between capturing texture intricacies and managing computational demands [71].

Although the function *graycoprops* (available in Image Processing Toolbox) successfully extracts four statistics specified in properties from the GLCMs - Contrast, Correlation, Energy, and Homogeneity - which were used, a few more parameters were also computed according to [72]. Each feature output vector has four dimensions, corresponding to the four-direction offset array previously mentioned. Table 5.2 provides a comprehensive overview of all extracted features [72]. To fully understand the data presented, the specific notation used to describe matrix entries and associated calculations is provided below:

Notation [72]

$p(i,j)$	(i,j)th entry in a normalized gray-tone spatial-dependence matrix.
N_g	Number of distinct gray levels in the quantized image.
μ_x, μ_y	Mean gray level intensities of p_x and p_y, respectively.
σ_x, σ_y	Standard deviations of p_x andp_y, respectively.
$H(X)$	Entropy of p_x.
$H(Y)$	Entropy of p_y.

$p_x(i) = \sum_{j=1}^{N_g} p(i,j)$ - marginal row probabilities.

$p_y(j) = \sum_{i=1}^{N_g} p(i,j)$ - marginal column probabilities.

$p_{x+y}(k) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)$ where $i+j=k$ and $k = 2,3,\ldots,2N_g$.

$p_{x-y}(k) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)$ where $|i-j|=k$ and $k = 0,1,\ldots,N_g-1$.

$$HXY = -\sum_i\sum_j p(i,j)log(p(i,j))$$

$$HXY1 = -\sum_i\sum_j p(i,j)log\{p_x(i)p_y(j)\}$$

$$HXY2 = -\sum_i\sum_j p_x(i)p_y(j)log\{p_x(i)p_y(j)\}$$

Table 5.2: Feature Description.

Index	Feature	Formula	Description
f_1	Autocorrelation	$\sum_i \sum_j (ij) p(i,j)$	Measures magnitude of the fineness and coarseness of texture
f_2	Contrast	$\sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \mid \lvert i-j \rvert = n \right\}$	Measures the local variations between pixels
f_3	Correlation 1	$\frac{HXY - HXY1}{max\{HX, HY\}}$	Estimates the combined probability occurrence of the indicated pixel pairs
f_4	Correlation 2	$\frac{\sum_i \sum_j (ij) - \mu_x \mu_y}{\sigma_x \sigma_y}$	Estimates the combined probability occurrence of the indicated pixel pairs
f_5	Cluster Prominence	$\sum_i \sum_j (i + j - \mu_x - \mu_y)^4 p(i,j)$	Measures the skewness and asymmetry of the GLCM
f_6	Cluster Shade	$\sum_i \sum_j (i + j - \mu_x - \mu_y)^3 p(i,j)$	Measures the skewness and uniformity of the GLCM
f_7	Dissimilarity	$\sum_i \sum_j \lvert i-j \rvert \cdot p(i,j)$	Measures local intensity variation defined as the mean absolute difference between the neighbouring pairs
f_8	Energy	$\sum_i \sum_j p(i,j)^2$	Specifies the sum of squared elements in the GLCM
f_9	Entropy	$-\sum_i \sum_j p(i,j) log(p(i,j))$	Assesses the randomness of an intensity image
f_{10}	Homogeneity 1	$\sum_i \sum_j \frac{p(i,j)}{1+\lvert i-j \rvert})$	Measures the nearness of the distribution of elements in the GLCM to the GLCM diagonal

Continued on next page

Table 5.2: Feature Description. (Continued)

Index	Feature	Formula	Description
f_{11}	Homogeneity 2	$\sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j)$	Measures the nearness of the distribution of elements in the GLCM to the GLCM diagonal
f_{12}	Maximum Probability	$\max_{i,j}(p(i,j))$	Occurrences of the most predominant pair of neighboring intensity values
f_{13}	The sum of Squares: Variance	$\sum_i \sum_j (i-\mu)^2 p(i,j)$	Measures the distribution of neighboring intensity level pairs about the mean intensity level in the GLCM
f_{14}	Sum Average	$\sum_{i=2}^{2N_g} i p_{x+y}(i)$	Measures the relationship between occurrences of pairs with lower intensity values and occurrences of pairs with higher intensity values
f_{15}	Sum Variance	$\sum_{i=2}^{2N_g} (i-f_{16})^2 p_{x+y}(i)$	Measures groupings of voxels with similar gray-level values
f_{16}	Sum Entropy	$-\sum_{i=2}^{2N_g} p_{x+y}(i) log\{p_{x+y}(i)\}$	Measures neighborhood intensity value differences
f_{17}	Difference Variance	variance of p_{x-y}	Measures the heterogeneity that places higher weights on differing intensity level pairs that deviate more from the mean.
f_{18}	Difference Entropy	$\sum_{i=0}^{N_g-1} p_{x-y}(i) log\{p_{x-y}(i)\}$	Measures the randomness/variability in neighbourhood intensity value differences

Continued on next page

Table 5.2: Feature Description. (Continued)

Index	Feature	Formula	Description
f_{19}	Information Measure of Correlation 1	$\frac{HXY-HXY1}{max\{HX,HY\}}$	Assesses the correlation between the probability distributions of i and j
f_{20}	Information Measure of Correlation 2	$1-exp[-2.0(HXY2-HXY)])^{\frac{1}{2}}$	Assesses the correlation between the probability distributions of i and j
f_{21}	Inverse Difference Normalized	$\sum_{k=0}^{N_g-1} \frac{p(x-y(k))}{1+\left(\frac{k}{N_g}\right)}$	Measures the local homogeneity of an image
f_{22}	Inverse Difference Moment Normalized	$\sum_{k=0}^{N_g-1} \frac{p(x-y(k))}{1+\left(\frac{k^2}{N_g^2}\right)}$	Measures the local homogeneity of an image

With the assistance of Matlab® 2023b, all the extracted feature information was organized in matrices by subject/feature per group (AD, CN, MCI) and per anatomical view, also considering all three views (axial, coronal, sagittal and 3 planes). Furthermore, a comprehensive matrix was constructed by combining the information extracted from all three groups and again per views (axial, coronal, sagittal and 3 planes).

Subsequently, the Z-score method was applied for normalization to each matrix. In Matlab, normalization using the Z-score method is performed by column, which ensures that each feature is treated consistently during normalization.

In statistical analysis, Z-scores, also known as standard scores, are a fundamental tool for standardizing data. Z-scores measure the distance of a data point from the mean in terms of the standard deviation. This process transforms the data into a standardized scale, facilitating comparison across different datasets with varying units and distributions. For a random variable X with mean μ and standard deviation σ, the z-score of a value x is as demonstrated in Equation 5.1 [73].

$$Z\text{-score} = \frac{x-\mu}{\sigma}. \tag{5.1}$$

5.5 Classification Framework

The normalized matrices were then uploaded to Python (version 3.9.12) to execute a set of Machine Learning models and generate discrimination reports.

5.5.1 Scikit-learn ML Models

The performance of the model in discriminating between pairs of study groups (*AD vs. CN*, *AD vs. MCI*, *CN vs. MCI*, and *All vs. All*) was evaluated using eighteen selected *Scikit-learn* machine learning models, employing a hold-out method. *Scikit-learn* [74] is a popular Python library for machine learning, offering simple tools for data analysis and modelling, including various supervised learning algorithms, as used in this work. Each algorithm processed either 2112 features when utilizing data from all three anatomical planes (22 features extracted, each with ($\times$) 4 offsets from ($\times$) 8 slices ($\times$) per 3 anatomical planes, for each group comparison) or 704 features when analyzing each plane individually (22 features extracted, each with ($\times$) 4 offsets from ($\times$) 8 slices ($\times$) per 1 anatomical plane, for each group comparison).

Table 5.3: *Scikit-learn* ML classifiers configurations.

Classifier	**Hyperparameters**
GaussianProcessClassifier (GauPro)	$1.0 \times \text{RBF}(1.0)$
LinearSVC (LinSVC)	Default parameters
SGDClassifier (SGD)	max_iter : 100, tol : 0.001
KNearestNeighborsClassifier (KNN)	Default parameters
LogisticRegression (LogReg)	solver : "lbfgs"
LogisticRegressionCV (LogRegCV)	cv : 3
BaggingClassifier (BaggC)	Default parameters
ExtraTreesClassifier (ExTreeC)	n_estimators : 300
RandomForestClassifier (RF)	max_depth : 5, n_estimators : 300, max_features : 1
GaussianNB (GauNB)	Default parameters
DecisionTreeClassifier (DeTreeC)	max_depth : 5
MLPClassifier (MLP)	α : 1, max_iter : 1000
AdaBoostClassifier (AdaBoost)	Default parameters
QuadraticDiscriminantAnalysis (QuaDis)	Default parameters
OneVsRestClassifier (OvsR)	random_state : 0
LightGBMClassifier (LGBM)	Default parameters
GradientBoostingClassifier (GradBoost)	Default parameters
SGDClassifier (SGD)	max_iter : 100, tol : 0.001

In the hold-out method, the dataset is divided randomly into a *train* and *test* set. The training set is utilized for model training, while the test set evaluates the model's performance on unseen data [75]. Holdout validation is the most common approach for evaluating ML models [76]. Specifically, 80% of the data was allocated for training, and the remaining 20% was used for testing, which is a common split.

Table 5.3 provides details of each classifier used, along with corresponding hyperparameters [74]. For this purpose, certain configurations were set based on existing literature, while default hyperparameters from *Scikit-learn* were also employed [77].

5.5.2 Feature Selection

In the initial phase, the model was deployed without feature selection, utilizing all eighteen classifiers. Following an analysis of the outcomes from this phase, a subset of seven classifiers demonstrating superior accuracy was identified. These selected classifiers were subsequently employed for feature selection, utilizing the F-score method.

The F-score method stands as a simple yet robust approach to feature selection. Specifically tailored through statistical principles, it proves particularly effective in scenarios involving imbalanced data with multiple class outputs. The F-score method assesses each feature individually, ranking them based on relevance. Equation 5.2 presents the F-score formula in which f_i corresponds to a given i-th feature, n_j indicates the total instances of class j and n the total instances of all classes, μ denotes the mean feature value across all classes, μ_j the mean feature value in the jth class j, σ_j the standard deviation of class j and c the number of output classes [78].

$$f_{score}(f_i) = \frac{\sum_j \frac{n_j}{c-1}(\mu_j - \mu)^2}{\frac{1}{n-c}\sum_j (n_j - 1)\sigma^2} \tag{5.2}$$

In the model's development stages, feature selection was systematically conducted by incrementally adding one feature per iteration, ranging from 2 to 2111 features (combined planes variant) or from 2 to 703 features when each plane was analyzed separately. This process was strictly performed using the training data to prevent any data leakage between the training and testing. This approach ensured the integrity and independence of the test data were preserved, thereby maintaining the validity of the model's testing phase. The features identified and selected during the training were subsequently utilized to assess the model's performance, with the test data remaining completely separate and uninvolved in the feature selection process.

5.5.3 Classification Metrics

Given that no metric alone can fully capture all the characteristics of a model, it is important to report multiple metrics to provide a comprehensive overview of a model's performance, particularly when addressing medical diagnostics [79]. The performance evaluation of the proposed model was conducted using ten metrics: $Accuracy$, $Specificity$, $Precision$, $Recall$, $F1\text{-}Score$, AUC (Area Under the Curve), $Kappa$, MCC (Matthews Correlation Coefficient), CSI (Critical Success Index), and $G-mean$.

Accuracy represents the number of corrected classified classes concerning all cases and can be defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{5.3}$$

where, a TP, TN, FP, and FN are, respectively, the true positives, true negatives, false positives, and false negatives [77].

- True Positive (TP) – Number of instances that are correctly classified as positive. For example, the number of patients actually having AD and being correctly diagnosed as having AD would count as true positives for the AD class.
- False Positive (FP) – Number of instances that were incorrectly classified as positive. For example, patients who do not have AD but are incorrectly diagnosed as having AD would be considered false positives for the AD class.
- True Negative (TN)– Number of instances that were correctly classified as negative. For instance, patients who do not have AD and are correctly predicted as not having AD count as true negatives for the AD class.
- False Negative (FN) – Number of instances that were incorrectly classified as negative. A false negative would occur if a patient actually has AD but is incorrectly diagnosed as not having it.

Precision, also known as positive predictive value, indicates the proportion of correctly classified positive cases among all cases predicted as positive. It can be defined as [77]:

$$Precision = \frac{TP}{TP+FP} \times 100\% \tag{5.4}$$

Recall, also known as sensitivity, represents the proportion of correctly predicted positive cases among the total number of actual positive cases, being defined as [77], [79]:

$$Recall = \frac{TP}{TP+FN} \times 100\% \tag{5.5}$$

Specificity is the negative class version of the *Recall* and denotes the rate of negative samples correctly classified. Specificity ranges from 0 to 1, where 1 indicates perfect prediction of the negative class, and 0 means all negative class samples are incorrectly predicted. It is defined as [79]:

$$Specificity = \frac{TN}{FN+TN} \tag{5.6}$$

The $F1$-*Score* is the harmonic average between Recall and Precision, thereby penalizing extreme values of either [79]. The corresponding equation is defined as [77]:

$$F1\text{-}Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\% \tag{5.7}$$

$Kappa$ normalises Accuracy by the possibility of agreement by chance. It is defined as: [77]

$$Kappa = \frac{2 \times (TP \times TN - FN \times FP)}{(TP+FP) \times (FP+TN) + (TP+FN) \times (FN+TN)} \tag{5.8}$$

MCC is particularly valuable for evaluating classification performance on imbalanced datasets. Ranging between 0 and 1, MCC achieves its worst value at 0 and its best at 1. It is calculated as [77]:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}} \tag{5.9}$$

The CSI offers a nuanced assessment of a binary classification model's effectiveness by taking into account both the accurate identification of positive instances and the capability to avoid false positives. The CSI equation can be expressed as follows [77]:

$$CSI = \frac{TP}{TP+FP+FN} \tag{5.10}$$

The Geometric Man ($Gmean$) is a metric that evaluates the balance in performance across all classes. A higher Gmean value indicates a reduced risk of overfitting, which means the model has been trained in a way that is specific to the training sample, compromising its ability to generalize to new data [80]. Gmean is mathematically defined as [77]:

$$Gmean = \sqrt{Recall \times Specificity} \tag{5.11}$$

The Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve assesses a model's ability to discriminate between positive and negative classes. It achieves this by comparing the true positive rate to the false positive rate across various classification thresholds. AUC values range from 0 to 1, with a perfect classifier yielding 1 and a random classifier yielding 0.5. Utilizing AUC provides a concise measure of a model's performance, enabling straightforward comparison between models and evaluation in scenarios with class imbalances [77].

CHAPTER 6

Results and Discussion

The focus of this chapter is to present and analyze the classification results derived from binary and multi-class classifications across three distinct study groups, each representing a different stage of the disease. The classification results serve as a crucial aspect in understanding the performance and effectiveness of the model in distinguishing between different AD stages. Moreover, a comparative discussion of the results obtained with and without feature selection is made. Finally, a juxtaposition with state-of-the-art studies is reviewed to contextualize these findings within the field of AD classification.

6.1 Classification Results

For every pairing of study group and study plane, Sections 6.1.1 and 6.1.2 present the classification metrics associated with the highest *Accuracy* achieved through cML algorithms, both with and without employing feature selection.

In scenarios where feature selection was not applied (Section 6.1.1), the classifiers that produced the best results in terms of *Accuracy* were documented. A total of seven classifiers were identified, and subsequently utilized for feature selection (Section 6.1.2). Additionally, in this case, the number of features that contributed to achieving the highest classification *Accuracy* was recorded.

6.1.1 Discrimination results without Feature Selection

The discrimination results from utilizing the complete feature set are presented in Tables 6.1 through 6.4.

- *CN vs. AD*: Referring to Table 6.1, the highest $Accuracy$ of 85.2% was achieved using information from the 3 planes, obtained through the cML algorithms *LinSVC* or *OvsR*. The lowest classification $Accuracy$ was 66.7%, obtained using the axial plane.

- *AD vs. MCI*: From Table 6.2, the highest classification $Accuracy$ of 98.5% was achieved from the 3 planes using the *LogReg* classifier. The axial plane exhibited the lowest classification $Accuracy$, recording a value of 64.7%.

- *CN vs. MCI*: As seen in Table 6.3, the most notable classification outcome was 88.9%, achieved from the coronal plane using the *ExTreeC* cML algorithm. The axial plane yielded the least favourable outcome, demonstrating an $Accuracy$ of 56.8%.

- *All vs. All*: Table 6.4 shows that the most significant classification $Accuracy$ of 82.2% was achieved through the cML algorithms *LinSVC* or *OvsR*. The minimum $Accuracy$ attained was 49.5%, again from the axial plane.

Table 6.1: Classification metrics for the pair *CN vs. AD*.

CN vs. AD	**Classifier**	**Accuracy**	**Recall**	**Precision**	**Specificity**	**F1**	**AUC**	**Kappa**	**MCC**	**CSI**	**Gmean**
3 Planes	*LinSVC / OvsR*	85.2%	77.3%	85.0%	85.3%	81.0%	83.9%	68.9%	69.1%	68.0%	85.1%
Axial Plane	*DeTreeC*	66.7%	10.5%	66.7%	66.7%	18.2%	53.8%	9.5%	16.0%	10.0%	66.7%
Coronal Plane	*LinSVC*	70.4%	52.4%	64.7%	73.0%	57.9%	67.1%	35.4%	35.9%	40.7%	68.7%
Sagittal Plane	*LogRegCV*	79.6%	57.9%	78.6%	80.0%	66.7%	74.7%	52.5%	53.7%	50.0%	79.3%
Best Result	*LinSVC / OvsR*	85.2%	77.3%	85.0%	85.3%	81.0%	83.9%	68.9%	69.1%	68.0%	85.1%

Table 6.2: Classification metrics for the pair *AD vs. MCI*.

AD vs. MCI	**Classifier**	**Accuracy**	**Recall**	**Precision**	**Specificity**	**F1**	**AUC**	**Kappa**	**MCC**	**CSI**	**Gmean**
3 Planes	*LogReg*	98.5%	94.1%	100.0%	98.1%	97.0%	97.1%	96.0%	96.1%	94.1%	99.0%
Axial Plane	*BaggC*	64.7%	40.0%	40.0%	75.0%	40.0%	57.5%	15.0%	15.0%	25.0%	54.8%
Coronal Plane	*ExTC*	94.1%	73.3%	100.0%	93.0%	84.6%	86.7%	81.1%	82.6%	73.3%	96.4%
Sagittal Plane	*LinSVC / OvsR*	95.6%	95.2%	90.9%	97.8%	93.0%	95.5%	89.8%	89.9%	87.0%	94.3%
Best Result	*LogReg*	98.5%	94.1%	100.0%	98.1%	97.0%	97.1%	96.0%	96.1%	94.1%	99.0%

Table 6.3: Classification metrics for the pair *CN vs. MCI*.

CN vs. MCI	**Classifier**	**Accuracy**	**Recall**	**Precision**	**Specificity**	**F1**	**AUC**	**Kappa**	**MCC**	**CSI**	**Gmean**
3 Planes	*LinSVC + OVsR*	80.2%	74.4%	82.9%	78.3%	78.4%	80.0%	60.3%	60.6%	64.4%	80.5%
Axial Plane	*BaggC*	56.8%	40.0%	50.0%	60.4%	44.4%	54.8%	9.8%	10.0%	28.6%	54.9%
Coronal Plane	*ExTreeC*	88.9%	83.8%	91.2%	87.2%	87.3%	88.5%	77.5%	77.7%	77.5%	89.2%
Sagittal Plane	*ExTreeC*	82.7%	80.6%	75.8%	87.5%	78.1%	82.3%	63.9%	63.9%	64.1%	81.4%
Best Result	*ExTreeC*	88.9%	83.8%	91.2%	87.2%	87.3%	88.5%	77.5%	77.7%	77.5%	89.2%

Table 6.4: Classification metrics for the pair *All vs. All*.

All vs. All	Classifier	Accuracy	Recall	Precision	Specificity	F1	AUC	Kappa	MCC	CSI	Gmean
3 Planes	*LinSVC / OvsR*	82.2%	82.2%	83.0%	89.9%	81.9%	89.8%	71.7%	72.3%	70.0%	86.2%
Axial Plane	*BaggC*	49.5%	49.5%	48.4%	74.5%	48.1%	62.2%	22.0%	22.4%	32.3%	60.0%
Coronal Plane	*ExTreeC*	65.3%	65.3%	62.5%	84.2%	63.6%	86.6%	45.8%	46.1%	50.4%	71.8%
Sagittal Plane	*ExTreeC*	68.3%	68.3%	68.1%	85.6%	66.3%	86.4%	49.8%	51.0%	52.2%	76.2%
Best Result	*LinSVC / OvsR*	82.2%	82.2%	83.0%	89.9%	81.9%	89.8%	71.7%	72.3%	70.0%	86.2%

6.1.2 Discrimination results with Feature Selection

The discrimination results obtained using the F-Score method for feature selection are presented in Tables 6.5 through 6.8.

- *CN vs. AD*: Referring to Table 6.5, the highest classification $Accuracy$ of 85.2% was achieved utilizing the sagittal plane, incorporating 491 selected features, and employing the *ExTreeC* classifier. The lowest classification $Accuracy$ was 72.2% using the axial plane.
- *AD vs. MCI*: From Table 6.6, the highest classification $Accuracy$ of 98.5% was attained through the ML algorithms *LinSVC* or *OvsR*, using 3 planes and 1379 features. The lowest classification $Accuracy$ attained was 80.9% using the axial plane.
- *CN vs. MCI*: From Table 6.7, the peak $Accuracy$ value of 95.1% was obtained using 1129 features selected from the 3 planes and the *ExTreeC* classifier. The minimum $Accuracy$ reached was 67.9% from the axial plane.
- *All vs. All*: As indicated in Table 6.8, the classification presented the highest $Accuracy$, 87.1%, from the 3 planes using 1366 features selected and the *LinSVC* classifier. The axial plane demonstrated the lowest classification $Accuracy$, registering 55.4%.

Table 6.5: Classification metrics for the pair *CN vs. AD* - FS.

CN vs. AD	# of Features	Classifier	Accuracy	Recall	Precision	Specificity	F1	AUC	Kappa	MCC	CSI	Gmean
3 Planes	943	*LogRegCV*	77.8%	70.8%	77.3%	78.1%	73.9%	77.1%	54.6%	54.8%	58.6%	77.7%
Axial Plane	204	*OvsR*	72.2%	42.9%	75.0%	71.4%	54.5%	66.9%	36.6%	39.6%	37.5%	73.2%
Coronal Plane	533	*LinSVC / OvsR*	79.6%	65.2%	83.3%	77.8%	73.2%	77.8%	57.1%	58.3%	57.7%	80.5%
Sagittal Plane	491	*ExTreeC*	85.2%	71.4%	88.2%	83.8%	78.9%	82.7%	67.7%	68.6%	65.2%	86.0%
Best Result	491	*ExTreeC*	85.2%	71.4%	88.2%	83.8%	78.9%	82.7%	67.7%	68.6%	65.2%	86.0%

Table 6.6: Classification metrics for the pair *AD vs. MCI* - FS.

AD vs. MCI	# of Features	Classifier	Accuracy	Recall	Precision	Specificity	F1	AUC	Kappa	MCC	CSI	Gmean
3 Planes	1379	*LinSVC / OvsR*	98.5%	94.1%	100.0%	98.1%	97.0%	97.1%	96.0%	96.1%	94.1%	99.0%
Axial Plane	8	*BaggC*	80.9%	50.0%	69.2%	83.6%	58.1%	71.0%	46.1%	47.1%	40.9%	76.1%
Coronal Plane	645	*ExTreeC*	91.2%	80.0%	95.2%	89.4%	87.0%	88.8%	80.4%	81.1%	76.9%	92.3%
Sagittal Plane	199	*ExTreeC*	95.6%	95.8%	92.0%	97.7%	93.9%	95.6%	90.4%	90.5%	88.5%	94.8%
Best Result	1379	*LinSVC / OvsR*	98.5%	94.1%	100.0%	98.1%	97.0%	97.1%	96.0%	96.1%	94.1%	99.0%

Table 6.7: Classification metrics for the pair *CN vs. MCI* - FS.

CN vs. MCI	# of Features	Classifier	Accuracy	Recall	Precision	Specificity	F1	AUC	Kappa	MCC	CSI	Gmean
3 Planes	1129	*ExTreeC*	95.1%	93.1%	93.1%	96.2%	93.1%	94.6%	89.3%	89.3%	87.1%	94.6%
Axial Plane	5	*ExTreeC*	67.9%	58.8%	62.5%	71.4%	60.6%	66.6%	33.6%	33.6%	43.5%	66.8%
Coronal Plane	606	*ExTreeC*	86.4%	75.0%	93.1%	82.7%	83.1%	85.3%	72.0%	73.1%	71.1%	87.7%
Sagittal Plane	509	*ExTreeC*	88.9%	75.9%	91.7%	87.7%	83.0%	86.0%	74.9%	75.6%	71.0%	89.7%
Best Result	1129	*ExTreeC*	95.1%	93.1%	93.1%	96.2%	93.1%	94.6%	89.3%	89.3%	87.1%	94.6%

Table 6.8: Classification metrics for the pair All vs. All - FS.

All vs. All	# of Features	Classifier	Accuracy	Recall	Precision	Specificity	F1	AUC	Kappa	MCC	CSI	Gmean
3 Planes	1366	*LinSVC*	87.1%	87.1%	87.7%	92.7%	87.3%	94.7%	80.9%	81.0%	78.2%	90.0%
Axial Plane	694	*BaggC*	55.4%	55.4%	57.0%	75.2%	55.2%	73.8%	30.6%	31.1%	39.7%	64.7%
Coronal Plane	495	*ExTreeC*	75.2%	75.2%	73.6%	89.5%	73.4%	87.2%	59.1%	60.1%	60.7%	81.1%
Sagittal Plane	608	*ExTreeC*	71.3%	71.3%	68.6%	89.5%	67.7%	89.3%	54.4%	55.7%	56.5%	78.1%
Best Result	1366	*LinearSVC*	87.1%	87.1%	87.7%	92.7%	87.3%	94.7%	80.9%	81.0%	78.2%	90.0%

6.1.3 Results Analysis and Discussion

From the results, it is possible to infer that feature selection overall enhances classification $Accuracy$. Although this metric is frequently utilized as a metric in medical ML applications, relying exclusively on it can be deceptive when class distributions are uneven [79]. This is because high $Accuracy$ might easily be achieved by predominantly classifying all samples into the most common class, which does not necessarily reflect true model performance [79]. Despite this, enhancements were also observed in other critical metrics, such as $Recall$, $F1$, AUC or $Gmean$, which were already performing well even before feature selection. Furthermore, the metrics initially showing residual values, for instance, $Kappa$, MCC or CSI, showed significant empowerment post-feature selection. This means that, by selecting features, it is possible to capture more relevant information which in turn improves the model's ability to distinguish between classes more effectively.

- $Recall$ showed significant improvements, indicating a reduction in FN. This metric is considered crucial for medical studies because it aims to miss as few positive cases as possible, which means achieving a high $Recall$ is essential [79]. In such a context, failing to detect positive cases of AD or incorrectly classifying its stages can have severe consequences. The improved Recall ensures that fewer cases of AD are missed, thereby supporting more accurate diagnoses and appropriate interventions.

- $F1$-$Score$ also increased, reflecting a balanced enhancement in not only identifying more real positive cases (higher $Recall$) but also making fewer false positive errors (higher $Precision$). In the context of AD diagnosis, where classification errors are to be avoided, this can prove model robustness.

- An enhanced AUC, as observed, can be especially beneficial in a clinical setting as it indicates that the model correctly classifies patients into the AD, MCI, or CN categories, distinguishing

effectively between varying states of cognitive health. It also helps better evaluate the model since class imbalances exist.

- The improvement of the *Gmean* metric can signify the model's success in medical settings due to suggesting a low risk of overfitting, ensuring that the model remains robust when encountering new data in real-world scenarios, consequently preventing misclassification that could lead to inappropriate and unsafe medical decisions.

The above stated coincides with the known advantages of feature selection - identifying and retaining only the essential or relevant features for the classification task, culminating not only in better classification results but also in reducing the dimensionality of the data, improving overall performance. Irrelevant features not only fail to contribute meaningfully to classification accuracy but can also impede the performance of the entire model [81].

Focusing on feature selection results, the two result values above 90.0% correspond to the pairs *CN vs. MCI* and *AD vs. MCI*. This is of high importance because, firstly, there is no active cure for AD to this day. As MCI is considered a transitional stage between normal cognitive aging and AD, monitoring individuals diagnosed with MCI allows for early detection of those who are more likely to progress to AD, enabling interventions and treatments to be initiated earlier. Thus, distinguishing between CN and MCI, and AD and MCI, can be helpful to clinicians and caregivers in the early development of personalized treatment plans, tailored to the individual's stage of cognitive decline.

To add, the *CN vs. MCI* and *AD vs. MCI* groups show extensively good performance across all metrics reported. For the pair *CN vs. MCI*, high $Specificity$ (96.2%) demonstrates the model's exceptional capability in correctly identifying CN individuals (the negative class), which highlights efficiency in avoiding false positive diagnoses and ensuring that individuals without MCI are rarely misclassified and avoiding inappropriate medical interventions for those not affected by MCI. Furthermore, a *Gmean* of 94.6% gives the assurance that $Specificity$ is well balanced with $Sensitivity$ ($Recall$), indicating that the model effectively manages classification errors (FP and FN). Additionally, the AUC score of 94.6% signifies the model's robust ability to discriminate between the CN and MCI classes, performing well even in scenarios where class proportions aren't even. Concerning the pair *AD vs. MCI*, a $Precision$ of 100.0%, paired with a $Recall$ of 94.1%, suggests a highly reliable model that is excellent at correctly identifying MCI cases without falsely labeling AD patients as MCI. The *Gmean* of 99.0% is extremely relevant in this case, enhancing confidence that, even in the presence of class imbalances, which happens for this study pair, the model does not disproportionately favor one class over the other.

Typically, the pair *CN vs. AD* would be revealing the highest classification performance since this groups represent the most significant anatomical differences in the brain [21], also supported by the most significant differences between the MMSE scores reported in Chapter 5 (Table 5.1), which relate to the subject's level of symptomatology. Nevertheless, it's important to note that accurately distinguishing between cognitively normal individuals and those with AD presents significant

challenges due to the complex and heterogeneous nature of the condition [82]. This complexity could explain why some metrics, for instance, $Kappa$ (67.7%), MCC (68.6%) and CSI (65.2%) indicate only moderate agreement between the model's classifications and the actual classifications. Despite these challenges, it has demonstrated considerable strength in $Precision$, showing that when the model predicts AD those predictions are correct about 88.2% of the time, which is crucial to avoid the consequences of FP, such as unnecessary testing, or treatment for individuals incorrectly diagnosed with AD. Moreover, a $Gmean$ of 86.0% suggests that the model maintains a good balance between $Recall$ and $Specificity$, essential for minimizing both types of classification errors, indicating that the model performs satisfactorily well in identifying both CN and AD conditions with low error rate.

Regarding the *All vs. All* pair, although the multi-class classification proved to be challenging, the model capability is affirmed by key metrics. Alongside an $Accuracy$ of 87.1%, the model achieved an $F1\text{-}Score$ of 87.3%, reflecting a well-tuned model that not only captures most positive cases ($Recall$) but also ensures that its predictions of the positive class are accurate ($Precision$), minimizing the risk of harmful FP. Concurrently, $Specificity$ (92.7%) and $Gmean$ (90.0%) show that the model correctly identifies negative cases and provides balanced performance across various class distinctions, suggesting a reduced risk of overfitting.

Examining the outcomes across various study planes in Section 6.1.2, it is evident that the most successful results stemmed from the utilization of the brain's three-dimensional representation (using all three planes). This outcome is expectable, since combining the coronal, axial, and sagittal planes provides a holistic perspective of the brain, enabling clearer visualization of critical structures for AD detection, namely the cerebral cortex, ventricles, and hippocampus, which are among the principal regions implicated in AD pathology as referenced in Chapters 2 and 3. Incorporating data from all three anatomical planes can offer complementary features, beneficial for classification endeavours, as evidenced by research findings indicating that multi-view models tend to exhibit superior accuracy compared to single-view models [83]. This multi-dimensional perspective likely empowers the classification model to capture a more comprehensive understanding of the structural changes associated with AD, ultimately enhancing classification accuracy.

For a better understanding of the classification results obtained by the proposed model, an outlook within existing studies and respective findings are summarized in Tables 6.9, 6.10 and 6.11. These provide information for like-wise studies along the ADNI database, alternative databases and also non-imaging modalities. Once again, $Accuracy$ is used to facilitate comparison of the results with those achieved by state-of-the-art methods, being aware of its restrictions.

Focusing on findings among the ADNI database (Table 6.9), it's worth noting that not all state-of-the-art methods attempted the three binary classifications conducted in the present work. Furthermore, only four of the state-of-art methods carried out the multi-class classification *All vs. All* and in that sense, the present study did overcome the results from [21] and [89] by 11.8% and 16.1%, respectively. Although it did not outperform [84] or [28], these studies are significantly different from the one proposed here when looking at the methodology and exam modality.

Indeed, the present work did not outperform the state-of-the-art results in what concerns the pair *CN vs. AD*, being among the lowest values shown for the ADNI database. One study showed the same classification $Accuracy$ of 85.2% while others surpassed by margins ranging from 2.3% [24] to 13.8% [85]. Nevertheless, most studies did not report other classification metrics, which can be misleading. The proposed model shows an array of classification metrics that indicate correctly classified majority of instances, while effectively discriminating between classes, being further prepared to be implemented.

On the other hand, in terms of the pair *CN vs. MCI*, the proposed model outperformed majority of the sMRI-based studies [22], [21], [23], [24], [28], [87], [88] and [89] by 9.5%, 6.9%, 2.2%, 15.9%, 17.1%, 22.7%, 18.2% and 5.9%, respectively. The ones that performed better than the current work did so by a maximum of 1.4% [86].

The pair that showed the higher classification $Accuracy$, *AD vs. MCI*, successfully outperformed all sMRI-based studies cited but also the ones relative to fMRI and PET, which proves the importance

Table 6.9: Comparison with previous studies within the ADNI database.

Study	Database	Exam Modality	Features Extracted	Best Classifier	Feature Selection	Accuracy
[22]	ADNI	sMRI	Spatial Information	Lightweight Neural Network	Not Applied	CN vs. AD - 95.0% CN vs. MCI - 85.6% AD vs. MCI - 87.5%
[84]	ADNI	sMRI	Hierarchical patterns, Textures and other Structural Features	2D CNN (ResNet-50v2)	Not Applied	CN vs. AD - 91.2% All vs. All- 90.3%
[85]	ADNI	sMRI	Statistical Features	Ensemble LR_SVM	K-Best	CN vs. AD - 99.0% CN vs. MCI - 96.0% AD vs. MCI - 85.0%
[21]	ADNI	sMRI	Statistical and Textural Features	BagT FGSVM QSVM SubKNN	F-score	CN vs. AD - 93.3% CN vs. MCI - 88.2% AD vs. MCI - 87.7% All vs. All - 75.3%
[26]	ADNI	fMRI	Graph Theoretical Measures	Linear SVM	MRMR	CN vs. AD - 96.8%
[86]	ADNI	fMRI	Correlation Matrix between different ROIs	SL-RELM	Adaptive Structure Learning	CN vs. AD - 95.4% CN vs. MCI - 96.5% AD vs. MCI - 98.4%
[23]	ADNI	sMRI	Multi-scale and Hierarchical Transformed Features	CNN (PKG-Net)	Not Applied	CN vs. AD - 94.3% CN vs. MCI - 92.9% AD vs. MCI - 92.1%
[24]	ADNI	sMRI	Cortical Volumetric Features	Fully Connected Neural Network	PCA	CN vs. AD - 87.5% CN vs. MCI - 79.2% AD vs. MCI - 83.3%
[28]	ADNI	Amyloid PET	3D Slice-Based Features	3D CNN	Not Applied	CN vs. AD - 92.9% CN vs. MCI - 78.0% AD vs. MCI - 87.2% All vs. All- 90.5%
[87]	ADNI	sMRI	Spatial Features	Ensemble-based on 2D CNN	Not Applied	CN vs. AD - 90.4% CN vs. MCI - 72.4% AD vs. MCI - 77.2%
[88]	ADNI	sMRI	Cortical Surface Labels	DNN	Not Applied	CN vs. AD - 85.2% CN vs. MCI - 76.9% AD vs. MCI - 72.7%
[89]	ADNI	sMRI	Voxel-Level GM Volume Density	CNN Ensemble Model	Not Applied	CN vs. AD - 92.8% CN vs. MCI - 89.2% AD vs. MCI - 81.5% All vs. All - 71%
Proposed Model	ADNI	sMRI	Statistical and Textural Features	ExTreeC LinSVC OvsR	F-score	CN vs. AD - 85.1% CN vs. MCI - 95.1% AD vs. MCI - 98.5% All vs. All - 87.1%

Table 6.10: Comparison with previous studies with different databases.

Study	Database	Exam Modality	Features Extracted	Best Classifier	Feature Selection	Accuracy
[90]	AIBL	sMRI	Voxel-based Features	SVM	Not Applied	CN vs. AD - 88.0%
[91]	OASIS	sMRI	Model Components	12-layer CNN	Not Applied	CN vs. AD - 97.8%
[92]	FHS	sMRI	Disease Probability Maps	CNN	Not Applied	CN vs. AD - 76.6%
[92]	NACC	sMRI	Disease Probability Maps	CNN	Not Applied	CN vs. AD - 81.8%
Proposed Model	ADNI	sMRI	Statistical and Textural Features	ExTreeC LinSVC OvsR	F-score	CN vs. AD - 85.1% CN vs. MCI - 95.1% AD vs. MCI - 98.5% All vs. All - 87.1%

Table 6.11: Comparison with previous studies using non-imaging modalities.

Study	Database	Exam Modality	Features Extracted	Best Classifier	Feature Selection	Accuracy
[25]	IRCCS	EEG	Frequency Domain Features	KNN	Information Gain Filter	CN vs. AD - 97.0% CN vs. MCI - 95.0% AD vs. MCI - 83.0% All vs. All - 75.0%
[93]	University Hospital Center of São João	EEG	Cepstral and Lacstral Distances	ANN	KW test	CN vs. ADM - 96.0% CN vs. MCI - 98.1% ADM-ADA vs. MCI - 93.9% All vs. All - 95.6%
[94]	University Hospital Center of São João	EEG	Relative Power, Spectral Ratios, Maxima, Minima and Zero Crossing Distances	ANN	KW test	CN vs. AD - 95.0% CN vs. MCI - 77.0% AD vs. MCI - 83.0% All vs. All - 90.0%
[27]	ADNI	Biomarkers	SNPs, gene and clinical data	SVM	Mutual Information	CN vs. MCI/AD - 95.0%
[29]	DementiaBank	Speech Signal	Statistical and Non-Linear Parameters	LogReg	Not Applied	CN vs. AD - 85.2%
Proposed Model	ADNI	sMRI	Statistical and Textural Features	ExTreeC LinSVC OvsR	F-score	CN vs. AD - 85.2% CN vs. MCI - 95.1% AD vs. MCI - 98.5% All vs. All - 87.1%

of the present work's findings. The results were higher with differences ranging from 0.10% [86] to 25.8% [88]. Even though the classes are imbalanced, due to a high number of MCI patients when compared with AD, the $Accuracy$ attained is accompanied by a $Gmean$ value of 99.0% in this pair, which shows it the model almost perfectly reliable when distinguishing between individuals with AD and those with MCI with minimal misclassifications. Additionally, such a high $Gmean$ suggests that the model generalizes well, avoiding overfitting to the training data. As already referred before, since MCI is considered a precursor to AD, this is a very important finding, opening doors to a more efficient early diagnosis in real-world settings.

Observing sMRI-based studies among other databases (Table 6.10), where only *CN vs. AD* accuracies were available, the proposed work outperformed the FHS and NACC [7] results obtained by CNNs with an 8.6% and 3.4% difference, respectively. Regarding the AIBL [90] and OASIS [91] databases, the proposed model underperformed by 2.8% and 12.6%, respectively. Nevertheless, it is relevant to note that overall these studies were conducted with smaller datasets. To add, the fact that the images' database employed is distinct, although relevant, adds a layer of difficulty and uncertainty when conducting comparisons.

When analysing other exam modalities utilized in the diagnosis of AD (Table 6.11), such as EEG, which utilizes the electrical signals produced by the brain's activity, the proposed model was able to

surpass the accuracies regarding the pair *AD vs. MCI* by 15.5% [25], [94] and 4.6% [93]. In the case of the *CN vs. MCI* pair, the model surpassed the accuracies reported by 18.1% [94] and by 0.1% [25], although it underperfomed by 3.0% compared to [93]. When considering *All vs. All*, the model outperformed [25] by 12.1%, but was outperformed by [93] and [94] by 8.5% and 2.9%, respectively. Again, it was unable to outperform all the *CN vs. AD* accuracies, being surpassed by 11.9% [25], 10.9% [93] and 9.9% [94]. Regarding a study conducted for biomarkers with the SVM algorithm, it provided a classification for *CN vs. MCI/AD* of 95.0% and thus similar to the proposed model for the pair *CN vs. MCI* but 9.8% higher than the proposed for the pair *CN vs. AD*. A recent study analysed speech signals to detect AD [29] and an $Accuracy$ of 85.2% was obtained for the pair *CN vs. AD*, which corresponded to the same value reached by the present work.

Although enriching, all these comparisons need to be carefully analyzed because different signals, image databases, feature selection methods and classifying means were employed in the studies.

CHAPTER 7

Conclusions and Future Perspectives

7.1 Conclusions

Alzheimer's Disease, as stated throughout this dissertation, poses significant challenges due to its incurability and elusive diagnostic process. Early detection is of great interest, allowing patients to delay the onset of debilitating symptoms that erode their quality of life and shorten their lifespans. This work aimed to introduce an artificial intelligence model that can detect AD on the stages CN, MCI, and AD itself using sMRI. By analyzing pre-processed brain images, 22 features computed from the GLCM were extracted. Initially, ML algorithms were trained using the complete feature set, followed by a refinement step where only the most discriminative features were utilized, thereby incorporating a feature selection process. The findings demonstrate that feature selection enables the improvement model performance, resulting not only in heightened accuracies but overall enhancement of other relevant classification metrics that support more reliable and effective decision-making in practical settings. Consequently, the algorithms achieved the following classification accuracies: 85.1% for *CN vs. AD*, 98.5% for *AD vs. MCI*, 95.1% for *CN vs. MCI*, and 87.1% for *All vs. All*. For the pair *AD vs. MCI*, the proposed model outperformed all of the cited state-of-the-art sMRI-based studies within the ADNI database. Considering that most studies did not present results for *All vs. All* classification, the proposed model added that variant and performed well. It successfully surpassed the results of a similar model proposed by [21], which also utilized statistical and textural features alongside feature selection with F-score and cML algorithms for classification. However, the referenced model operated on a significantly smaller dataset, it used a leave-one-out cross-validation procedure and the analysis was done under a Two-Dimensional Discrete Wavelet Transform (DWT2D) decomposition rather. In the classification matchups *AD vs. MCI*, *CN vs. MCI*, and *All vs. All*, the current work demonstrated superior performance, improving accuracy by 10.8%, 6.9%, and 11.8%

respectively. Although not presenting the best results for the pair *CN vs. AD* and underperforming when compared with the previously referenced work, this model's methodology is a step ahead closer to being ready for implementation in the market by being conducted using a hold-out validation method, splitting the dataset into two subsets: one for training the model and another for testing its performance, as referred in Chapter 5. The rationale behind this approach is related to accounting for the dataset's considerable size and simulating how the model would perform on unseen data, thus providing a more realistic assessment of its generalization capabilities. Furthermore, the hold-out method offers simplicity and computational efficiency when compared to other methods such as cross-validation [75]. Additionally, providing a wide array of classification metrics and not only focusing on $Accuracy$ provides a wider and deeper outlook on the model's performance, boosting confidence. Therefore, the potential of this work in facilitating the diagnosis of AD in real-world circumstances is reinforced.

7.2 Limitations and Future Developments

Although the wide set of metrics suggest that the proposed model can help in the diagnosis process, providing an additional option for medical practitioners to make more confident decisions, this study did encounter certain limitations, which could potentially be mitigated in future developments.

- **The imbalanced dataset, that poses difficulties to the classification step:** For future developments, an increase in balance between the number of subjects within groups, and even growing the dataset overall, can ensure a more reliable generalisation and thus a more robust and real-world prepared model.

- **Requiring up to 1379 features selected:** This indicates that the features extracted from the GLCMs contain a lot of relevant information, but also that the information necessary to make accurate predictions is dispersed across a wide array of features, posing challenges in terms of model complexity, computational efficiency, and potential overfitting. Further refinement and optimization of the feature selection process may help in reducing the number of necessary features while maintaining or even improving the accuracy of the model.

- **Using F-score for feature selection means does not furnish information on the combination of features [95]:** This is a known weakness of F-score, as it reveals the discriminating power of each feature independently from others, overlooking the potential benefits of feature inter-dependencies, which might be crucial for achieving higher accuracy and more robust model performance, for instance enabling a more effective classification between CN and AD. To mitigate this limitation, exploring more sophisticated feature selection techniques that evaluate feature sets collectively could be beneficial.

- **Using default/literature ML algorithm's hyperparameters:** A frequent obstacle in machine learning is how to tune the algorithms to achieve optimal performance. This is a complex and time-consuming task hence why default or literature settings were used. Nevertheless, algorithms offer flexibility by configuring parameters that are not directly learnt within estimators, the so called hyperparameters. While *Scikit-learn* provides a robust framework with adjustable hyperparameters, using default values might not exploit the full potential of the algorithms [96]. Future work could investigate the impact of optimizing hyperparameters to better suit data and purpose, through, for example, grid search methods.
- **Not discerning between stage's severities, as studied in other state-of-the-art works:** To enhance the discriminative capabilities of the ML model, future work could involve evaluating additional AD stages, such as advanced or moderate AD [93]. Moreover, predicting the conversion from MCI to AD would add valuable insights concerning disease progression evaluation [97]. Furthermore, working on distinguishing AD from other forms of dementia, such as Lewy Body Disease, Frontotemporal Dementia or Vascular Dementia could be a pertinent step forward. Incorporating these distinctions could lead to more precise diagnostics and ultimately improve management and treatment strategies for patients across the spectrum of diseases.

Bibliography

[1] A. Porsteinsson, R. Isaacson, S. Knox, M. Sabbagh, and I. Rubino. "Diagnosis of Early Alzheimer's Disease: Clinical Practice in 2021". In: *The Journal Of Prevention of Alzheimer's Disease* (2021), pp. 1–16. ISSN: 2274-5807. DOI: 10.14283/jpad.2021.23.

[2] N. Raza, A. Naseer, M. Tamoor, and K. Zafar. "Alzheimer Disease Classification through Transfer Learning Approach". In: *Diagnostics* 13.4 (Feb. 2023), p. 801. ISSN: 2075-4418. DOI: 10.3390/diagnostics13040801.

[3] P. Scheltens, B. De Strooper, M. Kivipelto, H. Holstege, G. Chételat, C. E. Teunissen, J. Cummings, and W. M. van der Flier. "Alzheimer's disease". In: *The Lancet* 397.10284 (Apr. 2021), pp. 1577–1590. ISSN: 0140-6736. DOI: 10.1016/s0140-6736(20)32205-4.

[4] World Health Organization. *World failing to address dementia challenge*. Visited on 21/01/2024. URL: \url{https://www.who.int/news/item/02-09-2021-world-failing-to-address-dementia-challenge}.

[5] Alzheimer Europe. *Prevalence of dementia in Europe*. Visited on 21-01-2024. URL: https://www.alzheimer-europe.org/dementia/prevalence-dementia-europe.

[6] A. A. Tahami Monfared, M. Byrnes, L. White, and Q. Zhang. "Alzheimer's Disease: Epidemiology and Clinical Progression". In: *Neurology and Therapy* 11.2 (Mar. 2022), pp. 553–569. ISSN: 2193-6536. DOI: 10.1007/s40120-022-00338-8.

[7] C. Qiu, M. Kivipelto, and E. von Strauss. "Epidemiology of Alzheimer's disease: occurrence, determinants, and strategies toward intervention". In: *Dialogues in Clinical Neuroscience* 11.2 (June 2009), pp. 111–128. ISSN: 1958-5969. DOI: 10.31887/dcns.2009.11.2/cqiu.

[8] M. B. Abubakar, K. O. Sanusi, A. Ugusman, W. Mohamed, H. Kamal, N. H. Ibrahim, C. S. Khoo, and J. Kumar. "Alzheimer's Disease: An Update and Insights Into Pathophysiology". In: *Frontiers in Aging Neuroscience* 14 (Mar. 2022). ISSN: 1663-4365. DOI: 10.3389/fnagi.2022.742408.

[9] Alzheimer's Association. "2023 Alzheimer's disease facts and figures". In: *Alzheimer's & Dementia* 19.4 (2023), pp. 1598–1695. DOI: https://doi.org/10.1002/alz.13016.

[10] NIH. *What Happens to the Brain in Alzheimer's Disease?* Visited on 21/01/2024. URL: https://www.nia.nih.gov/health/alzheimers-causes-and-risk-factors/what-happens-brain-alzheimers-disease.

[11] Z. Breijyeh and R. Karaman. "Comprehensive Review on Alzheimer's Disease: Causes and Treatment". In: *Molecules* 25.24 (2020). ISSN: 1420-3049. DOI: 10.3390/molecules25245789. URL: https://www.mdpi.com/1420-3049/25/24/5789.

[12] Z. Ismail, L. Agüera-Ortiz, H. Brodaty, A. Cieslak, et al. "The Mild Behavioral Impairment Checklist (MBI-C): A Rating Scale for Neuropsychiatric Symptoms in Pre-Dementia Populations". In: *Journal of Alzheimer's Disease* 56.3 (Feb. 2017), pp. 929–938. ISSN: 1875-8908. DOI: 10.3233/jad-160979.

[13] C. Bellenguez, F. Küçükali, I. E. Jansen, L. Kleineidam, et al. "New insights into the genetic etiology of Alzheimer's disease and related dementias". In: *Nature genetics* 54(4) (2022), 412–436. DOI: https://doi.org/10.1038/s41588-022-01024-z.

[14] NIH. *Alzheimer's Disease Genetics Fact Sheet*. Visited on 22/01/2024. URL: https://www.nia.nih.gov/health/genetics-and-family-history/alzheimers-disease-genetics-fact-sheet.

[15] Alzheimer Portugal. *Exames utilizados no Diagnóstico de Demência*. Visited on 22/01/2024. URL: https://alzheimerportugal.org/exames-utilizados-no-diagnostico-de-demencia/.

[16] W. M. van Oostveen and E. C. M. de Lange. "Imaging Techniques in Alzheimer's Disease: A Review of Applications in Early Diagnosis and Longitudinal Monitoring". In: *International Journal of Molecular Sciences* 22.4 (2021). ISSN: 1422-0067. DOI: 10.3390/ijms22042110. URL: https://www.mdpi.com/1422-0067/22/4/2110.

[17] V. K. Bairagi and S. M. Elgandelwar. "Early diagnosis of Alzheimer disease using EEG signals: the role of pre-processing". In: *International Journal of Biomedical Engineering and Technology* 41.4 (2023), p. 317. ISSN: 1752-6426. DOI: 10.1504/ijbet.2023.130834.

[18] P. Rossini, R. Di Iorio, F. Vecchio, M. Anfossi, et al. "Early diagnosis of Alzheimer's disease: the role of biomarkers including advanced EEG signal analysis. Report from the IFCN-sponsored panel of experts". In: *Clinical Neurophysiology* 131.6 (June 2020), pp. 1287–1310. ISSN: 1388-2457. DOI: 10.1016/j.clinph.2020.03.003.

[19] Alzheimer Portugal. *Intervenção Não-Farmacológica*. Visited on 30/01/2024. URL: https://alzheimerportugal.org/intervencao-nao-farmacologica/.

[20] C. E. Conti Filho, L. B. Loss, C. Marcolongo-Pereira, J. V. Rossoni Junior, et al. "Advances in Alzheimer's disease's pharmacological treatment". In: *Frontiers in Pharmacology* 14 (Jan. 2023). ISSN: 1663-9812. DOI: 10.3389/fphar.2023.1101452.

[21] J. Silva, B. C. Bispo, and P. M. Rodrigues. "Structural MRI Texture Analysis for Detecting Alzheimer's Disease". In: *Journal of Medical and Biological Engineering* 43.3 (Apr. 2023), pp. 227–238. ISSN: 2199-4757. DOI: 10.1007/s40846-023-00787-y.

[22] Q. Zhang, Y. Long, H. Cai, and Y.-W. Chen. "Lightweight neural network for Alzheimer's disease classification using multi-slice sMRI". In: *Magnetic Resonance Imaging* 107 (Apr. 2024), pp. 164–170. ISSN: 0730-725X. DOI: 10.1016/j.mri.2023.12.010.

[23] Z. Pei, Z. Wan, Y. Zhang, M. Wang, C. Leng, and Y.-H. Yang. "Multi-scale attention-based pseudo-3D convolution neural network for Alzheimer's disease diagnosis using structural MRI". In: *Pattern Recognition* 131 (Nov. 2022), p. 108825. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2022.108825.

[24] R. Prajapati and G.-R. Kwon. "A Binary Classifier Using Fully Connected Neural Network for Alzheimer's Disease Classification". In: *Journal of Multimedia Information System* 9.1 (Apr. 2022), pp. 21–32. ISSN: 2383-7632. DOI: 10.33851/jmis.2022.9.1.21.

[25] D. Pirrone, E. Weitschek, P. Di Paolo, S. De Salvo, and M. C. De Cola. "EEG Signal Processing and Supervised Machine Learning to Early Diagnose Alzheimer's Disease". In: *Applied Sciences* 12.11 (2022). ISSN: 2076-3417. DOI: 10.3390/app12115413. URL: https://www.mdpi.com/2076-3417/12/11/5413.

[26] L. Wang, J. Sheng, Q. Zhang, R. Zhou, Z. Li, Y. Xin, and Q. Zhang. "Functional Brain Network Measures for Alzheimer's Disease Classification". In: *IEEE Access* 11 (2023), pp. 111832–111845. ISSN: 2169-3536. DOI: 10.1109/access.2023.3323250.

[27] M. E. AlMansoori, S. Jemimah, F. Abuhantash, and A. AlShehhi. "Predicting early Alzheimer's with blood biomarkers and clinical features". In: *Scientific Reports* 14.1 (Mar. 2024). ISSN: 2045-2322. DOI: 10.1038/s41598-024-56489-1.

[28] N. Goenka and S. Tiwari. "Multi-class classification of Alzheimer's disease through distinct neuroimaging computational approaches using Florbetapir PET scans". In: *Evolving Systems* 14.5 (Oct. 2022), pp. 801–824. ISSN: 1868-6486. DOI: 10.1007/s12530-022-09467-9.

[29] M. Silva, P. Ribeiro, B. Bispo, and P. Rodrigues. "Detecção da Doença de Alzheimer através de Parâmetros Não-Lineares de Sinais de Fala". In: Jan. 2023. DOI: 10.14209/sbrt.2023.1570923829.

[30] 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 (Jan. 2012). ISSN: 2157-1422. DOI: 10.1101/cshperspect.a006213.

[31] H. A. Sharma. "MRI physics–basic principles". In: *Acta Neuropsychiatrica* 21.4 (Aug. 2009), pp. 200–201. ISSN: 1601-5215. DOI: 10.1111/j.1601-5215.2009.00404.x.

[32] G. Katti, S. Ara, and D. Shireen. "Magnetic resonance imaging (MRI) - A review". In: *Intl J Dental Clin* 3 (Mar. 2011). URL: https://www.researchgate.net/publication/279471369_Magnetic_resonance_imaging_MRI_-_A_review.

[33] V. S. Vassiliou, D. Cameron, S. K. Prasad, and P. D. Gatehouse. "Magnetic resonance imaging: Physics basics for the cardiologist". In: *JRSM Cardiovascular Disease* 7 (2018). PMID: 30128147. DOI: 10.1177/2048004018772237.

[34] MRI Master. *T1 vs T2 MRI*. Visited on 30/01/2024. URL: https://mrimaster.com/t1-vs-t2-mri/.

[35] S. E. Forshult. "Magnetic Resonance Imaging MRI – An Overview". In: (2007). ISSN: 1403-8099. URL: https://biomedical.gov.al/lit/n.pdf.

[36] Michigan State University Libraries. *Anatomical Planes*. Visited on 30/01/2024. URL: https://openbooks.lib.msu.edu/neuroscience/chapter/anatomical-terminology/.

[37] X. Hu, M. Meier, and J. Pruessner. "Challenges and opportunities of diagnostic markers of Alzheimer's disease based on structural magnetic resonance imaging". In: *Brain and Behavior* 13.3 (Feb. 2023). ISSN: 2162-3279. DOI: 10.1002/brb3.2925.

[38] G. B. Frisoni, N. C. Fox, C. R. Jack, P. Scheltens, and P. M. Thompson. "The clinical use of structural MRI in Alzheimer disease". In: *Nature Reviews Neurology* 6.2 (Feb. 2010), pp. 67–77. ISSN: 1759-4766. DOI: 10.1038/nrneurol.2009.215.

[39] A. Chandra, G. Dervenoulas, and M. Politis. "Magnetic resonance imaging in Alzheimer's disease and mild cognitive impairment". In: *Journal of Neurology* 266.6 (Aug. 2018), pp. 1293–1302. ISSN: 1432-1459. DOI: 10.1007/s00415-018-9016-3.

[40] IBM. *What is machine learning (ML)?* Visited on 01/02/2024. URL: https://www.ibm.com/topics/machine-learning.

[41] J. A. Nichols, H. W. Herbert Chan, and M. A. B. Baker. "Machine learning: applications of artificial intelligence to imaging and diagnosis". In: *Biophysical Reviews* 11.1 (Sept. 2018), pp. 111–118. ISSN: 1867-2469. DOI: 10.1007/s12551-018-0449-9.

[42] J.-H. Cai, Y. He, X.-L. Zhong, H. Lei, F. Wang, G.-H. Luo, H. Zhao, and J.-C. Liu. "Magnetic Resonance Texture Analysis in Alzheimer's disease". In: *Academic Radiology* 27.12 (Dec. 2020), pp. 1774–1783. ISSN: 1076-6332. DOI: 10.1016/j.acra.2020.01.006.

[43] I. El Naqa and M. J. Murphy. "What Is Machine Learning?" In: *Machine Learning in Radiation Oncology*. Springer International Publishing, 2015, pp. 3–11. ISBN: 9783319183053. DOI: 10.1007/978-3-319-18305-3_1.

[44] A. Coronato, M. Naeem, G. De Pietro, and G. Paragliola. "Reinforcement learning for intelligent healthcare applications: A survey". In: *Artificial Intelligence in Medicine* 109 (Sept. 2020), p. 101964. ISSN: 0933-3657. DOI: 10.1016/j.artmed.2020.101964.

[45] IBM. *Reinforcement Learning*. Visited on 02/02/2024. URL: https://www.ibm.com/topics/reinforcement-learning.

[46] I. H. Sarker. "Machine Learning: Algorithms, Real-World Applications and Research Directions". In: *SN Computer Science* 2.3 (Mar. 2021). ISSN: 2661-8907. DOI: 10.1007/s42979-021-00592-x.

[47] DataCamp. *Classification in Machine Learning: An Introduction*. Visited on 01/02/2024. URL: https://www.datacamp.com/blog/classification-machine-learning.

[48] Scikit-learn. *Support Vector Machines*. Visited on 02/02/2024. URL: https://scikit-learn.org/stable/modules/svm.html.

[49] S. Shobha and B. R. Karthikeyan. "Classification of Alzheimer's Disease using Transfer Learning and Support Vector Machine". In: *International Journal of Intelligent Systems and Applications in Engineering* 12 (2024), 498–508. URL: https://ijisae.org/index.php/IJISAE/article/view/4774.

[50] S. Ghosh, A. Dasgupta, and A. Swetapadma. "A Study on Support Vector Machine based Linear and Non-Linear Pattern Classification". In: *2019 International Conference on Intelligent Sustainable Systems (ICISS)*. IEEE, Feb. 2019. DOI: 10.1109/iss1.2019.8908018.

[51] Y. Ren, F. Hu, and H. Miao. "The optimization of kernel function and its parameters for SVM in well-logging". In: *2016 13th International Conference on Service Systems and Service Management (ICSSSM)*. IEEE, June 2016. DOI: 10.1109/icsssm.2016.7538563.

[52] V. Viswanatha, A. C. Ramachandra, R. Mohammed, V. Sai Kiran, and P. Kumthekar. "Support Vector Machine Implementation to Separate Linear and Non-Linear Dataset". In: *Saudi Journal of Engineering and Technology* 8.1 (Jan. 2023), pp. 4–15. ISSN: 2415-6264. DOI: 10.36348/sjet.2023.v08i01.002.

[53] A. Elen, S. Baş, and C. Közkurt. "An Adaptive Gaussian Kernel for Support Vector Machine". In: *Arabian Journal for Science and Engineering* 47.8 (Mar. 2022), pp. 10579–10588. ISSN: 2191-4281. DOI: 10.1007/s13369-022-06654-3.

[54] A. L. Yadav, K. Soni, and S. Khare. "Heart Diseases Prediction using Machine Learning". In: *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, July 2023. DOI: 10.1109/icccnt56998.2023.10306469.

[55] A. D. Arya, S. S. Verma, P. Chakarabarti, T. Chakrabarti, A. A. Elngar, A.-M. Kamali, and M. Nami. "A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease". In: *Brain Informatics* 10.1 (July 2023). ISSN: 2198-4026. DOI: 10.1186/s40708-023-00195-7.

[56] V. Nasteski. "An overview of the supervised machine learning methods". In: *HORIZONS.B* 4 (Dec. 2017), pp. 51–62. ISSN: 1857-9892. DOI: 10.20544/horizons.b.04.1.17.p05.

[57] Scikit-learn. *Decision Trees*. Visited on 02/02/2024. URL: https://scikit-learn.org/stable/modules/tree.html.

[58] G. Alfian, M. Syafrudin, I. Fahrurrozi, N. L. Fitriyani, F. T. D. Atmaji, T. Widodo, N. Bahiyah, F. Benes, and J. Rhee. "Predicting Breast Cancer from Risk Factors Using SVM and Extra-Trees-Based Feature Selection Method". In: *Computers* 11.9 (Sept. 2022), p. 136. ISSN: 2073-431X. DOI: 10.3390/computers11090136.

[59] A. Goudjil, M. K. Smail, L. Pichon, H. R. E. H. Bouchekara, and M. S. Javaid. "Wiring networks diagnosis using K-Nearest neighbour classifier and dynamic time warping". In: *Nondestructive Testing and Evaluation* (Jan. 2024), pp. 1–18. ISSN: 1477-2671. DOI: 10.1080/10589759.2024.2304710.

[60] B. Naderalvojoud and T. Hernandez-Boussard. "Improving machine learning with ensemble learning on observational healthcare data". In: *AMIA, Annual Symposium proceedings. AMIA Symposium* 2023 (Jan. 2024), pp. 521–529.

[61] X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma. "A survey on ensemble learning". In: *Frontiers of Computer Science* 14.2 (Aug. 2019), pp. 241–258. ISSN: 2095-2236. DOI: 10.1007/s11704-019-8208-z.

[62] Alzheimer's Disease Neuroimaging Initiative - ADNI. *Data Sharing and Publication Policy*. Visited on 15/05/2023. URL: https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_DSP_Policy.pdf.

[63] C. R. Jack, J. Barnes, M. A. Bernstein, B. J. Borowski, et al. "Magnetic resonance imaging in Alzheimer's Disease Neuroimaging Initiative 2". In: *Alzheimer's & Dementia* 11.7 (2015), pp. 740–756. DOI: https://doi.org/10.1016/j.jalz.2015.05.002.

[64] P. Kuhnke. *Structural Pre-Processing Batch*. Visited on 25/05/2023. URL: https://github.com/PhilKuhnke/fMRI_analysis/blob/main/1Preprocessing/SPM12/batch_structural_preproc_job.m.

[65] P. Kuhnke, M. Kiefer, and G. Hartwigsen. "Task-Dependent Recruitment of Modality-Specific and Multimodal Regions during Conceptual Processing". In: *Cerebral Cortex* 30.7 (Mar. 2020), pp. 3938–3959. ISSN: 1460-2199. DOI: 10.1093/cercor/bhaa010.

[66] A. Fatima, A. R. Shahid, B. Raza, T. M. Madni, and U. I. Janjua. "State-of-the-Art Traditional to the Machine- and Deep-Learning-Based Skull Stripping Techniques, Models, and Algorithms". In: *Journal of Digital Imaging* 33.6 (July 2020), pp. 1443–1464. ISSN: 1618-727X. DOI: 10.1007/s10278-020-00367-5.

[67] D. Arafa, H. E.-D. Moustafa, H. Ali, A. Ali-Eldin, and S. Saraya. "A deep learning framework for early diagnosis of Alzheimer's disease on MRI images". In: *Multimedia Tools and Applications* 83.2 (May 2023), pp. 3767–3799. ISSN: 1573-7721. DOI: 10.1007/s11042-023-15738-7.

[68] G. Kumar and P. K. Bhatia. "A Detailed Review of Feature Extraction in Image Processing Systems". In: *2014 Fourth International Conference on Advanced Computing amp; Communication Technologies*. IEEE, Feb. 2014. DOI: 10.1109/acct.2014.74.

[69] F. E. Shehata, M. Makkey, and S. A. Abdelrahman. "Classification of Brain Neuroimaging for Alzheimer's Disease Employing Principal Component Analysis". In: *Menoufia Journal of Electronic Engineering Research* 33.1 (Jan. 2024), pp. 31–38. ISSN: 2682-3535. DOI: 10.21608/mjeer.2023.232914.1079.

[70] M. A. Ozdemir, G. D. Ozdemir, and O. Guren. "Classification of COVID-19 electrocardiograms by using hexaxial feature mapping and deep learning". In: *BMC Medical Informatics and Decision Making* 21.1 (May 2021). ISSN: 1472-6947. DOI: 10.1186/s12911-021-01521-x.

[71] MathWorks Help Center. *graycomatrix*. Visited on 7/10/2023. URL: https://www.mathworks.com/help/images/ref/graycomatrix.html.

[72] A. Uppuluri. *GLCM texture features*. MATLAB Central File Exchange. Visited on 8/10/2023. URL: https://www.mathworks.com/matlabcentral/fileexchange/22187-glcm-texture-features.

[73] MathWorks Help Center. *Zscore*. Visited on 01/11/2023. URL: https://www.mathworks.com/help/stats/zscore.html?searchHighlight=zscore&s_tid=srchtitle_support_results_3_zscore.

[74] Scikit Learn - Supervised Learning. Visited on 15/12/2023. URL: https://scikit-learn.org/stable/supervised_learning.html#supervised-learning.

[75] J. Allibhai. *Hold-out vs. Cross-validation in Machine Learning*. Medium. Visited on 20/02/2024. URL: https://medium.com/@jaz1/holdout-vs-cross-validation-in-machine-learning-7637112d3f8f.

[76] F. Maleki, N. Muthukrishnan, K. Ovens, C. Reinhold, and R. Forghani. "Machine Learning Algorithm Validation". In: *Neuroimaging Clinics of North America* 30.4 (Nov. 2020), pp. 433–445. ISSN: 1052-5149. DOI: 10.1016/j.nic.2020.08.004.

[77] P. Ribeiro, J. Sá, D. Paiva, and P. M. Rodrigues. "Cardiovascular Diseases Diagnosis Using an ECG Multi-Band Non-Linear Machine Learning Framework Analysis". In: *Bioengineering* 11.1 (Jan. 2024), p. 58. ISSN: 2306-5354. DOI: 10.3390/bioengineering11010058.

[78] 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, Aug. 2019. DOI: 10.1109/cyberneticscom.2019.8875656.

[79] S. A. Hicks, I. Strümke, V. Thambawita, M. Hammou, M. A. Riegler, P. Halvorsen, and S. Parasa. "On evaluation metrics for medical applications of artificial intelligence". In: *Scientific Reports* 12.1 (Apr. 2022). ISSN: 2045-2322. DOI: 10.1038/s41598-022-09954-8.

[80] S. Mutasa, S. Sun, and R. Ha. "Understanding artificial intelligence based radiology studies: What is overfitting?" In: *Clinical Imaging* 65 (Sept. 2020), pp. 96–99. ISSN: 0899-7071. DOI: 10.1016/j.clinimag.2020.04.025.

[81] P. Dhal and C. Azad. "A comprehensive survey on feature selection in the various fields of machine learning". In: *Applied Intelligence* 52.4 (July 2021), pp. 4543–4581. ISSN: 1573-7497. DOI: 10.1007/s10489-021-02550-9.

[82] R. Duara and W. Barker. "Heterogeneity in Alzheimer's Disease Diagnosis and Progression Rates: Implications for Therapeutic Trials". In: *Neurotherapeutics* 19.1 (Jan. 2022), pp. 8–25. ISSN: 1878-7479. DOI: 10.1007/s13311-022-01185-z.

[83] A. Ebrahimi-Ghahnavieh, S. Luo, and R. Chiong. "Transfer Learning for Alzheimer's Disease Detection on MRI Images". In: *2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*. IEEE, July 2019. DOI: 10.1109/iciaict.2019.8784845.

[84] L. Srividhya, V. Sowmya, R. Vinayakumar, E. A. Gopalakrishnan, and K. P. Sowmya. "Deep learning-based approach for multi-stage diagnosis of Alzheimer's disease". In: *Multimedia Tools and Applications* 83.6 (July 2023), pp. 16799–16822. ISSN: 1573-7721. DOI: 10.1007/s11042-023-16026-0.

[85] A. Shukla, R. Tiwari, and S. Tiwari. "Structural biomarker-based Alzheimer's disease detection via ensemble learning techniques". In: *International Journal of Imaging Systems and Technology* 34.1 (Sept. 2023). ISSN: 1098-1098. DOI: 10.1002/ima.22967.

[86] R. K. Lama, J.-I. Kim, and G.-R. Kwon. "Classification of Alzheimer's Disease Based on Core-Large Scale Brain Network Using Multilayer Extreme Learning Machine". In: *Mathematics* 10.12 (June 2022), p. 1967. ISSN: 2227-7390. DOI: 10.3390/math10121967.

[87] W. Kang, L. Lin, B. Zhang, X. Shen, and S. Wu. "Multi-model and multi-slice ensemble learning architecture based on 2D convolutional neural networks for Alzheimer's disease diagnosis". In: *Computers in Biology and Medicine* 136 (Sept. 2021), p. 104678. ISSN: 0010-4825. DOI: 10.1016/j.compbiomed.2021.104678.

[88] R. Prajapati, U. Khatri, and G. R. Kwon. "An Efficient Deep Neural Network Binary Classifier for Alzheimer's Disease Classification". In: *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*. IEEE, Apr. 2021. DOI: 10.1109/icaiic51459.2021.9415212.

[89] E. Lee, J.-S. Choi, M. Kim, and H.-I. Suk. "Toward an interpretable Alzheimer's disease diagnostic model with regional abnormality representation via deep learning". In: *NeuroImage* 202 (Nov. 2019), p. 116113. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2019.116113.

[90] J. Wen, E. Thibeau-Sutre, M. Diaz-Melo, J. Samper-González, A. Routier, S. Bottani, D. Dormont, S. Durrleman, N. Burgos, and O. Colliot. "Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation". In: *Medical Image Analysis* 63 (July 2020), p. 101694. ISSN: 1361-8415. DOI: 10.1016/j.media.2020.101694.

[91] E. Hussain, M. Hasan, S. Hassan, T. Azmi, M. A. Rahman, and M. Z. Parvez. "Deep Learning Based Binary Classification for Alzheimer's Disease Detection using Brain MRI Images". In: *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. Nov. 2020, pp. 1115–1120. DOI: 10.1109/ICIEA48937.2020.9248213.

[92] S. Qiu, P. S. Joshi, M. I. Miller, C. Xue, et al. "Development and validation of an interpretable deep learning framework for Alzheimer's disease classification". In: *Brain* 143.6 (May 2020), pp. 1920–1933. ISSN: 1460-2156. DOI: 10.1093/brain/awaa137.

[93] 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 (Sept. 2021), pp. 3384–3395. ISSN: 2168-2208. DOI: 10.1109/jbhi.2021.3069789.

[94] 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 (Jan. 2018), pp. 40–59. ISSN: 2160-9551. DOI: 10.4018/ijrqeh.2018010104.

[95] Y.-W. Chen and C.-J. Lin. "Combining SVMs with Various Feature Selection Strategies". In: vol. 207. Nov. 2008, pp. 315–324. ISBN: 978-3-540-35487-1. DOI: 10.1007/978-3-540-35488-8_13.

[96] Scikit-Learn. *Tuning the hyper-parameters of an estimator*. Visited on 28/02/2024. URL: https://scikit-learn.org/stable/modules/grid_search.html.

[97] S. Park, C. H. Hong, D.-g. Lee, K. Park, and H. Shin. "Prospective classification of Alzheimer's disease conversion from mild cognitive impairment". In: *Neural Networks* 164 (July 2023), pp. 335–344. ISSN: 0893-6080. DOI: 10.1016/j.neunet.2023.04.018.