UNIVERSITÀ DEGLI STUDI DI BARI
Dipartimento di Fisica

Dottorato di Ricerca in Fisica – XXXI Ciclo
S.S.D. FIS/07

Multiscale analysis of brain connectome for the characterization of Alzheimer's disease

---

Eufemia LELLA

Supervisori:
Prof. Roberto BELLOTTI
Prof. Sebastiano STRAMAGLIA

Coordinatore:
Prof. Giuseppe IASELLI

ESAME FINALE 2018

"Science is but a perversion of itself unless it has as its ultimate goal the betterment of humanity"

Nikola Tesla

UNIVERSITY OF BARI

Abstract

Department of Physics

Doctor of Philosophy

Multiscale analysis of brain connectome for the characterization of Alzheimer's disease

by Eufemia LELLA

Alzheimer's disease (AD) is the most common neurodegenerative disorder and a growing health problem. Investigating useful *biomarkers* for early diagnosis, prognosis and response to therapy is a primary goal today of the research activity in neuroscience. Physics plays a key role in neuroscience research, as it provides theory and methods for brain imaging acquisition, brain modeling and data analysis. The complexity of the brain connectome, in fact, lends itself to studies at different scales, ranging from a local voxel-wise investigation of images of the brain to more global properties captured by complex network models reconstructed from these imaging data. A promising imaging technique tailored to such a multiscale investigation is diffusion tensor imaging (DTI). DTI, in fact, provides insights into the white matter micro-structural integrity and can thus help identify alterations in white matter regions due to AD. In this thesis, voxel-based and graph-based approaches to the problem of analyzing DTI images of the human brain for AD characterization are explored. The investigations are carried out with a two-fold purpose. On one hand, statistical-descriptive analyses are performed to provide clinical insights on the brain areas mostly involved in the degeneration processes triggered by the disease. On the other hand, machine learning algorithms are applied to investigate the predictive potential of meaningful features characterizing AD for making the discrimination between healthy and unhealthy subjects automatic. In particular, the application of an uncommon graph metric, the communicability metric, is explored to characterize brain networks and the connectivity alterations due to AD. The data used in the present thesis are based on samples of healthy control elderly people, AD patients and individuals with a mild cognitive impairment from the publicly available ADNI dataset. Encouraging results have been obtained, paving the way for the development of statistical and computerized methods, as low-cost and non-invasive tools to support the disease assessment in the daily clinical trials. In particular, further evidence that measures describing the communication between brain regions represent promising biomarkers for AD has been found.

Part of the research carried out for this thesis appears in the following works:

- Maggipinto, T., Bellotti, R., Amoroso, N., Diacono, D., Donvito, G., Lella, E., Monaco, A., Scelsi, M. A., Tangaro, S., "DTI measurements for Alzheimer's classification", Physics in Medicine & Biology, 62(6), 2361-2375 (2017)

- Lella, E., Amoroso, N., Bellotti, R., Diacono, D., La Rocca, M., Maggipinto, T., Monaco, A., Tangaro, S., “Machine learning for the assessment of Alzheimer’s disease through DTI", In: Applications of Digital Image Processing XL, Vol. 10396, p. 1039619 (2017)

- Lella, E., Amoroso, N., Lombardi, A., Maggipinto, T., Tangaro, S., Bellotti, R., “Communicability disruption in Alzheimer’s disease connectivity networks", Journal of Complex Networks (2018)

- Lella, E., Amoroso, N., Lombardi, A., Maggipinto, T., Tangaro, S., Bellotti, R., “A Classification Framework for Alzheimer’s Disease based on Graph Communicability", In: 2nd Workshop on GRaphs in biomediAl Image anaLysis (GRAIL), satellite event at MICCAI (2018)

Acknowledgements

I would like to express my special thanks to the many people who contributed to the work presented in this thesis and to the achievement of this important goal of my life. I would thank my adviser Prof. Roberto Bellotti who supervised my work for these three years, trusting me and treating me as a researcher from the very first days. My second supervisor, Prof. Sebastiano Stramaglia, gave me valuable suggestions for improving my research. I would also thank the other components of my research group for supporting my work, providing me with several precious contributions which improved the quality of the research: Sabina Tangaro, Tommaso Maggipinto, Domenico Diacono, Nicola Amoroso, Alfonso Monaco, Angela Lombardi and Andrea Tateo. In particular, I would thank Sabina Tangaro for encouraging my work a lot of times with stimuli and challenges. And I would thank Angela Lombardi for supporting me everyday as a colleague and as a friend; in these years together, we shared ideas, worries, joys, hopes, successes and laughs. I would express my special thank to Prof. Ernesto Estrada of the University of Strathclyde, Glasgow, for inspiring my research and fueling my love for it, with his lessons that I had the opportunity to follow during more than one PhD school. I also wish to thank Dr. Robert Smith of the Florey Institute of Neuroscience and Mental Health, Austin, for his valuable suggestions on the image processing pipeline concerning the brain connectivity reconstruction with MRtrix.

A very special thank to Gennaro Vessio for supporting me in every single day of these three years. No word could be enough to say thanks for everything; probably, without him, I never would have begun and I never would have finished my PhD experience. I'm infinitely grateful to my family for their support, and for having taught me the importance of values like hard work, humility, honesty, and love. I would also thank my special and unique friends Ciccio, Giovanna and Nicola for being always present despite distances: no spatial distance will never separate us. A special thank also to Stefania and Marica, for the relaxing and good time spent together. Thank to my lifelong friends Elena, Raffaella, Agnese, Francesco e Leonardo for being always present. A special thank to Franca, my second mom and best friend, for encouraging me in every difficult moment.

Contents

List of Figures

List of Tables

List of Abbreviations

AD	**A**lzheimer's **D**isease
AUC	**A**rea **U**nder the ROC **C**urve
DTI	**D**iffusion **T**ensor **I**maging
FA	**F**ractional **A**nisotropy
FDR	**F**alse **D**iscovery **R**ate
FSB	**F**eature **S**election **B**ias
FSL	**F**MRIB **S**oftware **L**ibrary
HC	**H**ealthy **C**ontrols
MCI	**M**ild **C**ognitive **I**mpairment
MD	**M**ean **D**iffusivity
MRI	**M**agnetic **R**esonance **I**maging
NB	**N**aïve **B**ayes
RF	**R**andom **F**orest
RFE	**R**ecursive **F**eature **E**limination
SVM	**S**upport **V**ector **M**achine
WM	**W**hite **M**atter

To my parents

Chapter 1

Introduction

1.1 Physics and neuroscience

Neuroscience is an increasingly important area of research that aims to understand how the brain works. It is an interdisciplinary field that integrates psychology, biology, chemistry and physics. In particular, physics plays a key role in neuroscience research providing crucial insights into the understanding of the brain from three points of view: brain imaging technology, data analysis and brain modeling. The first aspect is a technical and instrumental one: research in physics has led to the development of important neuroimaging techniques such as positron emission tomography (PET), single-photon emission computed tomography (SPECT), magnetoencephalography (MEG), electroencephalography (EEG), computed tomography (CT), structural magnetic resonance imaging (MRI), functional magnetic resonance imaging (fMRI), diffusion tensor imaging (DTI), and so on.

The second aspect concerns the methodological strategies. During the most of neuroscience history, this science was mainly qualitative, with little or no reliance on statistics. In the last 10 years, neuroscience is undergoing faster changes generating quantitative big-sample datasets. New methodological approaches are required due to the progressively increasing granularity of digitized brain data. There is a growing interest for data sharing, open access and building "big-data" repositories. Neuroscience is entering the era of large-scale data collection, curation, and collaboration with a pressing need for statistical models tailored for high dimensional inference. Rich datasets, in fact, can extend the spectrum of possible findings about the brain [1]. It's on this new perspective that physics takes action applying in neuroscience the computational and methodological strategies used in astrophysics, particle physics and high energy physics: statistical and mathematical frameworks, software development, simulation and modeling tools, machine learning procedures, distributed computing.

The third aspect concerns the brain modeling. The human brain can be modeled as a system with several regions, whose interactions produce complex behaviors [2]. To understand the mechanisms behind higher-level brain functions, a detailed study of the individual neural cells is clearly insufficient; global functional and structural properties of such a complex system need to be considered as emerging from the interactions between the parts of the entire system, which combine statistical randomness with regularity [3]. Since the mid 1990s, developments in our understanding of the physics of complex systems have led to the rise of network science as a transdisciplinary effort to characterize network structure and function. Complex networks provide a powerful tool to model components of the human brain and their interactions. They are suited to study brain connectivity and, in particular, connectivity alterations due to neurodegenerative diseases, in order to understand the mechanisms underlying such degeneration.

1.2 Motivations and objectives

In the research on degenerative brain disorders, the application of mathematical and technological frameworks has the aim to find new *biomarkers* for early disease diagnosis, prognosis and response to therapy. Among neurodegenerative diseases, Alzheimer's disease (AD) is the most widespread and most disabling. It is characterized by short-term memory loss in its early stages, followed by a progressive decline in other cognitive and behavioural functions as the disease advances [4]. There is evidence supporting the biological hypothesis that this decline is related to a disrupted connectivity among brain regions, caused by white matter (WM) degeneration [5–7].

Because of their homogeneous chemical composition, conventional magnetic resonance imaging is not able to highlight the structure of the WM fibres; therefore, it is not tailored to investigate the physical disconnections arising among them. Conversely, diffusion tensor imaging has emerged as an helpful technique that measures the water diffusion along the WM fibres, and can thus provide useful information regarding their integrity [8]. Fractional anisotropy (FA) and mean diffusivity (MD) are among the invariants derived from the diffusion tensor that are closely related to such integrity: the water diffusion along a healthy axon is highly anisotropic, being constrained almost completely to one direction, so high values of FA and low values of MD are able to describe non-pathological scenarios. In this view, diffusion anisotropy effects point out micro-structural changes related to the disease, which can complement the information coming from conventional MRI in investigating human brain atrophy.

In addition, when combined with tractography algorithms [9], DTI enables the reconstruction of the WM fibre tracts, providing a characterization of the physical connections of the brain that can be subsequently investigated through a graph theory-based approach. Traditional network metrics tailored to describe topological properties of the brain include nodal degree and strength and shortest path length.

A promising application of DTI measures consists in computing meaningful features, for instance the most representative voxels of FA and MD maps, or graph-based topological metrics based on the brain connectivity network, and then feeding these features into machine learning algorithms to automatize the discrimination between AD patients and healthy control (HC) subjects. Developing decision support systems can provide a quantitative, non-invasive and low-cost tool-support to the neuropsychological assessments performed by expert clinicians.

More importantly, such an approach can be employed for the AD classification at a prodromal stage. To this end, a special attention is devoted to mild cognitive impairment (MCI) signs, as a person with MCI is at a high risk of developing dementia of the Alzheimer's type [10]. MCI is characterized as a non-disabling disorder that represents an early state of abnormal cognitive function. Although not all MCI cases represent prodromal AD, an estimated 10–15% of MCI subjects enter the dementia spectrum every year [11]. Discriminating AD at such an early stage is crucial in the perspective of the future medical treatments to be administered and for improving the quality of life of the patient.

In this thesis, a multiscale analysis on DTI images of the human brain for AD characterization has been carried out. This analysis ranged from the most consolidated voxel-wise approach to the more recently explored approach based on graph and complex network theory. Thus the investigation ranged from the microstructural localized information provided by the voxel-wise analysis to the more global perspective of connectivity network analysis at the cortical and at sub-cortical level.

These scales of investigation have been carried from two different perspectives. On one hand, statistical-descriptive analyses have been conducted to provide clinical insights on the brain areas mostly involved in the degeneration processes triggered by the disease and to find connectivity patterns related to AD. On the other hand, machine learning algorithms have been applied to investigate the predictive potential of meaningful features characterizing AD for making the discrimination between healthy and unhealthy subjects automatic. The data used in the present thesis are based on samples of healthy control elderly people, AD patients and individuals with MCI from the publicly available ADNI dataset.

The contribution of the present thesis is therefore manifold. First, this thesis contributes by overcoming some main limitations which prevent the reliability of state-of-the-art results obtained with the voxel-based approach. In particular, we performed a comparative study between *nested* and *non-nested* feature selection procedures on the same set of data, as first attempt in the literature to measure the bias introduced by non-nested feature selection in the classification of DTI images. This study led to the definition of a new baseline for the HC/AD and the HC/MCI discrimination with FA and MD maps based features.

A network-based approach was then applied on the same dataset, in particular to investigate whether the graph *communicability* metric is tailored to describe the disruption of communication between brain regions caused by AD. The literature investigated this issue applying only traditional graph metrics, mostly based on shortest path length. Communicability, instead, is a less common graph measure, which quantifies the ease of communications between two nodes of a network by considering not only the shortest path, but all possible paths connecting them. For this reason, this measure may be a better discriminant factor for AD. To investigate this hypothesis, we evaluated if and to which extent the communicability metric positively impacts the performance of a classification algorithm for the AD/MCI classification. In particular, we performed a comparative study between classification models trained with traditional network metrics and models trained with communicability. To our best knowledge, this is the first time the communicability metric has been proposed to study the connectivity changes in AD and MCI brain networks.

The graph-based analysis has been carried out both at the cortical and at the sub-cortical level; in the latter case, the effectiveness of novel metrics derived from communicability has also been investigated.

1.3 Structure of the thesis

This thesis deals with different issues that are part of different scales of analysis. For instance, different approaches require dedicated image processing pipelines. Heterogeneous groups of subjects have been available depending on the image processing modality under study. Some statistical and machine learning techniques are better suited to some data or features, but may be less useful with other sources of information. Different clinical insights are also likely to be found. For these reasons, the materials and methods of the experiments carried out are not summarized in a unique chapter but are discussed only when pertinent.

The rest of the present thesis is thus organized as follows:

Chapter 2 provides background knowledge on Alzheimer's disease, diffusion tensor imaging, complex networks and data analysis techniques, with emphasis

on their application to the problem at hand, i.e. the assessment of (possibly early) signs and manifestations of the disease.

Chapter 3 deals with the application of the more classic voxel-based approach to address such a problem. The chapter reviews some open issue in the current research and is aimed at providing novel baseline results for the binary classification AD vs. HC and MCI vs. HC relying on FA and MD maps. The content of this chapter is mainly based on [12] and [13].

Chapter 4 focuses on the application of the more recently investigated approach based on the use of complex network theory. The chapter investigates the usefulness of an uncommon graph measure, namely communicability, to better describe the disruption of connectivity in the human brain due to AD. In particular, the cortical connectivity network is studied. This chapter is mainly based on [14] and [15].

Chapter 5 extends the experiment reported in the previous chapter by considering the connectivity alterations arising at the sub-cortical level. Some novel metrics, derived from communicability, are proposed for this aim.

Chapter 6 concludes the thesis and depicts future developments of the research conducted so far.

Chapter 2

Background

In order to better explain the research activity carried out, the present chapter provides some background knowledge on Alzheimer's disease and the traditional approaches to its assessment. Then, the need for new biological markers, particularly brain imaging, is emphasized. The main principles of Diffusion Tensor Imaging and the most consolidated approaches proposed for its application on AD characterization are described. The more recent approach based on complex networks is also explained. Finally, an overview of statistical and machine learning tools used for data analysis is provided.

2.1 Alzheimer's disease

Alzheimer's disease is a degenerative brain disorder, the most frequent cause of dementia [16]. It is characterized by a progressive decline in cognitive function, in particular the memory domain. According to the Alzheimer's Disease International, the worldwide federation of Alzheimer associations supporting people with dementia and their families, there were an estimated 46.8 million people worldwide living with dementia in 2015 and this number is believed to be almost double every 20 years, reaching 75 million in 2030 and 131.5 million in 2050. Much of the increase will be in developing countries. Already 58% of people with dementia live in low and middle income countries, but by 2050 this will rise to 68% (Fig. 2.1 from https://www.alz.co.uk/research/statistics)

According to [4], in 2017 an estimated 5.5 million Americans have Alzheimer's dementia. By mid-century, the number of people living with Alzheimer's dementia

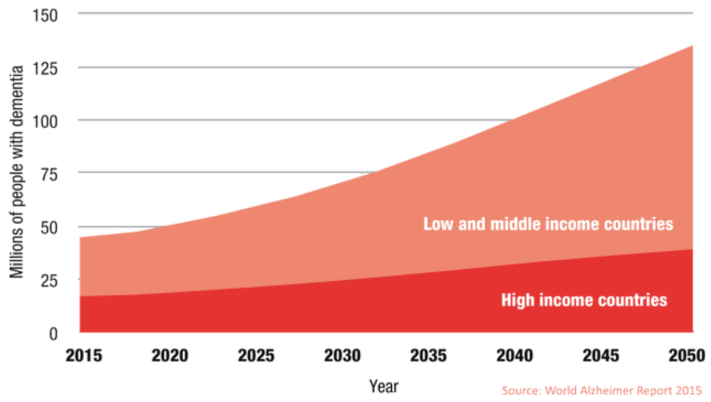


FIGURE 2.1: Worldwide projections of the number of people with dementia in low and middle income countries and in high income countries.

in the United States is projected to grow to 13.8 million. Today, someone in the country develops Alzheimer's dementia every 66 seconds. By 2050, one new case of Alzheimer's dementia is expected to develop every 33 seconds, resulting in nearly 1 million new cases per year. These worldwide projections show that AD is a really growing health problem, as well as a serious social issue.

One of the most common signs of AD is memory loss, especially forgetting recently learned information. Another manifestation is forgetting important dates or events. Some people experience difficulties in solving problems, changes in their ability to develop and follow a plan or to work with numbers; moreover, they may have trouble concentrating and they take much longer to do things than they did before. AD patients find it hard to complete daily tasks at home, at work or at leisure and they can get confused with time or place. Furthermore, people with Alzheimer's may have trouble understanding visual images and spatial relationships and problems with words in speaking or writing. Other signs of dementia are changes in judgment or decision-making and changes in mood and personality. AD patients can become confused, suspicious, depressed, fearful or anxious and they may remove themselves from hobbies, social activities, work projects or sport. People in the final stages of the disease are bed-bound and require around-the-clock care. Alzheimer's disease is ultimately fatal.

The difficulties in AD people occur because nerve cells (neurons), in parts of the brain involved in cognitive function, have been damaged or destroyed. In AD, neurons in other parts of the brain can be damaged or destroyed as well, including those that allow a person to carry out basic bodily functions.

The neuropathological hallmarks of AD consist in a strong cortical atrophy and in the accumulation in the cerebral cortex of senile plaques, neurofibrillary tangles and neuropil threads [17]. Fibrillary amyloid substance accumulates in senile plaques. The major subunit of the amyloid fibrils is the 4.2-kD amyloid-β peptide, known as Aβ because of its partial beta-pleated sheet structure. Neurofibrillary tangles contain paired helical filaments (PHFs) composed by the microtubule-associated protein tau [18]. As plaques and tangles form in particular brain regions, healthy neurons begin to work less efficiently and to lose their funcions and their capacity to communicate with each other, eventually dying off completely. Neuron death-off causes atrophy in the affected brain areas, mainly hippocampus which is essential in memory functions.

The pathomechanism of Aβ and tangle formation still remains unclear and it is still controversial whether tangles are a primary causative factor in disease or play a more peripheral role. Clinicopathological correlation studies have played a key role to generate hypotheses about the pathophysiology of the disease, by finding out that the amyloid plaque build-up occurs primarily before the beginning of cognitive deficits, while neurofibrillary tangles and neuron loss, particularly synaptic loss, occur parallel to the progression of cognitive decline [19]. Experimental evidence suggests that Aβ plaques begin to form many years before dementia and this progressive and gradual pathology offers a potential target for early intervention [20].

2.1.1 Stages of the disease

Changes in the brain due to AD begin years before any sign of the disease becomes manifest. This time period can last for years and it is called preclinical AD: it's a silent stage during which senile plaques are accumulating in the brain, but are not sufficient to cause remarkable symptoms. The disease progression goes through

three stages: mild Alzheimer's disease, moderate Alzheimer's disease and severe Alzheimer's disease (https://www.alz.org/alzheimers-dementia/stages).

The mild AD is the early stage in which the person is still able to do the normal activities such as driving and working, but she may have memories lapses regarding for example familiar words or the location of everyday objects. Friends and family begin to notice some difficulties, including:

- Problems remembering names, right words, information just read;
- Losing or misplacing a valuable object;
- Trouble with planning or organizing;
- Difficulties in performing tasks in social or work settings.

Moderate AD is the middle stage; typically the longest one as it can last for many years. During the moderate stage of Alzheimer's disease, individuals may have greater difficulty to perform tasks such as paying bills, but they may still remember important details about their life. Damage to nerve cells in the brain can make it difficult to express thoughts and perform routine tasks. At this stage, common symptoms of sick people are:

- Problems remembering events of personal history;
- Being unable to recall their own address or telephone number;
- Confusion about where they are or what day it is and increased risk of wandering and becoming lost;
- Changes in sleep patterns, such as sleeping during the day and becoming restless at night;
- Personality and behavioral changes, including suspiciousness and delusions or compulsive, repetitive behavior.

In the late stage, memory and cognitive capacity continue to decline and physical ability is severely impacted, for example the ability to control movements. Significant personality changes may take place and individuals need extensive help with daily activities. The main symptoms are:

- Losing awareness of recent experiences and of their surroundings;
- Increasing difficulty in communicating;
- Changes in physical abilities, such as walking, sitting and, eventually, swallowing;
- Vulnerability to infections, especially pneumonia.

A really important stage, in terms of early diagnosis, is the MCI state, which is intermediate between preclinical AD and dementia (Fig. 2.2). MCI is a condition in which an individual has mild but measurable changes in thinking abilities that are detectable to the person affected and to family members and friends, but do not affect the individual's ability to carry out everyday activities. The MCI state does not necessary evolve in AD. The formal definition of MCI is: "a syndrome defined as cognitive decline greater than expected for an individual's age and education level but that does not interfere notably with activities of daily life" [10]. There are several

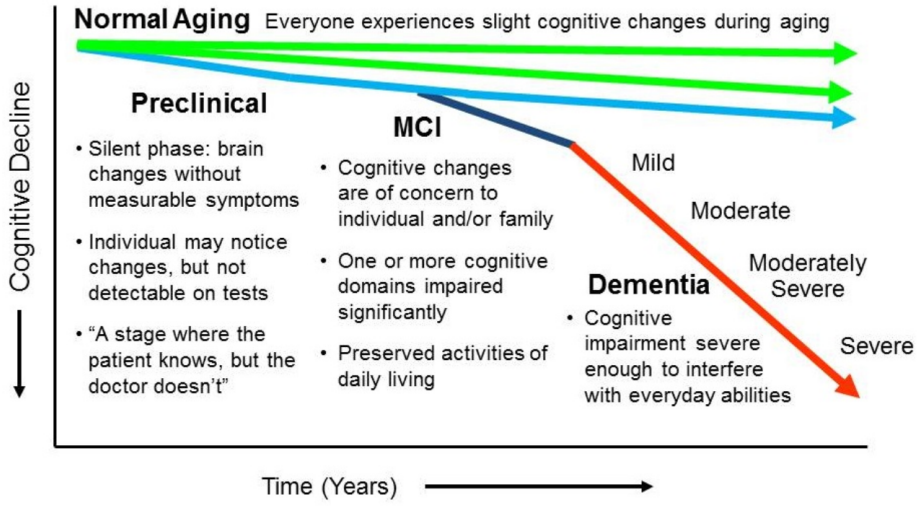


FIGURE 2.2: Progression from normal aging to Alzheimer's disease.

different sub-types of MCI, but the two major categories are amnestic (characterized by memory-related problems) and non-amnestic. It is difficult to estimate how many cases of MCI there are since many different definitions of MCI exist. According to the results of several studies, 15-20% of adults over the age of 65 meet the criteria for MCI [21]. People with MCI, especially MCI involving memory problems, are more likely to develop Alzheimer's or other dementias than people without MCI [22]. A systematic review of 32 studies found that an average of 32% of individuals with MCI developed Alzheimer's dementia in 5 years [23]. However, for unknown reasons, some people with MCI remain stable while some even improve to AD: today researchers and clinicians are not able to definitively predict if MCI will progress to dementia. Numerous studies on MCI are ongoing to identify those neuropsychological, neuropsychiatric and neuroimaging biomarkers that might indicate risk for further decline to AD, helping researchers to detect early brain changes and to test medications. The pharmacologic treatments available today for Alzheimer's dementia are not able to slow or stop the destruction of neurons that cause AD symptoms and make the disease fatal. In recent years, efforts to develop and validate AD biomarkers, including those detectable with brain imaging and in the blood and cerebrospinal fluid, have intensified. Such efforts are moving the focus from cognitive and functional symptoms to one that incorporates biomarkers. This new approach could promote diagnosis at an earlier stage of disease and lead to a more accurate understanding of AD prevalence and incidence.

2.1.2 Diagnosis and biomarkers

The evaluation of the patient's clinical status and his/her responsiveness to medication is typically achieved via a clinical workup which includes some core elements. A thorough medical history, where the neuropsychologist documents family history, previous illnesses, etc., is usually collected. The assessment of independent function and daily activities allows the examiner to focus on changes of the usual level of function of the individual. To make a clearer clinical picture, neuropsychological testing is also typically carried out, so that specific thinking skills can be evaluated through series of written tests. The most common presentation of AD dementia, in

fact, is the amnestic form, which involves impairment of episodic memory (i.e., the ability to learn and retain new information). However, some patients have initial involvement of other cognitive domains, such as language or visuospatial or executive functioning. Standard cognitive and functional tests include, among the others, those described in the following paragraphs.

Mini-Cog It is a short screening test composed by two parts: a 3-item recall test and a simple clock drawing test [24]. In addition to verifying short-term memory loss, the test evaluates visuo-spatial abilities, attention and executive functions. Mini-Cog is not a diagnostic test, but it is useful to identify subjects at risk of developing dementia and for which a more in-depth assessment is needed.

IADL scale This is an instrument to assess independent living skills [25]. More precisely, it provides information about 8 skills necessary for living in a community, for example food preparation, housekeeping and laundering. Individuals are scored in accordance with their highest level of functioning and on the need for assistance or supervision.

Mini-Mental State Examination It consists of a 30-point questionnaire including questions and problems in many areas: orientation to time and place; repetition of lists of words; attention and calculation; etc [26]. MMSE is one of the most widely used tests for assessing dementia. Any score greater than or equal to 24 points indicates a normal cognition. Below this, the score can indicate severe ($\leq$ 9 points), moderate (10-18 points) or mild (19-23 points) cognitive impairment.

Trail Making Test It requires the subject to connect a sequence of 25 consecutive targets. The test, in particular, is divided in two parts: in the first part, the targets are only numbers (1, 2, 3, ...) and the subject is asked to connect them in sequential order; in the second part, the subject alternates between numbers and letters (1, A, 2, B, ...). The test is tailored to explore different cognitive components, in particular attentional skills, visuo-motor planning and working memory [27].

Attentional matrices This is a cancellation test: three matrices are shown to the test taker and she is asked to cross out as fast as possible target numbers of either one, two or three digits [28]. The aim of the test is to evaluate the subject's ability to detect visual targets among several distractors.

These tests are used extensively today to assess cognitive impairment and they represent the current clinical standard. Diagnostic criteria for AD were first proposed more than thirty years ago and their clinical validity is acknowledged. Since their proposal in 1984, in fact, the key classification for the diagnosis of AD has been the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria. The original criteria served the neurologic field well, both for diagnosis in clinical practice and for inclusion of appropriate populations into clinical research. However, over the past few years, it has become evident that the diagnostic criteria should be revisited and updated to reflect recent advances in biomarker research [29]. Identifying accurate biomarkers for early and differential diagnosis, prognosis and response to therapy, in fact, is a primary goal of the research on neurodegenerative disorders today.

The quest for accurate and useful *in vivo* biomarkers for AD has evolved rapidly over the past decade, with advances in cerebrospinal fluid (CSF) markers and multiple types of neuroimaging [30]. They are briefly described below.

CSF biomarkers Due to the free transport between the brain and the CSF, levels of Aβ, total tau and p-tau in CSF reflect the metabolic processes in the brain and can thus be used to aid the accurate diagnosis of AD also at an early stage [31]. In MCI and AD, CSF levels of Aβ decrease while t-tau and p-tau increase compared to cognitevely normal subjects [32]. In addition, the combined evaluation of Aβ at different levels may further increase sensitivity and specificity in predicting progression from MCI to AD [33]. Unfortunately, the association between CSF biomarkers and the concentrations of deposited amyloid in plaques still remains unclear.

Structural MRI High-resolution MRI determines structural changes in the brain in vivo. Significant atrophy of the hippocampal formation, entorhinal cortex and parahippocampal gyrus can be demonstrated by MRI, even in the preclinical stages of AD, and can predict later conversion to AD with about 80% of accuracy, e.g. [34]. Hippocampal volumetry is the best-established structural biomarker for AD and appears to be suitable for risk stratification in MCI cohorts in treatment trials [35]. Nevertheless, this measure still involves a time-consuming process and a great deal of manual work, which may prevent its use as a routine diagnostic test in the next future. Due to the laborious nature of initial manual volumetric methods, various automated methods have been developed over the past years to demonstrate change in brain structure and morphology in AD patients more efficiently. The most commonly investigated method to date is voxel-based volumetry, which showed a reduction in the cortical gray matter in the region of the mediotemporal lobes and lateral and parietal association areas in AD patients, e.g. [36, 37].

Functional MRI Functional MRI allows one to measure brain activation during cognitive tasks, or during resting state, without requiring any contrast agent or radiation exposure to the patient. Many studies examined brain activation changes in MCI subjects compared to AD, for the development of a marker of early AD, e.g. [38]. A powerful approach consists in investigating changes in the functional connectivity between regions of an activated network [39]. The changes in functional connectivity preceded differences in brain activation between the MCI and healthy control group.

The network connectivity can also be investigated by means of Diffusion Tensor Imaging, which provides a measure of the structural integrity of the white matter tracts connecting regions of the brain. Since the present thesis deals specifically with this imaging modality, it is more in depth described in a dedicated paragraph.

Hopefully, these biomarkers will serve to improve understanding of the pathophysiologic processes of AD during life and to elucidate the link between the pathology and clinical manifestations of the disease. Much work remains to be done to standardize currently available biomarkers to enable optimal use in the clinical setting, to develop more sensitive and specific biomarkers for early diagnosis, to track progression and to monitor response to future disease-modifying therapy.

Nevertheless, despite the large number of promising results, biological markers of AD are at various stages of development and clinical evaluation, and have not generally been established in clinical routine so far. In order to approach this goal,

large-scale international controlled multicenter trials, such as ADNI, are currently being conducted in an attempt to systematically develop and validate feasible candidate biomarkers.

2.1.3 Alzheimer's Disease Neuroimaging Initiative

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a global research study that actively supports the investigation and development of medical treatments to slow or to stop the progression of AD. It's a multisite longitudinal study, in which researchers at 63 sites in the US and Canada collaborate to track the progression of AD in the human brain with clinical, imaging, genetic and biospecimen biomarkers through the process of normal aging, early mild cognitive impairment (EMCI), and late mild cognitive impairment (LMCI) to dementia or AD. The overall goal of ADNI is to find validate biomarkers for use in Alzheimer's disease diagnosis and clinical treatment trials.

ADNI has made a global impact on AD research, by developing standardized protocols to permit comparison of results from multiple centers and also by its data-sharing policy which makes available all ADNI data to qualified researchers worldwide. Researchers worldwide can submit an online data access request and can generally begin using ADNI data (which can be downloaded directly from the website `http://adni.loni.usc.edu/`), once the access request is accepted and the user is logged in. Several types of data from study volunteers are available:

- Clinical data including demographics, cognitive/neuropsychological information, diagnostic summary, exam dates of the follow-up visits;
- Medical history information, including adverse events, medication history, physical and neurological examination;
- Genetic data including genome-wide association study (GWAS) and whole genome sequenting data (WGS);
- Raw, pre- and post-processed neuroimaging data including MRI, fMRI, DTI and PET;
- Biospecimens, including blood, urine, and cerebrospinal fluid (CSF).

ADNI was funded in 2004, under the leadership of Dr. Michael W. Weiner, as a private-public partnership with $27 million contributed by 20 companies and two foundations (through the Foundation for the National Institutes of Health) and $40 million from the National Institute on Aging. The first phase of ADNI started in October 2004 and lasted 6 years. The ADNI1 study was the first ADNI study, designed to find more sensitive and accurate biomarkers for the early diagnosis and tracking of AD. The study collected and analyzed thousands of brain scans, genetic profiles, and blood and cerebrospinal fluid biomarkers. It included 400 subjects diagnosed with mild cognitive impairment, 200 subjects with the early AD, and 200 elderly control subjects.

ADNI was extended in 2009 during the ADNI GO phase, which extended the existing ADNI1 cohort with 200 new participants with early mild cognitive impairment. The aim of this phase was to asses biomarkers at an earlier stage of the disease. MR protocols were also adjusted in the ADNI GO phase. In 2011, ADNI2 began, including new participant groups: 150 elderly controls, 100 EMCI subjects, 150 late mild cognitive impairment subjects, and 150 mild AD patients. A new cohort, Significant Memory Concern (SMC), was also added in ADNI2 to address the

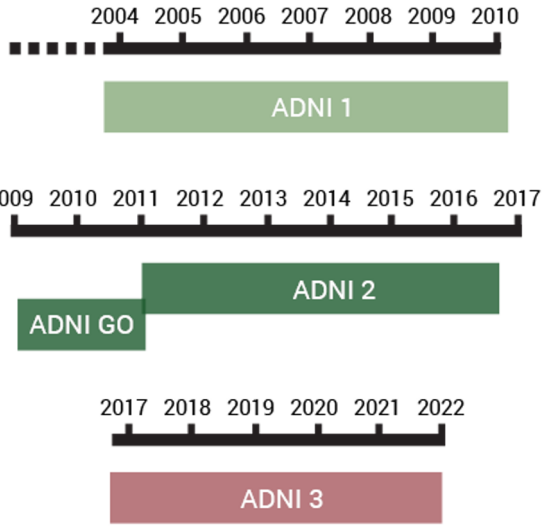


FIGURE 2.3: ADNI phases.

gap between healthy controls and MCI. ADNI3 began in 2016, with the goal to determine the relationships between clinical, cognitive, imaging, genetic, and biochemical biomarkers across the entire spectrum of AD. ADNI3 also adds brain scans that detect tau protein tangles.

Figure 2.3 illustrates the ADNI phases.

2.2 Diffusion Tensor Imaging (DTI)

Diffusion Tensor Imaging (DTI) is an MRI-based neuroimaging technique which estimates the location, orientation, and anisotropy of the brain's white matter tracts. The first steps towards the introduction of DTI date back to 1965, when Stejskal and Tanner measured the water diffusion coefficient using Nuclear Magnetic Resonance (NMR) and applying magnetic field gradients [40]. Compared to diffusion measurements carried out using chemical tracer, the method based on NMR has the advantage to be non-invasive. Moreover, using NMR the water diffusion along an arbitrary and fixed axis can be described. In fact, in order to describe water diffusing freely, no particular axis definition is needed since the diffusion is isotropic. On the contrary, the water diffusion in biological tissues is anisotropic: biological tissues, as muscles and brain, are made of fibres with coherent orientations and water tends to diffuse along the fiber direction. The macroscopic diffusion anisotropy is due to microscopic tissue heterogeneity. In the white matter of the brain, diffusion anisotropy is primarily caused by cellular membranes, with some contribution from myelination and the packing of the axons [41]. The water diffusion depends on the axis along which it is measured and, then, it is possible to identify the fibres orientation by measuring the diffusion along a great number of axes. The combination between diffusion measurements and MRI brought in the '90s to the introduction of a method to calculate, for each voxel of the 3D brain scan, a tensor describing the diffusive motion of water and its degree of anisotropy [8, 42]. This new imaging technique was thus named DTI. Pulsed magnetic field gradients are applied and this makes the magnetic resonance signal sensitive to the diffusion coefficient. The diffusion tensor describes the molecular mobility along the gradient axes and the

correlation between the directions. Using fitting algorithms, the diffusion tensor and its eigenvalues can be calculated for each voxel. Starting from the eigenvalues, it is possible to calculate some informative indexes which express the diffusivity and the degree of anisotropy of the diffusive motion of water. For these reasons, DTI is tailored to investigate the structural integrity of WM fibre tracts.

The importance of the introduction of DTI among neuroimaging techniques consists of two aspects. First of all, conventional MRI cannot reveal detailed anatomy of WM. In fact, MRI signal is related to the density of protons and so to the chemical composition of brain regions. For this reason, it not suitable to highlight the structure of the WM fibers, because of their homogeneous chemical composition. Moreover, after decades of anatomical studies on the human brain, we are still far from the full understanding of its connectivity. There is evidence supporting the biological hypothesis that the decline in neurodegenerative diseases, like AD, can be related to alterations in connectivity among brain regions [5–7]. As it will be explained in the next subsections, differently from the anatomical connectivity based on the correlation between the gray levels of regions and from the functional connectivity, DTI enables the reconstruction of the physical connections of the brain.

2.2.1 The physical principles of DTI

Diffusion is also known as "Brownian motion", referring to the constant random microscopic molecular motion due to heat. At a fixed temperature, the rate of diffusion of a particle can be described by the Einstein equation:

$$\left\langle r^2 \right\rangle \propto Dt, \tag{2.1}$$

where $\left\langle r^2 \right\rangle$ refers to the mean squared displacement of the particle, t is the diffusion time and D is the diffusion constant:

$$D = \frac{kT}{3\pi\eta a}, \tag{2.2}$$

being k the Boltzmann constant, T the temperature, η the fluid viscosity and a the particle diameter.

The signal intensity in conventional MRI is dominated by water concentration (proton density) and by signal relaxation time such as T1 and T2, whereas in DTI the signal intensity is made sensitive to the amount of water diffusion or diffusion constant. For this purpose, a pulsed magnetic field gradient is applied along the three axes x, y, z. In the magnet, the magnetic field, called the B_0 field, is pointing along the magnet bore, which is defined as the z axis. When an x, y, or z gradient is applied, the B_0 strength is modulated linearly along the axis (Fig. 2.4). The strength and polarity (positive and negative) of the gradient can be controlled and, typically, the gradients are turned on for only a short period of time (1–100 ms). The frequency ω of the MR signal is related to the B_0 field strength according to the relationship:

$$\omega = \gamma B_0, \tag{2.3}$$

being γ the gyromagnetic ratio (γ of proton is 2.675×10^8 rad/s/T or 42.58 MHz/T).

When the field gradient is applied, water molecules at different locations begin to resonate at different frequencies. Figure 2.5 shows what happens to MR signals when a pair of positive and negative gradients is applied. After the excitation RF pulse (time t_1), protons at different locations begin to emit MR signals at the same

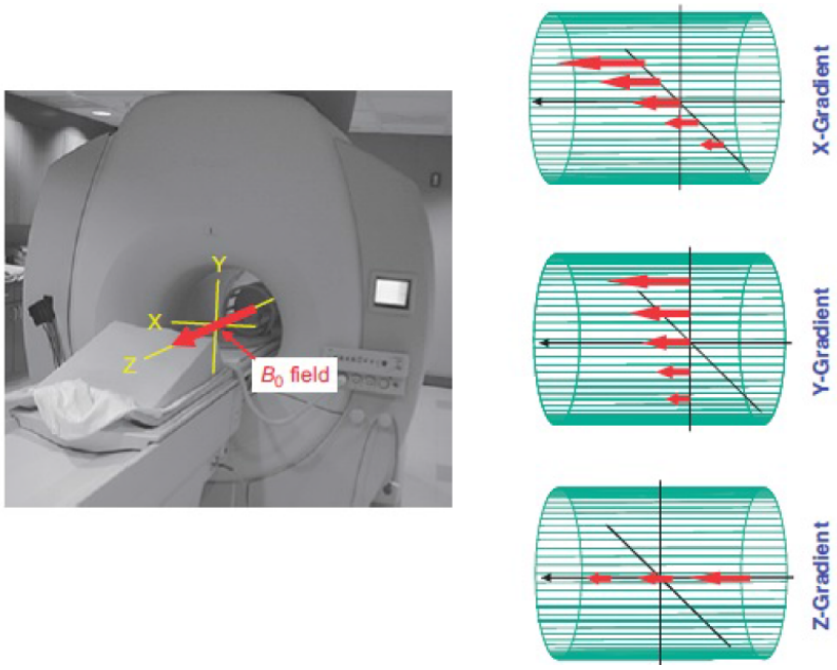


FIGURE 2.4: The main magnetic field, B_0, and orientations of x, y, and z field gradient [43].

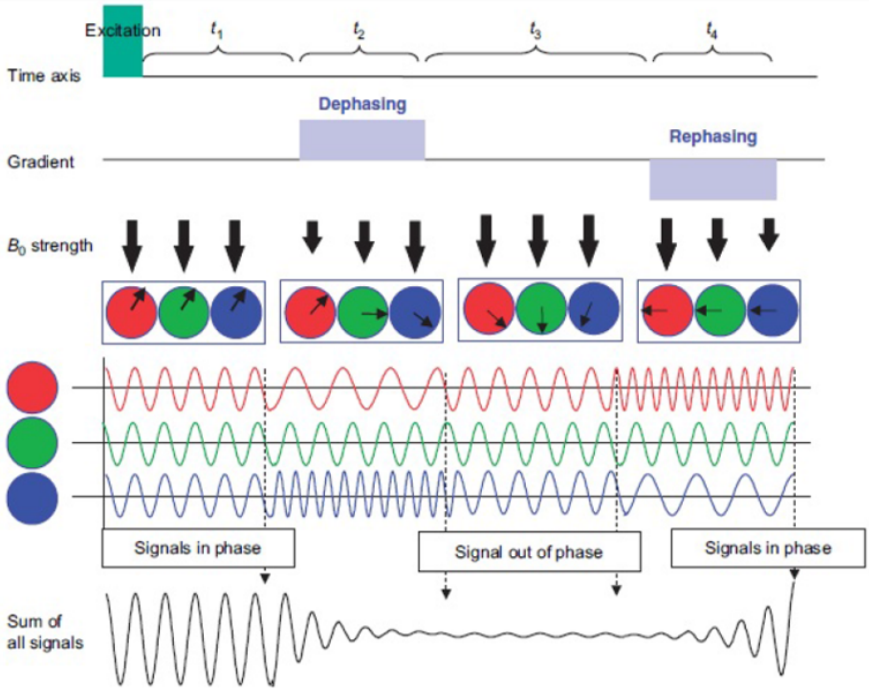


FIGURE 2.5: An example of a dephase-rephase experiment by gradient application [43].

frequency. During the first gradient application (t_2), protons see different magnetic field strengths and, depending on their locations, they start to resonate at different frequencies. In the example of the figure, the red proton sees a weaker magnetic field and rotates slower, while the blue proton sees a stronger field and rotates faster. When the gradient application is stopped, the system returns to a homogeneous B_0

(t_3) but the phases (locations of the rotating arrows) of the signals are no longer identical among the protons. As a consequence, there is a loss of overall signal (the bottom waveform), being it the vector sum of signals emitted from all water molecules. For this reason, the first gradient is called the "dephasing" gradient. During t_3, all protons resonate at the same frequency, but their signal phases remain dephased. The second gradient, applied during t_4, has the opposite polarity, so the red proton starts to rotate faster, joining other protons while the blue one starts to rotate slower. If the strength and length of the first and the second gradient are identical, the protons should regain the same phase at the end of the second gradient. For this reason, the second gradient is called the "rephasing" gradient. The resultant signal is now made sensitive to diffusion. This happens because there could be perfect refocusing only when water molecules do not change their locations in between the applications of the two dephase-rephase gradients. After the application of the rephrasing gradient, those molecules that have moved can be detected because of their different phases compared to the stationary molecules. Thus, DTI works by detecting the imperfect rephasing from the loss of signal intensity. The signal attenuation is S/S_0, where S_0 is the signal intensity without gradients and S is the signal with a pair of gradients.

One of the drawbacks of applying a pair of positive and negative gradient, i.e. a bipolar gradient, with the aim to introduce diffusion weightings into the MR signal, is that a fair amount of signal, due to T_2^* decay between the two gradient pulses (typically 20–40 ms in clinical scanners), can be lost. In order to refocus the T_2^*, a spin-echo sequence is most widely used: after the first 90° radio frequency (RF) pulse, a 180° refocusing RF pulse is employed to reverse the signal phase. In this case, the gradients have the same sign and thus are called "unipolar" gradients (Fig. 2.6).

The effect of diffusion on the MR signal is to attenuate the signal compared to the

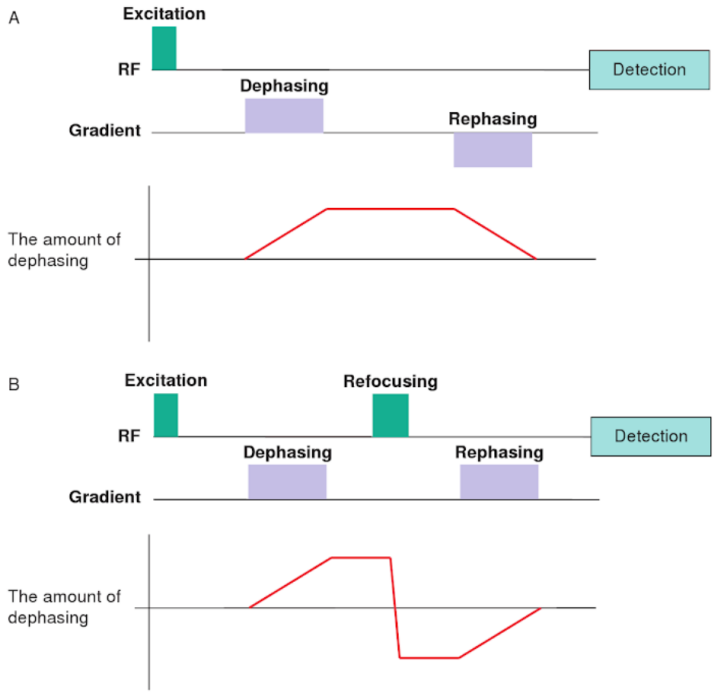


FIGURE 2.6: Comparison of bipolar gradients (A) and unipolar gradients (B) in a spin-echo sequence [43].

original signal with no gradients application. This signal attenuation S/S_0 depends on four parameters:

- The interval Δ between the two pulses (the longer Δ, the greater the chance for the water molecules to move around leading to more signal loss);
- The diffusion constant D (the higher D, the greater the chance for the water molecules to change their position within a fixed amount of time Δ);
- The strength G of gradients;
- The time length δ of gradients application.

Thus, we can write:

$$S/S_0 = f(\Delta, D, G, \delta). \tag{2.4}$$

Among these parameter, Δ, G, δ are experimental parameters that can be controlled; the signal intensities S and S_0 can be measured and thus it is possible to calculate D. Stejskal and Tanner, applying a spin-echo sequence 90°–180° and a pulsed magnetic gradient, derived the expression of the signal attenuation as a function of the four parameters [40]:

$$ln\left[\frac{S(TE))}{S_0}\right] = -\gamma^2\delta^2G^2D\Delta. \tag{2.5}$$

In [44], Tanner defines the effective diffusion coefficient D_{eff} as the temporal mean of D calculated on the echo interval TE, reformulating 2.5 as follows:

$$ln\left[\frac{S(TE))}{S_0}\right] = -bD_{eff} = -\gamma^2\delta^2G^2(\Delta - \delta/3)D_{eff}, \tag{2.6}$$

where b contains information about gradient pulses (timing, amplitude, shape).

In contrast to isotropic media, different effective diffusion constants can be observed in anisotropic media when diffusion gradients are applied in different directions. Therefore, diffusion can no longer be characterized by a single scalar coefficient, but requires a tensor, the effective diffusion tensor $\mathbf{D_{eff}}$, which fully describes the molecular mobility along the three spatial directions and the correlations between these directions:

$$\mathbf{D_{eff}} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix}. \tag{2.7}$$

This tensor is symmetric ($D_{ij} = D_{ji}$, with $i, j = x, y, z$). Equation 2.6 can thus be written:

$$ln\left[\frac{S(TE))}{S_0}\right] = -\mathbf{bD_{eff}}, \tag{2.8}$$

and Diffusion Tensor Imaging consists in the calculation of the diffusion tensor for each voxel of the 3D image.

For each $\mathbf{D_{eff}}$, it is possible to build a local orthogonal coordinate system, the principal coordinate axes along which diffusive fluxes are decoupled. Being $\mathbf{D_{eff}}$ positive and symmetric, its three eigenvectors ϵ_1, ϵ_2, ϵ_3 are orthogonal and the three positive eigenvalues λ_1, λ_2, λ_3 can be obtained by solving the equation:

$$\mathbf{D_{eff}E} = \mathbf{E\Lambda}, \tag{2.9}$$

being $\mathbf{E} = (\epsilon_1, \epsilon_2, \epsilon_3)$ and $\Lambda = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$.

In the brain anisotropic fibrous tissues, the $\mathbf{D_{eff}}$ main directions must be coincident with the orthotropic directions of the tissues. In particular, the eigenvector associated with the eigenvalue of higher value defines the fiber tract axis. In practice, unfortunately, measurements are made in the reference frame $[x, y, z]$ of the MRI scanner gradients, which usually does not coincide with the diffusion frame of the tissue. Therefore, one must also consider the coupling of the nondiagonal elements, b_{ij}, of the b matrix with the nondiagonal terms, D_{ij}, $(i \neq j)$, of the diffusion tensor (now expressed in the scanner frame), which reflect correlation between molecular displacements in perpendicular directions [8]. It is worth to note that, by using diffusion-encoding gradient pulses along one direction only, say x, signal attenuation not only depends on the diffusion effects along this direction but may also include contribution from other directions, say y and z. In order to determine the diffusion tensor fully, a collection of diffusion-weighted images along several gradient directions is needed, using diffusion-sensitized MRI pulse sequences such as echoplanar imaging (EPI). Being the diffusion tensor symmetric, measurements along only six directions are necessary (instead of nine), together with an image acquired without diffusion weighting ($b = 0$). Nevertheless, in clinical practice more than seven images are used, with different strength and orientation of the gradient, in order to obtain a better DTI fitting. In this case, the diffusion tensor is calculated by means of fitting algorithms.

The most intuitive way to understand the information provided by the diffusion tensor is to view it geometrically (Fig. 2.7). The tensor fits the angular variation of

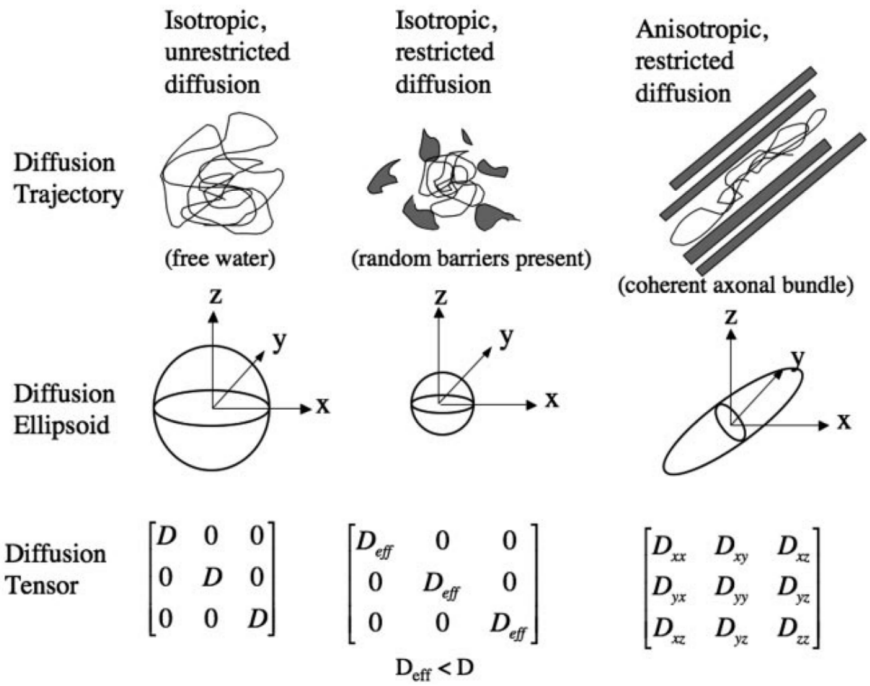


FIGURE 2.7: The figure shows the diffusion ellipsoids and tensors for isotropic unrestricted diffusion, isotropic restricted diffusion, and anisotropic restricted diffusion [45].

the diffusion coefficient values to the shape of a 3D ellipsoid. The diffusion ellipsoid is a shape defined by six variables describing the diffusion coefficient of water molecules in each direction at a particular time. In the case of isotropic diffusion, the diffusion ellipsoid becomes a sphere, because the diffusion coefficient in every direction is equal. Anisotropic diffusion is modeled with an elongated ellipsoid, expressing a greater mean diffusion distance along the longest axis of the ellipsoid. The elements of the tensor above the diagonal are always equal to those below the diagonal (Fig. 2.7, bottom row), a characteristic termed "conjugate symmetry", which reflects the antipodal symmetry of Brownian motion. The diagonal terms of the tensor indicate the magnitude of diffusivity in each of three orthogonal directions. In the case of anisotropic diffusion, the off-diagonal terms of the diffusion tensor indicate the magnitude of diffusion along one direction arising from a concentration gradient in an orthogonal direction.

2.2.2 Scalars derived from DTI data

From diffusion data, it is possible to gain information on tissue microstructure and architecture for each voxel, derived from the three eigenvalues of the diffusion tensor. The three eigenvectors and the eigenvalues λ_1, λ_2, λ_3 describe the directions and lengths of the three diffusion ellipsoid axes, respectively, in descending order of magnitude. The largest eigenvector, called the "primary eigenvector", and its associated eigenvalue λ_1 indicate, respectively, the direction and magnitude of greatest water diffusion. The primary eigenvector is important for fiber tractography algorithms because this vector indicates the orientation of axonal fiber bundles. Therefore, λ_1 is also termed "longitudinal diffusivity", since it specifies the rate of diffusion along the orientation of the fibers. The second and third eigenvectors are orthogonal to the primary eigenvector, and their associated eigenvalues λ_2 and λ_3 give the magnitude of diffusion in the plane transverse to the axonal bundles. Hence, the mean of λ_2 and λ_3 is also known as "radial diffusivity" (RD):

$$RD = \frac{\lambda_2 + \lambda_3}{2}. \tag{2.10}$$

The mean diffusivity (MD) characterizes the overall mean squared displacement of molecules (average ellipsoid size) and describes the directionally averaged diffusivity of water within a voxel. It is defined as the mean of the three eigenvalues:

$$MD = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}. \tag{2.11}$$

The fractional anisotropy (FA) and relative anisotropy (RA) indices measure the degree of directionality of intravoxel diffusivity:

$$FA = \sqrt{\frac{1}{2}} \frac{\sqrt{\left((\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2\right)}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}, \tag{2.12}$$

$$RA = \frac{\sqrt{\left((\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2\right)}}{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}. \tag{2.13}$$

When the primary eigenvalue is much larger than the second and third eigenvalues, anisotropy measures such as FA and RA will be high, indicating a preferred direction of diffusion. FA is basically a normalized variance of the eigenvalues. It is the most widely used anisotropy measure. Its name comes from the fact that it measures the fraction of the diffusion that is anisotropic. This can be thought of as the difference of the tensor ellipsoid's shape from that of a perfect sphere. FA value varies from zero, in the case of isotropic diffusion, up to a maximum of 1, indicating perfectly linear diffusion occurring only along the primary eigenvector.

2.2.3 The approaches on DTI data

DTI is a truly quantitative method providing a framework for acquisition, analysis and quantification of the diffusion properties of white matter. The diffusion coefficient is a physical parameter that directly reflects the physical properties of the tissues, in terms of the random translational movement of the molecules under study. In white matter study, DTI-derived scalar indexes provide information about directionally averaged diffusivity and degree of anisotropy pointing toward myelin fiber integrity. DTI has already shown its potential in the study of white matter architecture and integrity of the normal and diseased brains. Examples of brain diseases include multiple sclerosis [46], brain tumor [47, 48], schizophrenia [49], stroke [50] and so on. In particular, we are interested in the application of DTI in the study of AD. As neurodegeneration is accompanied by a progressive loss of barriers that restrict water molecule motion, MD increases pathologically, whereas FA decreases. Accordingly, DTI studies in AD patients have shown abnormally increased MD and reduced FA in Gray Matter (GM) and WM including regions as the cingulate gyrus, the genu and the splenium of corpus callosum, hippocampus, amygdala and thalamus [51, 52]. Diffusivity-based measures in this context are used for statistical studies addressed to investigate significant microstructural changes in AD and MCI compared to HC. The values of DTI-based indexes are also used as features to train classification models for the automatic diagnosis. In literature, three approaches are used to address these issues: region of interest (ROI)-based approach, voxel-based approach and network-based approach.

ROI-based approach DTI images are segmented using a parcellation scheme which relies on a standard brain atlas (for example, the AAL or the Harvard Oxford). Thus, the brain is parcelled into ROIs and the mean of the DTI measures are then calculated for each ROI, for example mean diffusivity, fractional anisotropy or some measures defined combining different diffusive indexes. The DTI scalar indexes averaged over each ROI are used as features for feeding machine learning algorithms to classify AD subjects also at early stages of the disease and for investigating WM integrity alterations. Several studies based on this approach have been conducted, e.g. [53, 54], also in multimodal analysis [55].

Voxel-based approach The DTI-based measures for each voxel derived from tensor fitting, constitute a map. The FA maps of a cohort of subjects are averaged and, using the Tract-Based Spatial Statistics algorithm described in the next subsection, a white matter "skeleton" is obtained, containing WM tracts common to all subjects. The FA map of each subject is projected onto the mean FA skeleton obtaining the FA skeleton of each subject. The FA or MD (or another index) values of the voxels belonging to the skeleton represent the features for feeding classification algorithms and for performing voxel-wise statistical analyses aimed at localizing brain changes

related to development, degeneration and disease. The classification performances obtained using this approach are the best in literature when using the single DTI modality [56, 57]. Nevertheless, these studies suffer from the so called *double dipping* due to the application of a feature selection algorithm *non nested* in the cross validation. The analysis reported in the next chapter will address this issue.

Network-based approach This approach on DTI data is the most recent and less investigated, especially for classification purposes. Indeed, an important potential application of DTI is the visualization of anatomic connections between different parts of the brain. By using DTI fiber tracking algorithms, which will be better explained in the next section, together with a parcellation scheme, the brain can be modeled has a network and its connectivity can be studied applying graph theory. Network measures can then be calculated to characterize the healthy brain and the connectivity alterations due to AD. The network measures are also the features that can be used to build classification models.

The analyses carried out in this thesis will consider the voxel-based and the network-based approaches.

2.3 Voxel-based analysis

FA has became the most widely used DTI-based index to provide information about anatomical connectivity in the brain, because it quantifies how strongly directional the local tract structure is. However, the use of standard registration algorithms and voxel-based morphometry (VBM), in which each subject's FA map is registered into a standard space before performing voxel-wise statistics, can compromise optimal voxel-wise statistical analysis. A solution to the problem of how to align FA images of multiple subjects, in a way that permits to extract features in order to reach valid conclusions from the voxel-wise analysis, was proposed by Smith et al. in 2006 [58] with the introduction of the Tract-Based Spatial Statistics (TBSS) algorithm. This method aims to improve the sensitivity, objectivity and interpretability of analysis of multi-subject diffusion imaging studies.

2.3.1 Tract-Based Spatial Statistics

There has been much debate about the strengths and limitations of VBM [59, 60]. Limitations include problems due to alignment inaccuracies, and the lack of a principled way for choosing smoothing extent. The TBSS method is fully automated and solves the alignment and smoothing problems investigating the "whole" brain. This is achieved by estimating a "group mean FA skeleton", which represents the centres of all fibre bundles that are generally common to the subjects involved in a study. After this step, each subject's FA map is then projected onto the mean FA skeleton so that each skeleton voxel takes the FA value from the local centre of the nearest relevant tract, in order to hopefully solve issues of alignment and correspondence [58]. TBSS is provided with the FSL software library. The main steps of TBSS can be summarized as follow:

- The FA images are eroded slightly and the end slices are zeroed (to remove likely outliers from the diffusion tensor fitting).

- A nonlinear registration is applied, aligning all FA images to a standard target. The target image used in the registrations can either be a pre-defined target, or can be automatically chosen to be the most "typical" subject in the study. In FSL, it is recommended to use the FMRIB58_FA standard-space image as the target for TBSS. This involves carrying out just one registration per subject and generally gives good alignment results.
- All subjects are affine-aligned into a standard space (usually MNI152). After this step, each subject's FA image has the nonlinear transform to the target and then the affine transform to the standard space is applied, resulting in a transformation of the original FA image into the standard space. The subjects' images are all merged into a single 4D and the mean of all FA images is created and fed into the FA skeletonisation step.
- The tract skeleton generation aims to represent all tracts which are "common" to all subjects. The skeleton will represent each tract as a single line (or surface) that run down the centre of the tract. Away from the centre surface or line, the FA values fall off gradually and become very low moving out of white matter. This skeletonisation step consists first in the estimation of the local surface perpendicular direction (at all voxels in the image) and then in a procedure of non-maximum suppression in this direction. In other words, a search is made along all voxels in the local "tract perpendicular direction", and the voxel with the highest FA is identified as the centre of the tract. The mean skeleton is thresholded to exclude voxels belonging to gray matter or cerebrospinal fluid, as well as voxels from the outermost part of the cortex, which are zones of greater inter-subject variability. A threshold value between 0.2 and 0.3 is recommended [58].
- Each subject's (aligned) FA image is projected onto the skeleton extracting the WM tracts for each subject. This last step of TBSS is best described in the following subsection.

2.3.2 Voxel-based feature extraction

Each subject's aligned FA image is projected onto the mean FA skeleton. An important issue in this step is to account for residual misalignments between subjects after the initial nonlinear registration. In order to address this issue, at each point in the skeleton, a given subject's FA image is searched in the perpendicular tract direction in order to find the maximum FA value and assign this value to the skeleton voxel. This actually allows to achieve alignment between the skeleton and this subject's FA image without needing perfect nonlinear pre-registration. Therefore, even if there is any systematic difference in exact tract location between groups of subjects, it will not affect the comparison of FA values between the groups in statistical analysis. By means of this procedure, for each subject a vector of features corresponding to the values of a diffusive map in the skeleton voxels is available, and it can be used for voxel-wise statistics across subjects and for building classification models.

2.4 Brain network analysis

The human brain is one of the most complex systems in nature. In principle, all possible combinations of its basic elements (neurons and synapses) should be considered to describe its static and dynamic properties. However, this approach would

be computationally unfeasible; moreover, it would ignore that these properties can emerge from the behaviour of large-scale brain areas. Complex networks provide a powerful framework to model the brain regions and their interactions.

2.4.1 Complex networks

Many real-world systems are composed by a large number of elements characterized by highly dynamic interconnections. Biological and chemical systems, social networks, the Internet and the World Wide Web are only a few examples.

Graph theory is usually considered the most appropriate framework for the mathematical treatment of such complex systems: a complex network is represented as a graph in which nodes are the elements of the system and edges represent the interactions between them [61]. If only the presence or absence of a connection between two elements is considered, the graph $G(N, L)$ consists of a set of N nodes connected by a set of L unweighted links. Then, a graph G can be completely described by the adjacency matrix A, which is an $N \times N$ square matrix whose elements a_{ij} are equal to 1 if the link l_{ij} exists, or 0 if the link is absent. If the weight w_{ij} of the link l_{ij} is taken into account, the graph $G(N, L, W)$ is fully described by the weights matrix W, which is an $N \times N$ square matrix whose element w_{ij} is the weight of the edge l_{ij}. A subgraph $G'(N', L')$ of $G(N, L)$ is a graph such that $N' \subseteq N$ and $L' \subseteq L$. The usual way to picture a graph is by drawing a dot for each node and joining two dots by a line if the two corresponding nodes are connected by a link.

A central concept in graph theory is the notion of reachability of two different nodes in a graph. In fact, two nodes that are not adjacent may be reachable from one to the other. A *walk* from node i to node j is an alternating sequence of nodes and edges starting from i and ending to j. The length of the walk is the number of edges in the sequence. A *path* is a walk in which no node is visited more than once. The walk of minimal length between two nodes is known as *shortest path*.

2.4.2 Network measures

Several traditional metrics used in graph theoretical analysis are based on the shortest path length d_{ij}, defined as the number of edges along the shortest path connecting node i to j [61]. For a weighted network, the length λ_{ij} of the edge l_{ij} is defined as $\lambda_{ij} = 1/w_{ij}$ and the weighted shortest path length d_{ij} is the smallest sum of the edge lengths throughout all the possible paths in the graph from i to j.

The edge betweenness is defined as the number of shortest paths between pairs of nodes that run through that edge [62]: it gives information about how the relation between two nodes is important for the "communication" between all nodes in the network. The concept of betweenness can also be extended to nodes, as a measure of node centrality, giving the importance of a node for the information flow across the network. More precisely, the betweenness b_i of a node i, is defined as:

$$b_i = \sum_{j,k \in N, j \neq k} \frac{n_{jk}(i)}{n_{jk}}, \tag{2.14}$$

where n_{jk} is the number of shortest paths connecting j and k, while $n_{jk}(i)$ is the number of shortest paths connecting j and k and passing through i.

Other traditional network metrics are defined without the use of the shortest path length. The degree k_i of a node i in an unweighted network is the number of

edges crossing i:

$$k_i = \sum_{j \in N} a_{ij}. \tag{2.15}$$

For a weighted network, the concept of degree is replaced by the concept of node strength s_i, defined as:

$$s_i = \sum_{j \in N} w_{ij}. \tag{2.16}$$

The clustering coefficient c_i of node i, as introduced by Watts and Strogatz for an unweighted network [63], is defined as the ratio between the actual number of edges (denoted by e_i) in the subgraph G_i of neighbors of i and $k_i(k_i-1)/2$, the maximum possible number of edges in G_i:

$$c_i = \frac{2e_i}{k_i(k_i-1)} = \frac{\sum_{j,m} a_{ij}a_{jm}a_{mi}}{k_i(k_i-1)}. \tag{2.17}$$

The clustering coefficient measures the density of links connecting a node with its neighbors. If the neighbors of a node are also interconnected, they form a cluster. For a weighted network, the clustering coefficient of a node i, $C^w(i)$, can be expressed as follows [64]:

$$C^w(i) = \frac{2}{k_i(k_i-1)} \sum_{j,k} (\bar{w}_{ij}\bar{w}_{jk}\bar{w}_{ki})^{1/3}. \tag{2.18}$$

Finally, the eigenvector centrality e_i of a node i is given by the sum of the values within the principal eigenvector e corresponding to direct neighbors, as defined by the adjacency matrix, then scaled by the proportionality factor $1/m$ where m denotes the largest eigenvalue:

$$e_i = \frac{1}{m} \sum_{j \in N} a_{ij}e_j. \tag{2.19}$$

The same definition holds for weighted graphs but the adjacency matrix is replaced by the weights matrix. The eigenvector centrality is an iterative centrality in which the influence of a node is determined by the number and influence of its neighbors.

During the last twenty years, the grown availability of large datasets, as well as the development of more powerful computing tools, have constituted a better framework to explore the topological properties of several network systems from the real world. This allowed researchers to study the topology of the interactions in a large variety of systems. The main outcome of this research was to reveal that, despite the inherent differences, most of real-world networks are characterized by the same topological properties, as for instance small characteristic path lengths and high clustering coefficients. These features make these networks radically different from regular lattices and random graphs, which are the standard models studied in mathematical graph theory.

2.4.3 From brain to networks

Neuroimaging provides the most effective way to map the human brain to a network. With the progression of advanced neuroimaging techniques, it is possible to map the brain as a complex network at the macro-scale level, consisting of a set of nodes (representing voxels, regions or sensors) and a set of connections between the nodes (representing WM pathways, structural, or functional pairwise relationships). Starting from neuroimaginga data, depending on the kind of data, technique and information to be studied, a specific pipeline is followed to get from brain to

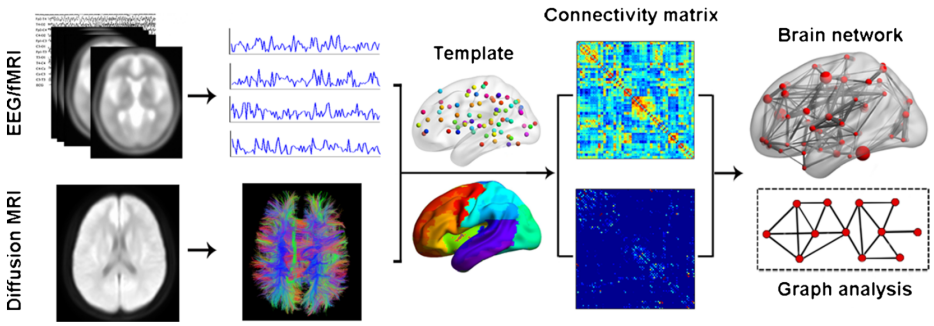


FIGURE 2.8: A flowchart of the construction and analysis of functional and structural brain networks [65].

network (Fig. 2.8): the time series from the EEG/fMRI data or the fiber pathways from diffusion MRI data are first extracted. The brain regions are then parcelled by structurally or functionally defined templates. The individual connectivity matrices are generated by considering the pairwise functional or structural associations between brain regions. Thus, the brain network is obtained and represented as a graph and its topological properties can be studied with graph theoretical approaches.

The first issue to be addressed is the definition of nodes and edges.

Nodes The first step consists in defining the nodes of the network. Anatomical parcellation schemes are commonly used to define a partition of the brain. Anatomical atlases, such as Anatomical Automatic Labeling [66] and Harvard-Oxford probabilistic atlas, are the most diffused in neuroscience studies. These partitioning methods define networks of moderate resolution (approximately 100 nodes). Large scale networks can be obtained by considering each voxel of the brain as a node. This method leads to very high-resolution networks (about 10^4 nodes) and the analysis of such networks requires extensive preprocessing to filter out the large amount of noise and data redundancy. In several fMRI studies a set of fixed regions of interest (ROIs) with coherent activation patterns are identified, e.g. [67]. In physiological recordings, sensors and electrodes naturally define the nodes of the network.

Edges There are many different ways to choose the edges of the brain network, depending on the kind of imaging data. In structural MRI connectivity, correlations in several morphometric features or some characteristics of the fiber tracts linking areas of the brain, are used to estimate connectivity matrices, in particular by computing the correlations in the thickness or volume of gray matter between two cortical areas [68]. In functional connectivity analysis, statistical dependencies between distributed neuronal units are commonly estimated by various measures such as correlation, covariance and spectral coherence between pairs of time series. Pairwise direct associations obtained from statistical modeling and other time series analysis are used to infer effective connectivity [69]. In DTI, the connectivity is defined by the number of the reconstructed axonal fibers linking two nodes; the fiber reconstruction is obtained by means of tractography.

Tractography DTI fiber tracking uses the diffusion tensor to track fibers along their whole length. It determines intervoxel connectivity on the basis of the anisotropic diffusion of water. Starting from a seed ROI, generally defined manually, the fiber tracking algorithm looks for adjacent voxels whose main diffusion direction is in the

continuity of the previous one. In each brain voxel, the dominant direction of axonal tracts can be assumed to be parallel to the primary eigenvector of the diffusion tensor. Fiber tracking uses the diffusion tensor of each voxel to follow an axonal tract in 3D from voxel to voxel through the human brain. In clinical practice, the most tracked fiber bundle is the cortico-spinal tract. However, fiber tracking can identify most of the brain's white matter tracts. DTI fiber tracking algorithms can be divided into deterministic and probabilistic methods. Fiber assignment by continuous tracking (FACT) is a deterministic method, which begins fiber trajectories from user-defined voxels [70]. Fiber trajectories are named "streamlines" and follow the primary eigenvector from voxel to voxel in three dimensions. When the fiber trajectory reaches the edge of the voxel, the direction of the trajectory is changed to match the primary eigenvector of the next voxel. Noise, patient movement, and distortion from imaging artifacts produce uncertainty in the orientation of the diffusion ellipsoid and are detrimental to deterministic streamline fiber tracking [71]. Probabilistic fiber tracking methods incorporate the expected uncertainty into the tracking algorithm and can be used to produce a connectivity metric for each voxel. The probability density function of the orientation of a neuronal fiber can be estimated with an empiric function based on FA [72], a Bayesian model [73], or bootstrap statistics [74]. Probabilistic DTI fiber tracking techniques tend to disperse trajectories more than deterministic methods and have the potential to delineate a greater portion of a white matter tract. However, the accuracy of these probabilistic methods is still limited by the information contained in the diffusion tensor and the method of constructing the probability density function. Another fundamental limitation of diffusion tractography is that it cannot distinguish antegrade from retrograde along a fiber pathway, due to the antipodal symmetry inherent to the diffusion process [45]. The diffusion tensor imaging model assumes that, in each voxel, there is a unique orientation of the fibers, the direction of which is represented by the tensor's main eigenvector [75]. This assumption is not valid in case of crossing fibers. The term "crossing fibers" generally refers to regions in which the fibers' orientation is not unique, i.e. when the fibers are interdigitating, brushing past each other, curving, bending or diverging [75]. Higher order models, such as constrained spherical deconvolution, have been developed in order to address the issue of crossing fibers. Functions useful to develop tractography pipelines are provided with software libraries, such as FSL (`https://fsl.fmrib.ox.ac.uk/fsl/fslwiki`) and MRtrix3 (`http://www.mrtrix.org/`).

2.5 Statistics and machine learning in neuroscience

The complexity and the amount of data at disposal calls for algorithmic, automated methods for data analysis. They are provided by statistical and machine learning techniques.

Statistical analysis of the human brain connectivity mainly involves mass univariate hypothesis testing about local connectivity properties, i.e. at nodal or edge level, to identify specific brain regions that are different between the groups under investigation (unhealthy and controls). Although, on one hand, performing the same hypothesis simultaneously on many elements within the brain allows one to localize accurately the brain areas significantly altered or most affected by the disease, on the other hand, it also leads to a multiple comparison problem. The number of comparisons can be reduced through a priori assumptions that limit the analysis

to a subset of areas. Alternatively, the problem of multiple comparisons can be tackled with statistical procedures such as false discovery rate (FDR) [76] or more recent correction techniques which take into account spatial correlation of brain areas [77].

While statistical approaches based on null hypothesis testing aim at finding in-sample statistical differences, machine learning and, in particular, classification models provide a way to predict whether a single subject, previously unseen, belongs to one of the groups under study. In addition, statistically significant single features are not likely to provide high predictive accuracy [78]; multivariate patterns, instead, that are those derived from specific combinations of features, may result in better predictions when testing different configurations of the feature space [79].

Generally speaking, machine learning represents a set of methods to automatically detect patterns in the data: these patterns are used to provide predictions on future data or to support decision making under uncertainty [80]. There exist many types of machine learning: since, in the present thesis, data are provided with labels stating the current diagnosis or stage of disease, we focus only on *supervised* approaches. In supervised learning, the goal is to learn a mapping from inputs x to outputs y, given a labeled set of input-output pairs (x_i, y_i), which constitutes the so-called training set. Each training input x_i can be a high dimensional vector of elements each representing a feature that describes x_i. When the label is a categorical variable from a finite set, i.e. $y_i \in \{1, \ldots, C\}$, the problem is known as a *classification* problem: this is the setting we are going to encounter. If $C = 2$, this is called binary classification; otherwise, it is a case of multiclass classification.

One way to formalize the problem is to assume there is a relationship between y and x, which can be written in the general form $y = f(x) + \epsilon$. The ϵ term is an irreducible error which typically expresses the noise in the data, so the problem reduces to a function approximation problem whose aim is to find an estimate $\hat{f}$ that best approximates the unknown function f. In this view, machine learning refers essentially to a set of algorithms to estimate f. The training data are used to "train" the algorithm how to approximate the true model of the data, and the final goal is to make predictions on the label of novel inputs, that are examples the algorithm has never seen before. In other words, the machine learner should be able to *generalize* beyond the sample of observations it has been given as input. Therefore, to evaluate the predictive power of the model the algorithm has learned, a separate test set is usually used. When the entire dataset is small, splitting validation strategies, such as k-fold cross-validation, are preferred.

In this thesis, some state-of-the-art supervised classification models have been used: they are described in the following paragraphs. Note that, despite they come to prominence in the last years, we did not make use of neural networks and deep learning techniques. This because of two main reasons. First, they typically require sets of data much more larger than those we have used. Second, these techniques work by constructing hierarchies of features at high level of abstraction starting from features at a very low level. The features automatically learned can be very difficult to interpret. Instead, one of our main goals is to design meaningful features that not only are amenable to learning but can help characterize and study the mechanisms and patterns underlying the degeneration process due to the disease.

Naïve Bayes This is a *generative* algorithm that directly models the class conditional distribution for each class $c \in C$ [81]. Predictions on unseen examples are

obtained by applying the Bayes rule and by outputting the class for which the following estimated probability is the largest:

$$\Pr(Y = c \mid X = x) = \frac{\Pr(X = x \mid Y = c)\Pr(Y = c)}{\Pr(X = x)}.$$

$\Pr(X = x)$ is simply a scale factor; $\Pr(Y = c)$ can be easily calculated as the fraction of the training examples that are in class c; $\Pr(X = x \mid Y = c)$ can be estimated in several ways, in particular, in Naïve Bayes (NB), a univariate Gaussian density for each class and conditional independence among classes are assumed. Despite its strong assumptions, which are seldom true, NB is remarkably successful in practice.

Support Vector Machines Support Vector Machines (SVMs) work by constructing a separating hyperplane between two classes so that the minimal distance from the closest data points of either classes is the largest [82]. Test examples are predicted to belong to a class based on which side of the hyperplane they fall. To mitigate overfitting, the *margin* of the hyperplane is chosen so that most of the training examples are separated correctly, while some of them are misclassified. To learn nonlinear decision boundaries, the data points are mapped to a higher dimensional space via a kernel function: a popular choice is the radial basis function (RBF) kernel. In the present thesis, we employed both the linear kernel and the RBF kernel. Note that the bias-variance trade-off of the algorithm depends on the fine tuning of the penalty parameter C and the kernel coefficient γ in the case of RBF kernel [81]. We set $C = 1$ and $\gamma = \frac{1}{n}$, with n being the number of features, which are values commonly found in the literature.

Random Forest Random Forest (RF) is an *ensemble* method for classification which builds a multitude of decision trees at training time and outputs the mode of the classes predicted by each individual tree at test time [83]. In particular, RF repeatedly (B times) selects a random sample with replacement from the training set and fits a decision tree to this sample. Every decision tree is built on a subset of randomly selected features: typically, for classification problems with m features, $\sqrt{m}$ features are chosen. The reason for doing this is to reduce the high correlation between similar trees. The final predictions are obtained via majority voting of the B single predictions. B is a free parameter: in the experiments reported in this thesis, we chose B equals to 300 or 500 trees, which represent typical sizes for a forest.

2.6 Computational infrastructure

All the analyses here presented have been done within the MATLAB (MathWorks) computing environment. In addition, it is worth remarking that these analyses required a huge computational burden with preprocessing and diffusion tensor fitting time of about one hour per subject, and tractography time of more then ten hours per subject. To this end, the elaborations have been carried out on the distributed computing infrastructure ReCaS-Bari computing farm (`https://www.recas-bari.it`). This data center has been built by the ReCaS project, funded by the Italian Ministry of Education, University and Research to the University of Bari and INFN (National Institute for Nuclear Physics), whose goal is to empower preexisting computing infrastructures located in Catania, Cosenza, Napoli and Bari. In particular, the data center offers 128 servers, 64 cores per server, for a total amount of 8192 new cores,

reaching 12000 cores with the old computing farm. Each new server hosts 256GB RAM, 4GB RAM per core per server. Additionally, it offers about 3.5PB of disk space and 2.5PB of tape space.

To implement our analysis on the distribute infrastructure, we used the LONI Pipeline (LP), one of the most used workflow manager for medical image processing developed by the Laboratory of Neuro Imaging (http://pipeline.loni.usc.edu/). LP is widely used by the scientific community since it has proved to be a convenient and powerful tool. In particular, XML resource description facilitates the integration of disparate resources and provides a natural and comprehensive mechanism to support data provenance. It also enables the broad dissemination of resource metadata descriptions via web-services and the constructive utilization of multidisciplinary expertise by experts, novice users and trainees. Our group developed a general approach to submit and monitor LP workflows on distributed infrastructures [84]. This framework is based on a meta-scheduler, the Job Submission Tool (JST) [85], that is able to submit jobs to different computing architectures, exposing to the end users only a simple Web Service interface based on the Representational State Transfer (REST) protocol.

Chapter 3

Assessment of local changes due to AD through voxel-wise analysis

3.1 The feature selection bias in voxel-based DTI studies

In recent years, DTI has revealed itself as a very promising imaging modality to discriminate between HC subjects, AD patients and subjects with MCI. An analysis approach commonly found in literature consists in the computation of FA and MD maps (or other diffusion indices), followed by the identification of the most representative voxels; these voxels are then fed into machine learning algorithms to automate the classification.

For the discrimination HC/AD, Mesrob et al. [53] adopted a Support Vector Machine (SVM) classifier and a region of interest (ROI)-based approach; Dyrba et al. [55] used a ROI-based approach and a multimodal SVM combining DTI indices with gray matter volume derived from structural MRI; Schouten et al. [86] used a ROI-based approach in combination with Elastic Net Regression. For the classification HC/MCI, Cui et al. [87] used subcortical volumetric features extracted using a segmentation algorithm together with FA values obtained for white matter regions of interest. Dyrba et al., in [54], used a ROI-based approach and SVMs on a multicentric dataset and apply variance reduction methods.

The best performances in literature for the HC/MCI classification, using a single DTI modality, can be found in Haller et al. [56] and O'Dwyer et al. [57]. In these works, a voxel-based approach is used considering as features the voxel intensities in the diffusion maps. However, as also remarked in [57], in each of the above mentioned works, the methodological procedure relies on an *a priori* feature selection performed on the entire dataset to be analyzed. This procedure, also known as *non-nested* feature selection, circular analysis or double dipping, chooses the most discriminative voxels by using also the test set, thus introducing a bias in the classification model. A non-nested feature selection necessarily leads to overestimate the numerical values of accuracy and area under the ROC curve (AUC). On the contrary, a nested feature selection is obtained when the selection procedure is performed blind to the test set.

The practice of double dipping and its dangers are well known to the statistics and computer science community, and have been extensively described in the literature [88, 89]. Although recommendations and best practices are available [90], the field of neuroimaging is still widely populated by studies that noticeably perform non-nested feature selection, claiming classification performances close to perfect accuracy. The effects of double dipping on classification performances in neuroimaging studies have been quantitatively assessed when dealing with functional brain data, such as fMRI [90] or MEG [91], and with data derived from structural T1-weighted MR imaging (cortical thickness) in [92]. However, some of the image

classification studies involving DTI cited above seem to be affected by such feature selection bias and this first analysis of the thesis is aimed at investigating to which extent the reported performances are inflated by its presence.

In the analysis described in this chapter, DTI images are used for classification tasks in AD; the main aim of this analysis is to perform a comparative study between nested and non-nested feature selection on the same data set. This is the first study attempting to measure the bias introduced by non-nested feature selection, from now onward feature selection bias (FSB), in the classification of DTI images with a fair comparison, i.e., measuring the effect on the same fixed data set. Secondly, a comparative study has carried out between different state-of-the-art classification models and their efficiency to provide a robust support to the AD diagnosis.

3.2 Data

The images analyzed for this study are diffusion-weighted scans of 150 subjects (50 HC, 50 AD patients and 50 MCI), both males and females, aged 55 to 90, from the ADNI-GO and ADNI-2 phases. Scans were randomly selected from baseline and follow-up study visits. HC subjects show no signs of depression, mild cognitive impairment or dementia; participants with AD are those who meet the NINCDS-ADRDA criteria for probable AD; MCI subjects have reported a subjective memory concern, but without any significant impairment in other cognitive domains: they substantially preserved everyday activities with no sign of dementia. Two MCI levels (early or late) are usually distinguished according to the Wechsler Memory Scale Logical Memory II. For this study, we used a balanced group of 25 early and 25 late MCI, but these labels were not taken into account in the classification tasks.

To evaluate the methodology presented in this chapter on an independent test set, a second different set of scans from ADNI was also considered, consisting of 40 HC, 40 MCI (22 early and 18 late) and 39 AD. This second test set included both male and female subjects, and was age-matched with the training sample.

Diffusion-weighted scans were acquired using a 3 T GE Medical Systems scanner with 41 gradient directions ($b = 1000\ \mathrm{s/mm^2}$); in addition to these, 5 images with negligible diffusion effects (B_0 images) were acquired as reference scans for subsequent analysis. Figure 3.1 shows an example of B_0 image while Fig. 3.2 shows an example of diffusion weighted image.

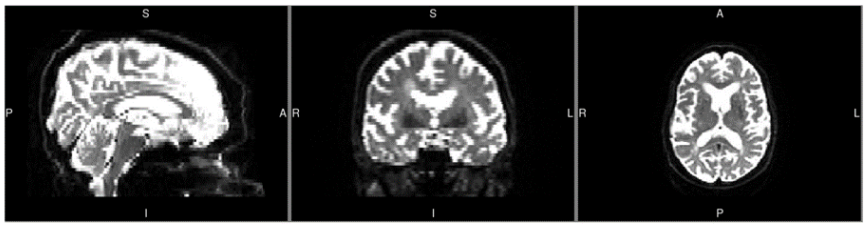


FIGURE 3.1: B_0 image example.

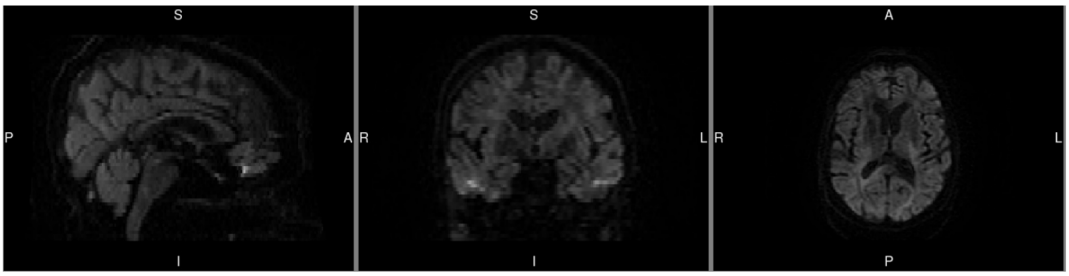


FIGURE 3.2: *Weighted* image example.

3.3 Methods

3.3.1 Image processing

Diffusion-weighted images were preprocessed using the FMRIB Diffusion Toolbox, included in the FSL software [93]. Preprocessing comprised:

1. Conversion to Nifti format;
2. Extraction of gradient directions and b-values;
3. Correction for eddy currents and head motion;
4. Skull-stripping using the Brain Extraction Tool (BET) (Fig. 3.3).

After preprocessing, a single diffusion tensor was fitted at each voxel in the image, using DTIfit. From the diffusion tensor, fractional anisotropy (FA) and mean diffusivity (MD) were then calculated. By definition, these two invariants are related to the eigenvalues of the diffusion tensor λ_1, λ_2, λ_3 by [8, 94]:

$$FA = \sqrt{\frac{1}{2}} \frac{\sqrt{\left((\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2\right)}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}, \tag{3.1}$$

$$MD = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}. \tag{3.2}$$

After diffusion tensor fitting, FA and MD maps need to be carefully aligned to a group-wise space before any voxel-wise statistical analysis is carried out; in addition to this, it is desirable to restrict the analysis only to voxels belonging to white matter fiber bundles. All this was achieved by means of the Tract-Based Spatial Statistics (TBSS) algorithm implemented in FSL [58]. TBSS performs the following steps:

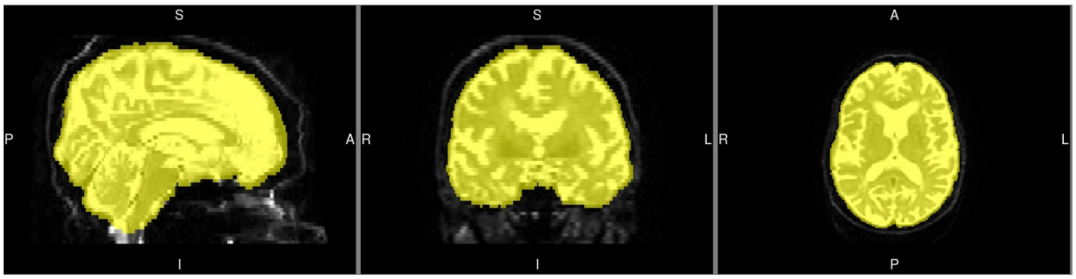
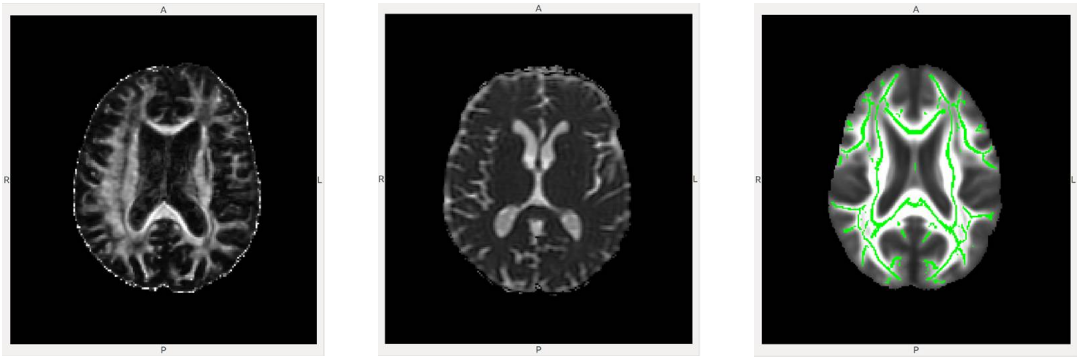


FIGURE 3.3: Binary mask obtained with BET overlaid on the B_0 image.

- Identify a common registration target (it can be either a mean FA template provided with the software or the most "representative" subject of the cohort) and apply nonlinear registration to align all subjects' FA maps to the selected target. The chosen target was the FMRIB58_FA standard-space FA template, generated by averaging 58 FA images from diffusion MRI data, in MNI152 space.
- After the nonlinear registration, the entire aligned dataset undergoes an affine transformation to bring it into $1 \times 1 \times 1$ mm^3 MNI152 space. Then, a mean FA image is created, averaging all the FA maps in the dataset, and the result is used to generate a mean FA skeleton of white matter fibre tracts common to all subjects. The mean skeleton is thresholded to exclude voxels belonging to gray matter or cerebrospinal fluid, as well as voxels from the outermost part of the cortex, which are zones of greater inter-subject variability. Figure. 3.4 shows an example of FA map (3.4a) and MD map (3.4b), and the FA skeleton mask overlapped onto the mean FA map (3.4c).



(A) Example of FA map (B) Example of MD map (C) Mean FA skeleton

FIGURE 3.4: From left to right: (a) a fractional anisotropy (FA) map and (b) a mean diffusivity (MD) map. For all subsequent analyses both maps are projected onto the mean FA skeleton (c).

3.3.2 Feature extraction, feature selection and classification

All subjects' FA images are projected onto the mean FA skeleton, achieving an alignment between subjects in the direction orthogonal to the fibre bundle orientation. This is the last step of TBSS and, as a result, the skeleton of main white matter fibre tracts was extracted from each subject, together with the corresponding values of FA and MD at each voxel in the skeleton. Approximately 120000 voxels for each subject map were projected onto the skeleton. These are the features to be used to train the classification models.

The following stage aimed at assessing which voxels are most significant for the purpose of discriminating HC from AD and MCI. It is important to note that it is not possible to rely on any assumption about the distribution of the test statistic under the null hypothesis; this implies that any statistical test has to be non-parametric. Wilcoxon rank sum test and the ReliefF algorithm were used both within a non-nested and nested approach. A Wilcoxon test compares the medians of the groups of data to determine if the samples come from the same population, and returns a p-value for the null hypothesis that samples are drawn from the same population [95], [96]. Then voxels are ranked and selected by thresholding on p-values. The basic

principle of ReliefF [97, 98] is to estimate features according to how well their values distinguish among data instances close to each other. Features are then ranked and sorted in order of decreasing importance.

For each classification task, fifteen reduced datasets were created by selecting an increasing number of most discriminating voxels, depending on the feature selection's output: 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 750, 1000, 2000 and 3000 voxels.

In this analysis, considering the profitability of using classification trees in the context of machine learning techniques applied to AD [99, 100], a Random Forest approach for the learning and classification phase was used. Since a few hundreds of samples represent the typical size of the forest, in this study a value equal to 300 for B was chosen. To determine the classification performance of the Random Forests classifier, a 100 times repeated 5-fold cross-validation for each reduced dataset was adopted. More precisely, every subject was shuffled into one of five folds from which one fold was selected as the test set, while the remaining folds form the training set. The subjects were stratified by diagnosis, such that each fold contained the same number of subjects from each diagnostic group. The classification process was repeated until each of the five folds was used as test set once. Finally, the full cross-validation procedure was repeated 100 times, using different permutations, to shuffle the subjects into the folds for a more general approximation of the performance.

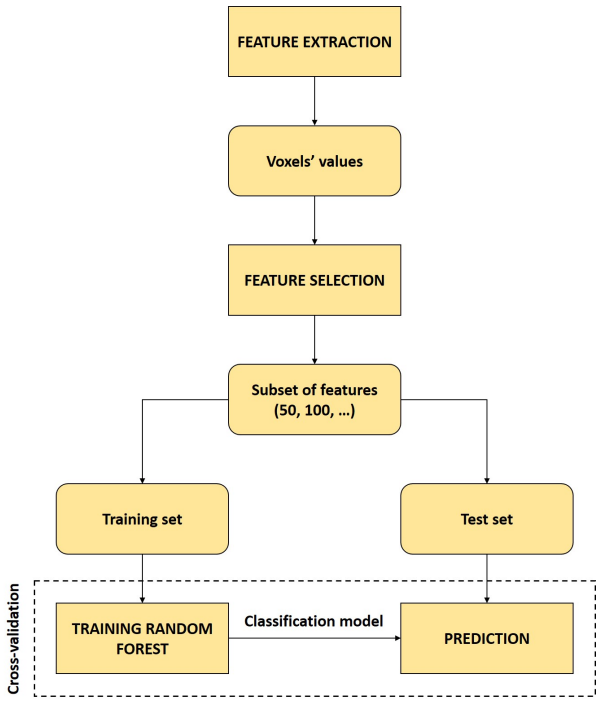
It is worth noting that the non-nested approach employed a feature selection on the entire dataset before the dataset was split (Fig. 3.5a). Conversely, in the nested approach (Fig. 3.5b), for each cross-validation round, the dataset was split into a training and test set, then the feature selection was applied on the training set blind to the test set. As measures of performance, the widely used accuracy and Area Under the ROC curve (AUC) were calculated.

3.4 Results

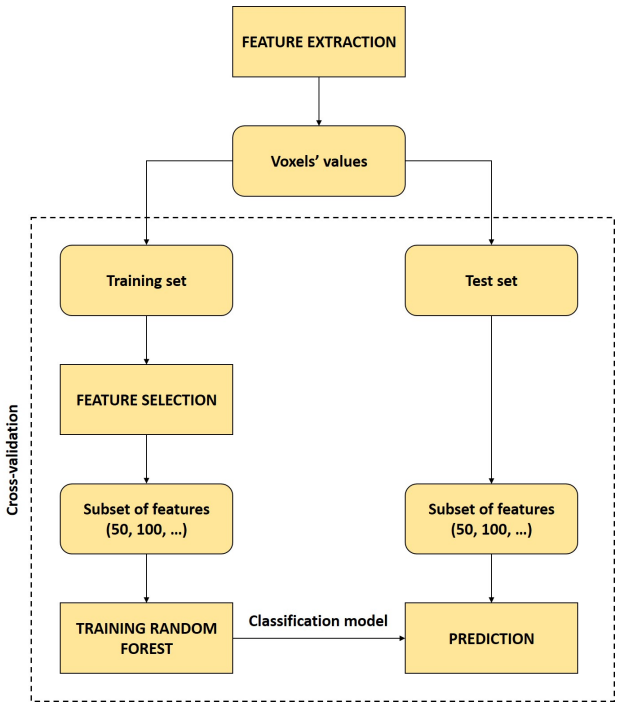
3.4.1 The feature selection bias effect

A primary question about the effects of excluding feature selection from cross validation procedures is whether or not the induced FSB is affected by the different kind of information employed, specifically FA and MD. Another question concerns the size of this effect. Besides, we also investigated whether or not the FSB was associated with the diagnosis, thus we separately studied the binary classification of HC/AD and HC/MCI. Finally, we included in our investigation two different feature selection techniques to assess whether the FSB effect could in some way depend on the methodology adopted to select the features. Mean AUCs for the classification involving both FA and MD measurements are plotted in Fig. 3.6 with both feature selection techniques.

It can be observed that switching from non-nested to nested feature selection, for the classification between HC and AD, accuracy considerably decreases from a maximum mean value of 0.87 to a maximum value of 0.75, while the best AUC drops from 0.96 to 0.84. It is worth noting that the best performance is obtained using ReliefF, but for both feature selection techniques a significant drop in performance is consistently seen. The performance decrease switching from non-nested to nested approach is more evident for the classification between HC and MCI: the best classification performance changes from 0.81 to 0.59 concerning accuracy, and from 0.90 to 0.65 concerning AUC.



(A) Non-nested approach



(B) Nested approach

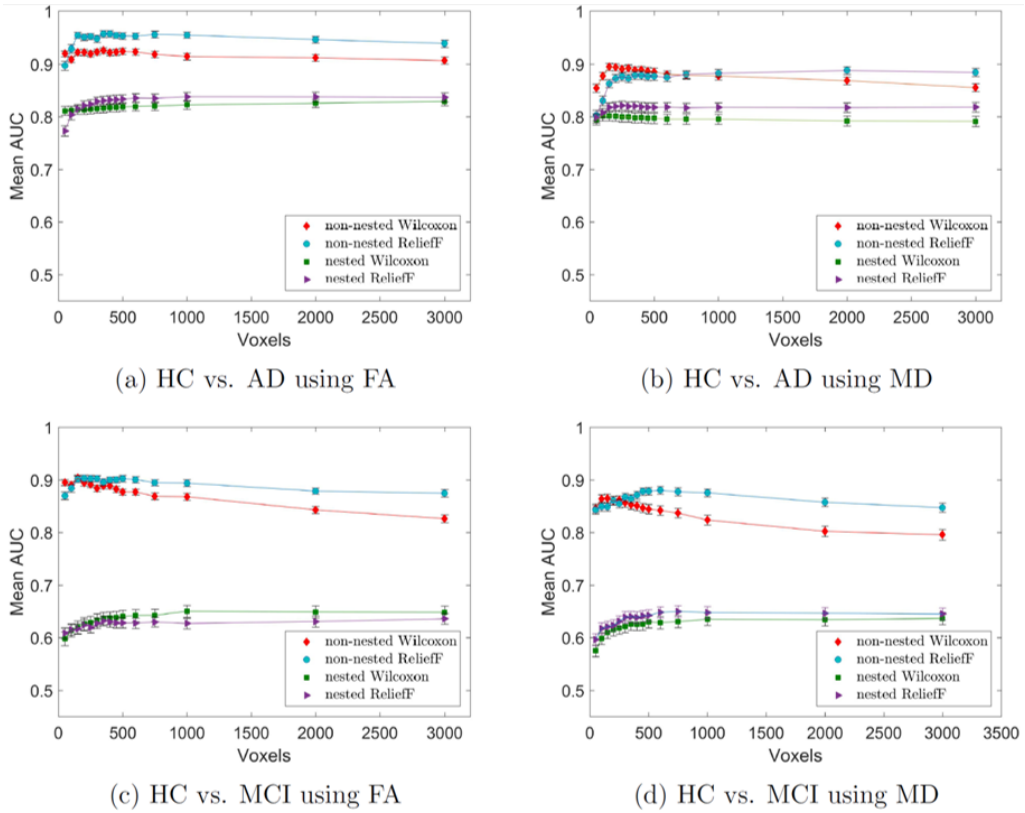
FIGURE 3.5: Flowcharts of the performed analyses: (a) non-nested feature selection and (b) nested feature selection. For readability, they only consider the steps following the feature extraction phase.

The same procedure was applied using MD. It is worth noting that moving from non-nested to nested feature selection, for the classification between HC and AD, best mean accuracy and AUC decrease respectively from 0.83 to 0.76 and from 0.90 to 0.82. For the discrimination HC/MCI the best accuracy falls from 0.79 to 0.60, while AUC decreases from 0.88 to 0.65. Again in this case, ReliefF performed better and the same performance deterioration detected for FA is clearly recognizable.

For each classification task and for each feature selection technique, the best performances in terms of mean accuracy and mean AUC are summarized in Table 3.1.

The Boxplot in Fig. 3.7 shows the distributions of the differences between the AUC values obtained in non-nested and nested best cases. It can be noticed that the FSB effect occurs regardless of the diffusion index (FA or MD) used for the classification and that this effect is more pronounced in the HC/MCI classification task.

A Wilcoxon rank sum test was performed to assess differences between the performance distributions with the nested and non-nested approach in a non-parametric fashion. Statistically significant differences (p-value < 0.01) were found between the median best performance obtained in the two cases (nested and non-nested) for all classification tasks and for both FA and MD. However, it must be noted that, for a given diffusion index (FA or MD), classification task (HC/AD or HC/MCI) and approach (nested or non-nested), the 100 measured performance metrics are not independent samples: all the 100 repetitions make use of the same images, and within each repetition there is substantial overlap among the training folds used for cross-validation. It has been shown that, in cases like the present one, no unbiased estimator exists for the variance of k-fold cross-validation [101]. The dependence of the samples and the impossibility to get an unbiased estimation of the variance violate the main assumption behind the use of standard parametric and non-parametric



(a) HC vs. AD using FA (b) HC vs. AD using MD

(c) HC vs. MCI using FA (d) HC vs. MCI using MD

FIGURE 3.6: Mean AUCs obtained varying the number of voxels.

TABLE 3.1: The first column refers to the classification task. Best average performances in terms of accuracy (Acc) and Area Under the ROC Curve (AUC) obtained in cross-validation with non-nested and nested feature selection are respectively reported in the second and third column; values are affected by a standard error of the mean approximately equal to 0.01 and a standard deviation approximately equal to 0.10. Non-nested feature selection always yields higher performances.

Classification	**Non-nested**	**Nested**
HC/AD with FA	Acc = 0.87	Acc = 0.75
	AUC = 0.96	AUC = 0.84
HC/MCI with FA	Acc = 0.81	Acc = 0.59
	AUC = 0.9	AUC = 0.65
HC/AD with MD	Acc = 0.83	Acc = 0.76
	AUC = 0.9	AUC = 0.82
HC/MCI with MD	Acc = 0.79	Acc = 0.6
	AUC = 0.88	AUC = 0.65

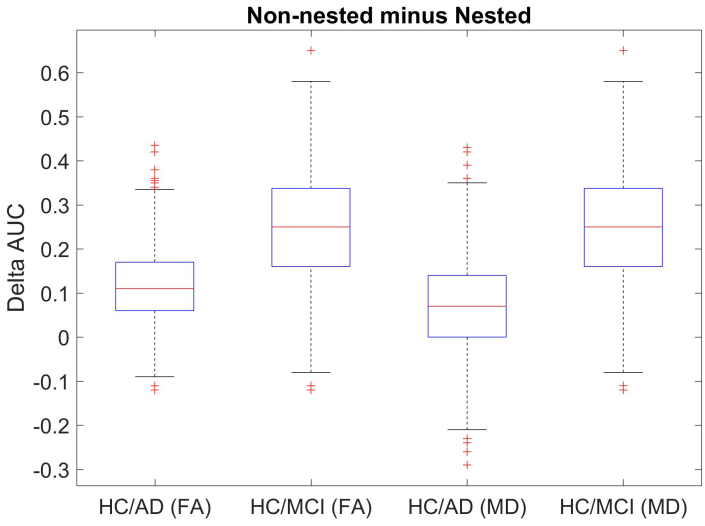


FIGURE 3.7: The distribution of the differences between the AUCs obtained in non-nested and nested best performances shows a consistent increment.

hypothesis tests. Therefore, we acknowledge the violation of the main assumption of hypothesis testing, and we warn the reader to use caution when interpreting the reported p-values.

The effectiveness of the voxels selected by the nested cross-validation in discriminating the diagnostic groups was then evaluated on a second independent set of images from the ADNI database, consisting of new scans of 40 HC, 40 MCI and 39 AD. We considered the classification tasks HC/AD and HC/MCI with FA and MD and adopted the classification tool obtained at the end of the training phase. In particular, we considered only those models constructed on the reduced sets of voxels corresponding to the best classification performance and by fixing the feature selection technique adopted, i.e. ReliefF.

TABLE 3.2: Comparison between best average performances, both in terms of accuracy (Acc) and Area Under the ROC Curve (AUC), on the training sample with nested feature selection and on the independent test sample. Independent test results (third column) are in good agreement with those obtained on the training set (training performances in the second column are affected by a standard deviation approximately equal to 0.10).

Classification	**Nested**	**Test (nested)**
HC/AD with FA	Acc = 0.75	Acc = 0.8
	AUC = 0.84	AUC = 0.91
HC/MCI with FA	Acc = 0.59	Acc = 0.56
	AUC = 0.65	AUC = 0.58
HC/AD with MD	Acc = 0.76	Acc = 0.73
	AUC = 0.82	AUC = 0.86
HC/MCI with MD	Acc = 0.6	Acc = 0.54
	AUC = 0.65	AUC = 0.6

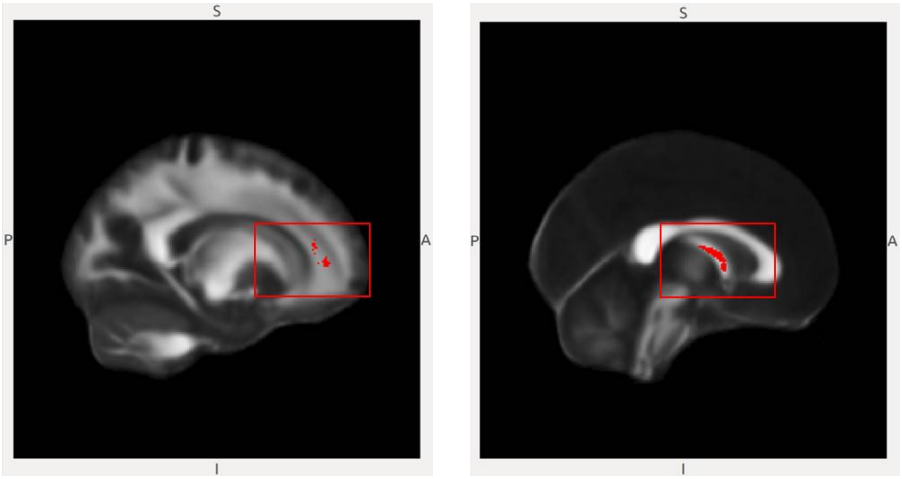
In order to confirm how the FSB impacts the reliability of estimated classification performances, the classification performances on the new data set were evaluated. To this end, we calculated the mean scores, indicating the average predicted class posterior probabilities obtained by all models; then we calculated accuracy and AUC accordingly. The results obtained are reported in the third column of Table 3.2. It can be noticed that they fall within one standard deviation of the corresponding mean value (second column).

3.4.2 Clinical findings: the disease-related brain regions

It is worth noting that the information coming from the voxel selection can be used to identify the most disease-related brain regions concerning the fiber integrity. Therefore, in the present analysis, it was also investigated whether the voxels selected during feature selection were localized in specific regions of interest of the brain.

For each classification task (HC/AD and HC/MCI) and for each feature selection technique (Wilcoxon and ReliefF), we considered the 1000 most discriminative voxels selected by the averaged nested feature selection. They are "averaged" in the sense that they are the voxels that were more frequently selected throughout all the 500 rounds of the entire nested cross-validation procedure. Two selected clusters of FA voxels are shown as an example in Fig. 3.8.

The position of the voxels derived from the average cross validation was then investigated. In order to carry out the disease-related-regions analysis, a combination of three atlases (HarvardOxford-Subcortical, JHU-ICBM-labels, JHU-ICBM-tracts) was used. More precisely, using the voxels selected from the FA maps, the comparison of HC and AD reveals differences predominantly in the Anterior Corona Radiata (bilateral but more widespread in the left hemisphere) but also in the Superior Longitudinal Fasciculus (more widespread in the left hemisphere), Fornix, Cingulum (Hippocampus), Forceps Major and Minor, Inferior Fronto Occipital Fasciculus (right), Cortospinal Tract, Anterior Thalamic Radiation, Uncinate Fasciculus (right, only with Wilcoxon), Superior Corona Radiata and External Capsule (only with ReliefF). In the comparison between HC and MCI, the FA changes are predominantly located in Forceps Minor, Superior Longitudinal Fasciculus, External Capsule (left)



(A) (B)

FIGURE 3.8: Clusters of voxels selected by ReliefF averaging all rounds of the nested feature selection (classification task HC/AD with FA): (a) voxels in the Anterior Corona Radiata (left); (b) voxels in the Fornix.

and, to a minor extent, in Inferior Fronto Occipital Fasciculus, Anterior Thalamic Radiation, Inferior Longitudinal Fasciculus, Cortical Spinal Tract, Fornix, Forceps Minor, Anterior Limb of Internal Capsule, Left Cerebral Cortex.

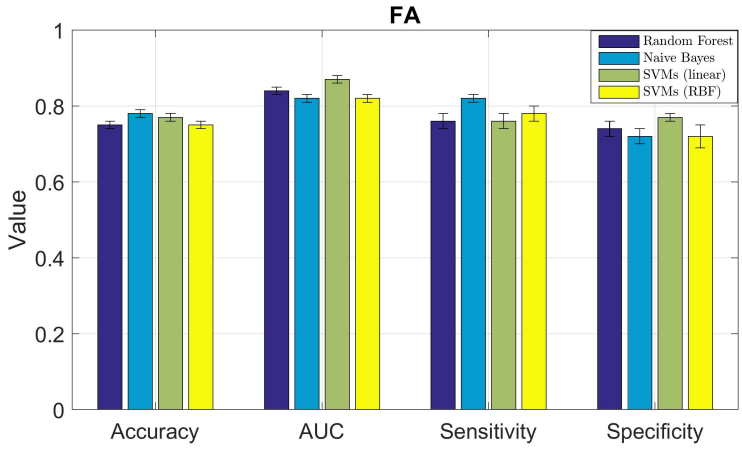
Concerning the voxels selected from the MD maps, comparing HC and AD, the predominant changes are localized in Fornix, Superior Longitudinal Fasciculus (more widespread in the left hemisphere in the case of Wilcoxon), Anterior Thalamic Radiation, Splenium and Body of Corpus Callosum, Inferior Longitudinal Fasciculus, Anterior Corona Radiata, Superior Corona Radiata (left). In the case of HC versus MCI, the MD differences are predominantly in Anterior Thalamic Radiation, Inferior Fronto Occipital Fasciculus (right), Forceps Major, Superior Longitudinal Fasciculus, Posterior Thalamic Radiation (right), Inferior Longitudinal Fasciculus, Fornix, Forceps Minor.

3.4.3 Classification model comparison

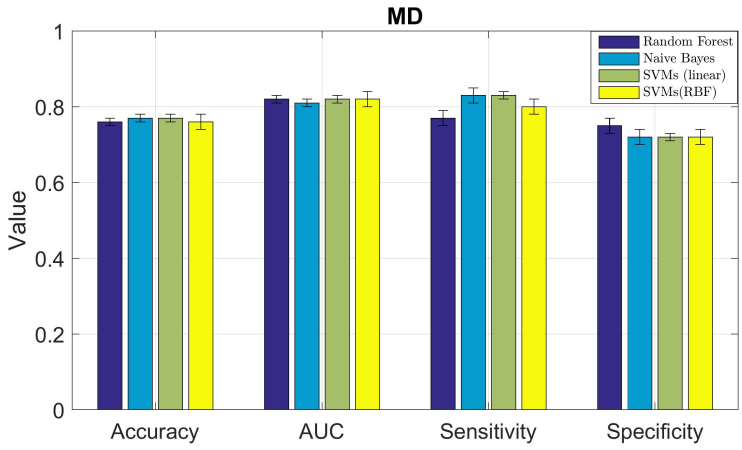
An additional analysis was carried out to explore how different supervised classification models provide a robust support to the diagnosis of AD patients. In particular, three state-of-the-art classification models, i.e. Random Forests, Naïve Bayes and Support Vector Machines were compared, and their accuracy in discriminating AD patients from HC subjects on a same dataset was evaluated. This evaluation was carried out considering the two most separated classes (HC/AD) and considering a subset of the initial dataset, composed by 40 HC and 40 AD in order to make data as balanced as possible. The classification procedure is the same described in Fig. 3.5b.

Table 3.3 summarizes in detail the best performances obtained by each classification model in terms of accuracy, Area Under the ROC Curve, sensitivity and specificity. While, an overview of these performances is provided in Fig. 3.9a and Fig. 3.9b. Note that each measure refers to the mean value obtained by averaging the performances achieved all over the cross-validation rounds.

The first observation that can be drawn is that comparable results were obtained by all classification models regardless of the choice of the diffusion index, specifically



(A) HC/AD with FA



(B) HC/AD with MD

FIGURE 3.9: Performance overview. Each bar's height indicates the mean value of the corresponding performance. The error bars indicate the standard errors.

TABLE 3.3: The first column refers to the classification task. Best average performances in terms of accuracy, Area Under the ROC Curve, sensitivity and specificity obtained in cross-validation are reported for each classification model. The higher values for these measures are in bold.

Task	Classifier	Accuracy	AUC	Specificity	Sensitivity
HC/AD with FA	Random Forest	0.75 ± 0.01	0.84 ± 0.01	0.76 ± 0.02	0.74 ± 0.02
	Naive Bayes	$\mathbf{0.78 \pm 0.01}$	0.82 ± 0.01	$\mathbf{0.82 \pm 0.01}$	0.73 ± 0.02
	SVMs (linear)	$\mathbf{0.77 \pm 0.01}$	$\mathbf{0.87 \pm 0.01}$	0.76 ± 0.02	$\mathbf{0.77 \pm 0.01}$
	SVMs (RBF)	0.75 ± 0.01	0.82 ± 0.01	0.79 ± 0.02	0.72 ± 0.03
HC/AD with MD	Random Forest	0.76 ± 0.01	0.82 ± 0.01	0.77 ± 0.02	$\mathbf{0.75 \pm 0.02}$
	Naive Bayes	$\mathbf{0.77 \pm 0.01}$	0.81 ± 0.01	$\mathbf{0.83 \pm 0.02}$	0.72 ± 0.02
	SVMs (linear)	$\mathbf{0.77 \pm 0.01}$	$\mathbf{0.83 \pm 0.01}$	$\mathbf{0.83 \pm 0.01}$	0.72 ± 0.01
	SVMs (RBF)	0.76 ± 0.02	0.82 ± 0.02	0.80 ± 0.02	0.72 ± 0.02

FA and MD. In other words, all models are suitable to be employed for such classification tasks. Although all classification models exhibited analogous behaviors, Naïve Bayes and linear SVMs achieved the best results in terms of mean accuracy (0.78 and 0.77 in the case of FA, 0.77 for both of them in the case of MD, respectively). Note that there is no statistically significant difference between the values of accuracy obtained with linear SVMs and Naïve Bayes using FA; whereas, these values are statically significant different from the values obtained using SVMs with RBF kernel and Random Forest (Wilcoxon–Mann-Whitney test at the significance level of 0.05). In the case of MD, instead, there is no statistically significant difference between the values of accuracy obtained with all classification models (Wilcoxon–Mann-Whitney test at the significance level of 0.05). In terms of AUC, the best performance was obtained only in the FA case: linear SVMs achieved a mean AUC equals to 0.87. Note that this result is statistically significant different with respect to the AUC values obtained with the other classification models (Wilcoxon–Mann-Whitney test at the significance level of 0.05). For the MD case, instead, the mean AUC values are almost comparable with no statistically significant difference.

Concerning sensitivity, the best performance in the FA case was obtained with Naïve Bayes; in the MD case, Random Forest exhibited the lower sensitivity but the higher specificity.

Another important question arising concerns the number of voxels/features necessary for the classification models to reach the best average performances. With respect to this issue, it is important to note that all the classification models reached the best performance at 1000 voxels for FA and 2000 voxels for MD. Then, the performances maintained the same level even when increasing the number of voxels taken into account.

Finally, another question concerns the agreement between the classification models. In fact, in principle, two distinct models could perform equally but misclassifying different subjects. In order to evaluate the information content provided by the algorithms and to assess the agreement between the model achieving the best performance, i.e. linear SVMs, and the other models, the relationships existing between the classification scores of the FA case was investigated. The mean classification scores for each subject, obtained throughout all the cross validation rounds with

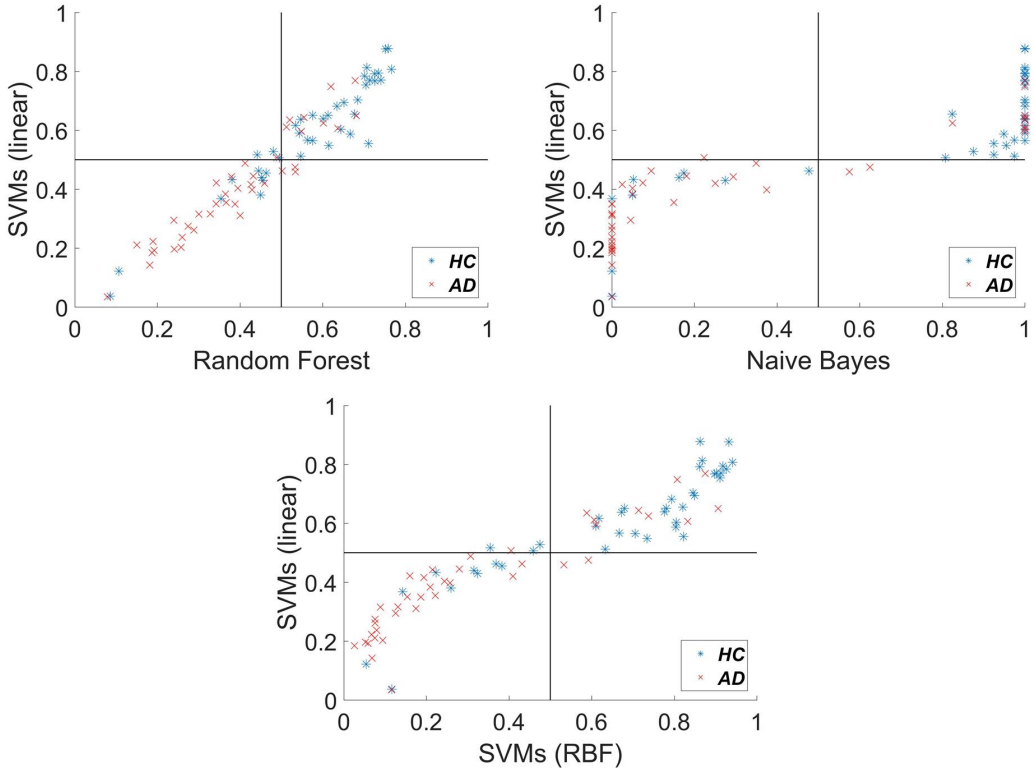


FIGURE 3.10: From left to right, the agreement between the scores obtained with linear Support Vector Machines and those obtained by Random Forest, Naïve Bayes and SVMs (RBF), considering only the FA case. The scores are densely distributed in the top right and bottom left quadrants, where their predictions agree. Looking at the top left and bottom right quadrants, instead, it can be noted that, when in disagreement, the linear SVMs and Naïve Bayes scores tend to be slightly more accurate than the other scores.

each classification model, were calculated: they are plotted in Fig. 3.10. The results attested the robustness of the nested feature selection employed: although the classification models are algorithmically different, in fact, they showed very similar findings. Looking at the top left and bottom right quadrants, it can be observed that, when in disagreement, the linear SVMs and Naïve Bayes scores are slightly more accurate than the scores of the other classifiers.

3.5 Discussion

In the analysis reported in this chapter, it has shown that:

(i) The use of non-nested feature selection techniques leads to overoptimistic classification performance;

(ii) The FSB is manifest both for FA and MD, thus it does not depend on the features adopted;

(iii) The FSB effect is more evident for the HC/MCI classification tasks;

(iv) The best performances in terms of accuracy for the HC/AD classification were obtained with (linear) Support Vector Machines and Naïve Bayes, but the results obtained with the different models are quite comparable confirming the robustness of the nested feature selection method.

The results obtained show that the voxel-based approach adopted in this study, without the bias introduced by the a priori feature selection, provides classification performances comparable with the ones obtained with other methodological procedures, except for the AUC achieved in the discrimination of HC vs. AD using FA. For the latter, the best accuracy is higher than the accuracy achieved by Mesrob et al. [53] and slightly lower than the value obtained by Schouten et al. [86]. Conversely, the AUC achieved is slightly higher than the one obtained by Schouten et al. [86]. For the classification HC/MCI it can be noticed that the accuracy and the AUC achieved with nested feature selection is lower than the one obtained in Cui et al. [87]; similarly, for the same classification task, the outcome is lower than the value obtained by Dyrba et al. [54].

If such detrimental effects on performance were somehow expected, it is worth noting that, as far as we know, no other study has measured this effect in the field of machine learning techniques applied to DTI data for AD. Furthermore, our findings regarding the significant regions for AD are consistent with several studies involving DTI, also when using other datasets than ADNI, thus reassuring about the informative content of the voxel-based approach from the clinical point of view.

Concerning the classification model comparison carried out on the HC/AD task, we found that (linear) Support Vector Machines provided the best performance in terms of Area Under the ROC Curve, i.e. 0.87, when Fractional Anisotropy maps were considered. While, the best performances in terms of accuracy were obtained with (linear) Support Vector Machines and Naïve Bayes, considering both FA and MD maps. In the FA case, although these classification models exhibited almost the same accuracy, they showed very different AUC values. This can be explained considering that SVMs exhibited a more balanced behavior in classifying positive and negative examples, i.e. sensitivity and specificity are almost the same; while, Naïve Bayes had a higher specificity but a lower sensitivity. Even so, all the classification models we here employed resulted in similar predictive accuracies; moreover, all of them showed a high level of agreement in discriminating the same subjects. This provides evidence of the robustness of the nested feature selection method we applied and attests the information content the features selected carried on.

In addition, looking at Fig. 3.10, it can be noted that Naïve Bayes tends to predict probabilities that are almost always either very close to zero or very close to one. This is a well-known behavior of the Naïve Bayes classifier to the machine learning community, which is mainly due to the simplifying assumption that features are conditionally independent given the class. Therefore the presence of the FSB in some studies using this approach is not detrimental to the anatomical and biological plausibility of the findings. In general, the existing literature provides evidence about the vulnerability of Fornix, Corpus Callosum and Cingulum to the early disease process involved in AD [102]. In particular, the white matter changes we found in the Fornix in all classification tasks (to a minor extent in the discrimination between HC and MCI using FA) have been reported in [103] and [104]. Indeed, FA reduction in the Fornix has been identified in the majority of whole-brain-TBSS studies applied to AD. Similarly, the predominant differences we observed in Cingulum, in the classification HC/AD using FA, are confirmed by looking, for example, at [105]

and [106]. Additionally, the changes we observed in the Splenium of Corpus Callosum, when classifying HC vs. AD using MD, have been reported in [107] and [105]. The most consistent results with our findings are those reported in [108], where significant changes have also been found in Uncinate Fasciculus, Inferior Longitudinal Fasciculus, Superior Longitudinal Fasciculus and Forceps Major, and in [109], which identified changes in Anterior Corona Radiata, Inferior Fronto Occipital Fasciculus and Forceps Minor. Finally, we remark that [109] also confirms the predominance of differences in the left hemisphere we found in our analysis.

3.6 Summary

In this chapter, the more established voxel-based approach has been applied to the discrimination problem at hand. The dangers of introducing a feature selection bias in the classification workflow have been highlighted and a new baseline for the binary classification task HC vs. AD and HC vs. MCI has been proposed based on FA and MD maps.

In the next chapter, we will move from the very punctual viewpoint provided by the voxel-wise analysis to the more global view provided by complex networks. In particular, we will consider properties of the brain related to the connectivity between cortical brain regions.

Chapter 4

Graph-based analysis at the cortical level

4.1 The idea of using the graph communicability to characterize AD

As already mentioned in Chapter 2, the most consolidated approaches on DTI data in AD studies are the voxel-based and the ROI-based. On the other hand, complex networks provide a powerful framework to model the human brain and its connectivity, in particular to study connectivity alterations due to neurodegenerive diseases. When applying tractography algorithms on DTI data, the brain can be modeled as a network whose nodes are the anatomical regions and whose edges are related to the fibre tracts connecting them. These brain networks can be subsequently investigated through a graph theory-based approach. Mostly, the literature applies the complex network approach on DTI data for revealing significant differences between the values of the network's measures observed in AD patients, also at early stages, against healthy control subjects by means of statistical analyses, e.g. [7, 110–112]. Very few studies, on the other hand, focused on applying DTI tractography, in combination with graph theory, to automatize the AD/HC discrimination through the use of machine learning algorithms [113]. Moreover, the literature investigated these issues applying only traditional network metrics.

The analysis presented in this chapter steams from the interest to investigate whether alternative, uncommon network metrics could serve as novel, potentially effective biomarkers for the AD diagnosis. Most of the traditional network measures are based on the shortest path connecting two nodes of a network [61, 114]. Their relevance rests on the idea that the communication between two nodes takes place through the shortest paths connecting them [115, 116]. But in many real-world networks information can spread along non-shortest paths [117, 118]. For example, in the gossip spreading in a social network, the information can flow back and forward several times before reaching the final destination. Consequently, shortest path length can miss important information on the network communicability as well as on finer structures of the network. There are two main problems in restricting the communication between two nodes through the shortest path. Imagine that a node wants to send a message to another node of the network: the sender may not know the global structure of the network. The first problem could be that the sender may not know which of the many routes connecting it with the destination is the shortest one (Fig. 4.1). The second problem could be that even if the sender knows the shortest path, it may not know a priori whether there are damaged edges in it (Fig. 4.2). Motivated by this consideration, Estrada and Hatano introduced in 2008 the *communicability* metric [119].

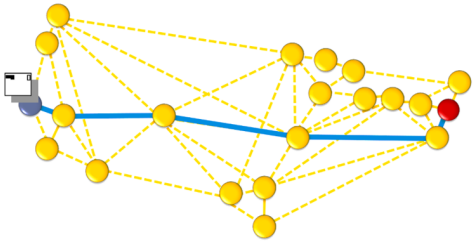


FIGURE 4.1: The figure shows a possible network configuration where the blue node wants to send a message to the red node. The sender node may knot know which is the shortest path (in blue color) to the intended destination.

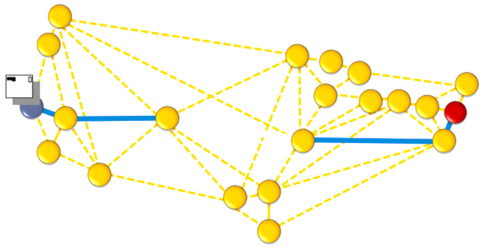


FIGURE 4.2: In the same network configuration, even if the blue node knows the shortest path to the red node, it may not known if there are damaged edges inside it.

4.1.1 Communicability

Estrada and Hatano [119] proposed a new concept of communicability, initially only for binary complex networks, defining the communicability between two nodes in a network as a function of the total number of walks connecting them, giving more importance to the shorter than to the longer ones (Fig. 4.3). If G is a graph of N nodes connected by edges and A is the $N \times N$ adjacency matrix, then:

$$(A^k)_{pq} := \sum_{r_1=1}^{N} \sum_{r_2=1}^{N} \cdots \sum_{r_{k-1}=1}^{N} a_{p,r_1} a_{r_1,r_2} a_{r_2,r_3} \cdots a_{r_{k-2},r_{k-1}} a_{r_{k-1},r_q} \tag{4.1}$$

counts the number of *walks of length* k starting at node p and ending at node q. The communicability between node p and node q is given by the total number of walks, weighted in decreasing order of their lengths, connecting the vertices p and q in a network G:

$$G_{pq} = \sum_{k=0}^{\infty} \frac{(A^k)_{pq}}{k!} = (e^A)_{pq}. \tag{4.2}$$

Equation 4.2 can be also rewritten in terms of the graph spectrum as:

$$G_{pq} = \sum_{j=1}^{n} \varphi_j(p) \varphi_j(q) e^{\lambda_j}, \tag{4.3}$$

where $\varphi_j(p)$ is the p-th element of the j-th orthonormal eigenvector of the adjacency matrix associated with the eigenvalue λ_j. As defined, the communicability between a pair of nodes in a network is bounded between zero and infinity, which are obtained for the two end nodes of an infinite linear chain and for a pair of nodes in an

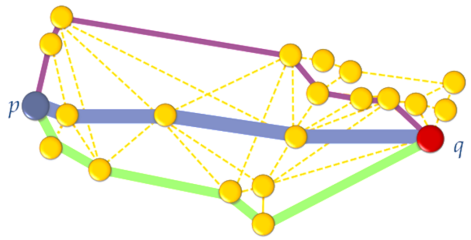


FIGURE 4.3: The figure shows three different walks, between the two nodes of the network, whose importance depends on the length of the walk.

infinite complete graph.

The concept of communicability was then extended to the weighted case by Crofts and Higham in [120] in the context of DTI connectivity matrices. The definition in eq. 4.2 is still valid but A is the $N \times N$ weighted matrix and the terms $a_{p,r_1}a_{r_1,r_2}a_{r_2,r_3} \cdots a_{r_{k-2},r_{k-1}}a_{r_{k-1},r_q}$ represent the weights of the walks $i \mapsto r_1$, $r_1 \mapsto r_2$, etc. In order to avoid the excessive influence of a node depending on its high weight, they introduced a normalization step dividing the weight a_{ij} by the product $\sqrt{s_i s_j}$, where s_i is the strength of node i. Therefore, the communicability between two nodes p and q in a weighted network is defined as:

$$G_{pq} = (exp(D^{-1/2}AD^{-1/2}))_{pq}, \tag{4.4}$$

where $D = diag(s_i)$ is the $N \times N$ diagonal strength matrix.

4.1.2 Why communicability could be suited for studying AD

By introducing the communicability metric, the communication between two nodes of a network is not limited only to the shortest path connecting them but occurs through all possible walks, giving more importance to the shortest then to the longer ones. This definition takes into account that the information in a network can flow back and forward several times before reaching the final destination. This metric reflects the network property of information to flow under a diffusion model and this aspect makes communicability really suited to study networks working in a diffusion-like process. Thus, it could be tailored to describe communication in brain networks, in particular in the case of DTI networks. Moreover, this metric provides a more global measure of the efficiency of communication in a network, considering all possible walks; therefore, it may be useful to characterize the brain connectivity alterations due to a disconnection syndrome related to an integrity loss of the communication routes.

In [120], communicability was successfully used for distinguishing local and global differences between stroke patients and controls, using DTI data. Similarly, communicability was used to detect subtle changes following stroke in the contralesional hemisphere [121]. Moreover, results obtained from the analysis of the human connectome showed advantages in using communicability over the conventional metrics in detecting highly connected nodes as well as subsets of nodes vulnerable to lesions [122]. Therefore, the idea that will be developed in this chapter and in the next one is that communicability may provide an insight into the structural properties of the brain which may be useful for emphasizing the topological differences between AD patients and HC subjects also at early stages.

The contribution of the analysis reported in this chapter is two-fold. On one hand, it will be investigated whether communicability is tailored to describe the disruption of communication between brain regions caused by AD. To this end, a statistical approach will be followed. On the other hand, it will be evaluated if and to which extent the communicability metric positively impacts the performance of a classification algorithm for the AD classification.

4.2 Data

Data used for this study were taken from the ADNI database. The images analyzed for this study belong to 122 subjects, both male and female. In accordance with the diagnosis, the subjects were grouped into 52 healthy controls (age 73 ± 6), 40 Alzheimer's disease patients (age 73 ± 8) and 30 MCI converter subjects (age 75 ± 6), i.e. MCI that converted to AD from 3 months to 5 years after the date of scan. Scans were randomly selected from baseline and follow-up study visits. As in the previous analysis, the diffusion-weighted scans were acquired using a 3-T GE Medical Systems scanner; more precisely, 46 separate images were acquired for each scan: 5 with negligible diffusion effects (B_0 images) and 41 diffusion-weighted images (b = 1000 s/mm^2).

4.3 Methods

4.3.1 Processing pipeline

For each subject, DICOM images were acquired from ADNI database. The *dcm2nii* software, within the *MRIcron* suite, was used to convert DICOM to NIFTI format. Then, the *FSL* software, and in particular its diffusion toolkit *FDT*, was employed for the subsequent processing pipeline. The preprocessing steps are the same described in the previous chapter. First, eddy current correction was performed to mitigate artifacts, such as enhanced background, image intensity loss and image blurring, caused by eddy currents and head motion. Second, Brain Extraction Tool (*BET*) was used in order to delete non-brain tissues from each subject scan.

After these preliminary steps, an affine registration of all scans was employed in order to spatially normalize the whole data set to the MNI152 standard space. Then, a single diffusion tensor was fitted at each voxel in every image by using *DTIfit*.

Finally, probabilistic tractography was performed using *ProbTrackX* [123] in order to obtain the connectivity matrix of each subject. More specifically, the Harvard-Oxford cortical atlas [124] was used, resulting in a brain parcellation of 96 regions, 48 for each hemisphere (Fig. 4.4). The final output consisted of a weighted connectivity matrix W whose elements w_{ij} corresponded to the strength of connectivity, i.e. the number of fibers, between region i and region j. The fundamental step of the whole image preprocessing was the fiber reconstruction. The *FDT* tool generates a probabilistic streamline or a sample of the distribution on the location of the true streamline. By taking many such samples, the histogram of the posterior distribution on the streamline location or the connectivity distribution is then built. Finally, the most probable traits connecting two regions are computed. It is worth noting that the intrinsic feature of the tractography algorithm to be probabilistic not always results in weighted symmetric connectivity matrices: to overcome this effect, we averaged the traits connecting region i to j and vice versa j to i to obtain a symmetric matrix for each subject. The overall preprocessing is depicted in Fig. 4.5.

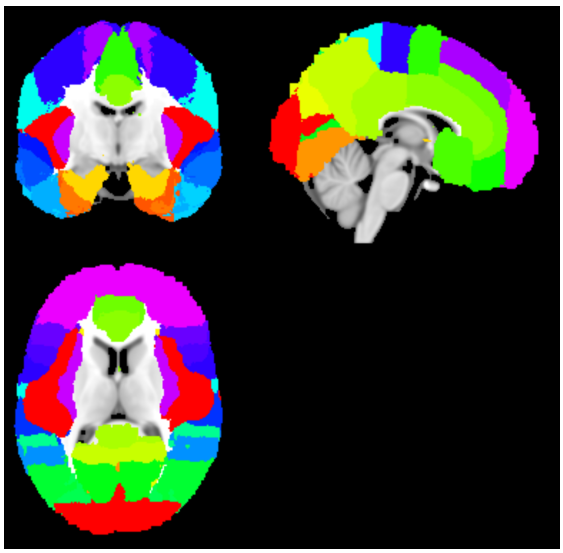


FIGURE 4.4: The figure shows a slice of three different views of the Harvard-Oxford cortical atlas.

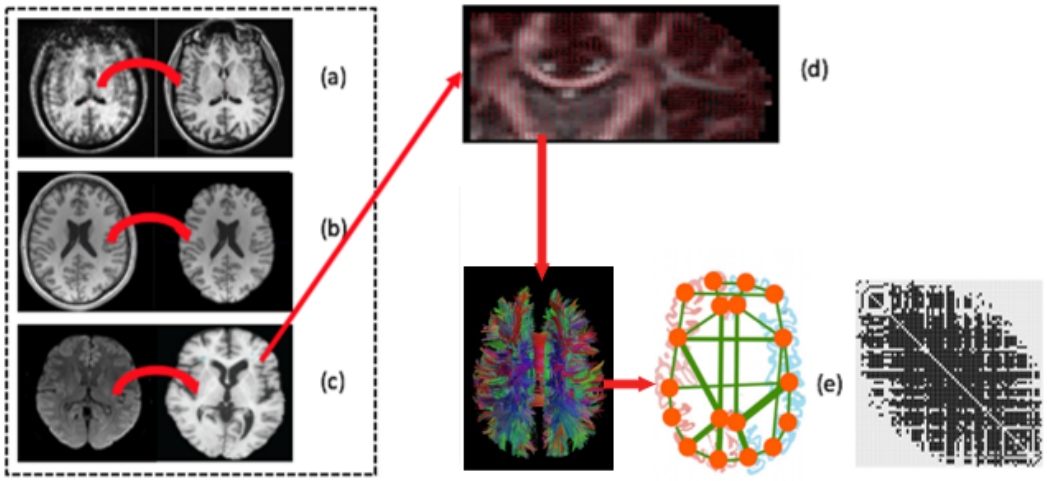


FIGURE 4.5: The figure shows the processing pipeline underwent by the brain DTI scans. The dotted box includes the dedicated image processing steps: (a) eddy correction, (b) brain extraction and (c) affine registration. For each voxel, the diffusion tensor was estimated (d), thus allowing the probabilistic fiber reconstruction (e). Using the Harvard-Oxford cortical atlas, the connectivity matrix derived from tractography was computed for each subject.

4.3.2 Statistical analysis on communicability group differences

The first goal of the present study was to assess whether communicability is suitable to describe the disruption of communication between brain regions due to AD. In other words, we investigated the information content provided by communicability from a biological perspective. To this end, a nonparametric rank sum test was performed over the communicability values of all node pairs, between the HC and AD matrices. The False Discovery Rate multiple testing correction was used to calculate the adjusted p-values. The node pairs with statistically significant different communicability (p-value < 0.05) were detected and an analysis was conducted to find if the regions involved in these connections are related, according to the literature, to the neurodegenerative processes of AD. Moreover, a nonparametric Kruskal-Wallis test was performed to compare the distibutions of the communicability values for the three groups HC/AD/MCI in order to find the node pairs with statistically significant different communicability scores according to the adjusted p-values.

4.3.3 Classification procedure

The second goal of the analysis we carried out was a comparative study, based on the same dataset, between classification models trained with traditional metrics and models trained with communicability. The learning and classification procedure was accomplished by using linear Support Vector Machines (SVMs): the classification workflow is depicted in Fig. 4.6.

The initial dataset consisted in the weighted and unweighted connectivity matrices obtained from the image preprocessing step described in Section 4.3.1. From these raw data the following features have been extracted, in accordance with the discussion provided in 2.4.2: edge betweenness; shortest path length; presence or absence of edge; and communicability. In the rest of the chapter, they are referred to as "connectivity features".

The next step concerns the validation of the classification procedure through a proper splitting of the dataset into training and test sets. Also in this analysis the subjects were stratified by diagnosis, so that each fold contained roughly the same number of subjects from each diagnostic group.

In order to reduce the dimensionality of the feature space and so to alleviate the problem of overfitting, a two-stage feature selection strategy was then applied. The first stage is an *ad-hoc* selection of features, customized to connectivity matrices. It consisted in the calculation of an healthy control binary mean matrix, i.e. a mask, onto which the matrices of all subjects were later projected to reduce the number of connectivity metrics to be considered for classification. For each training set in every cross validation iteration, the mean matrix of only the HC connectivity matrices was calculated. The HC mean matrix is a weighted matrix whose entries e_{ij} range from 0 to 1 and represent the frequency at which the corresponding edges occur among all HC matrices. The HC mean matrix was then thresholded in order to obtain the reference HC binary mean matrix to be used as mask. The threshold value was chosen at 0.7, using a binomial test (with $\alpha = 0.01$): if we consider an *a priori* probability of 0.5 that a link is present or not in a connectivity matrix, the binomial test established that a link is considered to be a "real" link if it occurs in more than 70% of all HC connectivity matrices. This feature selection strategy was motivated by the observation that, independently of the applied threshold, the healthy control mean matrix always shares more links than the unhealthy counterpart. In other words, HC matrices show a more stable topology; AD matrices, instead, show a greater

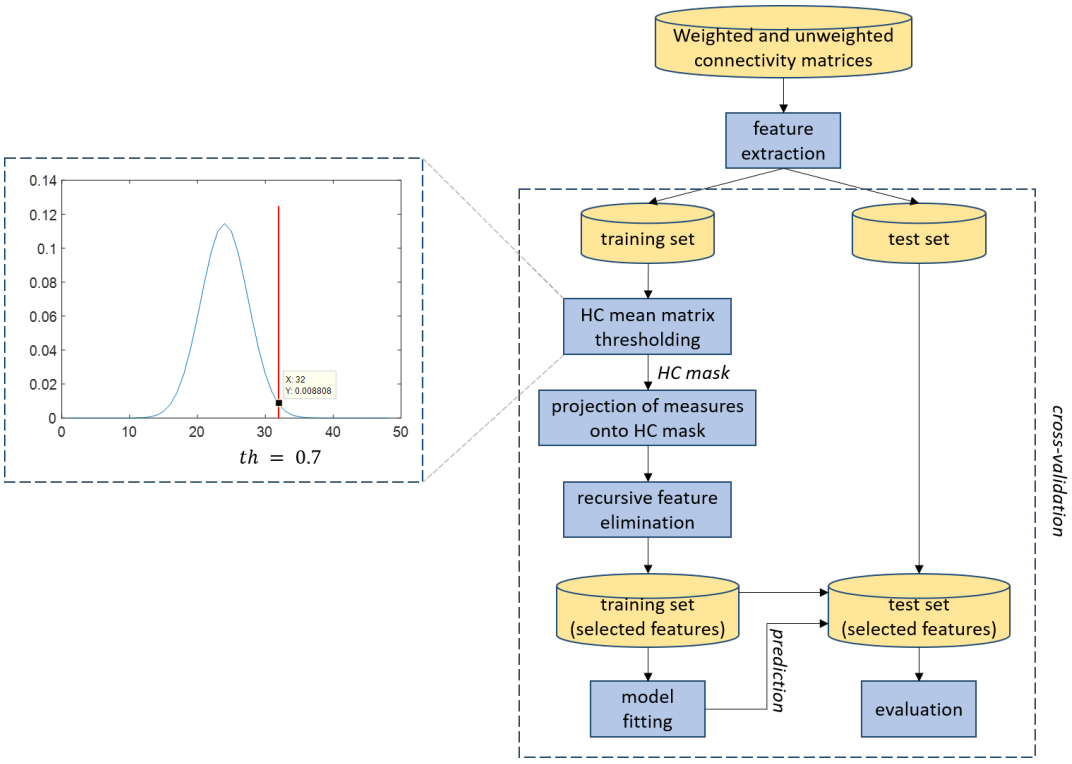


FIGURE 4.6: Classification workflow. For each cross-validation iteration, the whole dataset is split into a training and a test set. The training connectivity matrices are subjected to a two-stage feature selection. The first stage measures a mean healthy control matrix, to form a binary mask. The mask is binarized in accordance with a binomial test, which established to keep a link only if it occurs in more than 70% of the training HC matrices. Once all subjects' matrices are projected onto the mask, they are subjected to a recursive feature elimination. The output features are used to train the SVM model.

intra-variability due to the disrupted connectivity. This procedure would evaluate a significant and robust reference model to select the important links, as sampling of subjects considered in each round makes the definition of the set of significant links (i.e., the mask) robust with respect to outliers. Additionally, weak connections that can introduce noisy effects are filtered out in accordance with an objective statistical test with a strict significance threshold of the mask. This nested procedure of significant link selection was preferred to other thresholding methods such as fixing the same mean degree across all groups. Indeed, the latter procedure could be particularly problematic in "disconnection syndromes" or neurodegenerative diseases, where the assumption that node degree is the same between groups is likely to be invalid [125].

The second stage was a more conventional SVM recursive feature elimination (SVM-RFE) [126], *embedded* in the learning algorithm. The technique was applied to all the subjects' matrices obtained by the former feature selection. SVM-RFE uses criteria derived from the coefficients in the SVM models to assess features, then iteratively removes features having small criteria. In each iteration, a linear SVM model is trained: the feature with the smallest ranking criterion is removed as it has the least effect on classification. The remaining features are kept for the next iteration. The process is iteratively computed until all the features have been removed. The final outcome of the algorithm is a ranked feature list: feature selection is achieved

by choosing a group of top-ranked features. Depending on the output of SVM-RFE, subsets with an increasing number of features were evaluated, keeping only the one that obtained the best accuracy for the overall performance evaluation. Note that we used the implementation proposed in [127], which introduces a correlation bias reduction strategy that alleviates the problem of underestimating features that are highly correlated.

It is worth emphasizing that, in accordance with the findings reported in the previous chapter, we employed a *nested* feature selection strategy: for each cross-validation iteration, the two-stage feature selection was applied only on the training set blind to the test set. This methodological procedure avoids to introduce a *feature selection bias* which necessarily leads to overoptimistic results.

Once the dimensionality of data has been reduced, the SVM model can be fitted. SVMs are well known for their generalization ability and are particularly useful when the number of features, as in our case, is high [128].

4.4 Results

4.4.1 Significantly communicability differences involving disease related brain regions

Considering the HC and AD groups, 186 node pairs were found to have statistical significant different communicability (adjusted p-value < 0.05). Most of these pairs showed a reduced mean communicability in the AD subjects. Considering the HC and the AD distributions of the mean communicability values of the significant node pairs, the hypothesis of a decrease in the AD population median was tested performing a one-sided Wilcoxon rank sum test. A p-value of 1.35×10^{-7} was obtained: with enough evidence, it can be concluded that there is a negative shift in the median of the mean communicability values of the significant edges in AD compared to HC, at the 0.01 significance level, pointing out an overall communicability disruption due to AD. The differences between the HC and AD mean communicability values of the significant node pairs are shown in Fig. 4.7 by means of a glass brain: it is a 3D brain visualization in which nodes are localized exactly in the centroid position of the region of interest, in accordance with the Harvard-Oxford cortical atlas. For purposes of clarity, only the values in the 90th and 10th percentile of the communicability differences distribution are shown in the glass brain. The edges in the figure are a "nominal" representation of the difference of mean communicability, between HC and AD, of the two nodes they link: the edge color and the edge thickness are descriptive of the communicability difference of those brain region pairs between the two groups.

The brain regions mostly involved in all the 186 significant edges were found to be: Supramarginal Gyrus posterior and anterior division, Lateral Occipital Cortex, Angular Gyrus, Insular Cortex, Inferior and Superior Frontal Gyrus, Superior Parietal Lobule, Postcentral Gyrus, Middle Temporal Gyrus Temporoccipital part left, Lateral Occipital Cortex inferior division left, Frontal Pole left, Precentral Gyrus, Precuneous Cortex.

Considering the three groups HC, AD and MCI converter, 70 node pairs were found to be statistically significant different in communicability. Among these 70 pairs, 63 are in common with the 186 pairs previously found. The brain regions mostly involved in this case were found to be: Supramarginal Gyrus posterior and

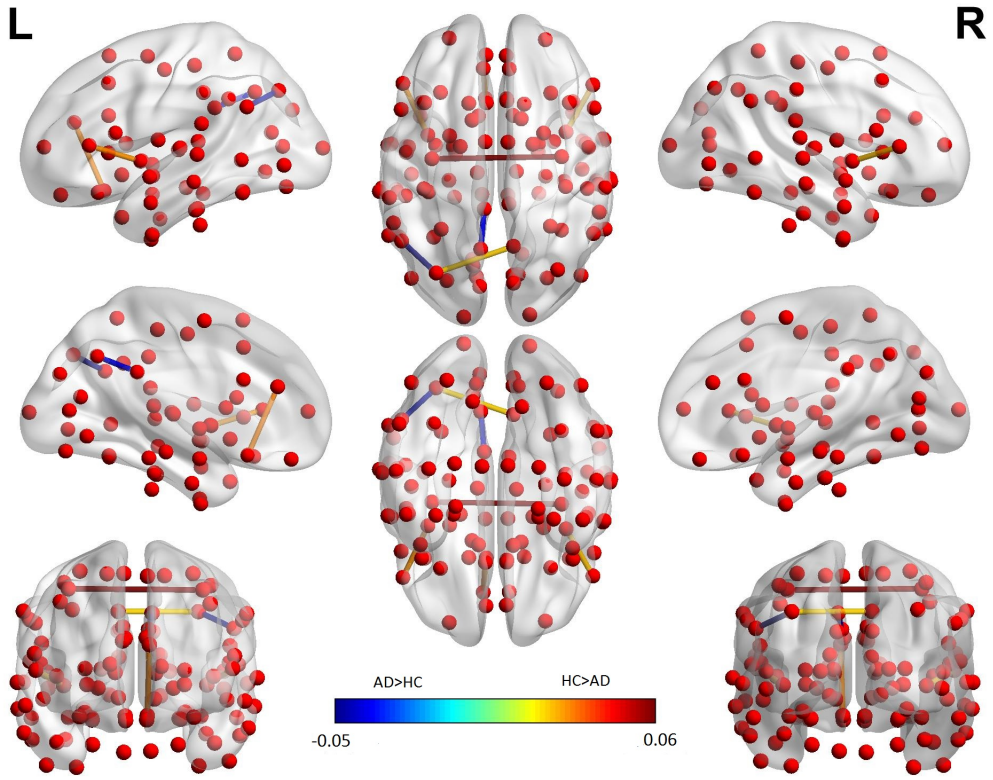


FIGURE 4.7: Glass brain visualization of the difference between the mean communicability values of the significant edges in HC and AD; the edge color and the edge thickness are descriptive of these values. Only the values in the 90th and 10th percentile of the distribution of the communicability differences are shown.

Supramarginal Gyrus anterior division, Angular Gyrus right, Insular Cortex, Precentral Gyrus, Postcentral Gyrus, Middle Frontal Gyrus, Superior Parietal Lobule right.

Additionally, a permutation test was performed to test the significance of the overlapping between the node pairs with statistical significant different communicability in the 3-class analysis (HC/AD/MCI) and the 186 node pairs with statistical significant different communicability in the 2-class analysis (HC/AD). The estimated p-value was found to be < 0.01, proving a significant percentage of overlapping links between the two tasks.

4.4.2 Classification comparison: communicability vs traditional metrics

HC/AD classification Classification performances, averaged over all the cross validation rounds, are expressed in terms of area under the ROC curve (AUC), accuracy, sensitivity and specificity. An overview of the best classification performance obtained by the network metrics individually is provided in Fig. 4.8. Note that each measure refers to the mean value obtained by averaging the performances achieved over all the cross-validation rounds. In the figure, each bar's height indicates the mean value of the corresponding performance; the error bars indicate the standard errors. The best performances were obtained using the communicability values as features (0.75 ± 0.02 of accuracy; 0.82 ± 0.02 of AUC; 0.69 ± 0.02 of

sensitivity; 0.80 ± 0.02 of specificity). Note that they are higher than the best performances obtained using as features the traditional connectivity metrics (0.69 ± 0.02 of accuracy; 0.74 ± 0.02 of AUC; 0.65 ± 0.03 of sensitivity; 0.78 ± 0.03 of specificity). It is worth remarking that the outperformance obtained when using communicability was tested by means of a non-parametric rank sum test. The performances achieved using communicability were found to be statistically significant different (p-value $<$ 0.0001) from those obtained using the traditional metrics, except for specificity in the case of binary Edge Betweenness.

Figure 4.9 shows an overview of the best performance obtained using, as features for training the classification models, all the binary connectivity measures together with or without weighted communicability. All the weighted measures together with or without weighted communicability are also reported. The inclusion of communicability, among the features used, significantly improved classification performances (p-value < 0.0001, except for specificity in the case of binary measures). Note that they are quite comparable to the performance achieved using only the communicability values as features.

Figure 4.10 shows the same results of Fig. 4.9, adding the nodal measures degree, strength, clustering coefficient, betweenness and eigenvector centrality as features. The best performances (0.77 ± 0.02 of accuracy; 0.81 ± 0.02 of AUC; 0.72 ± 0.02 of sensitivity; 0.81 ± 0.03 of specificity) have been obtained for the weighted matrices case. Also in this case the improvement of classification performances using communicability is statistically significant (p-value < 0.0001, except for specificity in the case of binary measures).

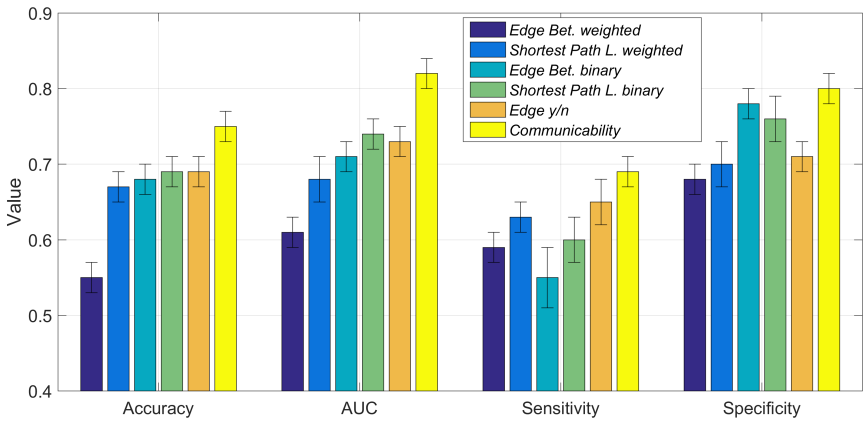


FIGURE 4.8: Overview of the performances achieved by using, as features for the HC/AD classification, the individual metrics computed on the weighted and binary connectivity matrices.

HC/AD/MCI classification The three class HC/AD/MCI performances, averaged over all the cross-validation rounds, are expressed in terms of accuracy and AUC, considering first AD as positive class and then MCI as positive class. In Fig. 4.11, the performances obtained using the previous traditional connectivity metrics as features and those obtained using communicability are compared. The best accuracy, i.e. 0.60 ± 0.02, was obtained using communicability and it is higher than the best one obtained by the models fitted on the weighted shortest path length (0.49 ± 0.02). The best AUC was achieved using communicability (0.79 ± 0.02 when considering AD as positive class, and 0.72 ± 0.02 when considering MCI as positive class), which

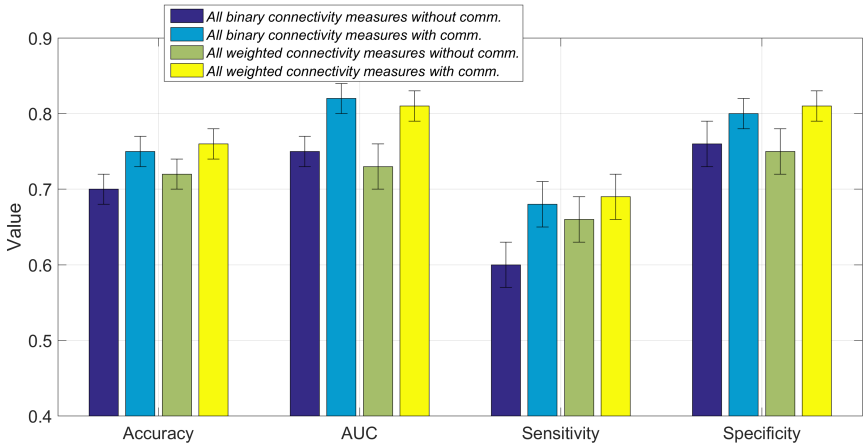


FIGURE 4.9: Performance overview of all the connectivity metrics, with and without communicability, computed on the unweighted and weighted connectivity matrices (HC/AD classification).

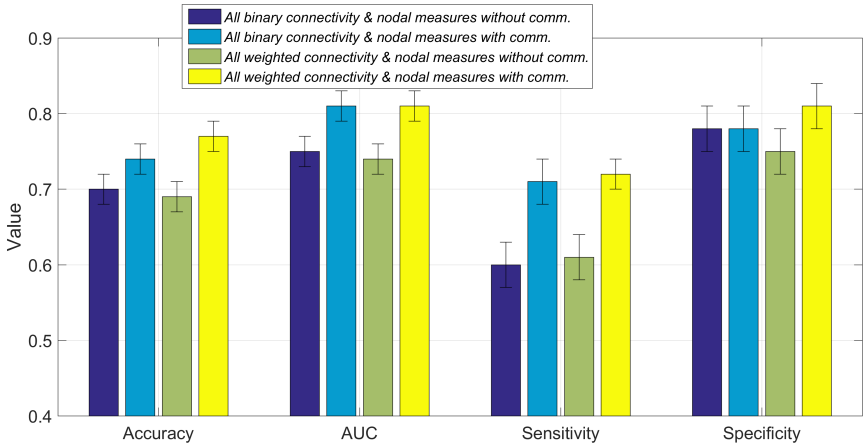


FIGURE 4.10: Performance overview of all the connectivity and nodal metrics, with and without communicability, calculated on the unweighted and weighted connectivity matrices (HC/AD classification).

was higher than the best performance reached with traditional metrics, considering the presence or absence of a link as features (0.70 ± 0.02 when considering AD as positive class, and 0.64 ± 0.03 when considering MCI as positive class). In addition, according to [129], a multiclass AUC was calculated through a pairwise comparison of classifiers, i.e. by using a one-vs-one strategy. Again, communicability outperformed the other metrics, as a multiclass AUC of 0.77 ± 0.02 was obtained against the best value of 0.68 ± 0.02 achieved with the traditional metrics. Also in the three-class classification, in all cases the classification performances obtained using communicability were found to be statistically significant different from those obtained with traditional network metrics (p-value < 0.0001).

4.5 Discussion

In this analysis, for the first time, the communicability metric was exploited with successful results to the problem of discovering connectivity differences in DTI brain

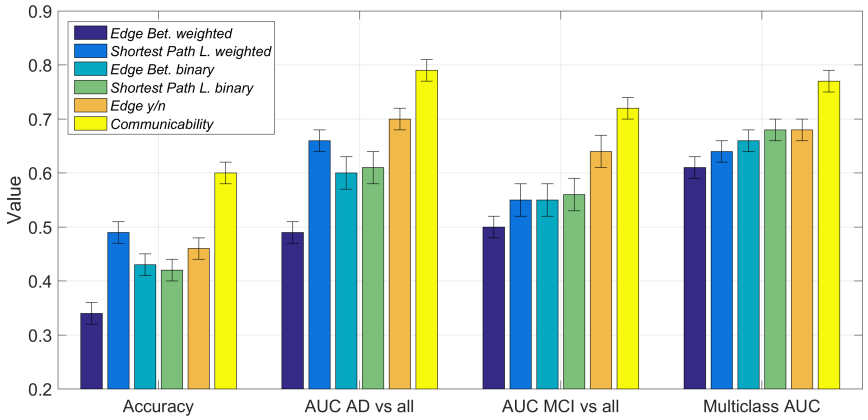


FIGURE 4.11: Performance overview of the individual metrics computed on the weighted and binary connectivity matrices (HC/AD/MCI classification).

networks of AD and MCI subjects. Communicability can be considered an alternative metric to traditional ones, mostly based on shortest path length. The aim of this study was to show the capability of this metric to highlight the connectivity changes among brain regions in patients with Alzheimer's disease, also at the early stages. The advantage of using this metric has been evaluated from two points of view:

(i) A statistical analysis pointed out pairs of brain regions with different communicability in AD and HC subjects and in AD, HC and MCI subjects;

(ii) A classification framework with different groups of features showed how communicability positively affects classification performance both for the HC/AD and the HC/AD/MCI discrimination.

First, 186 node pairs with statistical significant different communicability values between HC and AD subjects have been found, revealing a general communicability disruption among AD brains. For the three groups (HC, AD and MCI converter) statistical analysis, 70 node pairs were found to be statistically significant different in communicability, 63 in common with the 186 pairs previously found. The brain regions mostly involved in these pairs were found to be highly AD-related brain regions. It is well known the involvement of the Angular Gyrus in semantic processing, word reading and comprehension, number processing, default mode network, memory retrieval, attention and spatial cognition, reasoning and social cognition [130]. Also the language function impairment in AD patients is due to synaptic loss and dysfunction involving the Angular Gyrus [131]. The Supramarginal Gyrus is a cortical region of interest involved in impairments in verbal and semantic memory for AD [132], and an area of anatomical connectivity decrease related to the short-term memory dysfunction [133]. The Lateral Occipital Cortex, left worse than right, is a region of atrophy and hypo-metabolism in AD [134]. The regional atrophy of the Insular Cortex is associated with neuropsychiatric symptoms in AD patients [135] and pathologic changes within this area may play essential roles in AD symptoms like behavioral dyscontrol and visceral dysfunction related to autonomic instability and loss of the sense of self [136]. Moreover, left-Insula and left-Inferior Frontal Gyrus were recently found to be important regions to act on to protect memory performance against AD [137]. The Parietal Lobe, which is involved in many cognitive functions, including memory, is considered vulnerable to AD and a site of metabolic

changes and loss of WM integrity [138]. In fact, it is an AD-related area in preclinical dementia [139]. Also Precentral Gyrus is considered an AD-related brain region [140, 141]. Similarly, the medial Occipitotemporal and middle Temporal Gyri are sites affected by AD and atrophy in these areas may herald the presence of future AD among non demented individuals [142]. The Frontal Pole left and the Lateral Occipital Cortex were found to have altered white matter networks' properties in preclinical AD [111]. It is also well known the specific role of the Precuneous in self-processing during autobiographical memory retrieval [143]. Moreover the Middle Frontal Gyrus has been found to be a region of abnormal connectivity in MCI subjects [144].

Second, this analysis showed that communicability improves the overall classification performance compared to the traditional network metrics. This finding supports the idea that communicability is more apt than shortest path length to describe the efficiency of communication between brain regions when modeled as nodes of DTI networks. It is worth remarking that this study was the first using the communicability metric to classify HC, AD and MCI and to select disease-relevant biomarkers through the use of machine learning algorithms. Several other works using machine learning algorithms in this context mainly focused on two approaches to train the classification models: voxel-based, e.g. [145], and region of interest-based, e.g. [55].

4.6 Summary

In this chapter, the graph-based approach has been applied to the AD discrimination at two stages of disease progression: MCI converter and AD patient. The communicability metric has been introduced in this context as an alternative metric to traditional ones, mostly based on shortest path length, obtaining encouraging results. In the analysis described, a cortical parcellation scheme has been used for the estimated brain networks. This coarse anatomical scheme could overlook detailed patterns of connectivity, which may play a role in neurological diseases investigation. The next chapter will address this issue taking into account a different parcellation scheme and different reconstructions of the brain connectivity network to study more detailed patterns of connectivity.

Chapter 5

Graph-based analysis at the sub-cortical level

5.1 Sub-cortical network analysis in AD

In the analysis described in the previous chapter, we used the communicability metric with successful results to uncover connectivity differences in AD brain networks reconstructed at the cortical level and we developed a graph-based classification framework to distinguish unhealthy from healthy subjects automatically.

Interestingly, most of the existing literature, regarding AD discrimination using graph theory, focused on AD connectivity abnormalities at a relatively global level, such as the whole-brain level [146] and, as in our previous analysis, the cortex level [7, 147]. Much less research has been devoted to understanding the AD-related network changes at the subcortical structural level. In this analysis, the focus will be moved to the subcortical scale. The subcortical structural connectivity networks of 46 healthy control (HC) subjects and 40 AD patients will be analyzed. We have demonstrated that communicability, taking into account not only the shortest path but all possible paths connecting two nodes, revealed to be particularly suited to describe the alteration of communication between brain regions due to AD, also outperforming more classic network measures when used as feature of a supervised classification algorithm. Therefore, in this analysis this metric will be used to describe the connectivity between subcortical regions. We will measure the information content of subcortical connections in discriminating AD from normal subjects, in terms of communicability between the subcortical regions and between these regions and the rest of the network. This investigation is carried out again both with a statistical-descriptive approach and from a machine learning perspective.

5.2 Data

The scans processed for the present analysis belong to 86 subjects, 46 HC and 40 AD coming again from ADNI database. HC subjects show no sign of depression, mild cognitive impairment or dementia; participants with AD are those who meet the NINCDS/ADRDA criteria for probable AD. The diffusion-weighted scans were acquired using a 3-T GE Medical Systems scanner; more precisely, 46 separate images were acquired for each scan: 5 with negligible diffusion effects ($B0$ images) and 41 diffusion-weighted images ($b = 1000$ s/mm2). For each subject, the T1 anatomical scan was also used in order to perform tractography.

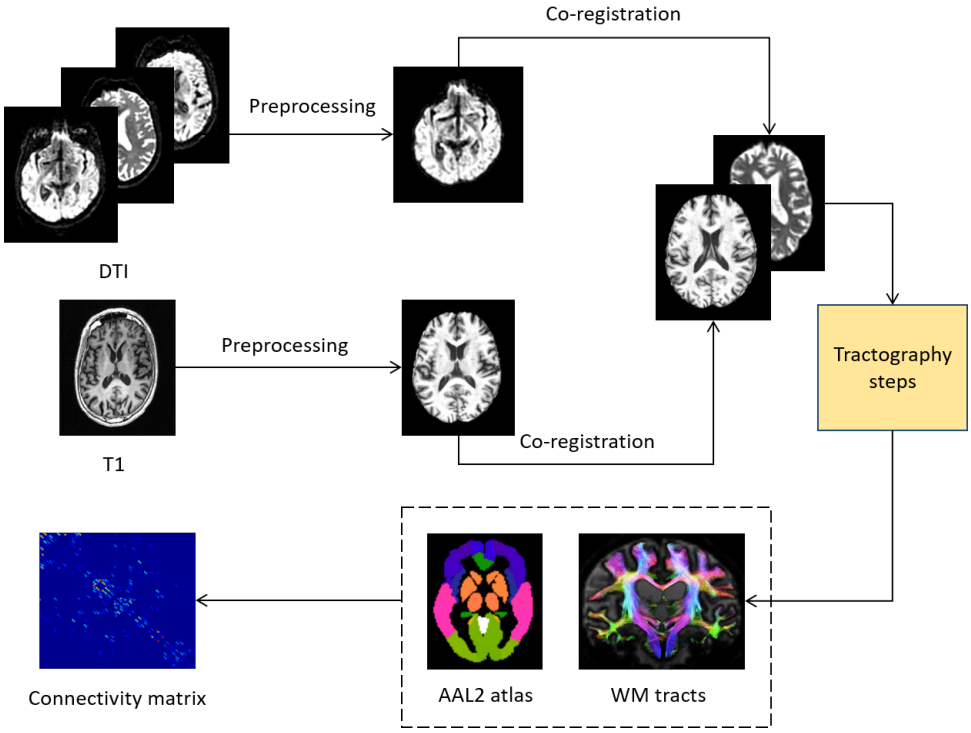


FIGURE 5.1: Main steps of the image processing pipeline.

5.3 Methods

5.3.1 Image processing and tractography

For each subject, the raw DICOM images were acquired from the ADNI database. The *dcm2nii* software, within the *MRIcron* suite, was used to convert the DICOM images into the NIFTI format. The NIFTI images were then organized in the standard BIDS format. The subsequent processing steps, from image preprocessing to structural connectome generation, were performed using tools provided by the *MRtrix3* software package (`http://mrtrix.org`) except for some preprocessing steps accomplished with the *FMRIB Software Library* (FSL) (`https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/`). The main steps of the whole procedure can be outlined as follow (see Fig. 5.1):

- Image preprocessing:
  - DTI preprocessing (denoising, brain extraction, eddy current and head motion correction, bias-field correction);
  - Anatomical T1 preprocessing (reorientation to the standard image MNI152, automatic cropping, bias-field correction, registration to the linear and non-linear standard space, brain-extraction);
- Structural connectome generation:
  - Inter-modal registration [148];
  - T1 tissue segmentation [149];
  - Spherical deconvolution;
  - Probabilistic tractography [150] utilizing dynamic seeding [151];

- Anatomically constrained tractography, which improves tractography reconstruction using anatomical information through a dynamic thresholding strategy [152];
- A spherical-deconvolution informed filtering of tractograms (SIFT2) [151];
- T1 parcellation using the AAL2 atlas [153] (120 regions), a revised version of the automated anatomical atlas (AAL);
- Robust structural connectome construction [154].

5.3.2 Communicability-based measures

The output of the image processing steps was a 120×120 weighted symmetric connectivity matrix W for each subject, whose entries w_{ij} corresponded to the number of fiber tracts connecting region i to region j. The matrices were normalized in the range $[0, 1]$. For each subject, the subcortical sub-network was extracted, including Hippocampus, Amygdala, Caudate, Putamen, Pallidum and Thalamus, both right and left, resulting in a 12×12 subcortical weighted connectivity matrix W_s.

In order to investigate the connectivity changes due to AD at the subcortical level, we analyzed the role of the subcortical regions in distinguishing the two groups under observation in terms of communicability.

As a first step, we calculated the communicability between pairs of subcortical regions, considering only the subcortical network W_s. Based on the definition provided in [120], and already introduced in the last chapter, the *weighted communicability* for each node pair p and q of the subcortical network W_s was calculated as follows:

$$G_{pq}^{sub} = \sum_{k=0}^{\infty} \frac{(M^k)_{pq}}{k!} = (e^M)_{pq}, \tag{5.1}$$

where $M = D^{-1/2} W_s D^{-1/2}$, being D the diagonal subcortical strength matrix. From this measure, we define the *intra strength communicability* for each node $i \in N_s$:

$$SC_i^{intra} = \sum_{j \in N_s} G_{ij}^{sub}, \tag{5.2}$$

which expresses the intensity of the total node connectivity with the other subcortical regions, in terms of communicability.

Then, in order to evaluate the extent of the inter-communication between the subcortical regions and the rest of the network, we calculated the whole communicability matrix G, whose elements are the communicability values between the node pairs of the whole network, i.e. the network including both cortical and subcortical regions. From this matrix, we define the *inter strength communicability* of a subcortical node i as the comunicability between i and only the cortical nodes. Being S the subset of the whole matrix indices corresponding to the subcortical regions, $\forall i \in S$ the inter strength communicability can be defined as:

$$SC_i^{inter} = \sum_{j \in \{N - \{S - i\}\}} G_{ij}, \tag{5.3}$$

which expresses the intensity of the total subcortical node connectivity, in terms of communicability, with the rest of the whole network, excluding the other subcortical regions.

5.3.3 Group-wise statistical analysis

For all subjects, the weighted sub-network communicability G_{pq}^{sub} was calculated for each node pair and a group-wise statistical analysis was performed in order to identify brain region pairs with statistically significant difference between HC and AD. In order to make the identification of the significantly different brain regions robust, permutation tests were performed to correct p-values and to address multiple comparison issues using random assignment of comparison groups. Permutation tests were performed by randomly assigning subjects to the two groups 10000 times. Differences were considered significant if they did not belong to 95% of the null distribution derived from the permutation tests (corrected p-value < 0.05).

The same statistical analysis was performed in order to find if any of the brain regions, within the 12 subcortical regions, has a statistically different intra strength communicability SC^{intra}, or a statistically different inter strength communicability SC^{inter} among the two groups HC and AD.

5.3.4 Classification

The second goal of the present analysis was to evaluate the discriminating power of the subcortical regions' communicability in distinguishing between HC and AD from a machine learning perspective. To this end, a supervised classification framework has been developed based on the random forest (RF) classifier. The classification procedure is quite the same used in the analysis described in the last chapter for the cortical networks. Since the dataset is small, the classification procedure has been validated through a 50-times repeated 10-fold cross-validation. The initial dataset consisted in the connectivity matrices obtained by the image processing steps described in Section 5.3.1. From these matrices, the communicability values for each node pair have been calculated: the values between the subcortical region pairs represent the features of the classification model to be trained. The examples were stratified by diagnosis to have approximately the same number of subjects from each diagnostic group in each fold. Within each cross-validation iteration, the training set was subjected to two feature selection steps in order to select the most relevant features. Note that, also in this case, nesting feature selection within cross-validation helps avoid a feature selection bias which may lead to overoptimistic results.

The first step is an *ad hoc* selection of features, customized to connectivity networks, already used in the cortical analysis. A mean matrix is calculated from the HC subjects of each training set, resulting in a weighted matrix whose entries e_{ij} range from 0 to 1 and represent the frequency at which the corresponding edges occur among the HC matrices. This matrix has then been thresholded, obtaining a binary matrix to be used as a mask. The matrices of all subjects were then projected onto this mask to reduce the number of features to be considered in the following step. Note that different selections of features can be obtained by varying the frequency threshold to build the mask. This procedure would evaluate a significant and robust reference model to select the important links. The second step was a more conventional RFE based on SVMs.

Finally, a RF classifier was trained on the selected features. In the present work, we set the number of trees $B = 500$, which is a common choice in the literature.

5.4 Results

5.4.1 Sub-cortical communicability patterns related to AD

Subcortical brain region pairs with significant (p-value < 0.05) group-wise differences in communicability were identified considering the G^{sub} matrix. In particular, 5 of 78 subcortical region pairs were found: (left Hippocampus, left Amygdala), (left Hippocampus, right Pallidum), (right Hippocampus, right Pallidum), (left Amygdala, right Caudate), (left Amygdala, right Pallidum). Figure 5.2 shows the relative difference between the HC mean communicability and the AD mean communicability for the 5 identified node pairs (for each node pair the difference between the HC mean communicability and the AD mean communicability is divided by the HC mean communicability). It can be observed an average disruption of communicability in AD subjects for all region pairs except for (left Amygdala, right Pallidum).

Concerning the intra strength communicability SC^{intra}, no statistically significant difference between HC and AD was found. Instead, the permutation test revealed that in terms of inter strength communicability SC^{inter} the regions with statistically difference at the 0.05 significance level are left and right Hippocampus and left and right Caudate (p-value = 0.005 for left Hippocampus, p-value = 0.01 for right Hippocampus, p-value = 0.001 for left Caudate, p-value < 0.001 for right Caudate in the permutation test). The absolute value of the relative difference between the mean HC inter strength communicability and the mean AD inter strength communicability for each subcortical region is depicted in Fig. 5.3 and is expressed in percentage: the regions with significant group-wise difference are marked with asterisks and the regions with negative difference between HC and AD are marked with a minus sign.

5.4.2 Classification performances

Following the procedure described in Section 5.3.4, the mean classification performances, averaged over all the cross-validation iterations, were measured in terms

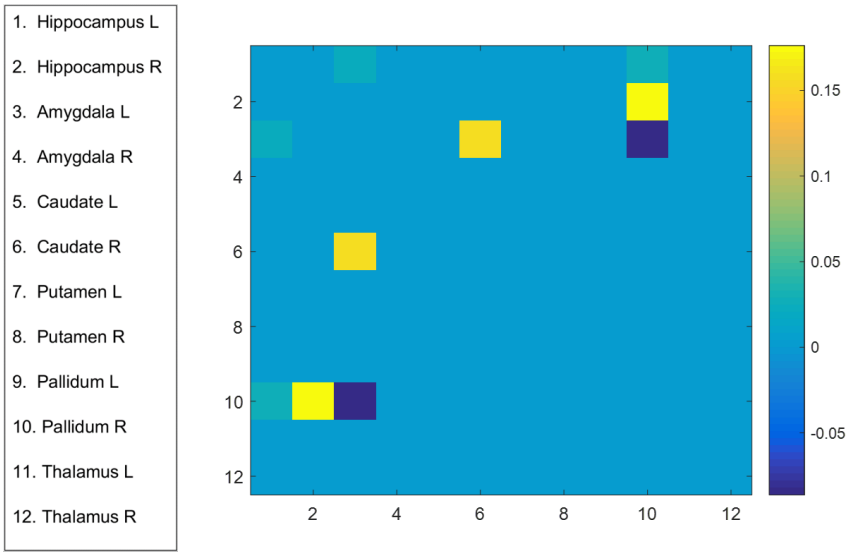


FIGURE 5.2: Heat map visualization of the relative differences between the mean communicability values of the significant edges in the HC and AD group in the case of the G^{sub} matrix. The edge color is descriptive of these values.

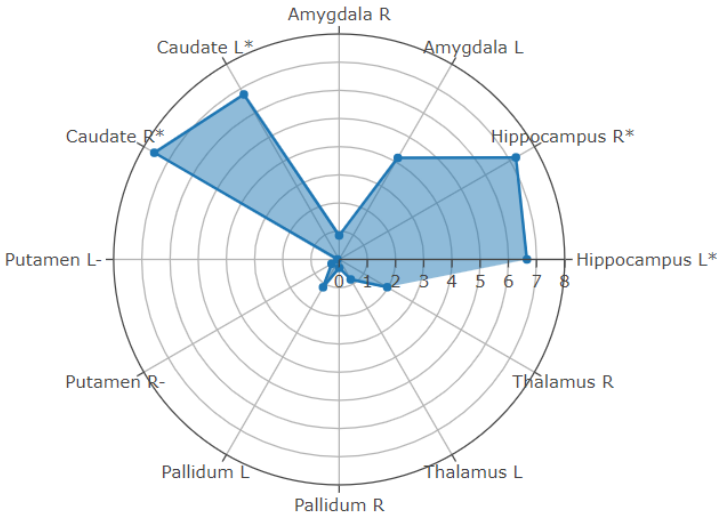


FIGURE 5.3: Radar plot which shows the group-wise difference of mean inter strength communicability (absolute value in percentage). The statistically significant different regions are marked with asterisks. The minus indicates the regions where the mean value is greater in AD than in HC.

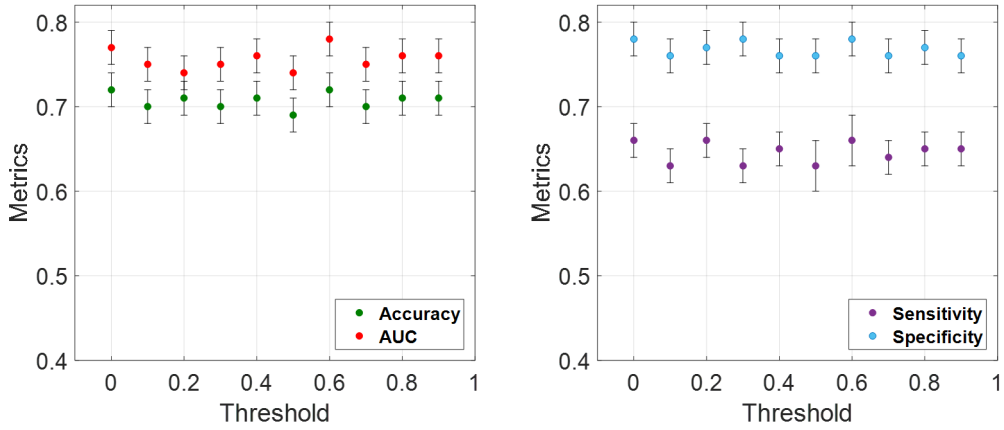


FIGURE 5.4: Classification performances at different threshold values.

of accuracy, area under the ROC curve (AUC), sensitivity and specificity. These performances quantifies how well the communicability between only the subcortical regions discriminates between AD patients and HC subjects. In order to find the optimal threshold for the mask calculation in the first feature selection step, a performance-driven approach was adopted: the performances were calculated for a threshold value varying from 0 to 0.9 with a step of 0.1. The results, depicted in Fig. 5.4, show that the best performances are obtained for a threshold value of 0.6, but it can be noticed that the performances remain quite stable for all the threshold values. The best classification performances are: 0.72 ± 0.02 of accuracy, 0.78 ± 0.02 of AUC, 0.78 ± 0.02 of specificity and 0.66 ± 0.03 of sensitivity.

5.5 Discussion

The aim of this analysis was to study the information content of the subcortical DTI sub-network in discriminating AD from HC subjects. To the best of our knowledge, there has been little research working on understanding the AD-related network changes at the subcortical structural level, especially on DTI data. The present analysis takes a step in this direction. Moreover, in this work the communicability metric has been used for the first time to identify the connectivity profile of the subcortical network. The information content of the subcortical regions in terms of communicability for the HC/AD discrimination was investigated from two points of view:

(i) A group-wise statistical analysis has been performed to find brain region pairs with significantly different communicability in the subcortical network. The same analysis has been conducted to identify subcortical regions with different intra and inter strength communicability, that are measures introduced to quantify the total intensity of subcortical nodes' connectivity, in terms of communicability with the other subcortical nodes and with the rest of the whole network.

(ii) A classification procedure has been adopted to investigate to which extent the communicability values between the subcortical regions are able to automatically discriminate between HC subjects and AD patients.

For what concerns the group-wise statistical analysis on the subcortical node pairs communicability, which was calculated starting from the subcortical weight sub-network, 5 subcortical region pairs have shown statistically significant different communicability between HC and AD, mostly involving Hippocampus, Amygdala but also Pallidum and Caudate. The key role played by these regions in AD is well known [155–159]. It is worth noting that in spite of the average communicability disruption in AD, the region pair (left Amygdala, right Pallidum) shows an increased communicability in AD compared to HC. This is not the first time some regions with greater communicability in patients compared to controls have been reported in a disconnection syndrome and using DTI data. In [121], some areas of greater communicability in stroke patients compared to controls were found also in the lesioned hemisphere. One possible interpretation of these results may be that the increased communicability in AD reflects adaptive changes in the white matter structure that have occurred secondary to the disease.

Another interesting result concerns the group-wise comparison of inter strength communicability. Hippocampus (both right and left) again play an important role showing a statistical significant difference, between HC and AD, together with the left and right Caudate. These results are in line with the literature, in particular with the evidence that one of the first events in AD is the disconnection of the hippocampal formation and neocortex [160]; moreover, neuropathological studies have documented the isolation of the hippocampal formation in AD, e.g. [161]. The connectivity patterns regarding the inter strength communicability of Hippocampus and Caudate should be more in-depth investigated analyzing their single connections with each cortical region.

Concerning classification, the communicability between the subcortical regions provided encouraging results (0.72 ± 0.02 of accuracy, 0.78 ± 0.02 of AUC, 0.78 ± 0.02 of specificity and 0.66 ± 0.03 of sensitivity). It is worth to note that these performances, obtained using only the 12×12 subcortical network, are quite comparable to those we obtained with the 96×96 cortical network in the previous chapter.

5.6 Summary

In this chapter, the graph-based analysis carried out in the previous chapter has been complemented by taking into account the brain regions at the subcortical level, which were not considered previously. Concerning the descriptive evaluation, novel insights have been found, in particular an altered communicability of the pathological group related mainly to Hippocampus, Amygdala and Caudate. This finding calls for further investigations. Concerning the classification performances, results quite comparable to those previously reported, have been achieved.

Chapter 6

Conclusion

The aim of this thesis was to investigate white matter alterations due to Alzheimer's disease through the analysis of diffusion tensor imaging data at different scales. More precisely, a voxel-based and a complex network-based approach (both at the cortical and sub-cortical level) have been applied.

Concerning the voxel-wise analysis, we preliminary highlighted how the best classification performances in the literature applying this approach suffer from the so-called feature selection bias. Instead, in the present thesis, we applied a nested feature selection strategy which allowed us to establish a new baseline, for the discrimination HC/AD and HC/MCI, for some of the most frequently used metrics for binary classification. In general, all the classification models we have employed resulted in similar predictive accuracy; moreover, all of them showed a high level of agreement in discriminating the same subjects. This provided evidence of the robustness of the nested feature selection method we applied and attested the information content the features selected carried on. This information content was evaluated by investigating whether the most discriminating voxels selected during the feature selection phase were localized in disease-related brain regions and the findings we observed confirmed this.

Concerning the graph-based approach, in this thesis, for the first time, the communicability metric has been exploited with successful results to the problem of discovering connectivity differences in DTI brain networks of AD and MCI subjects. Communicability can be considered an alternative metric to traditional ones, mostly based on shortest path length. At a first step, the importance of investigating a metric more suitable to describe diffusive processes was highlighted. Then, the application of communicability to DTI connectivity networks was firstly investigated with a statistical-descriptive aim: pairs of brain regions with statistical significant different communicability values in AD and HC subjects were found; also for the three class analysis between HC, AD and MCI subjects, pairs of brain regions, mostly overlapping with the previous case, were found to be statistical significant different in communicability. Second, the advantage of applying communicability was investigated from a quantitative point of view: it was demonstrated that, for both the two discrimination tasks HC/AD and HC/AD/MCI, using the communicability values as features for training classification models improves the performances achieved using traditional network measures instead. Although the main goal was to introduce a novel methodology in the DTI analysis for the study of AD, the clinical validity of our founding was verified by means of a comparison with literature results. In fact, it was demonstrated that communicability is really able to detect the brain connections mostly affected by the disease and to find the differences of network's communication in AD subjects also at early stages. The efficiency of this metric to uncover connectivity differences in AD brain networks brings to the conclusion that

communicability could be a powerful discriminant factor for more accurate AD diagnosis.

In particular, we reported results comparable to those obtained using the voxel-based approach. An increase in prediction performance has not been observed; however, the obtained results are encouraging and may pave the way for developing network-based classification models attaining a gain in terms of predictive accuracy. Indeed, from the prediction point of view, the potentiality of the network-based approach in this classification problem, as well as its limitations, have yet to be investigated. Currently, very few works used complex networks on DTI data, but no conclusive result has yet been obtained. Moreover, note that the complex network approach may provide a gain also in terms of inference, as novel insights and better understanding of the data may be revealed. The network approach, in fact, may complement the localized information provided by the voxel-wise analysis, by capturing more global patterns of alternation in the white matter connectivity structure due to AD. Indeed, in this thesis, the network approach enabled to uncover connectivity differences in specific brain regions and connections due to AD. Future work should investigate how to effectively combine the two different approaches to further improve prediction performances.

It is worth remarking that an important issue concerns the dependence of the adopted method on the network size. In the first experiment on communicability, coarse anatomical connectivities have been considered for the estimated brain networks. Connectivities obtained between relatively large regions patterns are more robust and reproducible; however, they could overlook detailed patterns of connectivity, which may play a key role in neurological diseases investigation. To address this issue, we extended our study by taking into account a different parcellation scheme and a different reconstruction of the brain connectivity network to study patterns of connectivity emerging at the sub-cortical level. Regarding the statistical analysis, some connectivity patterns, mainly involving Hippocampus, Amygdala, Caudate and Putamen were found. About the classification performances, results quite comparable to those obtained with the cortical connectivity matrices were achieved. There has been little research working on understanding the AD-related network changes at the subcortical structural level using DTI data. The present thesis took a step in this direction.

It is worth remarking that the subcortical analysis should be extended to a cohort of MCI subjects to study the communicability patterns in the early stages of AD. Future work should address how to better combine cortical and subcortical regions to develop more accurate classification models. Also the classification task involving the intermediate group of MCI subjects should be explored to support diagnosis at an early stage.

Finally, the integration of DTI to fMRI to investigate changes across a neural network has the potential to be a very powerful tool to aid the development of a biomarker for AD. Experiments need to be conducted on large samples of subjects using prospective and longitudinal study designs.

Bibliography

[1] D. Bzdok and B. Yeo. "The future of data analysis in the neurosciences". In: *arXiv preprint arXiv:1608.03465* (2016).

[2] O. Sporns, G. Tononi, and R. Kötter. "The human connectome: a structural description of the human brain". In: *PLoS computational biology* 1.4 (2005), e42.

[3] E. Bullmore and O. Sporns. "Complex brain networks: graph theoretical analysis of structural and functional systems". In: *Nature Reviews Neuroscience* 10.3 (2009), p. 186.

[4] A. Association. "2017 Alzheimer's disease facts and figures". In: ().

[5] S. E. Rose, F. Chen, J. B. Chalk, F. O. Zelaya, W. E. Strugnell, M. Benson, J. Semple, and D. M. Doddrell. "Loss of connectivity in Alzheimer's disease: an evaluation of white matter tract integrity with colour coded MR diffusion tensor imaging". In: *Journal of Neurology, Neurosurgery & Psychiatry* 69.4 (2000), pp. 528–530.

[6] D. Head, R. L. Buckner, J. S. Shimony, L. E. Williams, E. Akbudak, T. E. Conturo, M. McAvoy, J. C. Morris, and A. Z. Snyder. "Differential vulnerability of anterior white matter in nondemented aging with minimal acceleration in dementia of the Alzheimer type: evidence from diffusion tensor imaging". In: *Cerebral cortex* 14.4 (2004), pp. 410–423.

[7] C.-Y. Lo, P.-N. Wang, K.-H. Chou, J. Wang, Y. He, and C.-P. Lin. "Diffusion tensor tractography reveals abnormal topological organization in structural cortical networks in Alzheimer's disease". In: *Journal of Neuroscience* 30.50 (2010), pp. 16876–16885.

[8] P. J. Basser, J. Mattiello, and D. LeBihan. "MR diffusion tensor spectroscopy and imaging". In: *Biophysical journal* 66.1 (1994), p. 259.

[9] P. J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi. "In vivo fiber tractography using DT-MRI data". In: *Magnetic resonance in medicine* 44.4 (2000), pp. 625–632.

[10] S. Gauthier, B. Reisberg, M. Zaudig, R. C. Petersen, K. Ritchie, K. Broich, S. Belleville, H. Brodaty, D. Bennett, H. Chertkow, et al. "Mild cognitive impairment". In: *The Lancet* 367.9518 (2006), pp. 1262–1270.

[11] J Bischkopf, A Busse, and M. Angermeyer. "Mild cognitive impairment–a review of prevalence, incidence and outcome according to current approaches". In: *Acta Psychiatrica Scandinavica* 106.6 (2002), pp. 403–414.

[12] T. Maggipinto, R. Bellotti, N. Amoroso, D. Diacono, G. Donvito, E. Lella, A. Monaco, M. A. Scelsi, and S. Tangaro. "DTI measurements for Alzheimer's classification". In: *Physics in Medicine & Biology* 62.6 (2017), p. 2361.

[13] E. Lella, N. Amoroso, R. Bellotti, D. Diacono, M. La Rocca, T. Maggipinto, A. Monaco, and S. Tangaro. "Machine learning for the assessment of Alzheimer's disease through DTI". In: *Applications of Digital Image Processing XL*. Vol. 10396. International Society for Optics and Photonics. 2017, p. 1039619.

[14] E. Lella, N. Amoroso, A. Lombardi, T. Maggipinto, S. Tangaro, and R. Bellotti. "Communicability disruption in Alzheimer's disease connectivity networks". In: *Journal of Complex Networks* (2018).

[15] E. Lella, N. Amoroso, A. Lombardi, T. Maggipinto, S. Tangaro, and R. Bellotti. "A Classification Framework for Alzheimer's Disease based on Graph Communicability". In: *2nd Workshop on GRaphs in biomediAl Image anaLysis(GRAIL), satellite event at MICCAI* (2018).

[16] R. S. Wilson, E. Segawa, P. A. Boyle, S. E. Anagnos, L. P. Hizel, and D. A. Bennett. "The natural history of cognitive decline in Alzheimer's disease." In: *Psychology and aging* 27.4 (2012), p. 1008.

[17] J. Miklossy. "Chronic Inflammation and Amyloidogenesis in Alzheimer's Disease. Role of Spirochetes 1". In: *Journal of Alzheimer's disease* 13.4 (2008), pp. 381–391.

[18] M Goedert, C. Wischik, R. Crowther, J. Walker, and A Klug. "Cloning and sequencing of the cDNA encoding a core protein of the paired helical filament of Alzheimer disease: identification as the microtubule-associated protein tau". In: *Proceedings of the National Academy of Sciences* 85.11 (1988), pp. 4051–4055.

[19] A. Serrano-Pozo, M. P. Frosch, E. Masliah, and B. T. Hyman. "Neuropathological alterations in Alzheimer disease". In: *Cold Spring Harbor perspectives in medicine* 1.1 (2011), a006189.

[20] J. C Vickers, S. Mitew, A. Woodhouse, C. M Fernandez-Martos, M. T Kirkcaldie, A. J Canty, G. H McCormack, and A. E King. "Defining the earliest pathological changes of Alzheimer's disease". In: *Current Alzheimer Research* 13.3 (2016), pp. 281–287.

[21] R. Roberts and D. S. Knopman. "Classification and epidemiology of MCI". In: *Clinics in geriatric medicine* 29.4 (2013), pp. 753–772.

[22] A. J. Mitchell and M. Shiri-Feshki. "Rate of progression of mild cognitive impairment to dementia–meta-analysis of 41 robust inception cohort studies". In: *Acta Psychiatrica Scandinavica* 119.4 (2009), pp. 252–265.

[23] A. Ward, S. Tardiff, C. Dye, and H. M. Arrighi. "Rate of conversion from prodromal Alzheimer's disease to Alzheimer's dementia: a systematic review of the literature". In: *Dementia and geriatric cognitive disorders extra* 3.1 (2013), pp. 320–332.

[24] S. Borson, J. Scanlan, M. Brush, P. Vitaliano, and A. Dokmak. "The mini-cog: a cognitive 'vital signs' measure for dementia screening in multi-lingual elderly". In: *International Journal of Geriatric Psychiatry* 15.11 (2000), pp. 1021–1027.

[25] M. P. Lawton and E. M. Brody. "Assessment of older people: self-maintaining and instrumental activities of daily living". In: *The Gerontologist* 9.3_Part_1 (1969), pp. 179–186.

[26] M. F. Folstein, L. N. Robins, and J. E. Helzer. "The mini-mental state examination". In: *Archives of General Psychiatry* 40.7 (1983), pp. 812–812.

[27] R. M. Reitan. "Validity of the Trail Making Test as an indicator of organic brain damage". In: *Perceptual and Motor Skills* 8.3 (1958), pp. 271–276.

[28] S. Della Sala, M. Laiacona, H. Spinnler, and C. Ubezio. "A cancellation test: its reliability in assessing attentional deficits in Alzheimer's disease". In: *Psychological Medicine* 22.4 (1992), pp. 885–901.

[29] R. Sperling and K. Johnson. "Biomarkers of Alzheimer disease: current and future applications to diagnostic criteria". In: *CONTINUUM: Lifelong Learning in Neurology* 19.2, Dementia (2013), pp. 325–338.

[30] H. Hampel, K. Broich, Y. Hoessler, and J. Pantel. "Biological markers for early detection and pharmacological treatment of Alzheimer's disease". In: *Dialogues in clinical neuroscience* 11.2 (2009), p. 141.

[31] O. Hansson, H. Zetterberg, E. Vanmechelen, H. Vanderstichele, U. Andreasson, E. Londos, A. Wallin, L. Minthon, and K. Blennow. "Evaluation of plasma Aβ40 and Aβ42 as predictors of conversion to Alzheimer's disease in patients with mild cognitive impairment". In: *Neurobiology of aging* 31.3 (2010), pp. 357–367.

[32] M. Ewers, Z. Zhong, K. Bürger, A. Wallin, K. Blennow, S. J. Teipel, Y. Shen, and H. Hampel. "Increased CSF-BACE 1 activity is associated with ApoE-ε4 genotype in subjects with mild cognitive impairment and Alzheimer's disease". In: *Brain* 131.5 (2008), pp. 1252–1258.

[33] K. Höglund, O. Hansson, P. Buchhave, H Zetterberg, P. Lewczuk, E. Londos, K. Blennow, L. Minthon, and J. Wiltfang. "Prediction of Alzheimer's disease using a cerebrospinal fluid pattern of C-terminally truncated β-amyloid peptides". In: *Neurodegenerative Diseases* 5.5 (2008), pp. 268–276.

[34] C. R. Jack, R. C. Petersen, Y. C. Xu, P. C. O'Brien, G. E. Smith, R. J. Ivnik, B. F. Boeve, S. C. Waring, E. G. Tangalos, and E. Kokmen. "Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment". In: *Neurology* 52.7 (1999), pp. 1397–1397.

[35] M Ewers, S. Teipel, O Dietrich, S. Schönberg, F Jessen, R Heun, P Scheltens, L Van de Pol, N. Freymann, H.-J. Moeller, et al. "Multicenter assessment of reliability of cranial MRI". In: *Neurobiology of Aging* 27.8 (2006), pp. 1051–1059.

[36] J. Baron, G Chetelat, B Desgranges, G Perchey, B Landeau, V De La Sayette, and F Eustache. "In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease". In: *Neuroimage* 14.2 (2001), pp. 298–309.

[37] G. F. Busatto, G. E. Garrido, O. P. Almeida, C. C. Castro, C. H. Camargo, C. G. Cid, C. A. Buchpiguel, S. Furuie, and C. M. Bottino. "A voxel-based morphometry study of temporal lobe gray matter reductions in Alzheimer's disease". In: *Neurobiology of aging* 24.2 (2003), pp. 221–231.

[38] S. A. Rombouts, F. Barkhof, R. Goekoop, C. J. Stam, and P. Scheltens. "Altered resting state networks in mild cognitive impairment and mild Alzheimer's disease: an fMRI study". In: *Human brain mapping* 26.4 (2005), pp. 231–239.

[39] A. Bokde, P Lopez-Bayo, T Meindl, S Pechler, C Born, F Faltraco, S. Teipel, H.-J. Möller, and H Hampel. "Functional connectivity of the fusiform gyrus during a face-matching task in subjects with mild cognitive impairment". In: *Brain* 129.5 (2006), pp. 1113–1124.

[40] E. O. Stejskal and J. E. Tanner. "Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient". In: *The journal of chemical physics* 42.1 (1965), pp. 288–292.

[41] C. Beaulieu. "The basis of anisotropic water diffusion in the nervous system–a technical review". In: *NMR in Biomedicine* 15.7-8 (2002), pp. 435–455.

[42] P. J. Basser, J. Mattiello, and D. LeBihan. "Estimation of the effective self-diffusion tensor from the NMR spin echo". In: *Journal of Magnetic Resonance, Series B* 103.3 (1994), pp. 247–254.

[43] S. Mori. *Introduction to diffusion tensor imaging*. Elsevier, 2007.

[44] J. E. Tanner. "Transient diffusion in a system partitioned by permeable barriers. Application to NMR measurements with a pulsed field gradient". In: *The Journal of Chemical Physics* 69.4 (1978), pp. 1748–1754.

[45] P Mukherjee, J. Berman, S. Chung, C. Hess, and R. Henry. "Diffusion tensor MR imaging and fiber tractography: theoretic underpinnings". In: *American journal of neuroradiology* 29.4 (2008), pp. 632–641.

[46] M. Filippi, M Cercignani, M Inglese, M. Horsfield, and G Comi. "Diffusion tensor magnetic resonance imaging in multiple sclerosis". In: *Neurology* 56.3 (2001), pp. 304–311.

[47] C. A. Clark, T. R. Barrick, M. M. Murphy, and B. A. Bell. "White matter fiber tracking in patients with space-occupying lesions of the brain: a new technique for neurosurgical planning?" In: *Neuroimage* 20.3 (2003), pp. 1601–1608.

[48] B. P. Witwer, R. Moftakhar, K. M. Hasan, P. Deshmukh, V. Haughton, A. Field, K. Arfanakis, J. Noyes, C. H. Moritz, M. E. Meyerand, et al. "Diffusion-tensor imaging of white matter tracts in patients with cerebral neoplasm". In: *Journal of neurosurgery* 97.3 (2002), pp. 568–575.

[49] M. Kubicki, R. McCarley, C.-F. Westin, H.-J. Park, S. Maier, R. Kikinis, F. A. Jolesz, and M. E. Shenton. "A review of diffusion tensor imaging studies in schizophrenia". In: *Journal of psychiatric research* 41.1-2 (2007), pp. 15–30.

[50] D. J. Werring, A. T. Toosy, C. A. Clark, G. J. Parker, G. J. Barker, D. H. Miller, and A. J. Thompson. "Diffusion tensor imaging can detect and quantify corticospinal tract degeneration after stroke". In: *Journal of Neurology, Neurosurgery & Psychiatry* 69.2 (2000), pp. 269–272.

[51] S. E. Rose, A. L. Janke Phd, and J. B. Chalk. "Gray and white matter changes in Alzheimer's disease: a diffusion tensor imaging study". In: *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 27.1 (2008), pp. 20–26.

[52] M. J. Müller, D. Greverus, C. Weibrich, P. R. Dellani, A. Scheurich, P. Stoeter, and A. Fellgiebel. "Diagnostic utility of hippocampal size and mean diffusivity in amnestic MCI". In: *Neurobiology of aging* 28.3 (2007), pp. 398–403.

[53] L Mesrob, M Sarazin, V Hahn-Barma, L. C. D. Souza, B Dubois, P Gallinari, and S Kinkingnéhun. "DTI and structural MRI classification in Alzheimer's disease". In: *Adv. Mol. Imaging* 2 (2012), pp. 12–20.

[54] M. Dyrba, F. Barkhof, A. Fellgiebel, M. Filippi, L. Hausner, K. Hauenstein, T. Kirste, and S. J. Teipel. "Predicting Prodromal Alzheimer's Disease in Subjects with Mild Cognitive Impairment Using Machine Learning Classification of Multimodal Multicenter Diffusion-Tensor and Magnetic Resonance Imaging Data". In: *Journal of Neuroimaging* 25.5 (2015), pp. 738–747.

[55] M. Dyrba, M. Grothe, T. Kirste, and S. J. Teipel. "Multimodal analysis of functional and structural disconnection in Alzheimer's disease using multiple kernel SVM". In: *Human brain mapping* 36.6 (2015), pp. 2118–2131.

[56] S Haller, D Nguyen, C Rodriguez, J Emch, G Gold, A Bartsch, K. O. Lovblad, and P Giannakopoulos. "Individual prediction of cognitive decline in mild cognitive impairment using support vector machine-based analysis of diffusion tensor imaging data". In: *Journal of Alzheimer's Disease* 22.1 (2010), pp. 315–327.

[57] L O'Dwyer, F Lamberton, A. L. W. Bokde, M Ewers, Y. O. Faluyi, C Tanner, B Mazoyer, D O'Neill, M Bartley, D. R. Collins, T Coughlan, D Prvulovic, and H Hampel. "Using support vector machines with multiple indices of diffusion for automated classification of mild cognitive impairment". In: *PLoS ONE* 7.2 (2012).

[58] S. M. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T. E. Nichols, C. E. Mackay, K. E. Watkins, O. Ciccarelli, M. Z. Cader, P. M. Matthews, and T. E. Behrens. "Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data". In: *NeuroImage* 31.4 (2006), pp. 1487–1505. ISSN: 1053-8119.

[59] F. L. Bookstein. ""Voxel-based morphometry" should not be used with imperfectly registered images". In: *Neuroimage* 14.6 (2001), pp. 1454–1462.

[60] C. Davatzikos. "Why voxel-based morphometric analysis should be used with great caution when characterizing group differences". In: *Neuroimage* 23.1 (2004), pp. 17–20.

[61] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang. "Complex networks: Structure and dynamics". In: *Physics reports* 424.4 (2006), pp. 175–308.

[62] M. E. Newman and M. Girvan. "Finding and evaluating community structure in networks". In: *Physical Review E* 69.2 (2004), p. 026113.

[63] D. J. Watts and S. H. Strogatz. "Collective dynamics of 'small-world' networks". In: *nature* 393.6684 (1998), pp. 440–442.

[64] J.-P. Onnela, J. Saramäki, J. Kertész, and K. Kaski. "Intensity and coherence of motifs in weighted complex networks". In: *Physical Review E* 71.6 (2005), p. 065103.

[65] M. Cao, N. Shu, Q. Cao, Y. Wang, and Y. He. "Imaging functional and structural brain connectomics in attention-deficit / hyperactivity disorder". In: *Molecular neurobiology* 50.3 (2014), pp. 1111–1123.

[66] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain". In: *Neuroimage* 15.1 (2002), pp. 273–289.

[67] J. D. Power, A. L. Cohen, S. M. Nelson, G. S. Wig, K. A. Barnes, J. A. Church, A. C. Vogel, T. O. Laumann, F. M. Miezin, B. L. Schlaggar, et al. "Functional network organization of the human brain". In: *Neuron* 72.4 (2011), pp. 665–678.

[68] V. J. Wedeen, P. Hagmann, W.-Y. I. Tseng, T. G. Reese, and R. M. Weisskoff. "Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging". In: *Magnetic resonance in medicine* 54.6 (2005), pp. 1377–1386.

[69] K. Friston, C Buechel, G. Fink, J Morris, E Rolls, and R. J. Dolan. "Psychophysiological and modulatory interactions in neuroimaging". In: *Neuroimage* 6.3 (1997), pp. 218–229.

[70] S. Mori, B. J. Crain, V. P. Chacko, and P. C. Van Zijl. "Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging". In: *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 45.2 (1999), pp. 265–269.

[71] A. W. Anderson. "Theoretical analysis of the effects of noise on diffusion tensor imaging". In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 46.6 (2001), pp. 1174–1188.

[72] G. J. Parker, H. A. Haroon, and C. A. Wheeler-Kingshott. "A framework for a streamline-based probabilistic index of connectivity (PICo) using a structural interpretation of MRI diffusion measurements". In: *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 18.2 (2003), pp. 242–254.

[73] T. E. Behrens, M. W. Woolrich, M. Jenkinson, H. Johansen-Berg, R. G. Nunes, S. Clare, P. M. Matthews, J. M. Brady, and S. M. Smith. "Characterization and propagation of uncertainty in diffusion-weighted MR imaging". In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 50.5 (2003), pp. 1077–1088.

[74] M. Lazar and A. L. Alexander. "Bootstrap white matter tractography (BOOT-TRAC)". In: *NeuroImage* 24.2 (2005), pp. 524–532.

[75] S. Mori and J.-D. Tournier. *Introduction to diffusion tensor imaging: And higher order models*. Academic Press, 2013.

[76] Y. Benjamini and Y. Hochberg. "Controlling the false discovery rate: a practical and powerful approach to multiple testing". In: *Journal of the royal statistical society. Series B (Methodological)* (1995), pp. 289–300.

[77] A. Zalesky, A. Fornito, and E. T. Bullmore. "Network-based statistic: identifying differences in brain networks". In: *Neuroimage* 53.4 (2010), pp. 1197–1207.

[78] A. Lo, H. Chernoff, T. Zheng, and S.-H. Lo. "Why significant variables aren't automatically good predictors". In: *Proceedings of the National Academy of Sciences* 112.45 (2015), pp. 13892–13897.

[79] M. R. Arbabshirani, S. Plis, J. Sui, and V. D. Calhoun. "Single subject prediction of brain disorders in neuroimaging: promises and pitfalls". In: *NeuroImage* 145 (2017), pp. 137–165.

[80] K. P. Murphy. *Machine Learning: A Probabilistic Perspective. Adaptive Computation and Machine Learning*. 2012.

[81] T. Hastie, R Tibshirani, and J Friedman. *The elements of statistical learning*. Springer, 2009.

[82] V. Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013.

[83] L. Breiman. "Random forests". In: *Machine learning* 45.1 (2001), pp. 5–32.

[84] S Tangaro, N Amoroso, M Antonacci, M Boccardi, M Bocchetta, A Chincarini, D Diacono, G Donvito, R Errico, G. Frisoni, et al. "MRI analysis for hippocampus segmentation on a distributed infrastructure". In: *2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. IEEE. 2016, pp. 1–6.

[85] S. Vicario, B. Balech, G. Donvito, P. Notarangelo, and G. Pesole. "The BioVel Project: Robust phylogenetic workflows running on the GRID". In: *EMBnet. journal* 18.B (2012), pp–77.

[86] T. M. Schouten, M Koini, F de Vos, S Seiler, J van der Grond, A Lechner, A Hafkemeijer, C Möller, R Schmidt, M de Rooij, and S. A. R. B. Rombouts. "Combining anatomical, diffusion, and resting state functional magnetic resonance imaging for individual classification of mild and moderate Alzheimer disease". In: *NeuroImage:clinical* (2016).

[87] Y. Cui, W. Wen, D. M. Lipnicki, M. F. Beg, J. S. Jin, S. Luo, W. Zhu, N. A. Kochan, S. Reppermund, L. Zhuang, et al. "Automated detection of amnestic mild cognitive impairment in community-dwelling elderly adults: a combined spatial atrophy and white matter alteration approach". In: *Neuroimage* 59.2 (2012), pp. 1209–1217.

[88] N. Kriegeskorte, W. K. Simmons, P. S. Bellgowan, and C. I. Baker. "Circular analysis in systems neuroscience: the dangers of double dipping". In: *Nature neuroscience* 12.5 (2009), pp. 535–540.

[89] S. K. Singhi and H. Liu. "Feature subset selection bias for classification learning". In: *Proceedings of the 23rd international conference on Machine learning*. ACM. 2006, pp. 849–856.

[90] F. Pereira, T. Mitchell, and M. Botvinick. "Machine learning classifiers and fMRI: a tutorial overview". In: *Neuroimage* 45.1 (2009), S199–S209.

[91] E. Olivetti, A. Mognon, S. Greiner, and P. Avesani. "Brain decoding: biases in error estimation". In: *Brain Decoding: Pattern Recognition Challenges in Neuroimaging (WBD), 2010 First Workshop on*. IEEE. 2010, pp. 40–43.

[92] S. F. Eskildsen, P. Coupé, D. García-Lorenzo, V. Fonov, J. C. Pruessner, D. L. Collins, A. D. N. Initiative, et al. "Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning". In: *Neuroimage* 65 (2013), pp. 511–521.

[93] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith. "FSL". In: *NeuroImage* 62.2 (2012), pp. 782–790. ISSN: 1053-8119.

[94] D. Le Bihan, J. F. Mangin, C. Poupon, C. A. Clark, S. Pappata, N. Molko, and H. Chabriat. "Diffusion tensor imaging: concepts and applications". In: *Journal of Magnetic Resonance Imaging* 13.4 (2001), pp. 534–546.

[95] M. Hollander, D. A. Wolfe, and E. Chicken. *Nonparametric statistical methods*. John Wiley & Sons, 2013.

[96] E. Whitley and J. Ball. "Statistics review 6: Nonparametric methods". In: *Critical Care* 6.6 (2002), p. 1.

[97] K. Kira and L. A. Rendell. "The feature selection problem: Traditional methods and a new algorithm". In: *AAAI*. Vol. 2. 1992, pp. 129–134.

[98] I. Kononenko, E. Simec, and M. Robnik-Sikonja. "Overcoming the myopia of inductive learning algorithms with ReliefF". In: *Applied Intelligence* 7 (1997), pp. 39–55.

[99] D Salas-Gonzalez, J. Górriz, J Ramírez, M López, I Alvarez, F Segovia, R Chaves, and C. Puntonet. "Computer-aided diagnosis of Alzheimer's disease using support vector machines and classification trees". In: *Physics in Medicine and Biology* 55.10 (2010), p. 2807.

[100] A. Lebedev, E. Westman, G. Van Westen, M. Kramberger, A. Lundervold, D. Aarsland, H Soininen, I Kłoszewska, P Mecocci, M Tsolaki, et al. "Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness". In: *NeuroImage: Clinical* 6 (2014), pp. 115–125.

[101] Y. Bengio and Y. Grandvalet. "No unbiased estimator of the variance of k-fold cross-validation". In: *Journal of Machine Learning Research* 5.Sep (2004), pp. 1089–1105.

[102] J. Acosta-Cabronero and P. J. Nestor. "Diffusion tensor imaging in Alzheimer disease: insights into the limbic-diencephalic network and methodological considerations". In: *Frontiers in aging neuroscience* 6 (2014).

[103] K. Oishi and C. G. Lyketsos. "Alzheimer's disease and the fornix". In: *Frontiers in aging neuroscience* 6 (2014).

[104] M. A. Nowrangi and P. B. Rosenberg. "The fornix in mild cognitive impairment and Alzheimer's disease". In: *Frontiers in aging neuroscience* 7 (2015), p. 1.

[105] S. J. Teipel, R. Stahl, O. Dietrich, S. O. Schoenberg, R. Perneczky, A. L. Bokde, M. F. Reiser, H.-J. Möller, and H. Hampel. "Multivariate network analysis of fiber tract integrity in Alzheimer's disease". In: *Neuroimage* 34.3 (2007), pp. 985–995.

[106] F. Agosta, M. Pievani, S. Sala, C. Geroldi, S. Galluzzi, G. B. Frisoni, and M. Filippi. "White matter damage in Alzheimer disease and its relationship to gray matter atrophy". In: *Radiology* 258.3 (2011), pp. 853–863.

[107] R. Stahl, O. Dietrich, S. J. Teipel, H. Hampel, M. F. Reiser, and S. O. Schoenberg. "White Matter Damage in Alzheimer Disease and Mild Cognitive Impairment: Assessment with Diffusion-Tensor MR Imaging and Parallel Imaging Techniques". In: *Radiology* 243.2 (2007), pp. 483–492.

[108] N. H. Stricker, B. Schweinsburg, L. Delano-Wood, C. E. Wierenga, K. J. Bangen, K. Haaland, L. R. Frank, D. P. Salmon, and M. W. Bondi. "Decreased white matter integrity in late-myelinating fiber pathways in Alzheimer's disease supports retrogenesis". In: *Neuroimage* 45.1 (2009), pp. 10–16.

[109] G. Sousa Alves, L. O'Dwyer, A. Jurcoane, V. Oertel-Knöche, C. Knöchel, D. Prvulovic, F. Sudo, C. E. Alves, L. Valente, D. Moreira, F. Fußer, T. Karakaya, J. Pantel, E. Engelhardt, and J. Laks. "Different patterns of White Matter degeneration using multiple diffusion Indices and volumetric data in Mild Cognitive Impairment and Alzheimer patients". In: *PLoS ONE* 7.12 (2012).

[110] M. Pievani, F. Agosta, E. Pagani, E. Canu, S. Sala, M. Absinta, C. Geroldi, R. Ganzola, G. B. Frisoni, and M. Filippi. "Assessment of white matter tract damage in mild cognitive impairment and Alzheimer's disease". In: *Human brain mapping* 31.12 (2010), pp. 1862–1875.

[111] F. U. Fischer, D. Wolf, A. Scheurich, and A. Fellgiebel. "Altered whole-brain white matter networks in preclinical Alzheimer's disease". In: *NeuroImage: Clinical* 8 (2015), pp. 660–666.

[112] J. Rasero, C. Alonso-Montes, I. Diez, L. Olabarrieta-Landa, L. Remaki, I. Escudero, B. Mateos, P. Bonifazi, M. Fernandez, J. C. Arango-Lasprilla, et al. "Group-Level Progressive Alterations in Brain Connectivity Patterns Revealed by Diffusion-Tensor Brain Networks across Severity Stages in Alzheimer's Disease". In: *Frontiers in aging neuroscience* 9 (2017), p. 215.

[113] T. M. Schouten, M. Koini, F. de Vos, S. Seiler, M. de Rooij, A. Lechner, R. Schmidt, M. van den Heuvel, J. van der Grond, and S. A. Rombouts. "Individual classification of Alzheimer's disease with diffusion magnetic resonance imaging". In: *NeuroImage* 152 (2017), pp. 476–481.

[114] L. d. F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas. "Characterization of complex networks: A survey of measurements". In: *Advances in physics* 56.1 (2007), pp. 167–242.

[115] D. J. Ashton, T. C. Jarrett, and N. F. Johnson. "Effect of congestion costs on shortest paths through complex networks". In: *Physical review letters* 94.5 (2005), p. 058701.

[116] A. Trusina, M. Rosvall, and K. Sneppen. "Communication boundaries in networks". In: *Physical review letters* 94.23 (2005), p. 238701.

[117] S. P. Borgatti. "Centrality and network flow". In: *Social networks* 27.1 (2005), pp. 55–71.

[118] J. Hromkovič, R. Klasing, A. Pelc, P. Ruzicka, and W. Unger. *Dissemination of information in communication networks: broadcasting, gossiping, leader election, and fault-tolerance*. Springer Science & Business Media, 2005.

[119] E. Estrada and N. Hatano. "Communicability in complex networks". In: *Physical Review E* 77.3 (2008), p. 036111.

[120] J. J. Crofts and D. J. Higham. "A weighted communicability measure applied to complex brain networks". In: *Journal of the Royal Society Interface* (2009), rsif–2008.

[121] J. J. Crofts, D. J. Higham, R. Bosnell, S. Jbabdi, P. M. Matthews, T. Behrens, and H. Johansen-Berg. "Network analysis detects changes in the contralesional hemisphere following stroke". In: *Neuroimage* 54.1 (2011), pp. 161–169.

[122] J. Andreotti, K. Jann, L. Melie-Garcia, S. Giezendanner, E. Abela, R. Wiest, T. Dierks, and A. Federspiel. "Validation of Network Communicability Metrics for the Analysis of Brain Structural Networks". In: *PloS one* 9.12 (2014), e115503.

[123] T. E. Behrens, H. J. Berg, S. Jbabdi, M. F. Rushworth, and M. W. Woolrich. "Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?" In: *Neuroimage* 34.1 (2007), pp. 144–155.

[124] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, et al. "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest". In: *Neuroimage* 31.3 (2006), pp. 968–980.

[125] M. Drakesmith, K. Caeyenberghs, A Dutt, G Lewis, A. David, and D. K. Jones. "Overcoming the effects of false positives and threshold bias in graph theoretical analyses of neuroimaging data". In: *NeuroImage* 118 (2015), pp. 313–333.

[126] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. "Gene selection for cancer classification using support vector machines". In: *Machine learning* 46.1 (2002), pp. 389–422.

[127] K. Yan and D. Zhang. "Feature selection and analysis on correlated gas sensor data with recursive feature elimination". In: *Sensors and Actuators B: Chemical* 212 (2015), pp. 353–363.

[128] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*. Vol. 112. Springer, 2013.

[129] D. J. Hand and R. J. Till. "A simple generalisation of the area under the ROC curve for multiple class classification problems". In: *Machine learning* 45.2 (2001), pp. 171–186.

[130] M. L. Seghier. "The angular gyrus: multiple functions and multiple subdivisions". In: *The Neuroscientist* 19.1 (2013), pp. 43–61.

[131] H. J. Yun, K. Kwak, and J.-M. Lee. "Multimodal discrimination of Alzheimer's disease based on regional cortical atrophy and hypometabolism". In: *PloS one* 10.6 (2015), e0129250.

[132] J. Harasty, G. Halliday, J. Kril, and C Code. "Specific temporoparietal gyral atrophy reflects the pattern of language dissolution in Alzheimer's disease". In: *Brain* 122.4 (1999), pp. 675–686.

[133] H. Johansen-Berg and T. E. Behrens. *Diffusion MRI: from quantitative measurement to in vivo neuroanatomy*. Academic Press, 2013.

[134] D. F. Benson, R. J. Davis, and B. D. Snyder. "Posterior cortical atrophy". In: *Archives of neurology* 45.7 (1988), pp. 789–793.

[135] Y. Moon, W.-J. Moon, H. Kim, and S.-H. Han. "Regional atrophy of the insular cortex is associated with neuropsychiatric symptoms in Alzheimer's disease patients". In: *European neurology* 71.5-6 (2014), pp. 223–229.

[136] D. J. Bonthius, A. Solodkin, and G. W. Van Hoesen. "Pathology of the insular cortex in Alzheimer disease depends on cortical architecture". In: *Journal of Neuropathology & Experimental Neurology* 64.10 (2005), pp. 910–922.

[137] F. Lin, P. Ren, R. Y. Lo, B. P. Chapman, A. Jacobs, T. M. Baran, A. P. Porsteinsson, and J. J. Foxe. "Insula and Inferior Frontal Gyrus' Activities Protect Memory Performance Against Alzheimer's Disease Pathology in Old Age". In: *Journal of Alzheimer's Disease* 55.2 (2017), pp. 669–678.

[138] H. I. Jacobs, M. P. Van Boxtel, J. Jolles, F. R. Verhey, and H. B. Uylings. "Parietal cortex matters in Alzheimer's disease: an overview of structural, functional and metabolic findings". In: *Neuroscience & Biobehavioral Reviews* 36.1 (2012), pp. 297–309.

[139] H. I. Jacobs, M. P. Van Boxtel, H. B. Uylings, E. H. Gronenschild, F. R. Verhey, and J. Jolles. "Atrophy of the parietal lobe in preclinical dementia". In: *Brain and cognition* 75.2 (2011), pp. 154–163.

[140] Y. Zhang, Z. Dong, P. Phillips, S. Wang, G. Ji, J. Yang, and T.-F. Yuan. "Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning". In: *Frontiers in Computational Neuroscience* 9 (2015).

[141] E. Canu, D. G. McLaren, M. E. Fitzgerald, B. B. Bendlin, G. Zoccatelli, F. Alessandrini, F. B. Pizzini, G. K. Ricciardi, A. Beltramello, S. C. Johnson, et al. "Mapping the structural brain changes in Alzheimer's disease: the independent contribution of two imaging modalities". In: *Journal of Alzheimer's Disease* 26.s3 (2011), pp. 263–274.

[142] A Convit, J De Asis, M. De Leon, C. Tarshish, S De Santi, and H Rusinek. "Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer's disease". In: *Neurobiology of aging* 21.1 (2000), pp. 19–26.

[143] J. J. Summerfield, D. Hassabis, and E. A. Maguire. "Cortical midline involvement in autobiographical memory". In: *Neuroimage* 44.3 (2009), pp. 1188–1200.

[144] M. Li, G. Zheng, Y. Zheng, Z. Xiong, R. Xia, W. Zhou, Q. Wang, S. Liang, J. Tao, and L. Chen. "Alterations in resting-state functional connectivity of the default mode network in amnestic mild cognitive impairment: an fMRI study". In: *BMC medical imaging* 17.1 (2017), p. 48.

[145] L. O'Dwyer, F. Lamberton, A. L. Bokde, M. Ewers, Y. O. Faluyi, C. Tanner, B. Mazoyer, D. O'Neill, M. Bartley, D. R. Collins, et al. "Using support vector machines with multiple indices of diffusion for automated classification of mild cognitive impairment". In: *PloS one* 7.2 (2012), e32441.

[146] E. H. Seo, D. Y. Lee, J.-M. Lee, J.-S. Park, B. K. Sohn, D. S. Lee, Y. M. Choe, and J. I. Woo. "Whole-brain functional networks in cognitively normal, mild cognitive impairment, and Alzheimer's disease". In: *PloS one* 8.1 (2013), e53922.

[147] Y. He, Z. Chen, and A. Evans. "Structural insights into aberrant topological patterns of large-scale cortical networks in Alzheimer's disease". In: *Journal of Neuroscience* 28.18 (2008), pp. 4756–4766.

[148] C. Bhushan, J. P. Haldar, S. Choi, A. A. Joshi, D. W. Shattuck, and R. M. Leahy. "Co-registration and distortion correction of diffusion and anatomical images based on inverse contrast normalization". In: *Neuroimage* 115 (2015), pp. 269–280.

[149] Y. Zhang, M. Brady, and S. Smith. "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm". In: *IEEE transactions on medical imaging* 20.1 (2001), pp. 45–57.

[150] J. D. Tournier, F. Calamante, and A. Connelly. "Improved probabilistic streamlines tractography by 2nd order integration over fibre orientation distributions". In: *Proceedings of the international society for magnetic resonance in medicine*. Vol. 18. 2010, p. 1670.

[151] R. E. Smith, J.-D. Tournier, F. Calamante, and A. Connelly. "SIFT2: Enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography". In: *Neuroimage* 119 (2015), pp. 338–351.

[152] R. E. Smith, J.-D. Tournier, F. Calamante, and A. Connelly. "Anatomically-constrained tractography: improved diffusion MRI streamlines tractography through effective use of anatomical information". In: *Neuroimage* 62.3 (2012), pp. 1924–1938.

[153] E. T. Rolls, M. Joliot, and N. Tzourio-Mazoyer. "Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas". In: *Neuroimage* 122 (2015), pp. 1–5.

[154] R. E. Smith, J.-D. Tournier, F. Calamante, and A. Connelly. "The effects of SIFT on the reproducibility and biological accuracy of the structural connectome". In: *Neuroimage* 104 (2015), pp. 253–265.

[155] G. Allen, H. Barnard, R. McColl, A. L. Hester, J. A. Fields, M. F. Weiner, W. K. Ringe, A. M. Lipton, M. Brooker, E. McDonald, et al. "Reduced hippocampal functional connectivity in Alzheimer disease". In: *Archives of neurology* 64.10 (2007), pp. 1482–1487.

[156] C.-A. Cuénod, A. Denys, J.-L. Michot, P. Jehenson, F. Forette, D. Kaplan, A. Syrota, and F. Boller. "Amygdala atrophy in Alzheimer's disease: an in vivo magnetic resonance imaging study". In: *Archives of neurology* 50.9 (1993), pp. 941–945.

[157] C. L. Grady, M. L. Furey, P. Pietrini, B. Horwitz, and S. I. Rapoport. "Altered brain functional connectivity and impaired short-term memory in Alzheimer's disease". In: *Brain* 124.4 (2001), pp. 739–756.

[158] S. A. Rombouts, F. Barkhof, M. P. Witter, and P. Scheltens. "Unbiased whole-brain analysis of gray matter loss in Alzheimer's disease". In: *Neuroscience letters* 285.3 (2000), pp. 231–233.

[159] N. S. Ryan, S. Keihaninejad, T. J. Shakespeare, M. Lehmann, S. J. Crutch, I. B. Malone, J. S. Thornton, L. Mancini, H. Hyare, T. Yousry, et al. "Magnetic resonance imaging evidence for presymptomatic change in thalamus and caudate in familial Alzheimer's disease". In: *Brain* 136.5 (2013), pp. 1399–1414.

[160] L. Wang, Y. Zang, Y. He, M. Liang, X. Zhang, L. Tian, T. Wu, T. Jiang, and K. Li. "Changes in hippocampal connectivity in the early stages of Alzheimer's disease: evidence from resting state fMRI". In: *Neuroimage* 31.2 (2006), pp. 496–504.

[161] X. Delbeuck, M. Van der Linden, and F. Collette. "Alzheimer'disease as a disconnection syndrome?" In: *Neuropsychology review* 13.2 (2003), pp. 79–92.