UNIVERSITÀ DEGLI STUDI DI BARI
DIPARTIMENTO INTERATENEO DI FISICA
"Michelangelo Merlin"

DOTTORATO DI RICERCA IN FISICA XXX CICLO
Settore Scientifico Disciplinare FIS/07

Complex network-based quantitative methods applied to the study of neurodegenerative disease

Tutore: Prof. Roberto Bellotti
Coordinatore: Prof. Gaetano Scamarcio

Dottoranda: Marianna La Rocca

ANNO ACCADEMICO 2017-2018

CONTENTS

ABSTRACT

Technological advances have been leading to a progressive increase of information and data in different fields. This is especially true in the context of neuroscience where the creation of large repositories of medical images and metadata is making ever more pressing the necessity to produce quantitative automated methods and physical models to manage and analyze the great amount of data of complex matter (Big Data). The development of most of these tools is aimed at studying early diagnosis of neurodegenerative diseases such as Alzheimer and Parkinson. Magnetic resonance imaging (MRI) along with complex network is currently one of the most widely adopted techniques for detection of structural changes in neurological diseases. Network-based models of the brain have shown that both local and global topological properties can reveal patterns of disease propagation. However, these intra-subject descriptions do not exploit the whole information context, accessible through inter-subject comparisons. In this thesis a novel approach, which models inter-subject similarities with a multiplex network was implemented. Firstly, It was developed a physical brain model studying structural brain connectivity through network measures. A network for each MRI scan was defined considering as nodes, the sub-volumes (patches) the images are divided into and as links, Pearson's pairwise correlation between patches. Thus, each brain was represented through a feature vector encoding the relationships among brain regions. Then machine learning techniques were used for feature selection and learning model construction in order to distinguish patients from controls. On an independent dataset multiplex networks were able to correctly segregate from normal controls, Alzheimer's disease patients and subjects with mild cognitive impairment that will convert to Alzheimer (cMCI), with an accuracy respectively of 0.86 ± 0.01 and 0.84 ± 0.01. The method also showed that it is possible to underline illness effects, without a supervised segmentation by parcellating the brain in equal volumes with a size within the range of $[2250, 3200]$ mm^3. Extracting meaningful structures and data to unveil the underlying base of knowledge of a system, has always been a challenging task in social physical and life sciences. Therefore, in this thesis it was also shown how some salient networks can be extracted from the original networks and how these

new networks can emphasize the presence of the disease, besides reduce significantly data complexity and computational burden. As to Parkinson study, the experimental results demonstrated that the features obtained from multiplex networks are able to accurately (0.88 ± 0.06) discriminate PD from controls. Also in this case a privileged scale ($125\ mm^3$) for studying PD disease exists, however exploring the informative content of more scales leads to a significant improvement of the performances in the discrimination between disease and healthy subjects. Besides It was proved as the combination of the clinical and network features (summarized in a unique score) give an added value to classification accuracy (0.93 ± 0.04). A direct comparison to conventional methods on the same datasets showed that the multiplex network approach is more accurate. It also exhibits a higher sensitivity in terms of region number connected to the disease with respect to a standard voxel-based morphometry. This method has twofold potential applications: it is a reliable tool for clinical trials, especially on subjects at high risk of PD and AD conversion, and it can be used to study the disease signature of other neurodegenerative pathologies. The strong point of the methodologies and models, developed in this doctorate work, is their great versatility. Indeed in future, they, in principle, can be applied, without any restriction, to other contexts, other pathologies or to medical images acquired in different times and modalities.

1

INTRODUCTION

1.1 PHYSICS IMPORTACE IN NEUROSCIENCE

Neuroscience is a relatively new research field aimed at understanding the physiology, biochemistry, pharmacology, functions and structure of the vertebrate brain especially of the human healthy and diseased brain [1]. The progress in this scientific area has led to spectacular technological developments such as positron emission tomography (PET), Single-photon emission computed tomography (SPECT), functional and structural magnetic resonance imaging, transcranial magnetic stimulation (TMS), diffusion tensor imaging (DTI), magnetoencephalography (MEG), electroencephalography (EEG) and so on. Neuroscience is undergoing faster changes than ever before. In the last decade, neuroscience produced quantitative datasets of unparalleled breadth and size. With the growing data availability and information granularity, an important question is bringing out more and more: "How will the great data quantity affect data analysis practices?". The answer to this challeging question can be summarized in two words: quantitative neuroscience. It is really in this new perspective that physicists assume a strategical role. Indeed, all the computational and methodological strategies used so far in astrophysics, particle physics and high eneregy physics are becoming the new reality of the neuroscience. Besides the great interest in data sharing, open access, "big data" repository construction and large amounts of data to manage and analyze is a further confirm that neuroscience has a pressing need for statistical approaches, matemathical frameworks and sophisticated tools to simulate, model and study the brain. Thus, the new reference figures in neuroscience world have to be able of managing these instruments including machine learning, complex networks, computer programming, distributed multicore processing, cloud computing, and advanced visualization. In this respect, physicists seem to be particularly suitable and the bond between neuroscience and physics seems to be now inevitable and destined to increasingly consolidate itself.

1.2 THE NEED OF QUANTITATIVE BIOMARKERS

In the world, population growth has gone hand in hand with a progressive increase in the number of older people. In the developed countries, life expectancy is now rising well above 80 years and the number of people over 60 years is expected to rise from 841 million in 2013 to more than 2 billion in 2050. As populations get older, age-related neurodegenerative diseases such as Alzheimer's Disease (AD) and Parkinson's Disease (PD) have become more common [2]. Although in older people the most causes of death are still cardiovascular diseases and cancer, Alzheimer's and Parkinson's diseases are among the top ten mortal and age-related illness that cannot yet be cured or slowed significantly. The increase of these incurable neurodegenerative disorders is likely to have a dramatic impact on individuals, families and societies, unless effective means to reduce the incidence and progression of these diseases are discovered. The individual, social and financial burden for assisting these disabled patients is becoming step by step higher. In 2050 the economic toll is expected to rise to about one trillion of dollars per year in the USA alone [3]. Therefore, it is very important to find specific and efficient treatments. Currently, many potential disease-modifying therapies are being developed and evaluated at the preclinical stage, and will lead to clinical trials in the immediate future for which biomarkers are urgently required. Indeed, a major goal of current research in neurodegenerative diseases is to improve early detection of disease in order to test new treatments when they can be truly effective. For example in the context of Alzheimer's disease, it is important to find biomarker for identifying mild cognitive impairment subjects that will convert in AD subjects. Another fundamental neuroscience aim is to understand mechanisms responsible for disease and unveil pathological patterns in order to improve experimental therapies, speed medical research and limit pains and costs related to brain diseases. For making the first task possible, it is necessary to define computational algorithm and classification model for extracting unique disease signatures and combining information and sources of heterogeneous nature (numerical and textual), from several sites and times including both medical images and meta-data such as age, sex, gender, clinical scores, genetic features and so on. For the second task, it is very important to develop models of the healthy brain to simulate disease. However, in both cases a computationally and scientifically sophisticated and comprehensive approach is needed to find Alzheimer and Parkinson disease biomarkers.

Many different approaches are used to identify biomarkers such as imaging, neurophysiological and cognitive testing. Biomarkers are measurable quantities that are indicative of the presence or progress of some underlying disease.

The ideal biomarker, in addition to be easy to quantify and measure, has to be reproducible, not subject to wide variation in the general population and unaffected by co-morbid factors. For evaluation of therapies the biomarker needs to change linearly with disease progression and closely correlate with established clinico-pathological parameters of the disease. it is unrealistic to expect that a single biomarker will fulfill all the criteria, it is most likely that a combination of biomarker will be needed for early diagnosis and similarly for evaluation of disease progression for therapeutic trials [4, 5]. For example in Figure 1 and Figure 2, some biomarkers associated respectively to AD and PD are reported. These figures show also as biomarkers can detect disease signatures before the diagnosis can be made and in some cases before any symptoms become apparent.

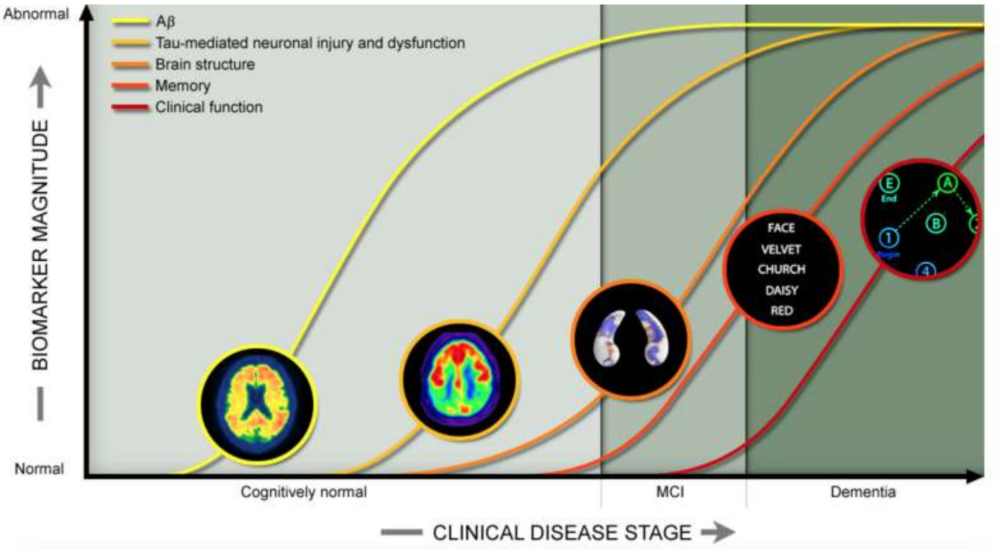

Fig. 1: Curves of changes caused by five dementia biomarkers: Amyloid beta scan detected in cerebrospinal fluid and amyloid PET scan (yellow line); neurodegeneration detected by the increase, in cerebrospinal fluid, of tau species and synaptic dysfunction, measured by FDG-PET (dark yellow line); cerebral atrophy and loss of neurons measured by magnetic resonance in particular in hippocampus, caudate nucleus and medial temporal lobe (orange line); memory loss measured by cognitive assessment (dark orange line); general cognitive decline measured by cognitive assessment (red line). The first three lines are given by biomarkers that can be observed before a diagnosis of dementia. The last two lines are the classic indicators of AD diagnosis [6].

Event though there are techniques able to earlier identify pathological changes, Magnetic Resonance Imaging (MRI) is a modality commonly applied in medical clinical and research practice to provide specific anatomical features to be

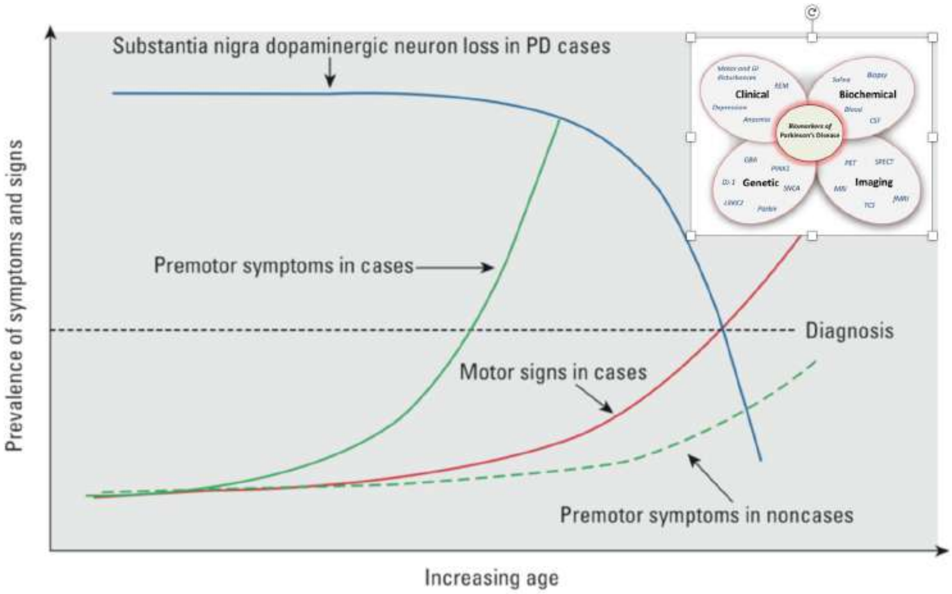

Fig. 2: Development of premotor symptoms a neurodegenerative signs among persons who will or will not develop PD in lifetime. The solid green line represent the presence of premotor symptoms in PD subjects, and the dashed green one in normal aging subjects. The red line represents motor signs among future PD patients. The blue line represents the loss of dopaminergic neurons in the substantia-nigra pars-compacta of PD patient. At the top right, there are some interesting candidates to provide PD indicators although for this disease, no specific biomarkers can be recommended in clinical practice yet. Clinical, laboratory, imaging, and genetic data need to be carefully combined to accurately predict disease status and progression [7, 8].

recognized and assessed for diagnostic purposes or to probe tissue physiological processes and pathologies. Compared to other approaches such as x-ray, PET, and SPECT, MRI are less invasive and does not use ionizing radiation that could be harmful, while provides distinct soft tissue contrast for anatomical identification. Conventional neuroimaging techniques, using MRI to find neurodegenerative disease biomarkers, can be divided into two principal categories:

- region-of-interest technique (ROI)
- Voxel-based morphometry (VBM)

ROI techniques usually compute a volume or other anatomical characteristics (surface, thickness and so on) for each brain structure, performing manual or automated segmentations along consecutive sections of a subject's MRI.

Reliability in outlining the structure boundaries is ensured by a formalized anatomical protocol with specific rules to guide image analysts in segmenting the structure. Automated algorithms such as FreeSurfer [9] have also been developed to segment these structures automatically on MRI. However reliable automated segmentations are not was so easy to obtain because image contrast is often low and is even affected by the disease process. The a priori knowledge that these anatomical regions are involved in AD makes ROI-based approach a powerful method, but it does not permit a detailed investigation and the discovery of new regions connected to the disease. However, it was thanks to this technique that the hippocampus and entorhinal cortex were established as the most important imaging biomarkers in AD [10].

Voxel-based morphometry (VBM) is an image analysis method that can detect multiple areas of cortical and subcortical atrophy. VBM was soon implemented in the widely-used Statistical Parametric Mapping software package [11]. The standard VBM method divide each subject's 3D brain MRI scan into individual maps that are representative of gray and white matter and CSF tissue classes, after registering all subjects's gray matter images to a common template, the results are averaged, smoothed through a filter and group differences related to cognitive or diagnostic scores are assessed by fitting a statistical model at each voxel location of the image. Statistical differences are then depicted as a spatial map. This identifies voxels where tissue density are statistically associated to the diagnosis or cognitive scores after pruning confounding factors such as age, sex, educational level and so on. However, VBM has some limitations due to its sensitivity to the natural cortical variability of the subjects and registration errors [12].

Currently, finding a novel strategy for building an intermediate model, that can take the advantages and discard the drawbacks of both ROI and VBM approaches, could be fundamental. In this regard, complex networks models have become the primary tool in modern neuroscience research supported by different neuroimaging techniques. A quantitative analysis of network organization could help us to get insight into the brain processing principles at micro- and macro-scales. Furthermore, these networks ease the analysis of structural and functional changes due to neurodegenerative diseases and subsequently the design of disease modifying therapies. Novel mathematical tools such as graph theoretical techniques can capture the brain connectivity. New descriptors of complex networks are able to quantify induced changes in topology or network organization or serve as theory-driven biomarkers to be used in disease prediction at the level of the individual subject.

1.3 A SUMMARY

The present work has the aim of using physics and its methodologies to innovatively adapt emerging graph theories and machine learning techniques to the neuroscience field in order to understand complex brain nature, study connectivity in diseased and healthy brain networks, investigate regions in graph networks responsible for disease and identify pathological signatures for early characterization and prediction of neurodegenerative diseases. In particular, the focus was addressed on the two most common neurodegenerative diseases: Alzheimer's and Parkinson's disease. In Chapter 2, the principal characteristics of these two diseases will be described, illustrating in detail the respective diagnostic criteria and the anatomical regions worst affected by atrophy and finally, the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Parkinson Progression Marker Initiative (PPMI), the two most important initiatives born with the goal of finding valid biomarkers for early diagnosis of AD and PD. Data originally processed and analyzed in this work belongs really to the previous international initiatives. One of the imaging techniques more used for detecting atrophy is the structural magnetic resonance imaging, especially after the introduction of high magnetic resonance scanners that improve image quality raising the signal to noise ratio. In Chapter 3, working principles of MRI will be depicted. Chapter 4 will be focused on how quantitative analysis of complex network organization could help us to understand the brain diseases. It will be discussed with particular attention the potential value added to neuroscience by the application of contemporary complex network theory according to a statistical physics understanding of graph theory. Complex network theory can be thought of as a subfield of statistical physics for structurally disordered, dynamically heterogeneous systems with non-trivial organization; indeed the statistical physics approach aims at explaining observable macroscopic behavior of a given system as coming to light from the complex interactions of a vast number of microscopic components. In the first part of the chapter, the reasons for studying complex networks and the history of complex networks will be briefly presented. Then, the principal network metrics, network models and network topologies will be introduced. In the second part of the fourth chapter, the whole workflow performed to extract important network features to early characterize and predict Alzheimer and Parkinson diseases will be exposed. Data and their preprocessing will be described, great space will be given to the network node, links and weight definition. Moreover, it will be shown in detail the development of the innovative approaches of the multiplex networks and the salient networks for capturing, from MRI scans, the most important information about the two diseases examined, as well as the feature selection and the machine learning techniques used to evaluate these methodologies. In Chapter 5 results are disclosed. Performances obtained from the normal controls vs patient classification and from the normal controls vs AD converter mild cognitive impairment subject classification, will be reported. Besides it will be shown that these performances obtained blind to the subject clinical state distinguish significantly the classes and are in line and in some cases even outperform literature results. In this chapter, it will be also underlined the advantage offered by the proposed network approaches in terms of number of regions connected to the diseases. Indeed, it is worthwhile to notice these anatomical regions are found through a totally unsupervised segmentation. Finally, it will be highlighted as the original methodologies presented in this work are able to provide a privileged scale for exploring Alzheimer's and Parkinson's disease and they allow a more insight into healthy and diseased brain connectivity through a study about network topology. Lastly in Chapter 6, the goal and the results of the work will be summarized and discussed, and conclusions and future research perspectives will be drawn.

2

ALZHEIMER'S AND PARKINSON'S DISEASE

2.1 alzheimer's disease

Alzheimer's disease (AD), the most common form of dementia, is characterized by a slow, progressive decline of cortical functions, especially cognition and memory. The pathological hallmarks of AD consists of cortical atrophy with an accumulation of senile plaques and neurofibrillary tangles in the cerebral cortex. The neurofibrillary tangles are found in the cytoplasm of neurons in the entorhinal cortex. There are two different kinds of plaques, neuritic and diffuse. Neuritic plaques are spherical structures that contain neurites, which are surrounded by an abnormal protein known as amyloid. Diffuse plaques lack neurites and have an amorphous appearance. Both types of plaques are found in the neocortex of the brain. As the number of plaques and tangles increases, healthy neurons begin to function less effectively. The neurons gradually lose their ability to communicate and consequently die, resulting in an overall shrinkage of brain tissue. Neuron death, particularly in the hippocampus, restricts the patient's ability to form new memories. According to Baraak [13], Alzheimer cascade begins in the temporal lobe in transentorhinal region (preclinical or presymptomatic phase), proceeds into cortical and subcortical components of the limbic system (mild cognitive impairment phase), and eventually extends into association areas of the neocortex (dementia) leading primarily to atrophy of the frontal and temporal association cortices. Scientists and clinicians describe preclinical phase as the silent stage during which senile plaques are accumulating in the brain, but are insufficient to cause noticeable symptoms. In mild cognitive impairment phase, the decline becomes noticeable to the family and impairment is significant but it does not interfere with everyday activities. Finally in dementia phase, significant loss of intellectual ability occurs, affecting memory plus one or more other cognitive ability, and the impairment interferes with everyday functioning. Preclinical phase can begins also 20 years before, AD occurs and a mild cognitive impairment subject can convert to AD also 10 ten years.

Once AD manifested, it can last for decades. There are three main stages of the disease, each with its own challenges and symptoms [14]. There are three main stages of the disease, each with its own challenges and symptoms. By identifying the current stage of the disease, physicians can predict what symptoms can be expected in the future and possible courses of treatment. Each case of AD presents with a unique set of symptoms, varying in severity. *Early stage* usually lasts from 2 to 4 years and begins when the disease is first diagnosed. In this stage family and friends may start to realize that there has been a decline in the patient's cognitive ability. Common symptoms at this stage include:

- Difficulty retaining new information
- Difficulty to solve problems and take decisions. Patients may start to have trouble managing finances or other instrumental activities of daily living.
- Personality changes. The person may begin to withdraw socially or show lack of motivation.
- Difficulty to express thoughts item Misplacing belongings or getting lost. The patient may have difficulty to navigate in familiar surroundings.

Moderate stage lasts from 2 to 10 years, this is the longest stage of the disease. Patients often experience increased difficulty with memory and may need help with activities of daily living. Recurrent symptoms, during this stage, include

- Increase of poor judgment and confusion. The patient may begin to confuse family members, lose orientation to time and place, and may begin wandering, making it unsafe for them to be left alone.
- Difficulty to complete complex tasks, including many of the instrumental activities of daily living, such as managing finances, grocery shopping, planning, and organization.
- Memory loss becomes more important. Patients may begin to forget details of their personal history. item Significant personality changes. The person may become withdrawn from social interactions and develop unusually high suspicions of caregivers.

Severe stage lasts from 1 to 3 years. In this final stage of the disease, cognitive capacity continues to decline and physical ability is severely impacted. Patient's family is not more able to care for the patient, therefore patients often spends his days in nursing home or other long term care facility placement. Common symptoms appearing in this stage are:

- Loss of ability to communicate. The patient may still speak short phrases, but are unable to carry on a coherent conversation.
- Reliance on others for personal care, such as eating, bathing, dressing, and toileting. Many patients become incontinent.
- Inability to function physically. The person may be unable to walk or sit independently. Muscles may become rigid and swallowing can eventually be impaired.

In Fig. 3 cognitive decline as a function of the time is represented. In particular, the normal aging progression and the disease evolution from the preclinical phase up to all the final stages of the Alzheimer's disease are reported.

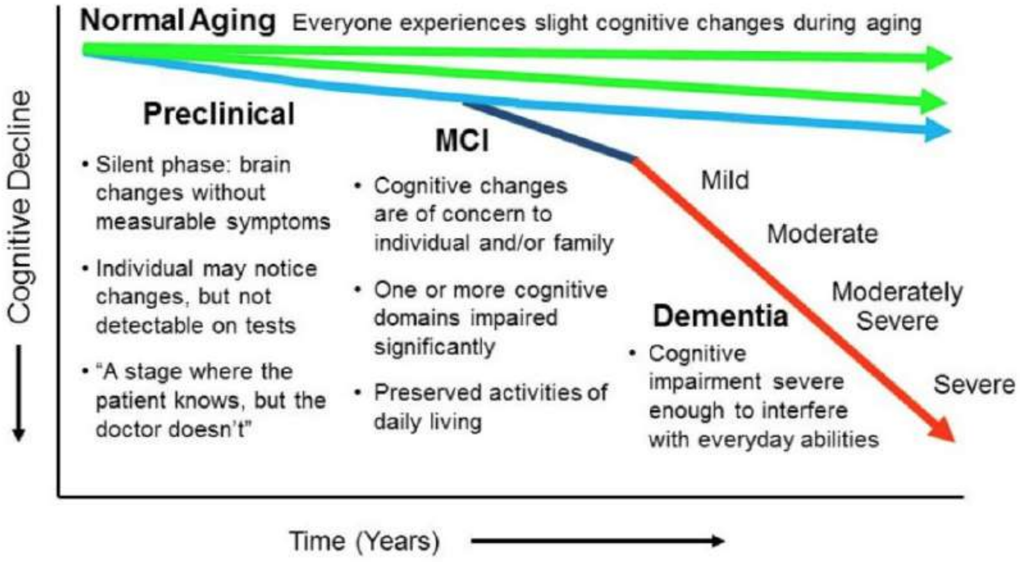

Fig. 3: Progression of the normal aging and Alzheimer's disease [15].

The greatest risk factor for developing Alzheimer's disease is age. Most cases of AD are seen in older adults, ages 65 years or above. A connection has been also found between a gene called Apolipoprotein E (ApoE) and the development of the disease. There may be a link between educational level and the risk of developing AD. Indeed people with fewer years of education seem to be at a higher risk. There is a strong edge between cardiovascular and brain health because damages to blood vessels in the brain, results in less blood flow and possible brain tissue death.

2.1.1 *Diagnosis*

The only method of definitively diagnosing AD is a brain autopsy. However, mental and behavioral tests and physical examinations allow physicians to make an accurate diagnosis of AD in 90% of cases [16].
For many years, The National Institute of Neurological and Communicative Disorders and Stroke along with the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria [17] and The Diagnostic and Statistical Manual of Mental Disorders last edition (DSM-IV-TR) criteria [18] were among the most adopted for diagnosing AD. Both criteria require deficits in memory and at least one other cognitive domain like:

- the ability to speak or understand spoken or written language;
- the ability to recognize or identify objects;
- the ability to perform motor activities;
- the ability to think abstractly, make sound judgments and carry out complex tasks.

DSM-IV-TR criteria additionally stipulate that the cognitive impairment must have an impact on social life or activities of daily living. NINCDS-ADRDA criteria classify the diagnosis as definite (typical symptoms and histologic confirmation), probable (typical symptoms without histologic confirmation) or possible (atypical symptoms but no alternative diagnosis at least equally likely).

The first important information to know, for applying the previous criteria and obtaining a diagnosis, is the patient's history. Studying the various details about patient's life, the physician will determine what symptoms are present, when they began, and how they have progressed over time. The family history of illness is also fundamental. The physician will perform a physical examination, including blood tests and urinalysis. This is done to exclude other potential causes of dementia, such as hormone imbalance, vitamin deficiency, and urinary tract infections. Brain scans may also be performed to exclude tumors, cerebrovascular accidents, traumatic brain injury, and infections. These scans are also useful in identifying the characteristic tangles and plaques seen in AD. Neuropsychological examinations may be used to identify cognitive symptoms. The most commonly administered test is the Mini-Mental State Exam (MMSE). The physician begins by asking a series of questions designed to test the patient's ability to recall and name a list of objects, perform simple arithmetic, and follow instructions. The patient is then assigned a score out of 30 possible points, with a score of less than 12 indicating severe dementia. AD patient's scores typically decrease 2 to 4 points every year [19]. The physician may also use the Alzheimer's disease Assessment Scale (ADAS) to measure the severity of the disease. The ADAS evaluates the patient's orientation, memory, reasoning and language on a scale of 0 to 70. A higher score represents a higher level of cognitive impairment. The cognitive portion of the ADAS is sensitive to a wide array of symptoms and assesses many cognitive skills, including spoken language ability, recall of instructions, ability to find correct words, following commands, and orientation to surroundings and time. In addition to mental tests, the doctor may perform a neurological exam to assess the function of the patient's brain and nervous system. This exam will test reflexes, coordination and balance, sensation, muscle strength, speech, and eye function [20].

DSM-IV-TR and NINCDS-ADRDA criteria are considered as quite reliable reaching a sensitivity of 81% and a specificity of 70% [21]. However they allow a diagnosis only at late stages of the disease, as it is broadly accepted that the onset of the pathologic process occurs years before the onset of clinical symptoms [22]. Therefore, criteria for mild cognitive impairment (MCI) were estabilished [23], a state in which activities of daily living are basically unimpaired. However, MCI is an unclear phase and precedes also dementia of other types than dementia. This led to a revision of the NINCDS-ADRDA criteria proposing the concept of prodromal AD [24], essentially a subtype of MCI. In order to diagnose prodromal AD, these criteria require a consistent episodic memory deficit plus at least one supportive biomarker. In 2011, the National Institute on Aging-Alzheimer's Association (NIA-AA) workgroup also published new recommendations concerning the definition and diagnosis of preclinical stages of AD [25, 26], recognizing the importance of biomarkers for the early diagnosis of AD and dementia of other causes. Biomarkers can be defined as variables (physiological, biochemical or anatomical) that can be measured invivo and that indicate specific features of disease-related pathological changes. Regarding AD, some biomarkers earned most of the attention and have been studied widely over the last years: the abnormal levels of tau neurofibrillary tangles and beta-amyloid plaques detected through the analysis of the cerebrospinal fluid (CSF) or the PET (positron emission tomography) images and the structural MRI (magnetic resonance imaging) measures of cerebral atrophy [27]. CSF is extracted through a lumbar puncture. In PET technique, the patient is injected with a radio-tracer that can chemically bond with the substance to examine. The tracer after a certain time decays emitting a positron that is indirectly detected through the measure of gamma rays produced by the annihilation of the positron with an electron. In MRI, spin hydrogen atoms that naturally exist within the brain were aligned in the direction of magnetic field applied, without causing any chemical changes in the tissues. When the hydrogen atoms return to their usual alignment, they emit different amounts of energy that vary according to the type of body tissue from which they come. The magnetic resonance scanner captures this energy and creates a picture of the tissues scanned based on this information. PET and CSF are very invasive and the resolution produced on the image is also limited by the type of radio-tracer used, nevertheless these methods can be an effective choice for diagnosing AD because it can detect multiple metabolic processes and the researchers concluded that the addition of these tests to the clinical diagnostic criteria would increase the accuracy of diagnosis. For MRI, the space inside the machine is very small, it may make hard to examine a claustrophobic patient. Then, if a patient has metallic objects inside of their body, they cannot use the MRI system due to the strong magnetic field. However, MRI procedure is painless and can detect very minute abnormalities without the radiation exposure of an X-ray. The resulting image also has high spatial resolution and some scans showed brain loss before the disease was full-blown, indicating that this technique would be effective for early diagnosis of the disease.

2.1.2 *Alzheimer Disease Neuroimaging Initiative*

The ADNI is the most important ongoing multicenter project designed to find the best predictors for Alzheimer's disease (AD) and help worldwide research projects to detect more reliably treatment effects in clinical trials. It was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a 67 million of dollars, 6 year public-private partnership. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is M. W. Weiner, master graduated, director of the Magnetic Resonance Operative Unit of the Veterans Affairs Medical Center and Professor at the University of California in San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations. At the beginning subjects have been recruited from over 50 sites across the U.S.A and Canada. The initial goal of ADNI was to recruit 800 voluntary subjects. Subsequently, ADNI has been followed by ADNI-GO and ADNI-2 and so far these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. In Fig. 4 it is possible to observe the different ADNI phases over time.

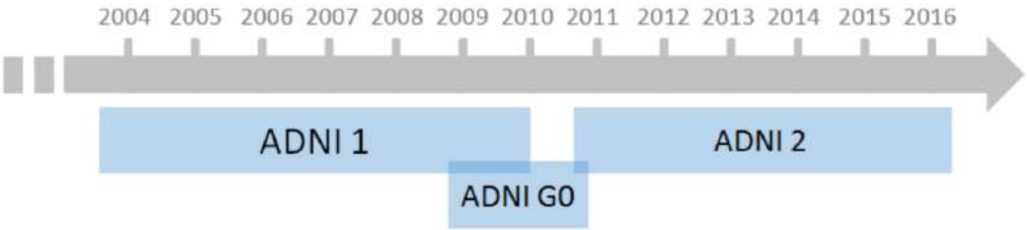

Fig. 4: Timeline of the three ADNI phases: ADNI1, ADNI-GO and ADNI2.

In Table 1 number of ADNI participants in each phase and types of data are reported. In particular with respect to ADNI1, ADNI-GO and ADNI2 added further AD and normal subjects and two new categories: early MCI (EMCI) and late (LMCI) that characterize respectively individuals with very mild memory impairment and with more advanced impairment as defined by the Petersen criteria [28]. Data related to these subjects include structural MRI (1.5T and recently 3T), functional MRI (fMRI), Diffusion Tensor Imaging (DTI), PET with Pittsburgh compound B (PIB) and with 2-fluorodeoxy-D-glucose (FDG) and florbetapir (18F-AV-45) tracers, CSF biomarkers and a variety of Neuropsychological Measurements (NM).

ADNI is also aimed at freely providing to the scientific community, checked data obtained over time, developing technical standards for longitudinal neuroimage studies, determining the optimal methodology for the acquisition and preliminary MRI analysis, validating images and bio-markers also taking into account clinical and psychometric investigations related to them. Data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. Researchers all over the world can obtain free access to the data of the ADNI through acceptance of the Data Use Agreement and the submission of an online application form. ADNI data can be downloaded directly from the website [29], once the access request is accepted and the user is logged in,

	Normal	EMCI	MCI	LMCI	AD	NM	MRI	fMRI	DTI	FDG	AV45	PIB	Biosamples
ADNI 1	200	–	400	–	200	✓	✓			✓		✓	✓
ADNI GO	↓	200	↓	–	↓	✓	✓	✓	✓	✓	✓		✓
ADNI 2	150	150	↓	150	200	✓	✓	✓	✓	✓	✓		✓

Table 1: Number of participants of each ADNI phase and types of data including structural MRI (1.5T and recently 3T), functional MRI (fMRI), Diffusion Tensor Imaging (DTI), PET with Pittsburgh compound B (PIB) and with 2-fluorodeoxy-D-glucose (FDG) and florbetapir (18F-AV-45) tracers, CSF biomarkers and a variety of Neuropsychological Measurements (NM). Subjects recruited over time belong to five categories: Alzheimer's disease (AD), mild cognitive impairment (MCI), late mild cognitive impairment (LMCI), early mild cognitive impairment (EMCI) and normal subjects.

allowing the user to acquire the so called 'study data', along with raw imaging data and genetic data, including genome-wide association study (GWAS) and gene expression results. The study data consists of the following multiple types of data:

- Information on the clinical assessment, including the diagnosed disease status and neuropsychological battery results;
- biospecimen results, such as CSF proteins concentration, ApoE genotype and so on;
- enrollment information, incorporating clinical verification, inclusion/exclusion criteria confirmation, exam dates of the follow-up visits and other registry information;
- genotype outcomes, with the respective methodology description;
- post-processed imaging data, along with acquisition and quality informations;
- medical history information, including adverse events, medication history, physical and neurological examination;
- subject characteristics, such as the Family History or subject demographics.

ADNI website has also an user interface from which it is possible to plot and visualize different cohort characteristics. Besides ADNI allows the user to immediately extract information through ADNIMERGE package available in multiple software like the version used in this thesis implemented in R [30].

2.2 PARKINSON'S DISEASE

Parkinson's disease (PD) is the second most common form of dementia after Alzheimer. Parkinson is a heterogeneous progressive neurological disorder, firstly described almost two centuries ago, basically related with early death of dopaminergic neurons in the substantia nigra and characterized by both motor and non-motor features [31, 32]. It is recognized that age is the greatest risk factor for PD, its incidence reaches a maximum at about 80 years of age. The slow progression is one of the most important features of PD. The disease course can be roughly separated in three phases. The first *prodromal* or *premotor*, phase can last up to 20 years and accounts for symptoms such as: impaired olfaction, constipation, depression, rapid eye movement sleep behavior disorder (RBD) and excessive daytime sleepiness (EDS) [33, 34, 35]. In fact, these symptoms could double the individual's risk of developing the disease. The second phase is called *early* and is characterized by insurgence of typical bradykinesia, tremor and fatigue; then in the texttitadvanced stages, psychosis, dysphagia, freezing of gait, falls and postural instability move in[36, 37]. In Fig. 5 the principal symptoms, during the different phases of the disease progression, are reported.

The basic phenomenon in the genesis of PD is a dopaminergic neuronal loss in the substantia nigra in the basal ganglia. Abnormal aggregates of protein called Lewy bodies can be observed inside the degenerating neurons. The dopaminergic neurons control the function of the extrapyramidal system that processes the movement information from the cortex to the striatum and returns it through the thalamus back to the cortex. In PD, the control of the extrapyramidal system is disturbed and the feedback from the striatum to the cortex is modified. The abnormalities in the function of basal ganglia lead to the motor symptoms of PD [38].

Although the exact cause of PD is still unknown, some risk factors have been identified. The risk factors include genetic and environmental risk factors like toxins, viruses and unhealthy food. The two most important risk factors are an increasing age and a positive family history of PD (in $15-20\%$ of cases). PD by itself is not fatal, it leads to physical deficits that predispose to certain diseases such as pneumonia, falls and resulting complications.

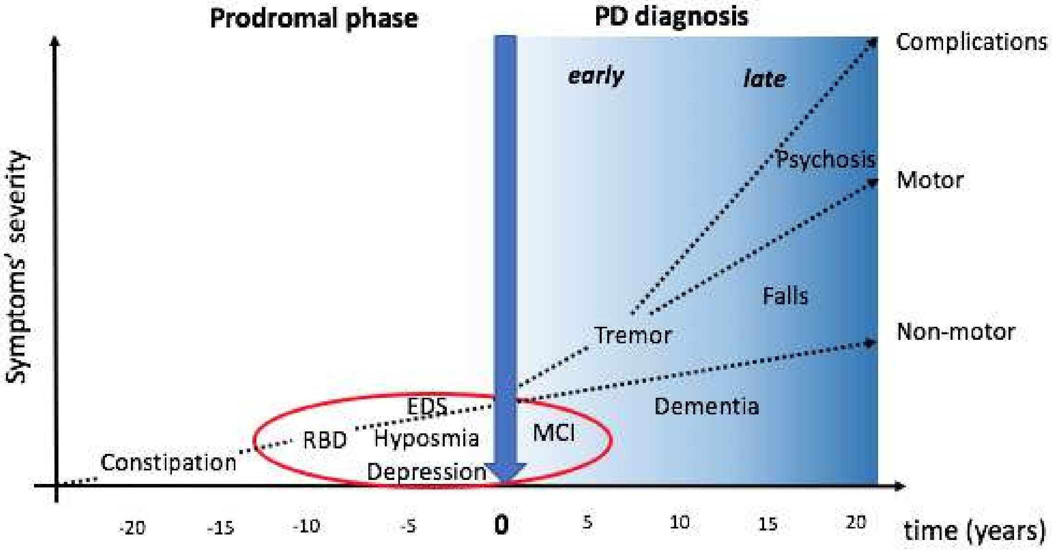

Fig. 5: PD diagnosis is related to the onset of motor symptoms (time 0). The symptoms characterizing the prodromal phase and the years immediately following the diagnosis are: constipation, depression, Rapid Eye Movement sleep behavior disorder (RBD), excessive daytime sleepiness (EDS), Hyposmia (a reduced ability to smell and to detect odors) and mild cognitive impairment (MCI). Accordingly, these symptoms are usually enclosed in models trying to forecast the disease onset.

2.2.1 *Diagnosis*

The diagnosis of PD is based on the presence of clinical symptoms and on the response to anti-parkinsonian medication. The diagnostic accuracy is 75% according to some clinicopathological studies from the UK and Canada, and can be as low as 70% at the early stages of the disease [39]. In specialized units, the diagnostic accuracy can be higher. Several diagnostic criteria have been provided for increasing the diagnostic accuracy. In the criteria, the symptoms have been divided into inclusionary and exclusionary symptoms. The diagnosis requires that two of the four primary symptoms are present

A significant reason for the low diagnostic accuracy is formed by other diseases with similar symptoms than PD. One of these diseases is the essential tremor (ET), which is ten times more common than PD and often confused with PD. Other PD like disorders include: multiple system atrophy, progressive supranuclear palsy, corticobasal degeneration and Lewy body dementia. Imaging techniques, such as the positron emission tomography (PET), single photon emission computed tomography (SPECT) and magnetic resonance imaging (MRI), can help to some extent in differentiating PD from other similar diseases. However, the imaging methods are costly and some of them are not widely available [40, 41].

The motor impairment and the efficacy of treatment in PD are often evaluated using standardized rating scales. The most widely used clinical rating scale is the Unified Parkinson's disease rating scale , which consists of four parts:

- mentation, behavior and mood;
- activities of daily living;
- motor examination;
- complications of therapy.

It has been criticized that the currently used UPDRS is confusing for capturing non-motor elements of PD and therefore the Movement Disorder Society has sponsored a revision of the UPDRS. The established clinical diagnostic criteria in combination with the clinical rating scales are currently the golden standard in the diagnostics of PD and in tracking the disease progression [42, 43].

It is known that the average latency between the onset of prodromal and motor symptoms is about $12 - 14$ years [44]. Thus, it would be of paramount importance the development of diagnostic strategies able to detect the disease in its prodromal phase and outline efficient markers. As shown in [45], a prominent role in early diagnosis should be played by RBD, EDS, Hyposmia, depression and mild cognitive impairment (MCI) which are typical of the non-motor, prodromal or very early, PD phase.

For what concerns imaging, candidate markers include positron emission tomography (PET) [46, 47] or single photon emission computed tomography (SPECT) [48] methods. In fact, these methodologies can accurately detect PD, are very effective in the motor phase, and they may support the diagnosis as well as monitor disease severity and progression. However, these methodologies are based on the detection of substantial losses of dopaminergic neurons, for example in the substantia nigra, whilst it would be desirable to diagnose the disease before this degeneration has occurred in order to enable early diagnosis before the onset of motor symptoms. With this regard, it should be taken into account that PD patients could lose up to 80% of dopamine before symptoms appear [49].

MRI markers could monitor structural changes in the brain and suggest increased risk for PD or be employed for differential diagnosis of PD syndromes [50, 51]. For example, voxel based morphometry (VBM) has revealed significant gray matter reductions in PD patients with dementia [52]; MRI has also shown progressive atrophy in PD, an effect already detectable in the early stage of the disease [53]. Thus, it would seem that there is still room to define effective MRI markers which outline the disease process before the death of dopaminergic neurons has triggered irreversible damages.

Several works have investigated the applicability of machine learning strategies to MRI data with fluctuating outcomes. However, [54] demonstrated that VBM features combined with diffusion tensor imaging can effectively distinguish PD patients from subjects with progressive supranuclear palsy. More recently, a synergistic paradigm combining Kohonen self organizing map and SVM claimed that MRI features can reach accurate classification performances including Subjects Without Evidence of Dopaminergic Deficiency (SWEDD) [55]. Feature selection strategies seem to play a relevant role to define imaging markers accurately distinguishing PD patients from controls [56]. These different approaches share a common feature, all of them rely on the supervised selection of PD-related regions of interest to obtain statistically significant associations between anatomy and clinical phenotype. Although these approaches reach excellent results, they can be limiting as they prevent the investigation of novel brain regions. Finally, transcranial sonography is another promising approach whose clinical applicability is still controversial. It detects subtantia-nigra hyper-echogenicity in PD patients reflecting increases in local iron content and microglial activation. However, there are concerns regarding the specificity and sensitivity of this technique in PD because substantia-nigra echogenicity is a finding that may also change over time in a normal individual [57, 58].

Although much progress has been made in identifying and evaluating biomarkers for PD, none of them is widely available or clinically used for PD. The validation of the biomarkers takes time and a large number of regulatory requirements need to be considered before a new method can be accepted as a clinical tool. It is clear that a single marker will not be able to allow accurate diagnosis and monitor disease progression. Instead, a combination of different markers should provide a more reasonable approach. The established clinical diagnostic criteria in combination with the clinical rating scales are currently the golden standard in the diagnostics of PD and in tracking the disease progression [59].

2.2.2 *Parkinson's Progression Markers Initiative*

Parkinson's Progression Markers Initiative (PPMI) is the largest-scale international project aimed at discovering novel biomarkers for Parkinson's disease and its progression in order to enhance the comprehension of the disease and eventually help the development of disease modifying therapies.

The study was launched in 2002 by The Michael J. Fox Foundation for Parkinson's Research (MJFF) that has been an essential driver of PD biomarker development efforts, funding nearly 90 million of dollars in biomarker research. Through this funding, MJFF has enabled researchers to identify several promising biomarker candidates. Recognizing that biomarker verification is traditionally an inefficient, lengthy and costly process, MJFF has worked with leading experts in the field and invested 60 million to develop a solution that would accelerate verification of PD biomarkers. PPMI is the resulting collaboration between researchers, industry, government and study participants and involves 21 clinical sites in the United States, Europe and Australia.

In the program, 423 patients at different stages of PD, 196 healthy controls, 64 SWEDD subjects and 65 prodromal subjects who have a diagnosis of hyposmia or REM sleep behavior disorder (RBD) are included from more than 30 centers throughout the world. Inclusion criteria required that subjects must have at least two of the following symptoms: resting tremor, bradykinesia, rigidity or either asymmetric resting tremor or asymmetric bradykinesia. In addition, the subjects had to have a pathological SPECT scan and I or II Hoehn and Yahr stage at baseline [60], in other words, he had to be in an initial stage of the disease.

Exclusion criteria were atypical PD syndromes due to drugs or metabolic disorders, encephalitis, or other degenerative diseases. In addition, it was required that the subject was not taking levodopa or other Parkinson's disease medication. Subjects satisfying this exclusion criteria are called *de novo*.

Similarly to ADNI, each researchers, prior registration and acceptance, can download from PPMI website [61] different imaging data (MRI,SPECT, DTI, DWI and PET) and all kinds of clinical, genetic, demographic information and biospecimen like CSF. Among the most clinical scores that PPMI provides, there is the *Montreal Cognitive Assessment* (MoCA), score obtained by a neuropsychological test battery developed to assess major cognitive domains affected by PD. Visuospatial function was evaluated using the 15-item version of the *Benton's Judgment of Line Orientation Test*, which examines the ability of a subject to estimate angular relationships between line segments by visually matching angled line pairs to 11 numbered radii forming a semi-circle. Verbal memory was assessed using the *Hopkins Verbal Learning Test-Revised* (HVLT-R) that consists in reading the words aloud and then asking the individual to verbally repeat the list of words immediately and after a delay of $20-25$ minutes and to identify the words from a verbal presentation. Executive functions were evaluated using some semantic fluency tests. Attention was assessed by the *Letter-Number Sequencing Test* (LNST) where a combination of numbers and letters is read to the subject who has to recall the numbers, first in ascending order and then the letters in alphabetical order. *The Symbol Digit Modalities Test* (SDMT) was also used to assess attention, in this assessment, specific numbers has to be paired with geometric figures based on a reference key within 90 seconds. Besides it is possibile to find the *Movement Disorder Society Unified Parkinson's Disease Rating* (MDS-UPDRS) score that takes into account motor symptoms, *Epworth Sleepiness Scale* (ESS), the Geriatric Depression Scale (GDS) and as previously disclosed the RBD.

3

THE KEY PRINCIPLES OF THE MAGNETIC RESONANCE IMAGING

3.1 AN OVERVIEW

In the next sections, we will introduce the foundation of the Magnetic Resonance Imaging (MRI) and how it works. We will show that MRI constitutes an high resolution and non-invasive means to investigate the three-dimensional architecture of the human brain. Magnetic Resonance Imaging (MRI) is the most popular application of *Nuclear Magnetic Resonance* (NMR), for medical diagnosis. NMR was simultaneously described by Felix Bloch [62] at Stanford University and by Edward Mills Purcell [63] at Harvard University in 1946. The basic principle behind NMR is the characteristic response of a nucleus with spin like H^1, after the effect of the alignment between its spin and a very strong external magnetic field. MRI allows one to acquire non-invasively 3D images at high spatial resolution. Various modalities can be obtained with the same device such as detailed anatomy (structural MRI), functional activity (functional MRI), water-molecules diffusion (diffusion weighted MRI), blood flow measurements (perfusion MRI), distribution of various metabolites (MR Spectroscopy) and also blood vessels (MR Angiography).

3.2 PHYSICAL MODEL

Each particle in an atomic nucleus possesses a fundamental property called a spin. Spin refers to presence of angular momentum in each individual particle. Spins can be positive and negative. Only nuclei with a non-zero spin (nuclei with an unpaired protons or neutrons) are interesting from the NMR point of view because spins of opposite signs cancel out. Nuclei with spin behave as small magnetic dipoles with a random orientation 6.

The result is that that the magnetization of a high number of nuclei cancels out on the average and the net magnetization is zero. In the following text, the attention is restricted to hydrogen nuclei H^1 with only one unpaired proton. MRI is primarily based on the hydrogen atoms because living tissue consists

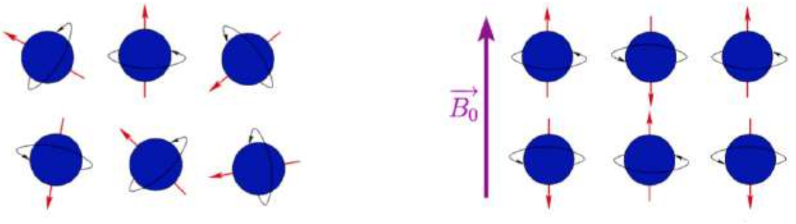

Fig. 6: Random directions of spins in the absence of an external magnetic field (left) and aligned spins in the presence of an external magnetic field (right). The spin rotation occurs within a cone around field direction.

mostly of water and fat which contain many hydrogen atoms. Therefore, approximately 63% of atoms in human body are hydrogen atoms. A magnetic dipole can be seen as a current loop, therefore in a magnetic field $\vec{B}$ it undergoes a Lorentz force given by:

$$d\vec{F} = I d\vec{l} \wedge \vec{B} \tag{1}$$

where I is the module of the current across the loop and $d\vec{l}$ is the infinitesimal loop length having the current direction. This force generates a torque $d\vec{\tau}$ that causes the rotation of the loop:

$$d\vec{\tau} = \vec{r} \wedge d\vec{F} \tag{2}$$

where r is the radius vector of the loop. Integrating and writing the module of the torque we have:

$$\tau = r \; sin\theta \; IlB \tag{3}$$

Thus, it is possible to define the magnetic moment $\vec{\mu}$ the vector perpendicular to the loops and having module equal to $r \; sin\theta \; Il$. Therefore, the torque can be written in terms of magnetic field:

$$\vec{\tau} = \vec{\mu} \wedge \vec{B} \tag{4}$$

However the magnetic moment of a particle, with charge q, mass m and distance r from the rotation axis, is:

$$\vec{\mu} \;=\; I\vec{A} \;=\; \frac{qv}{2\pi r}\,\pi r^2 \vec{u}_n \tag{5}$$

where $A = \pi r^2$ is the area of the circumference described by the particle, $I = \frac{qv}{2\pi r}$ by definition is the variation of charge in the time unit and $\hat{u}_n$ represents the versor perpendicular to the area A. As, the angular moment, by definition, is $L = rmv$ (in the perpendicular situation), we can write:

$$\vec{\mu} = \frac{q}{2m}\vec{L} \tag{6}$$

where $\gamma = \frac{q}{2m}$ is called the gyromagnetic ratio and depends on the particle or nucleus. Therefore, we can obtain:

$$\vec{\mu} = \gamma\,\vec{L} \tag{7}$$

The equation $\vec{\mu} = \gamma\,\vec{L}$ along with $\vec{\tau} = \vec{\mu} \wedge \vec{B}$ completely determines the motion of the magnetic moment $\vec{\mu}$

To summarize the interaction between a magnetic field and a particle with spin has the effect to align particle magnetic moment to the field direction through a torque in order to minimize system potential energy:

$$\tau = -\frac{dU}{d\theta} \tag{8}$$

Taking into account that the torque is the variation of the angular moment over time we can write:

$$\frac{d\vec{L}}{dt} = \tau = \vec{\mu} \wedge \vec{B} = \gamma\,\vec{L} \wedge \vec{B} \tag{9}$$

or its equivalent form:

$$\frac{d\vec{\mu}}{dt} = \gamma\frac{d\vec{L}}{dt} = \gamma\vec{\mu} \wedge \vec{B} \tag{10}$$

The equation just retrieved is:

$$\frac{d\vec{\mu}}{dt} = \gamma\vec{\mu} \wedge \vec{B} \tag{11}$$

It describes the motion of precession of the magnetic moment of a nucleus around the applied static magnetic field. If we consider the angle, between the magnetic field $\vec{B}$ and the magnetic moment $\vec{\mu}$, equal to ϕ, it is possible to find, from geometrical considerations 7, the infinitesimal magnetic moment variation $d\vec{\mu}$ corresponding to an infinitesimal angle variation $d\theta$:

$$|d\vec{\mu}| = \mu \, sin\phi \, |d\theta| \tag{12}$$

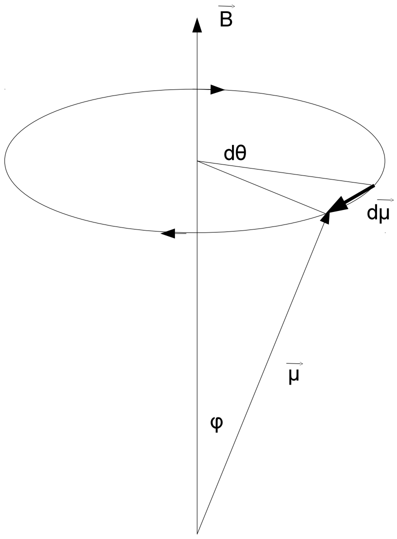

Fig. 7: Clockwise precession of a spin around the magnetic field $\vec{B}$ direction.

By substituting in the motion equation, we can obtain the angular speed or the angular spatial position over time called *Larmor frequency*:

$$|d\vec{\mu}| = \mu sin\phi \, |d\theta| = \gamma|\vec{\mu} \sin\phi \vec{B}|dt \Rightarrow \omega \equiv \left|\frac{d\theta}{dt}\right| = \gamma B \tag{13}$$

The motion equation, once written in terms of Larmor frequency, decomposed into the three cartesian components, and taken the second derivate, become the equation of an harmonic oscillator with angular speed ω. Therefore, it is possible solve this differential equation obtaining:

$$\mu(t) = \mu_{+}(0)e^{-i\omega_0 t} \tag{14}$$

$$\phi_0(t) = -\omega_0 t + \phi_0(0) \tag{15}$$

In conclusion the coupling between magnetic moment and an external static magnetic field is an instantaneous rotation of the moment about the field itself.

At a macroscopic level, a single magnetization vector representing the spin of all the particles, was considered. The net resulting magnetization $\vec{M}$, for example in 1 mm^3 or *voxel* where there are about 10^15 protons, is the sum of all the elementary moments:

$$\vec{M} = \frac{1}{V} \sum_{protons} \vec{\mu}_i \tag{16}$$

Assuming an uniform distribution of the dipoles orientations in a given volume, the net magnetization will be null. However, under the action of a static magnetic field $\vec{B}$ particles will align themselves with the direction of the field and have a magnetization parallel to $\vec{B}$ at equilibrium. Nevertheless, amplitude of $\vec{M}$ represents only a small fraction of what it would have been if all the particles were aligned in the same direction. Indeed, from the laws of thermodynamics, the number of spins following the orientation imposed by $\vec{B}$ (low energy state, called spin-up) slightly outnumbers the amount of spins anti-parallel to the external field (high energy state, called spin-down). The difference is small and given by the ratio:

$$\frac{N^-}{N^+} = e^{-\frac{E_{\vec{B}}}{kT}} \tag{17}$$

where N^- and N^+ are respectively the number of spins in the upper and lower states, k is the Boltzmann constant and T the temperature in Kelvin. Applying the Boltzmann relation, one can estimate that, at the ambient temperature and within a 1.5 T field, there is a difference of 10 in favor of low energy protons among a total of 1 million protons.

The net magnetization $\vec{M}$, as shown in 8 can be decomposed into two components:

- a longitudinal component $\vec{M_z}$, parallel to $\vec{B}$
- a transverse component $\vec{M_{xy}}$, orthogonal to $\vec{B}$

At equilibrium, after a sufficient exposure time to $\vec{B}$, the transverse component $\vec{M_{xy}}$ vanishes. All the individual spins are indeed precessing, but they are all out of phase with each other [64].

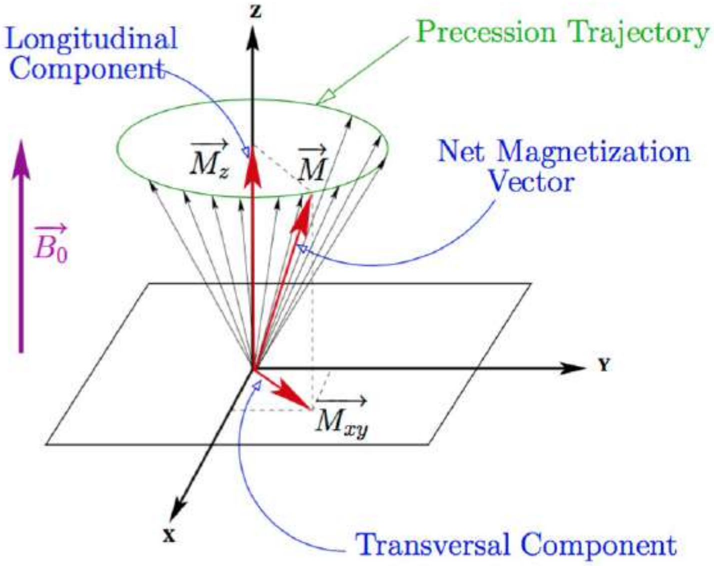

Fig. 8: The net magnetization vector $\vec{M}$, decomposed into a longitudinal component M_z and a transverse component M_{xy}

3.3 EXCITATION PHASE

By applying an oscillating electromagnetic (radio-frequency) pulse containing also Larmor frquency, two phenomena happens in the body to be examined. Firstly, the difference in the number of atoms between the two energy states was perturbed. Indeed, particles absorb energy and their spins flip into high energy state, because a much weaker field than $\vec{B}$ at the Larmor frequency of the targeted nuclei was applied through a rotating reference frame orthogonal to $\vec{B}$. As a consequence longitudinal magnetic field decreases. In addition, the radio frequency pulse pushes protons to synchronize their spins together. This exposure to the electromagnetic pulse causes the net magnetization to spiral away from $\vec{B}$. $\vec{M}$ rotates away from the longitudinal position in an amount proportional to the duration of the pulse. A pulse of 90 degrees would zero out the longitudinal component of M (see 9), while a 180 degrees pulse, or inversion pulse, completely inverts the longitudinal component through an excess of antiparallel spins.

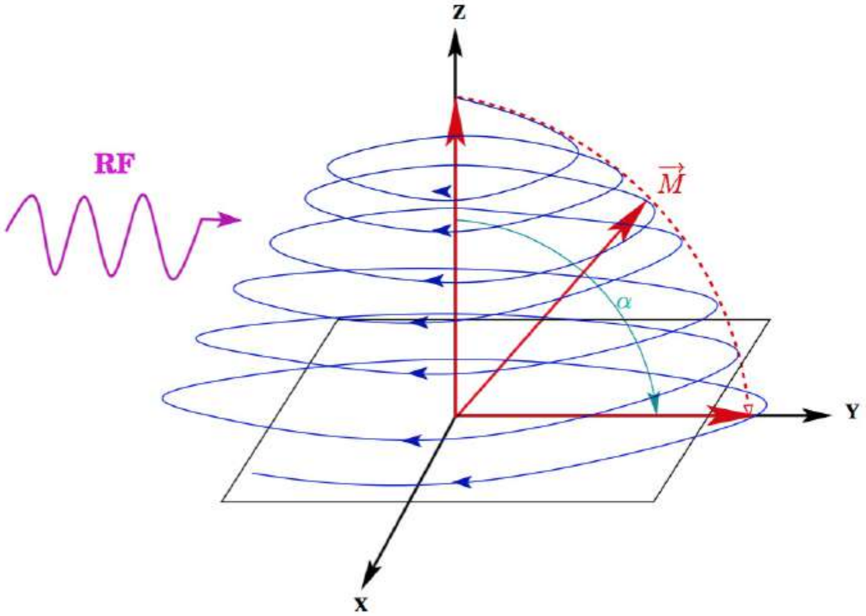

Fig. 9: Excitation phase: the energy given by the RF pulse flips the net magnetization vector of an angle α (here $\alpha = 90$).

3.4 RELAXATION TIMES

By removing the radio-frequency pulse, we can observe two different types of relaxation processes: *spin lattice relaxation* (T_1) and textspin lattice relaxation (T_2).

The spin lattice relaxation is based on the energy exchange between protons and surrounding molecules. This energy dissipation is characterized by the restoration of the longitudinal component that increases again to its equilibrium value. This recovery process is modeled by an exponential function characterized by a time constant T_1, the period for the longitudinal magnetization to recover 63% of its equilibrium value. For a 90 degree excitation pulse, we have:

$$M_z = M_0(1 - e^{-\frac{t}{T_1}}) \tag{18}$$

Spin-spin relaxation refers to the loss of net magnetization in the transverse plane related to protons dephasing. Spins do not only give up their energy to surrounding lattice molecules but also to other neighboring non excited spins. This process is also modeled by an exponential function characterized by another time constant T_2, which corresponds to the period for the transversal component to loose 63% of its value just after the RF pulse

$$M_z = M_0(e^{-\frac{t}{T_2}}) \tag{19}$$

This dephasing is actually further increased by local magnetic field inhomogeneities, since the Larmor frequency will also be nonuniform throughout the region. A time constant slightly different to T_2, T_2^*, is therefore used. The different biological tissues are characterized by respective T_1 and T_2 values, as shown in Table **??**. The intensities of the magnetic resonance images comes from these values. These relaxation times allow tissues with similar density to be differentiated and depend on different factors: matter aggregation, molecule mobility, molecular weight and temperature.

Tissue	T_1**(ms)**	T_2**(ms)**
gray matter (GM)	950	100
white matter (WM)	600	80
muscle	900	50
cerebrospinal fluid (CSF)	4500	2200
fat	250	60
arterial blood	1200	200
venous blood	1200	100

Table 2: The relaxation times T_1 and T_2 for different types of human tissues with an external fields of about 10^-2

3.5 THE BLOCH EQUATIONS

As the magnetic moment also its motion equation can be extended to the macroscopic case obtaining:

$$\frac{d\vec{M}}{dt} = \gamma\vec{M} \wedge \vec{B} \tag{20}$$

Nevertheless, the previous equation neglects the interactions among protons. For having a motion equation as reliable as possible, it is necessary consider that protons naturally tend to align with the external field through an exchange of energy. Moreover, when a radio-frequency pulse is applied protons absorb energy and approximately assume the same phase. In 1946, to take into account these relaxation phenomena, the Bloch equation was introduced by the homonym physicist, Felix Bloch:

$$\frac{d\vec{M}}{dt} = \gamma\vec{M} \wedge \vec{B} + \frac{1}{T_1}(M_0 - M_z)\hat{z} - \frac{1}{T_2}\vec{M}_{xy} \tag{21}$$

where the second and the third term describe respectively that the value of the longitudinal magnetization M_z returns to the initial equilibrium value M_0 with with time constant T_1 and the transverse magnetization M_{xy}, in the plane xy, exponentially drops with time constant T_2 [65].

3.6 IMAGE CONSTRUCTION THROUGH PULSE SEQUENCE

The signal detection from magnetic resonance is based on Faraday's law of electromagnetic induction:

$$\epsilon = -\frac{d\Phi}{dt} \tag{22}$$

where Φ is the flux of the magnetic field through a coil and ϵ the inducted electromotive force. Indeed as a current generates a magnetic field, variations of the magnetic field produce in a coil a current that can be measured. Once the RF pulse is turned off, the same coils, that has produced the excitation of the system, are used to measure the resonance magnetic signal. Different radiofrequency pulse sequences were developed to indirectly measure the relaxation times.

A pulse sequence is a series of RF pulses and magnetic field gradients applied to a sample to produce a specific form of MR signal. It is indeed possible to encode and thus recover the MR signal from specific regions in the volume of interest by means of RF and linear gradients applied along the 3 spatial directions. A first gradient G_z in the $\vec{B}$ magnetic field direction results in a linear intensity variation of the magnetic field that can be used to select a slice. In this case, a slice is a plane orthogonal to $\vec{B}$ with a typical thickness of $1-10$ mm. The spins of a given slice are thus characterized by a specific Larmor frequency. After the RF pulse at the frequency related to the target slice, two transient gradients are applied to encode the x and y dimensions in the slice plane. A first gradient G_y in the y direction induces a phase shift related to the position along the y axis: this is the *phase encoding*. A second gradient G_x in the remaining x direction is applied, leading to a precession frequency variation along the x axis: this is the *frequency encoding*. This process performs an acquisition of the plane data in the frequency space. For each selected slice, an inverse Fourier transform finally maps these data back into the 2D spatial domain.

Gradient Echo is the simplest sequence that consists in the repetition of two basic RF pulse with a delay called *Repetition Time* R_T. Most sequences often comprise additional RF pulses following the slice selection one, to partially refocus the transverse magnetization and produce an echo, leading to a more reliable measure. *Spin-Echo* is the application of a 90 degree pulse followed by a 180 degree pulse after the half of a time called *Echo Time* thais by definition the time spins need to return to the initial phase after a RF pulse. The second pulse refocuses the transverse magnetization and removes local field inhomogeneities dephasing, allowing the measure of the T_2 time. Another sequence is the *Inversion Recovery* which includes a 180 degree pulse followed, after a time called *Inversion Time* (T_I), by a 90 degree pulse that permit the T_1 measure. The choice of the specific pulse sequence parameters (T_R, T_E, T_I) finally determines the image contrast. Two distinct tissues may, for instance, have similar T_1 values but distinct T_2 values, thus the choice of the sequence depends on the information of interest.

3.7 NOISE AND CONTRAST

The quantitative general expression to evaluate the noise affecting a signal s is given by the *signal-to-noise* ratio (SNR).

$$\mathrm{SNR} = \frac{s}{\sigma} \tag{23}$$

with σ that is an estimate of the noise. Two aspects are fundamental in MRI: to accurately measure the signal and discriminate among the different tissues.

The principal noise to consider is the thermal gaussian noise causing uncertainties in the coil and in the sample:

$$\sigma_{thermal} \sim 4kTR\beta_w \tag{24}$$

where R is the effective resistance of the coil, the body and the electrons and β_w is the bandwidth of the noise-voltage detecting system. This noise can be reduced by averaging over several acquisitions

In MRI, improving resolution lowers the SNR. Usually, a better SNR is preferred to an high resolution because several benefits derive from a low resolution such as artifacts reduction, overcoming field inhomogeneities, In the discrimination of different signals in presence of noise, an important concept is that of *contrast-to-noise* ratio (CNR). In the case of MRI, if two tissues correspond to signals s_A and s_B their contrast C is by definition:

$$C \equiv s_A - s_B \tag{25}$$

and therefore the (CNR) is:

$$CNR = C_{AB} \; = \; SNR_A - SNR_B \tag{26}$$

MRI presents three fundamental types of contrast:

- proton density ρ_0;
- T_1 weighted;
- T_2 weighted.

The contrast between two tissues, depends on the relaxation times and times T_E and T_R introduced in the previous sections, according to the following expression:

$$C_{AB} \; = \; \rho_{0,A}(1 \; - \; e^{-T_R/T_{1_A}})e^{-T_E/T^*_{2_A}} - \rho_{0,B}(1 \; - \; e^{-T_R/T_{1_B}})e^{-T_E/T^*_{2_B}} \tag{27}$$

Considering the following limits:

$$T_E \ll T^*_{2_{A,B}} \; \Rightarrow e^{T_E/T^*_2} \rightarrow 1 \tag{28}$$

$$T_R \gg T_{1_{A,B}} \; \Rightarrow e^{T_R/T_1} \rightarrow 0 \tag{29}$$

Proton Density contrast,$C_{AB} = \simeq \rho_{0,A} \; - \; \rho_{0,B}$, is obtained minimizing T_E and T_R. T_1-*weighted contrast*, $C_A B = S_A(T_R) - S_B(T_R)$, was retrieved minimizing T_E. Normal soft tissues are characterized by quite different T_1 values, therefore T_1-weighted images are a powerful method to individuate different tissues. Finally, T_2-*weighted contrast*, $C_A B = S_A(T_E) - S_B(T_E)$, is found maximizing T_R.

If several tissues are present it is useful to adopt two different values for T_R and T_E depending on the contrast considered. The optimal values can be obtained graphically by plotting $C_A B$ between two tissues as a function of T_R or T_E respectively T_R and T_E. The general appearance of the three contrast types is shown in figure 10.

It is worthwhile to note that proton density images are in general interpreted as a count of the proton number in tissues. T_1 and T_2 weighted images appear to have reversed intensities, in particular CSF is brighter in T_2 [66].

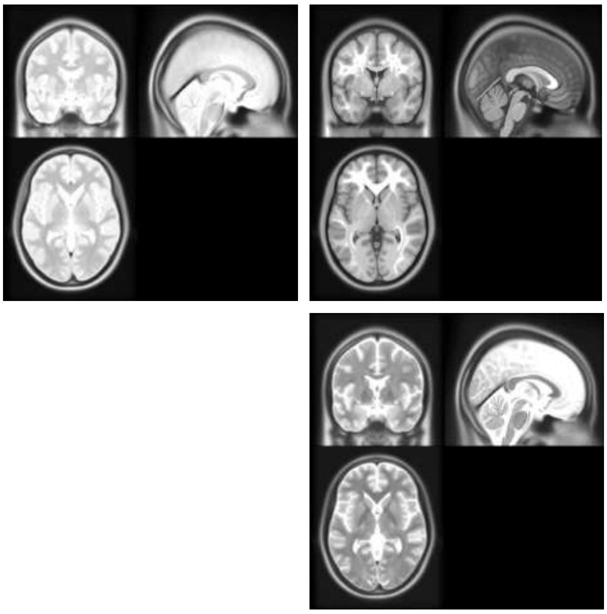

Fig. 10: The figure shows a comparison among (in clockwise order): a Proton density, a T_1-weighted and a T_2-weighted brain scan from ICBM.

4

BRAIN CONNECTIVITY ANALYSIS

4.1 COMPLEX NETWORKS

A network models pairwise relations between objects from a certain collection. These objects are represented by circles called *nodes* or *vertices* and their relations are represented by lines called *links* or *edges*. A concept rather abstract but for this very reason able to adapt to the study of any kind of phenomena as the articles, that have been working for more than a decade, on social, biological, computer and medical networks, show. In this paragraph, it is explained why it is important studying complex networks and the principal challenging goals that complex network scientists have undertaken to reach in the next future. Moreover, the different steps that led to the explosive interest in networks are described.

4.1.1 *Why studying complex networks?*

For a long time physicists studied the components of the nature singularly overlooking to explore the collective system behavior. Soon, they realize, a study taking into account only the singular unit behavior will necessarily be limited. For example, to really understand how the water state transformation from liquid to solid happens, it is necessary to consider the topological modification of the interaction network of water molecules. The whole universe is arranged by all types of systems and researchers are interested in what the systems are composed of, how their components interact together, how they behave and so on. However, the universe elements are often too complicated to be fully understood, since a component in a system might in turn be a complex system. To make the problems easier, scientists consider a given system as a network, by presenting the components as the nodes and the relationships as the links. In everyday life we are surrounded by networks, brain network guides our behavior and our metabolic network offers us energy. Cells are networks of molecules connected by biochemical reactions. Society is a networks of people connected by friendship, familial and professional relationships. Food webs and ecosystems can be considered as networks of predator-prey relations. When we use our phone or laptop to access to our virtual friend network of Facebook, electricity is supplied by the electric power grid to our devices and in the same time we are on the Internet network where World Wide Web pages are connected by hyper links. If we travel by plane we are using one of the many transportation networks, answering the phone we are on the communication network. Networks are everywhere, even the language can be considered a network, made up of words connected by semantic relationships.

In the last decade complex networks have increasingly been attracting attention for the interesting practical applications that could provide in different fields: medicine, business, computing, physics, mass transportation, power distribution, law, and so on. Many vaccination companies are very interested in identifying individuals that have many social ties compared to the average because to treat strategical nodes of the social network, first, allows a greater effectiveness. Studying the interactions among genes, proteins and other molecules, regulating cell activity in organisms, is useful to monitore drug effects. In the business world, understanding how companies, industries, global market and people are connected each other could promote economy and prevent financial failures. As to the World Wide Web, finding out which are the most critical nodes would help to know whether the network is vulnerable to virus attacks and sabotages. In the case of the brain networks, dealt with in this thesis project, exploring the difference between diseased and healthy brain networks could be very important to better understand the disease under investigation and to develop automated systems in view of its early diagnosis.

4.1.2 *Goal of complex network research*

Researchers have defined various metrics to characterize networks from different perspectives and several models to describe network topologies. However, taking into account all properties to investigate each real network could be too complicated and time consuming also because many real networks can be used to address a same issue from different point of views or to solve different problems with the same resources or even they can have different sizes and densities or belong to fields that need of completely different scientific interpretations. It is not clear which metrics are the most suitable and essential for characterizing a particular network. Thus, current complex network research is interested in how to select a set of representative metrics to efficiently characterize a real networks. Scientists are also investing many efforts to understand which combination of models might better describe a specific real network topology. Moreover how to define nodes and links in the most appropriate way is still an open problem. One of the most important organizations in this ambit: the Center for Complex Network Research (CCNR), directed by Professor Barabási summarizes its objectives reporting they focus on how networks emerge, what they look like, how they evolve and how networks impact on understanding of complex systems [67].

4.1.3 *History of complex network*

Network theory comes from graph theory that has its origins in the seven bridges of Königsberg problem asking to find a walk through the city that would cross each of the seven red bridges once, and only once. Every bridge must be crossed completely every time. The starting point and the ending point can be any mainland. The map of Königsberg city, with seven bridges joining the central islet with the mainlands is represented in Fig. 11.

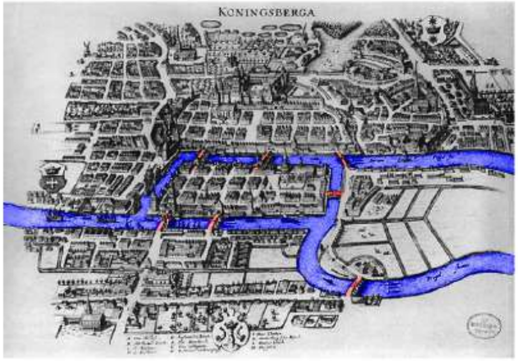

Fig. 11: Map of Königsberg in Euler's time, with seven bridges joining the mainlands.

The answer to this question was given by Euler in 1736. Euler simplified the question using the graph, represented in Fig. 12, and he proved that the problem has no solution.

Indeed, to fulfill the requirement, every node in the graph, except possibly the starting and the ending nodes visited,should be connected by an even number of bridges. Instead, the number of links passing through the four nodes in the considers graph are all odd [68].

Since 1960s, thanks to the work of mathematicians Paul Erdós and Alfred Rényi, network theory approached the graph theory bringing out, for the first time, the common features of these two disciplines and the important properties on both sides. They introduce the probability p indicating level of connec-

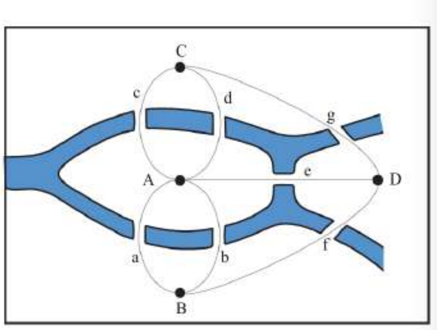

Fig. 12: Euler's graph abstraction of the seven bridges of Königsberg problem.

tion of the networks that is the number of edges. When $p \to 0$ the network becomes a cloud of disconnected nodes whereas in the opposite situation, when $p \to 1$, the outcome graph will be complete that is every node is connected to every others. For intermediate values of p, the network will have approximately $pn(n-1) \setminus 2$ links on average and nodes are connected to randomly selected neighbors [69].

One of the first works able to highlight the presence of networks in everyday life was, in 1977, that of Stanley Milgram, a sociologist at Harvard University. Curious about small-world phenomenon, the same that leaves us astonished when during a conversation with a stranger we find out that we have a common friend, Milgram decided to study the average number of acquaintanceship that separate two perfect strangers. To do this he gave a letter, without any address, to several Kansas and Nebraska citizens, asking them to send the letter to a Massachussets businessman. Given the distance between these States, the probability that senders knew personally the recipient were very low. Therefore, the only way to accomplish the mission was to pass the letter from hand to hand through the network of personal friendships. The result was surprising. To get the letters to destination were necessary, on average, only five brokers [70].

Despite these first attempts to broaden the spectrum of applications, up to the 1990s complex network were subject of study of two restricted groups of researchers. On the one hand mathematicians, little interested in the the applications of networks in everyday life and on the other hand the sociologists unaware about formal results obtained by the first in the graph theory.

The study of networks has a long history, with roots in graph theory and sociology, the modern chapter of network science emerged with an explosive start only during the first decade of the 21st century after the most two cited works respectively of the mathematicians Watts e Strogatz on the small-world networks [71] and the physicists Barabási and Albert on the scale-free networks [72]. Indeed in the past, there were not the tools to map these networks. It was difficult to manage of the huge amount of data behind them. The Internet revolution, bringing fast and effective data sharing methods and cheap digital storage, substantially changed our capacity to collect, assemble, share, and analyze data relative to real networks. This technological advance gave rise to different initiatives. For example, the CAIDA project [73] that offered the first large-scale maps of the Internet; the hundreds of millions of dollars spent by biologists to experimentally map out protein-protein interactions in human cells; the efforts made by social network companies, like Facebook, Twitter, or LinkedIn, to develop accurate repositories of our friendships and professional connections or the Connectome project of the US National Institute of Health that aims to systematically trace the neural connections in mammalian brains [74]. All these factors at the end of the 20th century were the reason of the boom of the network science studies.

4.2 CHARACTERISTICS OF COMPLEX NETWORKS

An advantage of representing a system through the complex networks is to be able to study the main characteristics and compare them with those of other networks using quantitative indices. Such a study can be done through the property characterization of the individual nodes or by measuring the global quantities relative to the whole network such as the size or the overall edge number. In this paragraph the principal metrics used to characterize the complex networks will be described. Besides, the most important and used models, proposed during the last decade to describe network topologies, will be reviewed. Indeed, the topological study of the networks is fundamental with a view to develop network models that are as close as possible to the real networks.

4.2.1 *Graph definition*

A graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ is by definition a couple of two sets $\mathcal{N}$ and $\mathcal{E}$, respectively representing the nodes or vertices and the edges or links of the graph. In particular, $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$ and $\mathcal{E} = \left\{(u_i, u_j), i \neq j\right\}$ when the loops, obtained for $i = j$ are not considered. A compact graph representation can be given through the relative *adjacency matrix* A (30) whose elements a_{ij}, with $i, j = (1, ..., N)$, are defined for each edge $e \in \mathcal{E}$. Adjacency matrix is a squared matrix of dimension N having binary ($a_{ij} \in \{0, 1\}$) elements for *undirected graphs*, negative or positive binary elements $a_{ij} \in \{-1, 0, 1\}$ for *directed graphs* and continuous and real ones for *weighted graphs* that can be directed and undirected.

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \tag{30}$$

While the binary networks indicate only if there is or not a connection, directed networks specify the direction for moving from one node to another. An example is the network of sellers and buyers who are part of a market where links are represented by direction of the use of goods or money among individuals. A weighted graph $\mathcal{G} = (\mathcal{N}, \mathcal{E}, \mathcal{W})$ consists of the previously illustrated sets $\mathcal{N}$ and $\mathcal{E}$ with the addition of a weight set $\mathcal{W}$ of which the elements w_{ij} taking into account the intensity of the connections within the graph, *i. e.* the node similarity. Examples of weighted network are telephone or transport networks where each link brings with it an additional information, such as the time two users i and j are on the phone or the flow of goods and passengers traveling along a certain route. A figurative representation of these typologies is shown in Figure. 13.

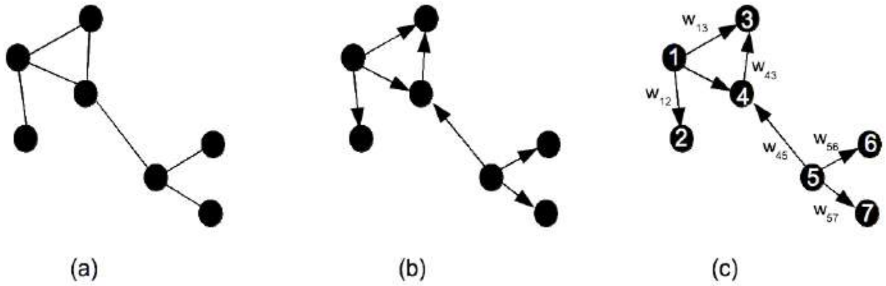

Fig. 13: The figure exemplifies how networks are characterized according to the nature of the interrelationships among the nodes. From left to right: undirected (a), directed (b) and directed weighted (c) graphs.

Another important aspect is the distinction between *static* and *dynamic* networks. The first describes closed systems with a constant number of nodes. Instead the second is used to describe open systems where the number of elements change over the time.

A *subgraph* $\mathcal{G}' = (\mathcal{N}', \mathcal{E}')$ of a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ is a graph such that $\mathcal{E}' \subseteq \mathcal{E}$ and $\mathcal{N} \subseteq \mathcal{N}'$. If $\mathcal{G}'$ contains all links of $\mathcal{G}$ that join two nodes in $\mathcal{N}'$, then $\mathcal{G}'$ is said to be the subgraph induced by $\mathcal{N}'$ and is denoted as $\mathcal{G}' = \mathcal{G}\,[\mathcal{N}']$.

A graph is said to be *connected* if, for every pair of distinct nodes i and j , there is a link sequence from i to j visiting a node only once, otherwise it is said *unconnected* or *disconnected* . A component of the graph is a *maximally connected induced subgraph*. A *giant component* is a component whose size is of the same order as N.

Given the degree matrix K whose diagonal elements are equal to k_i and the remaining elements are null, for each graph $\mathcal{G}$ it is possible define the Laplacian $\mathcal{L} = \{L_{ij}\}$ as the subtraction of K and A so that:

$$k_i = \begin{cases} k_i & \text{if i=j} \\ -1 & \text{if i and j are connected} \\ 0 & \text{otherwise} \end{cases} \tag{31}$$

Laplacian is symmetric by construction and its eigenvalues are all real and non-negative so that $0 \leqslant \lambda_1 \leqslant \cdots \leqslant \lambda_k \leqslant \cdots \leqslant \lambda_N$. The set of L eigenvalues is what is generally referred as *spectrum* of the graph $\mathcal{G}$ itself. Graph spectra and its energy, defined as the sum of the eigenvalues, have interesting properties concerning the connectivity of the graph. For example, the second eigenvalue of the Laplacian, also known as Fiedler eigenvalue, is related to the number of connected components, thus it is also known as algebraic connectivity [75].

4.2.2 *Graph metrics*

Network metrics can be divided into two principal categories: centrality measures and community measures. The former provide information on the fundamental elements of a system, the second ones provide an idea of the internal structure of a graph, establishing, for example, if, within a node group, there are different communities or if there are groups that interact with some rather than with others.

Centrality measures

The *degree* k_i of a node i is the number of edges incident with the node, and is defined in terms of the adjacency matrix A as:

$$k_i = \sum_{j=1}^{N} a_{ij} \tag{32}$$

If the graph is directed, the degree of the node has two components: the number of outgoing links $k_i^{out} = \sum_{j=1}^{N} a_{ij}$ and the number of ingoing links $k_i^{in} = \sum_{j=1}^{N} a_{ij}$. The total degree is then defined as $k_i = k_i^{out} \pm k_i^{in}$. The most basic topological characterization of a graph can be obtained in terms of the *degree distribution* $P(k)$ given by the fraction of nodes $\frac{N_k}{N}$ having degree k. As previously done this definition, along with the next about graph measures, can be extended to the directed graphs.
If there are vertices in a network that have a number of connections greater than the other, these are called *hubs* and their characterization coincides with the study of main nodes of the system.
To switch from *local* to *global* properties, it is necessary to average the local measures over all the elements of the network. Following this procedure the average degree $< k_i >$ can be computed.

$$< k >= \frac{1}{N} \sum_{i=1}^{N} k_i \tag{33}$$

Total link number E can be written in terms of node degree:

$$E = \frac{1}{2} \sum_{i=1}^{N} k_i \tag{34}$$

where the factor $1/2$ takes into account that in the sum, each link is counted two times. Finding the sum from the previous equation and replacing it in the average degree, it is possible to have the following expression:

$$< k >= \frac{2E}{N} \tag{35}$$

thus fixed N, the network is called *dense* if it have an high average degree and many connections otherwise the network is called *sparse*.

The extension of the node degree for weighted networks is the *strength* s_i of a node i:

$$s_i = \sum_{j=1}^{N} w_{ij} \tag{36}$$

Another important centrality measure is the *inverse participation ratio* Y_i of a node i:

$$Y_i = \sum_{j=1}^{N} \left(\frac{w_{ij}}{s_i} \right)^2 \tag{37}$$

Strength measurements denote which nodes are more relevant within a network. Inverse participation ratio attains the heterogeneity of the weight distribution within each layer [76].

A *walk* from node i to node j is a succession of links joining two vertices i and j. The length of the walk is defined as the number of edges in the sequence. A *trail* is a walk in which no edge is repeated. A *path* is a walk in which no node is visited more than once. To introduce other important centrality measures, it is necessary to consider the concept of path. In binary networks, a path is just the sum of links separating two points. Whereas in the weighted networks, the distance between two nodes i and j is defined as the reciprocal of the weight w_{ij}.

$$l_{ij} = \frac{1}{w_{ij}} \tag{38}$$

Therefore a path is the sum of the distances between the nodes that separate the two extreme nodes. In a network the *shortest path length* between two node assumes a strategical role, indeed in brain networks these paths allow fast communications and save system resource. It is useful to represent all the shortest path lengths of a graph $\mathcal{G}$ as a matrix $\mathcal{D}$ in which the entry d_{ij} is the length of the geodesic from node i to node j. The maximum value of d_{ij} is called the *diameter* of the graph. A measure of the typical separation between two nodes in the graph is given by the average shortest path length

$$L = \frac{1}{N(N-1)} \sum_{\substack{j,i \in \mathcal{N}, \\ i \neq j}} d_{ij} \tag{39}$$

Other two standard measures of node centrality are the *closeness* defined as the inverse of the average distance from all other nodes and the betweenness b_i of a node i given by the following expression:

$$b_i = \sum_{\substack{j,k \in \mathcal{N}, \\ j \neq k}} \frac{n_{jk}(i)}{n_{jk}} \tag{40}$$

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. Finally, The edge betweenness is defined as the number of shortest paths between pairs of nodes that run through that edge.

Community measures

It is quite likely to find in many real systems dense subgroups where if two vertices have both a link with a third node then they are connected each other too. The two principal community measures used to describe this behavior are the*transitivity* T and the *clustering coefficient* C. The clustering coefficient c_i of a node i is defined as the ratio between e_i, the effective link number of the neighbors subgraph of i, and $k_i(k_i-1)/2$, the maximum possible number of edges in this subgraph.

$$c_i = \frac{2e_i}{k_i(k_i-1)} = \frac{\sum_{j,m} a_{ij}a_{jm}a_{mi}}{k_i(k_i-1)} \tag{41}$$

The clustering coefficient of the graph is then given by the average of c_i over all the nodes, being a global measure of the network:

$$C =< c_i >= \frac{1}{N}\sum_{i=1}^{N} c_i \tag{42}$$

An alternative possibility is to use the transitivity defined as the fraction of connected triples of nodes which form triangles:

$$T = \frac{3 \times \text{number of triangles}}{\text{number of connected triples of nodes}} \tag{43}$$

The factor 3 compensates for the fact that each complete triangle of three nodes contributes three connected triples, one centered on each of the three nodes.

The clustering coefficient and the transitivity, highlight the tendency of the network elements to group into densely interconnected communities. Nevertheless, these metrics are not able to identify the communities in the network. Recognizing network communities, defined as groups of nodes such that there is a higher density of edges within groups than between them, is important to better understand a system, compare it with others and eventually predict its future evolution. Algorithm to find communities can be divided into *agglomerative* methods, which proceed by series of fusions of the N objects into groups, and *divisive* methods, which separate N objects successively into finer groupings. Agglomerative technique developed by Johnson [77] includes the following steps:(i) Given N items, computing similarities between them; (ii) Finding the most similar pair of items and merging them into a single cluster; (iii) measuring similarities between the new cluster and each of the old items; (iv) Repeat steps 2 and 3 until all items are clustered into a single cluster of size N.

The divisive method of Girvan and Newman [78] comprises the following stages: (i) Calculating the betweenness score for each of the edges; (ii) Removing the edge with the highest score (iii) Analyzing the new components obtained (iv) Go back to point 1 until all the edges are removed and the system breaks up into N non-connected nodes.

Therefore to obtain a community representation of a complex network, it is necessary to chose a level of aggregation or decomposition. Since, for each network, there is not a single possible grouping, it is useful to have a measure to determine which is the optimal subdivision. The quantity that satisfies this request is a global property of the network called *modularity*. Indeed, the level of decomposition or aggregation with the highest modularity corresponds to the modular structure sought. The modularity of a community i is indicated with Q and is given by the next formula:

$$Q = \sum_{i} (e_{ii} - a_i^2) \tag{44}$$

where $\sum_i e_{ii}$ is the fraction of edges in the network that connect vertices in the same community and $a_i = (\sum_i e_{ij})$ give the fraction of edges that connect to vertices in community i. Values approaching $Q = 1$, indicate a strong community structure because the connection number within each subgroup is very different from the link number with the other modules; conversely, $Q = 0$ shows a poor community network.

To investigate the relationships and the hierarchies between different communities, the *participation coefficient* P_i of a node i can be employed computing the coming expression:

$$P_i = 1 - \sum_{c \in C} \left(\frac{k_i(c)}{K_i} \right)^2 \tag{45}$$

with C the set of modules,$k_i(c)$ the degree of node i in module c and k_i the degree node i in the overall network. A node with low participation has high internal connectivity and great control on the information flow within the relative module. Whereas an high participation node has great external connectivity and plays a strategical role in the information transfer to the rest of the network [79].

4.2.3 *Complex network topologies*

System topology modeling is a fundamental element that is extensively used in many fields of science. The main reasons for using modeling in network science are twofold. First, it formalizes a system in order to being able to use mathematical and analytical tools to describe properties in a precise way. Furthermore, is widely used as a prediction and description tool for the real network behavior.

Regular Graphs

Regular graphs are simple and ideal network models. They are characterized by a ordered schemes of connections between the nodes such as rings, grids or lattices. Their peculiar properties is that each node has the same degree k thus it is connected to the same number of neighbors clockwise and to an equal number in the opposite direction as it is possible observe in Fig. 14. Therefore the clustering coefficient and the average path length of these kind of graphs is very high. Indeed, two adjacent linked nodes tend to be connected to the same neighbors. Whereas, to reach a point starting from that diametrically opposite, a distance growing as the node number increases is to be trodden.

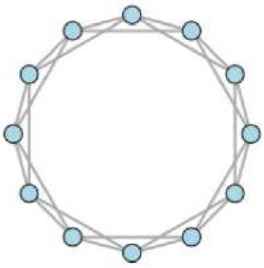

Fig. 14: Example of regular graph where each node has the same degree k

Random graphs

The systematic study of random graphs was initiated by Erdós and Rényi [69] with the original purpose of studying, by means of probabilistic methods, the properties of graphs as a function of the increasing number of random connections. In their first article, Erdós and Rényi proposed a model to generate random graphs $\mathcal{G}(\mathcal{N}, \mathcal{L})$, with N nodes and L links called Erdós-Rényi (ER) graphs. Starting with N disconnected nodes, the model randomly selects couples of nodes and connects them (prohibiting multiple connections), until the number of edges is equals to L. Since there are various possible realizations of a graph $\mathcal{G}(\mathcal{N}, \mathcal{L})$, each one has probability:

$$P(\mathcal{G}) = \binom{N^2}{L}^{-1} \tag{46}$$

An alternative model for ER random graphs consists in connecting each couple of nodes with a probability $0 < p < 1$ independent from other edges. This procedure defines a different ensemble of graphs $\mathcal{G}(\mathcal{N}, p)$ having N nodes connected with probability p and a different average number of links $L = \frac{N(N-1)}{2}p = \binom{N}{2}p$ where $\binom{N}{2}$ is the total number of edges in a complete graph. Thus, each graph with L edges appears with probability:

$$P(\mathcal{G}) = p^L(1-p)^{\binom{N}{2}-L} \tag{47}$$

Besides, considering that the average degree k can be written in terms of link and node number, it is possible to obtain $k = p(N-1) = O(pN)$ for large N. These degree coincides in the limit of large N with the degree $2L/N$ of the first model [80].

In the second model for each pair of nodes the presence or absence of a link is nothing more than the success and failure rate of a binomial distribution, therefore the probability that a given node has degree k is:

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \tag{48}$$

ER models are good predictors of the diameter and the small average path length observed in the real networks. On the contrary they are not able to explain the clustering coefficient and the degree distribution of the highly clustered real systems. Indeed real networks have clustering coefficient that are even several orders of magnitude greater than their edge densities, while ER graphs have a low clustering equal to the probability two nodes are connected and thus a clustering value comparable with edge density of the network.

For large N, the connectedness properties and the structure of ER graph changes for different values of the degree:

- If $k < 1$ then a graph $\mathcal{G}(\mathcal{N}, p)$ will almost surely have no connected components of size larger than O(log(N)).
- If $k = 1$, the critical average degree, a graph $\mathcal{G}(\mathcal{N}, p)$ will almost surely have a largest component whose size is of order $N^{2/3}$.
- If $k > 1$, then a graph $\mathcal{G}(\mathcal{N}, p)$ will almost surely have a unique giant component containing a positive fraction of the vertices. No other component will contain more than O(log(N)) vertices.
- If $c < \ln N$ then a graph $\mathcal{G}(\mathcal{N}, p)$ will almost surely contain isolated vertices, and thus be disconnected.
- If $c > \ln N$ then a graph $\mathcal{G}(\mathcal{N}, p)$ will almost surely be connected.

Small-World networks

The research of a network model able to represent real world continued with Wattz and Strogatz [71]. They introduced a new model known as small world. The starting point is a N node and k degree ring, in which each node is symmetrically connected to its $2k/N$ nearest neighbors. Then, for every node, each link connected to a clockwise neighbor is rewired to a randomly chosen node with a probability p_{rew}, and preserved with probability $1 - p_{rew}$. In Fig. 15 it is possible notice that for $p_{rew} = 0$ network retains its regular lattice topology, while for $p_{rew} = 1$ the model produces a random graph with the constraint that each node has a minimum connectivity $k_{min} = k/N$ all lattice-like features have disappeared. Intermediate values of $p_{rew} = 0$ result in networks that consist of a mixture of random and regular connections.

Also the study of the network properties as a function of the rewiring probability is very interesting. Indeed, in Fig. 16 it is shown clustering coefficient $C(p_{rew})$ and the average path length $L(p_{rew})$, both normalized by their values for the regular network $P(0)$ and $L(0)$. Note that there is a broad range for the rewiring probability p where networks have clustering that is similar to that of the regular network, and a path length that is similar to that of the random network. Within this range, networks exhibit small-world attributes. The

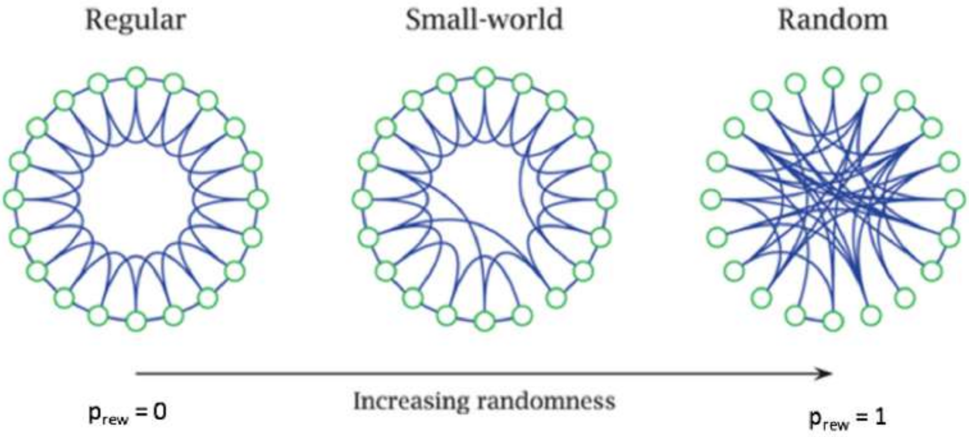

Fig. 15: Random rewiring procedure for moving from a regular ring lattice to a random network, without changing the number of nodes or edges in the graph.

two previous network metrics were computed by Watts and Strogatz performing 20 random realizations of the rewiring process with networks consisting of 1000 nodes and degree 10. Alternative procedures to construct small-world networks exist and are based on adding edges instead of rewiring them [81]. Small world properties have been widely found in food web, the World Wide Web, power grid networks, brain networks and many other systems. Even though small-world model is considerably more adherent to real networks than random graphs, it has some limitations: it do not reproduce the dynamics with which a real network evolves and the degree distribution that in many systems, such as the brain, is a power-law and indicates the presence of hubs.

Scale-Free networks

The graphs able to model networks with degree distribution are called scale-free. In many real world networks, the degree distribution follows a power law $P(k) \sim ck^{-\gamma}$ where c is a constant and γ is a positive exponent, the value of which determines different behaviors of the network [82]:

- If $1 < \gamma < 2$ the regime is abnormal and the number of the links to the largest hub grows faster than the size of the network. This means that for N large enough, hub degree exceeds the total number of nodes in the network. Because of these network abnormalities scale-free networks in this regime can not exist.

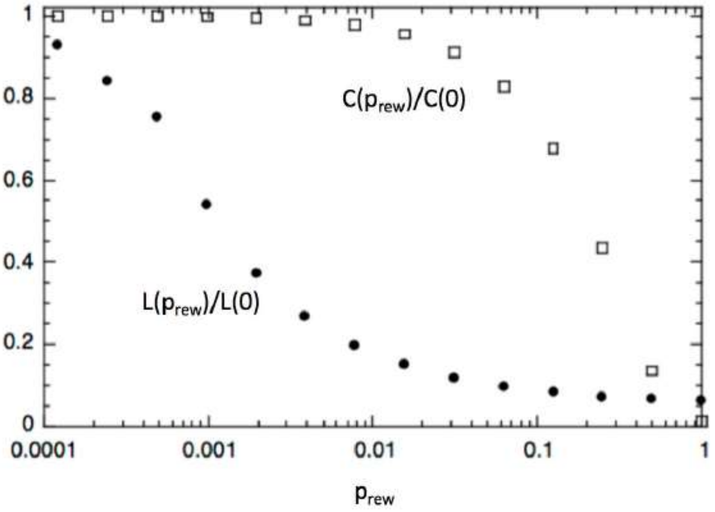

Fig. 16: Characteristic path length $L(p_{rew})$ and clustering coefficient $C(p_{rew})$ both normalized by their values for the regular network $P(0)$ and $L(0)$ for 20 randomly rewired graphs having 1000 nodes and degree 10.

- If $2 < \gamma < 3$, the regime is scale-free and the network has small-world characteristics and a great number of hubs.

- If $\gamma > 3$, the regime is random and it is difficult to distinguish the properties of a scale-free network from those of a random one because for great values of γ the distribution $P(k)$ decays enough quickly to make the hubs smaller and less numerous.

Having a scale-free network, thus a network degree distribution that has a decaying tail in the power law means that the vast majority of nodes have low degree and that there exist few nodes, the so-called hubs, that have an extremely high connectivity as it is possible to notice in Fig. 17.

Such networks have been named scale-free because, at all scales, network shape remains substantially unchanged and power-laws have the property of having the same functional form apart from a multiplicative factor. This is extremely important in statistical physics relatively to the study of fractals and phase transitions.

Even though one might expect a limited influence of the hubs within real networks because of their small cardinality, they play a strategical role. For example in brain networks, hubs and their connections are particularly vulnerable to several neurodegenerative diseases.

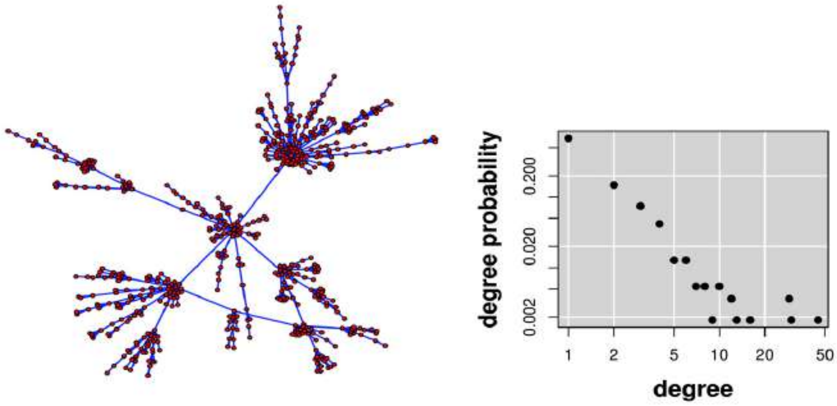

Fig. 17: On the left, the representation of a 549 node scale-free network simulated with the Software R [30]. The network shows the striking fractal quality to appear similar at various levels of scale as when considering the shape of a tree, then the shape of its major branches and so on until the structure of individual leaves. On the right, the degree distribution of this scale-free network in logarithmic scale.

There are two types of scale-free models available in the literature. The former creates static scale-free networks and is simply generated as a special case of random graphs with a given degree distribution [83]. The second one, developed by Barabási and Albert, is the most used and realistic because creates evolving scale-free networks taking into account structural properties of the network [72]. This latter model is based on two basic ingredients: growth and preferential attachment. The growth is obtained considering a network of few nodes N_0 and at each time step a new node with connectivity $N < N_0$ is added to the network. The preferential attachment is based on the concept that the more connected a node is, the more likely it is to receive new links. The new node j links to an existing node i in the network with a probability Π which is function of the connectivity degree k of node i

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \tag{49}$$

This model over the time has seen some improvements. Indeed, in many real networks such as the World Wide Web often some new nodes are able to have a connection number far above the oldest hubs. Therefore Barabási along with Bianconi, to take into account this behavior contrary to preferential attachment, introduced the *fitness term* that is the ability of a node to attract new network elements [84].

The fact that a network is scale free is a sufficient, but not a necessary condition for being small-world. Indeed, there are small-world networks that do not have the characteristic scale-free behavior because of two effects that limit the preferential attachment. The former is the node aging because often in real networks there are nodes that go on to contribute to the network statistic but they do not receive more links. The second effect concerns the impossibility of an hub to effectively manage all the possible connections.

4.3 BRAIN CONNECTIVITY MODEL

In literature, various studies based on the brain description through complex networks have shown that graph theory is a valid mathematical tool to detect alterations due to neurodegenerative diseases [85]. The existing approaches, that apply the theory of graphs to neuroimages, have the limit not to be able to collectively study what happens to the same nodes as their interactions change. However, in the context of complex networks in the last two years, the concept of multiplex network have been developing. Multiplex networks are innovative investigation tools able to provide an information among networks with fixed nodes and variable connections that determine different layers. In fields other than neuroscience, it was shown that the content of a multiplex network is greater than the sum of the information contained in the individual parts [86]. To apply this mathematical framework to structural magnetic resonance images, the first issue to be addressed, was the definition of nodes, connections, and layers of the multiple network. In fact, there is no standard method for building a brain network and, in principle, any quantity associated with neuroimages can be used to define a complex network as it is possible to observe in Fig. 18 and in Fig. 19.

Depending on the type of data and the information to be studied, nodes can be anatomical regions of interest (ROI), single voxel (volumetric units of $1\mathrm{mm}^3$), acquisition sites or electrodes, if considering the study of functional magnetic resonances, electroencephalograms and magnetoencephalograms. Besides, there are also many different ways to choose the edges, they can be similarity measurements or even real physical entities such as the number of fibers linking two regions in tractographies extracted from diffusion tensor imaging (DTI).

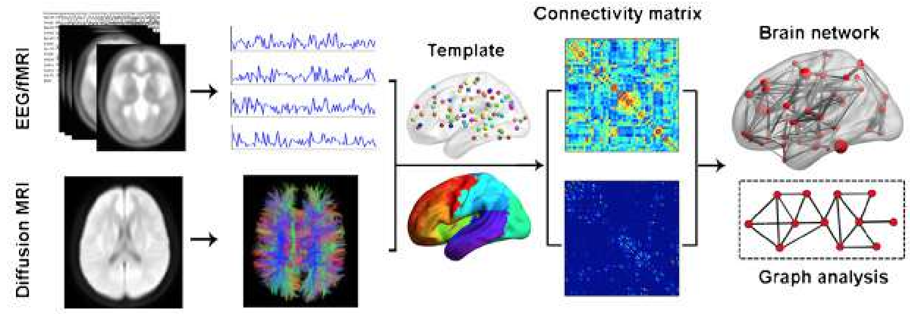

Fig. 18: Representation of the different necessary steps to explore brain connectivity starting to Diffusion magnetic resonance images and electroencephalograms. After node and edge definition and the subsequent brain network construction defining edges and nodes, it is possible to represent every network with its relative adjacency matrix in order to extract and analyze the properties of the associated graph [87].

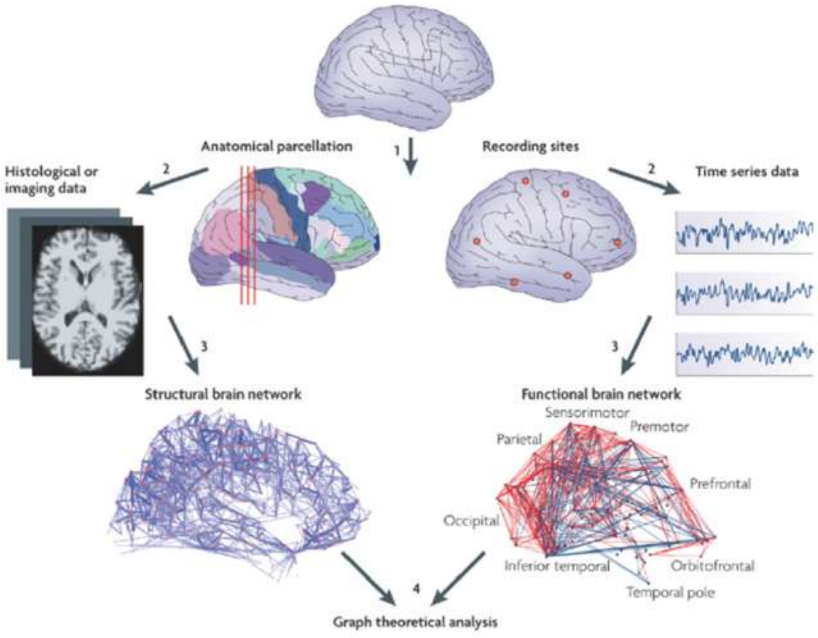

Fig. 19: Representation of the stages used to investigate brain connectivity starting to structural and functional magnetic resonance images. Once network nodes and edges are defined, it is possible to build and consequently study a brain network [88].

4.3.1 *Materials*

As to the study of the Alzheimer's disease, it was used for training a set $\mathcal{D}_{\mathtt{train}}$ of 67 MRI scans, composed of 29 normal controls (NC) and 38 Alzheimer's disease (AD) subjects, from the Alzheimer's Disease Neuroimaging Initiative (ADNI). These subjects belonged to a larger benchmark dataset selected in order to obtain a compact yet representative sample of ADNI [89]. An independent test set of 148 subjects $\mathcal{D}_{\mathtt{test}}$, composed by 52 NC, 48 AD and 48 subjects with mild cognitive impairment converting to AD (cMCI) was also employed. $\mathcal{D}_{\mathtt{test}}$ subjects were randomly chosen within the whole ADNI in order to match the demographic characteristics of training subjects. The training sample (67) and the test sample (148) are of sufficient size for the construction of robust classification models [90, 91]. cMCI subjects had converted to AD in a time range of $[30, 108]$ months subsequent to the initial assessment. All 215 participants underwent whole-brain MRI at 34 different sites. Both 1.5 T and 3.0 T scans were included in $\mathcal{D}_{\mathtt{train}}$ and $\mathcal{D}_{\mathtt{test}}$.

ADNI images consisted of T1-weighted MP-RAGE MRI brain scans, acquired with a gradient echo 3D technique and a resolution of $1 \times 1 \times 1$ mm^3; as a consequence from now on, voxels and mm^3 will be interchangeably used. Some clinical and demographic information, including the Mini Mental State Examination (MMSE) score, age, years of education and sex for the $\mathcal{D}_{\mathtt{train}}$ and $\mathcal{D}_{\mathtt{test}}$, is detailed in Table 3. MMSE is a 30-point questionnaire that is used extensively in clinical and research settings to measure cognitive impairment, low to very low scores correlate closely with the presence of Alzheimer's dementia. Except for MMSE scores, there were no significant differences among the three groups.

For studying Parkinson's disease (PD), in this thesis, data from the Parkinson's progression markers initiative (PPMI) database were used. MRI data obtained by the PPMI for this study consisted of MPRAGE T1 brain scans from 3T SIEMENS MAGNETOM Trio scanners. Images were acquired with the following parameters: repetition time 2300 ms, echo time 2.98 ms, flip angle 9° and voxel size $1 \times 1 \times 1\ mm^3$. Data from PPMI comes from different worldwide sites, along with structural MRI it is possible to find demographic and clinical metadata, such as age, gender and cognitive scores. It is worth noting that PD patients are de novo patients in that they are unmedicated, an important aspect, as PD therapies could not have the desired effect of modifying the possible markers. More importantly, PD patients enrolled are mostly at the first stages of the disease, according to the Hoehn and Yahr scale [60]; in fact, [92] explains that 98% of the subjects affected by the disease is in stages 1 and 2 (over 5), corresponding to mild inconvenient but not disabling symptoms.

	$\mathcal{D}_{\textbf{train}}$		$\mathcal{D}_{\textbf{test}}$			**Total**
Disease status	AD (38)	NC (29)	AD (48)	NC (52)	cMCI (48)	215
Female/male	18/20	13/16	22/26	25/27	21/27	99/116
Age (years)	74 ± 8	75 ± 6	78 ± 6	75 ± 6	76 ± 6	76 ± 6
Education (years)	15 ± 3	17 ± 3	15 ± 3	16 ± 3	15 ± 3	16 ± 3
MMSE	23 ± 2	29 ± 1	24 ± 2	29 ± 1	27 ± 2	26 ± 2
n sites	23	19	26	29	18	34

Table 3: Group size and sex information are reported for each class. The table also provides age, years of education and MMSE (mean and standard deviation). The disease status reported is as assessed at baseline visit. MMSE score resulted statistically different for all groups with a p-value < 0.01 except between $\mathcal{D}_{\mathtt{train}}$ and $\mathcal{D}_{\mathtt{test}}$ normal controls and between $\mathcal{D}_{\mathtt{train}}$ and $\mathcal{D}_{\mathtt{test}}$ AD patients.

The database consisted of two populations including respectively 107 male and 62 female NC, for a total of 169 subjects, and 243 male and 131 female patients, for an overall amount of 374 PD subjects. The populations are matched for age (60.2 ± 11.5 for NC and 61.6 ± 9.8 for PD). In the following Table 4 the baseline values for the Epworth Sleepiness Scale (ESS), the Geriatric Depression Scale (GDS), the Montreal Cognitive Assessment (MoCA), the Movement Disorder Society Unified Parkinson's Disease Rating (MDS-UPDRS) and the Rapid Eye Movement sleep behavior disorder (RBD) are enlisted.

$\mathcal{D}_{\textbf{PPMI}}$

Diagnosis	**Age**	**ESS**	**GDS**	**MDS-UPDRS**	**MoCA**	**RBD**
PD (374)	61.6 ± 9.8	3 [2, 6]	2 [1, 3]	29 [18, 41]	27 [26, 29]	5 [3, 8]
NC (169)	60.2 ± 11.5	2 [1, 4]**	1 [0, 2]**	27 [17, 37]*	28 [27, 29]**	5 [3, 8]

Table 4: Demographic and clinical information. Mean age and standard deviation are given, for other indicators with asymmetric distributions medians and interquartile ranges are preferred. Significant differences between normal controls (NC) and Parkinson's disease (PD) are reported with the Kruskal-Wallis p-value (* for $p < 0.05$ and ** for $p < 0.01$).

ESS is a standardized simple measure for sleep propensity [93]. Daytime sleepiness is usually associated to sleep disorders, but it can also be a symptom of prodromal PD phase along with RBD. Like constipation and olfactory disturbance, RBD can precede the development of the motor signs of Parkinson's disease and longitudinal data suggest that RBD heralds the onset of motor symptoms in up to 40% of patients [35]. RBD is measured according to the screening questionnaire proposed in [94]. ESS and RBD scores lie within ranges of normalcy for both NC and PD subjects, however PD patients show a small but significant increment in ESS (Kruskal-Wallis p-value < 0.01).

GDS score is based on a questionnaire of 30 items with binary outputs [95]. For each affirmative answer 1 point is scored; healthy people should score 5 ± 4, mildly depressed and very depressed people 15 ± 6 and 23 ± 5 respectively. In particular, for the present study, the shorter form was used [96]. Accordingly, the subjects of the study were not mildly either very depressed, nevertheless PD patients showed a significantly higher GDS score (Kruskal-Wallis p-value < 0.01).

The motor symptoms are taken into account by the Movement Disorder Society Unified Parkinson's Disease Rating (MDS-UPDRS) [43]. It is important to recall how the MDS-PD criteria assign a central role to motor symptoms to define clinical PD [97]; however, non-motor manifestations are present in most patients so that this feature can play a fundamental role for diagnosis, but it cannot capture the complexity of this heterogeneous disease. As expected, MDS-UPDRS significantly distinguished (Kruskal-Wallis p-value < 0.05) the NC and PD cohorts.

Mild cognitive impairment (MCI) is a symptom commonly found in PD patients; it usually occurs with the onset of motor symptoms and may be a harbinger of dementia. MCI condition could be related to early PD symptoms, such as RBD, in any case it is known that it may be found in up to 80% of long term PD patients [98]. The Montreal Cognitive Assessment (MoCA) index is the preferred measure for accurate screening of cognition in PD [99]. A final MoCA score of 26 and above is considered normal: PD and NC cohorts of this study resulted normal on average with 27.0 ± 2.3 and 28 ± 1 respectively. A small but significant difference was observed (Kruskal-Wallis p-value < 0.01).

4.3.2 *Image processing*

To perform a suitable medical image study, a preliminary processing, commonly called *image registration*, is necessary. The term 'image registration' indicates the spatial alignment between images of the same or different subjects, acquired with the same or different modalities, and also the registration of images with the coordinate system of a treatment device or tracked localizer. Automated saving time tools for the registration, are fundamental to make the best possible use of the medical images in terms of speed and efficiency. Registration process can be used for multiple goals: combining information from several imaging modalities; investigating changes in shape, size or image intensities over time, coordinating preoperative images and surgical plans and matching an individual's anatomy to a standardized atlas.

A indispensable preparation step for the registration is the brain extraction that remove non-brain tissue with highly variable contrast such as scalp, marrow and so on. The word registration is used with two slightly different meanings. The first meaning indicates a transformation that can relate the position of features in an image or coordinate space with the position of the corresponding feature in another image or coordinate space. The second meaning of registration denotes not only relates the position of corresponding features but enables also the comparison of the intensity at those corresponding positions. To reliably overlay two images or subtract one from another, it is necessary to know the more complete transformation that maps both position and associated intensity value from an image A to an image B.

Registration algorithms that make use of geometrical features in the images such as points, lines and surfaces, identify features like sets of image points that correspond to the same physical entity visible in both images. Then, they iteratively determine the spatial transformation and infer from it the global transformation when the algorithm converge.

Registration algorithms that are based on image intensity values work differently. They iteratively determine the image transformation that optimizes a voxel similarity measure, modifying image until the *cost function* associated to the similarity metric reaches a minimum. It is worthwhile to note that the metric choice has a strategical role to achieve satisfactory results. Another optimization technique for the image registration is the interpolation. It is a numerical analysis tool that for an image interpolates the point values of a range , dividing them into more sub-ranges and choosing for each of them a polynomial that best interpolates them. Next, the continuity conditions of the first derivatives for consecutive polynomials are imposed. Also if for evaluating the cost function, that quantifies the goodness of the registration, is necessary only the site intensity, an interpolation of the intensities in the several positions can be useful.

The spatial mapping between two 3-D magnetic resonance images is in general defined as a function $\mathcal{T}$ which maps a vector $\vec{x}$ representing the spatial coordinates of a point in a new reference frame coordinates $\vec{x}'$:

$$\mathcal{T} : \vec{x} \in \mathbb{R}^3 \mapsto \mathcal{T}(\vec{x}) = \vec{x}' \qquad (50)$$

In general the input data is also referred as *moving space/image* whilst the arrival space is called *target space/image*.

The rigid transformation is the simplest and is characterized by six degrees of freedom: three translations and three rotations. The key characteristic of a rigid body transformation is that all distances are preserved. Example of rotation and translation are depicted in Fig. 20.

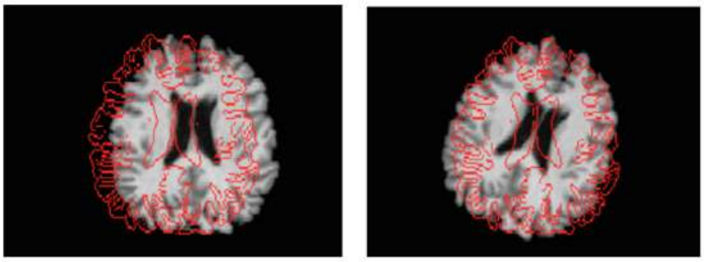

Fig. 20: The figure shows the transformations that compose the rigid registration: translation on the left and rotation on the right [100].

In this type of registration, the intensity information can be used to calculate the images center of mass or the momentum distributions. In the first case the moving and the target images are registered shifting the moving image to align the centers of mass while in the second, the first moments are aligned. In general this second procedure is used, because of its generality in that it does not make any assumption about the physical origins of the moving and target images. After images are rigidly registered, a second registration was usually performed to tackle deformations due to heart and respiratory cycles, and in general to any position change. A transformation with 12 degree of freedom that includes scaling and skews as well as the rigid body parameters is referred to as affine, and has the important characteristics that all line parallelism is preserved but their lengths and angles can change. In Fig. 21 transformations that characterize this registration are represented.

An affine registration can add to the previous rigid transformation three degrees of freedom through an anisotropic dilation, that is a different scale factor along the three directions (x, y, z) and also other three degree of freedom through a distortion. Therefore, a rigid body transformation can usefully be considered as a special case of affine, in which the scaling values are all unity and the skews all zero.

Also if affine transformation is not really a linear registration because it is composition of linear transformations with translations, all the the transformations, with a number of degree of freedom greater than 12, are called non linear registrations. A non linear registration can be considered a warp field that

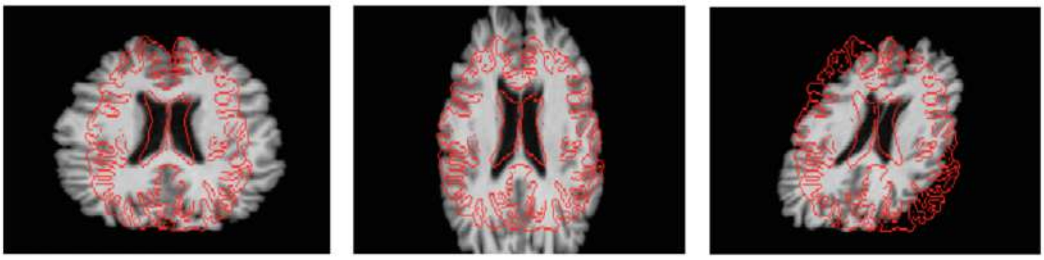

Fig. 21: The figure depicts an example of scaling along x axis (on the left), along y axes (in the middle) and an example of skew (on the right) [100].

records the displacement vectors for each voxel of the image that needs to to be aligned to the target image, as shown in Fig. **??**. In this kind of registrations it is very important to use optimizations based on the cost function which is the combination of a term due to the similarity measure and a regularization term. The term of regularization gives a high cost to unwanted transformations, such as an excessive stretching or a too accentuated local curvature.

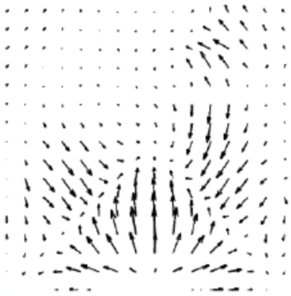

Fig. 22: Deformation field displayed as displacement vectors along a plane [100].

After images are registered to a template or a reference atlas they all share common features that can be therefore analyzed in a statistical framework [101].

4.3.3 *Node and link design*

Standard image processing procedure was carried out with the Oxford FMRIB library FSL [102]. Firstly, MRI scan intensity differences were normalized, then intra-cranial regions were extracted with the FSL Brain Extraction Tool (BET). Secondly, spatial normalization was performed to co-register the different images into the common coordinate space MNI152 [103] of size of $197 \times 233 \times 189$

mm^3 and resolution of $1 \times 1 \times 1$ mm^3; an affine registration, with default configuration, was performed with the FSL Linear Registration Tool (FLIRT). Linear registration is preferred to a non-linear for the methodology to be robust to subtle local differences, due for example to subject morphological variability, or small registration failures. Finally, using the template brain coordinates, the brain of each subject into the two hemispheres was automatically segmented and, starting from the medial longitudinal fissure plane, was uniformly divided each hemisphere in an equal number of rectangular ($l_1 \times l_2 \times l_3$) boxes called patches or supervoxels and covering the whole brain, see Fig. 23. From now on, the terms patches and supervoxels will be interchangeably used. Only those patches overlapping with the template brain for more than 10% were kept.

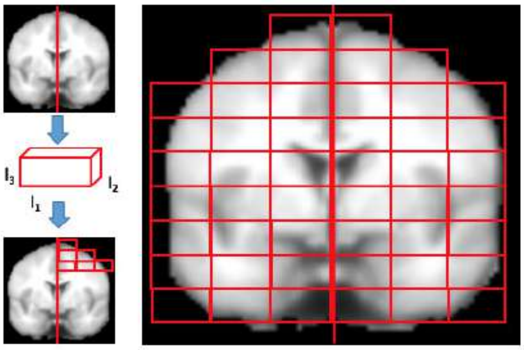

Fig. 23: The figure qualitatively shows how MRI brain scans are segmented in rectangular boxes of dimensions $l_1 \times l_2 \times l_3$ called patches or supervoxels. Firstly, the brains normalized to MNI152 template are divided in left and right hemispheres using the medial longitudinal fissure, then the patch dimensions are set and finally the brain is segmented. Only patches overlapping the brain for at least the 10% of their content are kept, others are discarded.

For each subject, an undirected weighted network having as nodes the patches and as edges the pairwise Pearson's correlation among patches, was built. Patches do not necessarily pinpoint a predefined anatomical area, as it happens for a Region Of Interest (ROI) based approach. Indeed, patches can be chosen with an intermediate size V, considering that too large regions would not be able to distinguish, from the natural inter-subject variability, the subtle disease effects, often diffused in different points of a region. On the other hand, too small patches could be considerably affected by registration noise as in the voxel-based approaches. However, to investigate the robustness of the representation and its meaningfulness, the size V of supervoxels was varied from a minimum of 1000 to a maximum of 4000 for AD (see Section 5.2.2) case and from 125 to 8000 voxels for PD case (see Section 5.5.1). Edges represent the similarity grade between patches and are defined ,in the proposed approach, in terms of Pearson's correlation coefficient r_{ij} with $i, j = (1, ..., N)$:

$$r_{i,j} = \frac{\sum_{k=1}^{V}(s_i^k - \bar{s_i})(s_j^k - \bar{s_j})}{\sqrt{\sum_{k=1}^{V}(s_i^k - \bar{s_i})^2}\sqrt{\sum_{k=1}^{V}(s_j^k - \bar{s_j})^2}} \tag{51}$$

The numerator is the sum over the product of the voxels intensities s_i^k and s_j^k at each voxel position k after subtraction of the supervoxel average values and the denominator is the product of standard deviations of s_i and s_j gray-level distributions.

As AD is a neurodegenerative disease it affects the brain involving a moderate or a severe grade of atrophy. Thus, Pearson's correlation, as other similarity metrics used in structural imaging, can be a suitable choice to model this effect. It is fast to implement and compute, simple to understand and interpret and it does not require any scaling or centering of the supervoxels as it is intrinsically normalized. This measure emphasizes how similar two patches are in two distinct ways; first of all, comparing the gray-level intensities of voxels, correlation gives a measure of how much the quantities of gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) are equally distributed in the examined patches. In addition, as this comparison is performed voxel-by-voxel, correlation measures how similar the spatial distribution of matter is in different patches. Two patches are highly correlated if, voxel by voxel, their gray levels are similar and therefore if, spatially the distributions of white matter, gray matter and cerebrospinal fluid (corresponding to voxel intensities of decreasing intensity) are similar as well.

In the case discussed here, it is worth noting that negative correlations can be found, for example, between patches in which gray matter and white matter undergo a left-right inversion. Distinguishing positive and negative correlations would have included in the model a left-right bias. Therefore, it was decided to consider undirected networks keeping intact the informative content about structural modifications as those for example caused by atrophy, see Fig. 24. Atrophy effects presented are of course to be intended subtler that those magnified in the illustration, nevertheless this kind of effects should be detectable with a network description.

Correlation is a similarity criterion that associates correspondent voxels within supervoxels taking spatial relationship between voxels into account. To investigate the importance of preserving spatial voxel correspondence, different sim-

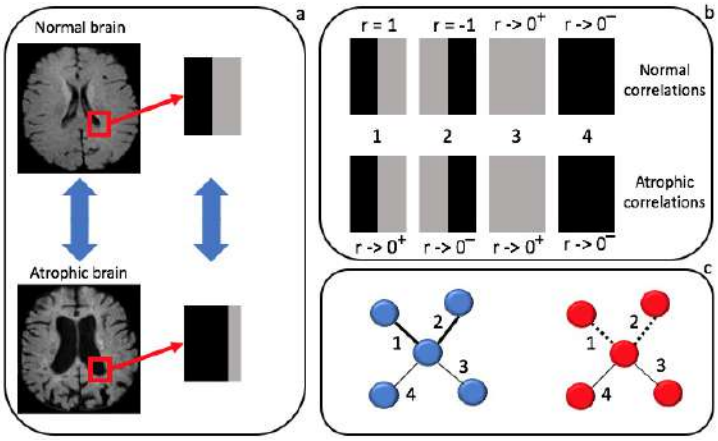

Fig. 24: (a) Brain morphological changes occur in localized regions and affect the spatial distribution of gray level intensities. For illustration, atrophy increases the cerebrospinal fluid (CSF) volume at the expenses of gray matter (GM) in panel. (b) Pearson's correlation of these two patches is computed against: (1) a patch with a symmetric distribution of GM and CSF; (2) an anti-symmetric patch mimicking left-right inversion; (3) a pure GM patch; (4) a pure CSF patch. (c) In atrophic brains (red) connections (1) and (2) disappear (dotted lines) while they remain strong connections in normal brains (blue).

ilarity metrics were explored in Section 5.2.1. This experiment demonstrated Pearson's correlation was the optimal choice.

Edges can be weighted or unweighted. Unweighted network topology is easier to study and interpret, and has computational advantages. If in several cases the decision to binarize a weighted network with a suitable threshold could be appropriate, this would seem a forced decision in our case, with the patch similarity being an intrinsically continuous measure. On the other hand, weighted networks can include weak relationships that might be spurious and introduce noise into the graph. Therefore, it was decided to threshold the networks by setting to 0 all connections whose absolute correlation was less than moderate ($|r| < .3$), in order to exclude noisy interrelationships in the model, and reducing as much as possible the loss of important links. For higher correlations, weights were kept in the model, thus resulting in a weighted undirected network representation for each subject:

$$w_{ij} = \begin{cases} 0, & \text{if } |r_{ij}| \leqslant 0.3 \\ r_{ij}, & \text{otherwise} \end{cases} \tag{52}$$

An investigation on how the threshold affects the method ability to detect diseased patterns was carried out in Section 5.2.1.

4.3.4 *Multiplex network construction*

As previously disclosed, graph theory provides tools to concisely quantify the properties of complex networks that describe interrelationships (represented by edges) between the objects of interest (represented by nodes). A general criterion defining how to best choose and use nodes and their connectivity function has not been yet found, this is in particular true for brain networks and neuroimaging data. In the present thesis work, the collection of the undirected weighted networks constructed for each subject were used and analyzed within a multiplex network framework. Multiplex network $\boldsymbol{\mathcal{G}} = \{\mathcal{G}_1, \mathcal{G}_2, ..., \mathcal{G}_\alpha, ..., \mathcal{G}_M\}$ was, for the present case, a collection of single subject weighted networks $\mathcal{G}_\alpha = (N, \mathcal{E}_\alpha, \mathcal{W}_\alpha)$ (see Figure. 25 for a pictorial representation) sharing a common number of nodes N, while the set of links $\mathcal{E}_\alpha$ changes depending on the layer α [104]. According to this notation, each network $\mathcal{G}_\alpha$ can also be represented by the corresponding adjacency matrix $A_\alpha = a_{ij}^{\alpha}$.

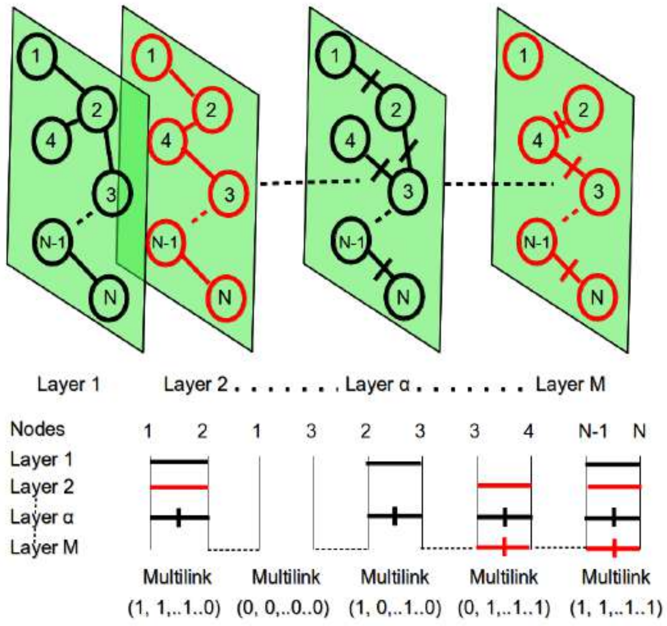

Fig. 25: At the top: the multiplex network with M layers and N nodes. At the bottom: the representation of multilinks for the different pairs of network nodes. Within each layer different nodes can be connected with a link and a specific weight. This context information is then used to detect different patterns.

To summarize the proposed model was a multiplex composed of M weighted undirected networks: each of them representing an MRI brain scan and including N nodes given by the supervoxels. For each layer interrelationships were described by $\mathcal{W}_\alpha = \{w^\alpha_{ij}\}$ in which w_{ij} were given in terms of Pearson's correlation r_{ij}.

4.3.5 *Network feature extraction*

In a multiplex, it is possible to introduce several topological characteristics that are usually adopted to describe a complex network [105]. In this approach, following indicators: the strength s^α_i and the inverse participation ratio $(y^\alpha_i)^{-1}$ of a node i in a layer α were employed:

$$s^\alpha_i = \sum_{j=1}^{N} w^\alpha_{ij} \tag{53}$$

$$y^\alpha_i = \sum_{j=1}^{N} \left(\frac{w^\alpha_{ij}}{s^\alpha_i}\right)^2 \tag{54}$$

Strength measurements denote which nodes are more relevant within the network describing a single layer (*i. e.* a subject) of the multiplex. Inverse participation ratio attains the heterogeneity of the weight distribution within each layer.

Along with these two measurements, to detect variations in brain connectivity among nodes with the same degree k, which is the number of connections existing upon a node, the conditional means of strength $s(k)^\alpha$ and inverse participation $(y(k)^\alpha)^{-1}$ for each subject and for each degree k ranging from 1 to the size N of the network, were evaluated:

$$s(k)^\alpha = \frac{1}{N_k} \sum_{i=1}^{N} s^\alpha_i \delta(k^\alpha_i, k); \tag{55}$$

$$y(k)^\alpha = \frac{1}{N_k} \sum_{i=1}^{N} y^\alpha_i \delta(k^\alpha_i, k); \tag{56}$$

where N_k is node number having degree k and δ is the Kronecker function that is 1 when a nodes i has degree k and 0 otherwise.Therefore, they help to understand how weights are distributed within each layer, thus, for example, distinguishing whether, on average, the weights of central nodes and less connected nodes are identically distributed or not. Several studies have already pointed out, especially with group-wise single layer approaches [106], how these features can describe significant differences among healthy and diseased subjects.

However, it is reasonable to assume that further evidence of significant differences between subjects, can arise from the context information provided by the multiplex framework. Accordingly, this information content was exploited by considering the aggregate adjacency matrix $A^{multi} = a_{ij}^{multi}$ where:

$$a_{ij}^{multi} = \{1 \text{ if } \exists \alpha | w_{ij}^{\alpha} > 0 \quad \wedge \quad 0 \quad otherwise\} \tag{57}$$

The matrix A^{multi} naturally allowed the introduction of the previous measurements within a global perspective. In fact, it was possible to compute for each node an aggregated degree and then use it to weight the previously defined strength and inverse participation. Analogously, A^{multi} was used to define the aggregate degree for each node and then re-computing the previous single-layer quantities. In this way in the description of each node the information produced by the whole multiplex was introduced.

In other words, to better capture inter-subject variations, it was considered the degree distribution of the whole training cohort k^{multi}, an array whose elements k_i^{multi} indicate the number of links connected to a node i over all the training subjects, thus the previous 4 quantities s, Y^{-1}, $s(k)$ and $Y^{-1}(k)$ were obtained by weighing them on the global degree k^{multi} and thus according to an overall perspective:

$$S_i^{\alpha} = \sum_{j=1}^{N} w_{ij} k_i^{multi}; \tag{58}$$

$$Y_i^{\alpha} = \sum_{j=1}^{N} \left(\frac{w_{ij}}{s_i}^{\alpha} \right)^2 k_i^{multi}; \tag{59}$$

$$S^{\alpha}(k) = \frac{1}{N_k} \sum_{i=1}^{N} S_i^{\alpha} \delta(k_i, k); \tag{60}$$

$$Y^{\alpha}(k) = \frac{1}{N_k}\sum_{i=1}^{N} S_i^{\alpha}\delta(k_i, k). \tag{61}$$

In conclusion each network was described by 8N features, 4N single layer and 4N multiplex features, resulting in a $M \times 8N$ feature representation. In particular, s, y^{-1}, S and Y^{-1} were computed for each node and $s(k)$, $y^{-1}(k)$, $S(k)$ and $Y^{-1}(k)$ for each degree. It is worthwhile to note that this characterization was independent from the clinical status of the subjects as the multiplex had been built blindly to diagnosis. This base of knowledge was then investigated with supervised machine learning models to extract specific disease effect patterns.

4.3.6 *Meaningful network extraction method*

Extracting meaningful structures and data, thus unveiling the underlying base of knowledge, is a challenging task in social physical and life sciences. Numerous studies across a different range of anatomical parts of the brain, scales and modalities have found that networks may show a behavior outcome of a combination of both regularity and randomness [107]. In fact it is known that brain networks can exhibit, at multiple levels, both small-world and scale-free properties in order to optimize brain organization and robustness respectively [108]. A network is called scale free when degree distribution follows a power law distribution and therefore presents most of the nodes with only a limited number of connections and a small number of so called hub nodes with a large number of connections. A network is considered small-world if it is highly clustered locally and has a small separation globally, these two tendencies are measured respectively by clustering coefficient and average path length.

Recently, many investigations highlighted that scale-free and small-world networks might be important to study for understanding how AD affects and modifies human brain networks. In fact it was shown that AD networks, if compared with control networks, moves mainly away from a small-world network becoming either more random or more regular. However nothing conclusive can be stated yet. Besides many studies found that hub nodes and their connections are particularly resistant to random failures and vulnerable to targeted attacks, these latter often are associated to the propagation, in a network, of a disease, such as AD [109].

Based on these motivations, a study aimed at determining a quantitative criterion to let the two main topologies of the brain networks, emerge [110, 111].

As a stating point, the detest of complex networks previously described and built from MRI, was used. To reduce complex networks to their core components is a common practice to explore a wide range of threshold values and then, according to specific application requirements, choose the most suitable one. A possible constraint to adopt for this search concerns the network topology which should be, especially for biological networks, small-world and scale-free [112]. Accordingly, it was explored how the topology of each network changed when disregarding edges below a particular Pearson's correlation value r_{thr}. To investigate scale-free topology, for each subject the agreement between the resulting degree distribution with a power-law fit in terms of adjusted R-squared statistics, was measured. Then, for quantifying and knowing how much the brain network of each subject was small-world as the threshold changes, the small-worldness indicator, given by the following expression, was studied:

$$SW = \frac{1}{M} \sum_{m=1}^{M} \left(\frac{C_m}{C_m^r} \Big/ \frac{L_m}{L_m^r} \right) \tag{62}$$

where C_m and L_m, and C_m^r and L_m^r are the clustering coefficient and the average shortest path length respectively of each network m to examine, and of the reference random graph with the same node number N and the same link probability p given by the ratio between mean degree $\bar{k}$ and N [113].

4.3.7 *Salient network construction*

High salience skeleton was extracted for each scale-free and small-world network obtained properly thresholding the original supervoxel correlation networks to focus on the relevant nodes and connections (hubs and highways) of the derived networks in order to emphasize the presence of the disease, besides reducing significantly data complexity and computational requirements. This is particularly important with the view to study through the multiplex networks, neurodegenerative pathologies characterized by smaller optimal scales and as consequence exhibiting a high computational load. The only network stringent condition required to calculate the high salience skeletons is just a scale-free topology.

The salience indicator was born to furnish an overall network description from a node-specific perspective and define a consensus among nodes on the importance of each edge within the network [114]. Given the set of weights $W = \{w_1, w_2, ..., w_{N'}\}$, with N' being the number of nodes remaining after the network scale-free transformation, the pairwise distance matrix D is defined. $D_{i,j}$ elements are simply the correlation reciprocal values.

$$D_{ij} = \frac{1}{w_{ij}} \tag{63}$$

Therefore, D captures the intuitive notion for which strongly correlated nodes should be closer than weakly correlated ones. Accordingly, in this way it is also possible to introduce a path length definition: if two supervoxels i and j are connected through a path p consisting of k steps, being them the terminal nodes of p, the length of p is simply the weight sum of the edges belonging to it. It is possible to define several path connecting i and j, in general it will not be true that the paths with minimum number of steps have also minimum length. Thus, throughout this paper the shortest path has to be considered as the path whose weight sum is maximum.

It has to be noted that according to this definition, the shortest path uniqueness is not assured. On the contrary, a typical situation is that multiple shortest paths can be found connecting two generic nodes, in that case all paths sharing the minimum length have to be taken into account. The collection of all links belonging to the shortest paths connecting a generic node n to all the remaining nodes of the network define what is generally called the shortest path tree $T(n)$ which represents, in fact, the most effective routes linking the node n to the network.

$$T_{ij} = \begin{cases} 1 & \text{if ij link } \exists \text{ in the shortest path from node } n = (1, .., N') \\ 0 & \text{if ij link } \nexists \text{ in the shortest path from node } n = (1, .., N') \end{cases} \tag{64}$$

One can calculate the shortest path trees for all the nodes of the network, then for a generic edge (i, j) connecting the nodes i and j salience $s_{i,j}$ is the indicator accounting how many times the edge belong to a shortest path tree in respect to the total number of shortest path trees. According to this definition, $s_{i,j}$ takes into account the fraction of shortest path trees including the edge between i and j.

$$S = \frac{1}{N} \sum_{n=1}^{N'} T(n) \tag{65}$$

Fig. 26 shows in detail the procedure adopted to get high salient skeletons from raw data. As $T(n)$ reflects the set of most efficient paths to the rest of the network from the perspective of the reference node n, $s_{i,j}$ is a consensus variable defined by the ensemble of root nodes. If $s_{i,j} = 1$, then link (i,j) is essential for all reference nodes, if $s_{i,j} = 0$, (i,j) is not a fundamental link for the network.

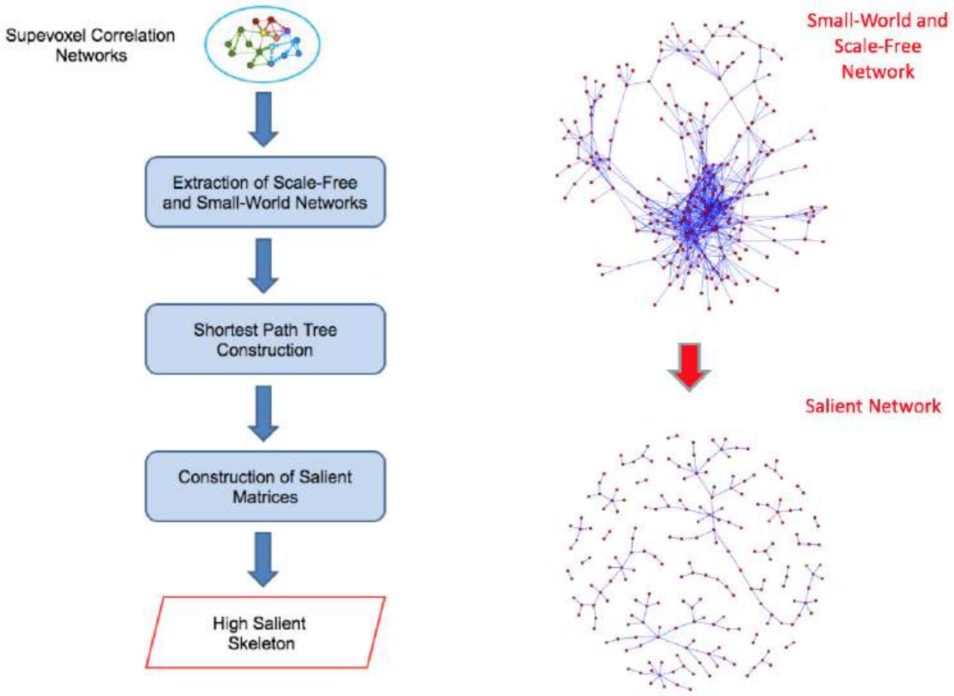

Fig. 26: This Figure shows in detail the procedure adopted to get to high salient skeletons from raw data. a) Scale free networks with power-law distributed weights were extracted from each initial subject network; b) Links participating at least once in the shortest paths, starting from a fixed reference node, were recorded in the shortest path tree matrix; c) Shortest path tree for each reference node were added up to obtain, for each subject, salience matrix, whose values $s_{i,j} = 1$ represent high salient skeleton

With respect of brain networks it is manifest that a salience picture can be of great importance. As measuring the efficiency of a path in a transport networks gives fundamental insight on the backbone structure of the network itself, analogously we expect that in brain structural networks this approach could emphasize which brain regions are mostly affected by atrophy caused by AD. One of the most important properties salience indicator has is its bimodal distribution with peaks on 0 and 1 values, as shown in Fig. 27. This makes natural finding what is called the salience skeleton, *i. e.* the backbone structure including all the most efficient links of the networks.

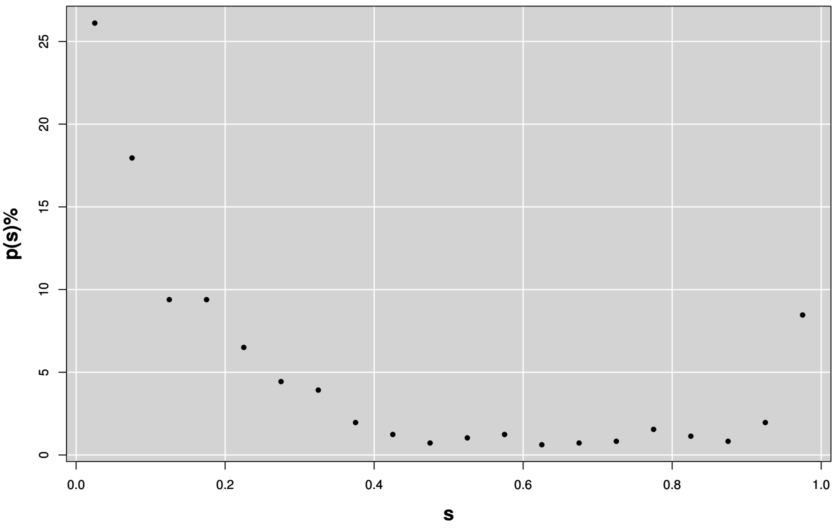

Fig. 27: The figure shows salience percentage frequency of a brain network as example. Saliency values are gathered on 0 and 1 thus it is possible to detect the salient link independently from a particular threshold value. In this network the link fraction contributing to the high salient skeleton is of the 8.46%.

Fig. 27 is an example of salience percentage frequency for a brain network. The salience skeleton is by definition the network containing only salient links. As the saliency distribution is bimodal there is no need of finding a particular threshold, but the skeleton stays pretty much unchanged for all threshold values between 0.1 and 0.9. For the present work the threshold 0.5 was used.

4.3.8 *Learning and validation algorithm*

In this section, the machine learning techniques used in this work will be reviewed. Then, cross-validation and classification processes to evaluate informative content provided by the complex networks and discriminate controls and patients both for Alzheimer and for Parkinson, will be explained. Besides it will be illustrated a learning framework that combine network and clinical to increase classification accuracy in order to obtain an useful diagnosis support tool for PD disease. Finally, it will be described how the evaluation and identification of the anatomical regions connected to the two disease was performed.

Machine learning techniques

The aim of supervised learning algorithms is not to exactly learn the representation of the training data, but rather to develop a statistical model able to describe heterogeneous data variability of which can preclude the possibility to obtain a deterministic description. In the context of machine learning is important to minimize two sources of error that prevent the generalization beyond the training set: the bias and the variance. These two terms along with the irreducible error due to the noise give the so-called *generalization error*. The bias is an error caused from wrong assumptions in the learning algorithm. High bias can induce an algorithm to miss the important relations between features and target outputs (underfitting). The variance is error due to sensitivity to small fluctuations in the training set. High variance can induce an algorithm to model the random noise in the training data, rather than the intended outputs (overfitting). The optimal balance between bias and variance can be reached through the tuning of the the model parameters values. For example removing superfluous parameters (pruning) or involving the addition of a penalty term to the error function through which the model is learn (regularization). When the learning is based on a number of features far grater than sample number, overfitting is due to the so-called overtraining. Learning models where the target variable can take a discrete set of values are called classification algorithms, decision trees where the target variable can take continuous values (typically real numbers) are called regression trees algorithms [115].

In this thesis work, to select the important network feature and evaluate their informative power, in the discrimination of patients from controls, Random Forest (RF) classifier was used. This choice was dictated by different reasons:

- excellence in accuracy.
- Effectiveness over large dataset. item capacity of dealing with a great number of features without the necessity to reduce them.
- estimation of the important features.
- Robustness even in case of missing data.
- Saving of the forests to use them again on new data.

Decision trees are based on the following idea: perform predictions using a sequence of simple decisions. In fact, it model consists of an ensemble of (binary) decisions organized in a hierarchical fashion. Hence, a decision tree is composed of a set of nodes and a set of directed edges. Each node encodes a decision, the node at the top of a tree is called root and its decision is based on the value of the most discriminative feature, nodes at the bottom are called leaves and represent class labels. Tree branches represent conjunctions of features that lead to a certain class label from the root node. An observation can traverse the tree downward, following a unique path (brunch) that is determined by decisions taken at each traversed node, until it reaches a leaf as illustrated on Fig. 28.

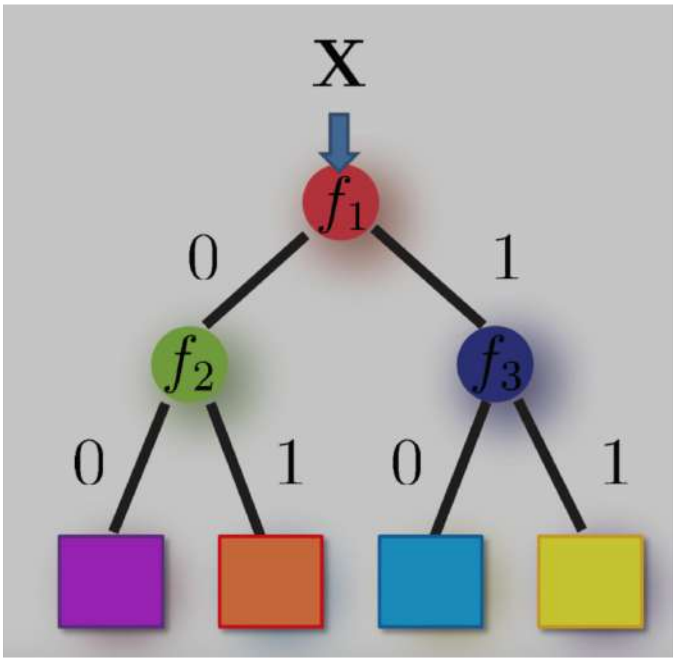

Fig. 28: a decision tree is a directed graph, where each node has a decision function and edges join the root node to the leaves that represent the target classes.

At each level, a node takes a decision basing on the most discriminating variables remained. the number of decisional levels of a tree is called tree depth. During the learning phase, the data that reached a given leaf is used to model the posterior distribution. During the test phase, these posterior distributions permit to make predictions about new unseen observations reaching a given leaf. In Random Forest as explained by Brieman [116], given the set $S = \{S_1, ..S_N\}$ of the available data, k bootstrap, with replacement, are formed obtaining k new sets used to grow k trees. This strategy to create k independent trees is called *bootstrap aggregating* or *bagging*. Then, each tree is grown by chaising at random a subset of features. The optimal number f of features randomly selected, that remains constant for all the process of growing, is $f = \sqrt(n)$ with n total feature number. This value f is chosen in order to minimize the out-of-bag error that is the prediction error obtained using the built learning model to validate the data left out of the training set. Finally, overall out of bag error as well as classification accuracy is obtained averaging over all subject sets that for each forest tree were left-out of the training. The two most important degrees of freedom are: (i) the number of trees and (ii), the tree depth. In Fig. 29 out-of-bag error is represented as a function respectively of the tree depth and the number of trees. Increasing the number of trees averages out noisy predictions, and thus leads to a monotonic decrease of the prediction error. The maximal allowed depth of the tree is a crucial parameter that needs to be optimized as it directly affects the generalization ability of each tree. Indeed, while a short tree will not be very confident in its prediction (underfitting), a very deep tree may not be able to generalize suffering of overfitting. Therefore, the tree start to explain too well the training data, e.g. by fitting noisy features. For this reason, the prediction error curve decreases with the tree depth until it reaches a minimum and then increases again. This minimum corresponds to the optimal tree depth, providing a good modeling of the observations and a great generalization.

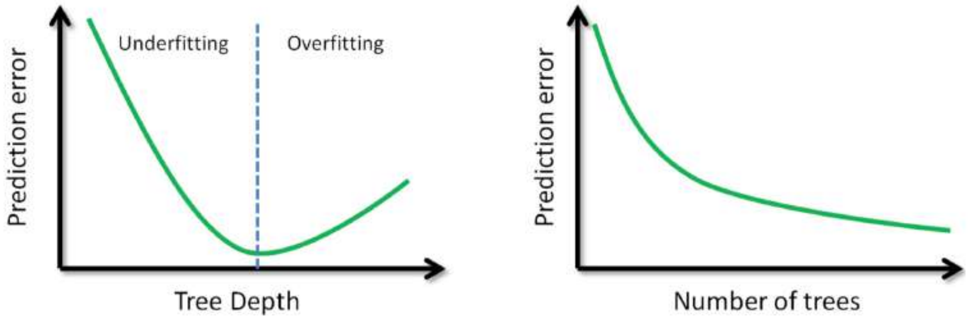

Fig. 29: Tree depth and number of trees are the two most important parame-ters. While increasing the number of trees correspond to a decreasing in the prediction error, tree depth needs to be carefully tuned as it controls the generalization ability of the forests.

However in this work, Random Forest results were compared also with other supervised learning techniques such as Naive Bayes classifiers, Support vector machine and Neural Networks.

Naive Bayes classifiers assign the most probable class to a given example by using the Bayes theorem. The name originates from the simplifying assumption that the feature values $f_1, f_2, .., f_n$ are considered to be conditionally independent given the class c_k. Therefore, according to Bayes theorem:

$$p(C_k \mid f_1, \ldots, f_n) = \frac{1}{Z} p(C_k) \prod_{i=1}^{n} p(f_i \mid C_k) \tag{66}$$

where the evidence $Z = p(\mathbf{x})$ is a scaling factor dependent only on $f_1, f_2, .., f_n$ in other words a constant if the values of the feature variables are known. Given the previous assumption, the classifier assigns a class label cto y as follows:

$$y = \underset{k \in \{1,\ldots,K\}}{\operatorname{argmax}}\ p(C_k) \prod_{i=1}^{n} p(f_i \mid C_k). \tag{67}$$

where K is the number of classes, f_i is the i-th feature, c_k is the k-class, $P(c_k)$ is the prior probability of c_k and $P(f_i \mid c_k)$ is the conditional probability of the i-th feature given class c_k. Naive Bayes involves a learning step in which the various $P(c_k)$ and $P(f_i \mid c_k)$ terms are estimated on the basis of their frequencies over the training data or in the case of features with continuos values $P(f_i \mid c_k)$ is obtained using the frequencies of a certain condition. Although its simple assumption is rarely true, Naive Bayes is remarkably successful in practice [117].

Support Vector Machines (SVM), given a representation of a data point as a N-dimensional vector, aims at finding a separating hyperplane in a high $(N-1)$-dimensional space so that the distance between the hyperplane and the closest training data points of each class is maximized. This hyperplane is called the maximum margin hyperplane. Intuitively, an SVM model is a representation of the examples as points in a space, mapped so that the examples of two different classes are clearly divided by a gap (the maximum margin hyperplane) that provides the greatest separation between the classes. New examples are then predicted to belong to a class based on which side of the gap they fall. Various types of dividing classification surfaces can be realized by applying a kernel function, such as linear, polynomial, Gaussian Radial Basis Function (RBF), or hyperbolic tangent. SVM are well known for their generalization ability and are particularly useful when the number of features, is much more higher then the number of data points [118].

Artificial Neural networks (ANN) are organized into layers: input, hidden and output. Input layer, consists of a set of neurons representing the input features. Output layer responds basing on the information learned. In between the input and output layers are one or more hidden layers, which form the majority of the artificial brain. Information flows through a neural network in two ways. When it's learning or operating normally (after being trained), patterns of information are fed into the network via the input layers, which trigger the layers of hidden layers, and these in turn arrive at the output layers. This common design is called a *feedforward network*. Each layer receives inputs from the layer to its left, and the inputs are multiplied by the weights of the connections that they travel along. Every layer adds up all the inputs that it receives in this way and if the sum $f_1w_1 + f_2w_2 + .. + f_nw_n$ is more than a certain threshold value, the layer triggers the next layer on its right until reaching the output layer.

$$o = \Phi(\sum_{j}^{n} w_j f_j + b) \tag{68}$$

where o is the output signal, Φ is the activation function, n is the number of connections to the perceptron, w_i is the weight associated with the i-th connection and f_i is the value of the i-th input connection, b represents the threshold.

A key characteristic in the learning of neural networks is *backpropagation*, this stage compares the output a network produces with the output it was meant to produce, and uses the difference between them to modify the weights of the connections between the layers in the network, working from the output layers through the hidden ones to the input layers-going backward. in other words, with the backpropagation the network learns, reducing the difference between actual and intended output through the minimization of the so-called *error function*.

$$E = \frac{1}{2} \sum_{i}^{p} \|\mathbf{o}_i - \mathbf{t}_i\|^2 \tag{69}$$

where $\mathbf{o}_i$ is the output i-simo in general different from the target i-esimo $\mathbf{t}_i$, p is the number of ordered pairs of vectors $\{(\mathbf{f}_1, \mathbf{t}_1), .., (\mathbf{f}_p, \mathbf{f}_p)\}$ respectively of n-features and m target, which are called the input and output patterns [119].

As described previously, the multiplex characterization of structural MRI scans yielded a matrix representation which could be used to feed machine learning models and thus unveil discriminating patterns for the examined disease.
The number of features f, involved in the proposed approach, could easily reach values ranging from 10^3 to 10^4 outnumbering the number of the available training samples. Therefore, to prevent overtraining issues yielded by the curse of dimensionality and assess the multiplex framework a feature selection was necessary. A flowchart of the whole feature selection method is represented in Fig. 30.

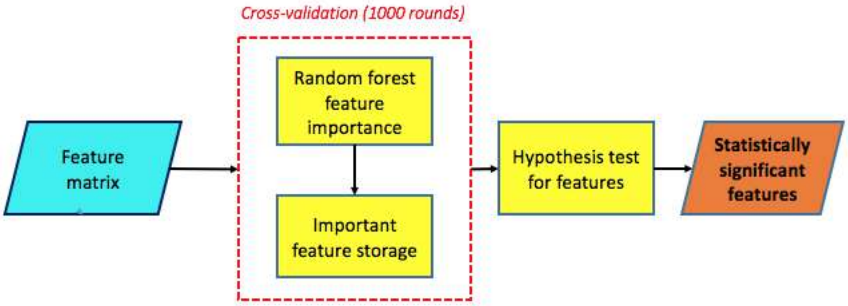

Fig. 30: A flowchart of the feature selection methodology: the features, stored in a matrix, are used to train a random forest model, this model provides a feature important estimation; the procedure is cross-validated with a 5-fold for 1000 times, at each round taking into account the selected feature. Finally, a statistical test of hypothesis establishes which features have been selected a significant number of times.

A 5-fold cross-validation feature importance selection was performed within a wrapper-based strategy. $\mathcal{A}_{\text{train}}$ was randomly divided in a training and a validation test. For each cross-validation round, a multiplex model on training subjects was built, then the related training multiplex features and the overall matrix $\mathcal{A}_{\text{train}}$ were computed . For test subjects, single layer features were straightforwardly computed. Features accounting the whole multiplex structure were in turn computed considering the training multiplex network degree k^{multi}. The reason for this choice can be justified considering the perturbation induced by the addition of one layer is small.
For each cross-validation round, a first Random Forest classifier was trained and the most important features were selected. Features were chosen in order to exceed in importance the third quartile of the normalized importance distributions computed in terms of mean accuracy decrease and mean decrease

in node impurities over all trees. In particular, impurity is obtained through Gini index that measures heterogeneity of the node classes and is given by $\sum_{i=1}^{K} p_i(1-p_i)$ with K class number p_i probability of an item with label i and $1-p_i$ probability of a mistake in categorizing that item. A pure node has Gini index 0 and consists of a single class, instead it reaches the maximum value of (K-1)/K when all the classes are different [120]. An example of the normalized importance distribution and the chosen threshold for a cross-validation round is reported in Fig. 31.

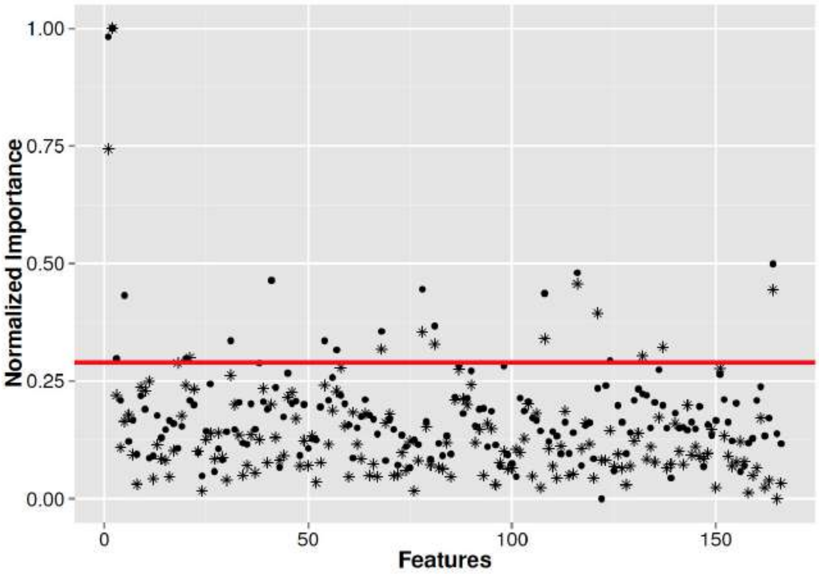

Fig. 31: The figure shows the features selected according to mean decrease in accuracy (circles) and Gini index (asterisks). The red line represents the third quartile of the overall importance distribution.

The selected features were stored for later use and used to train a second Random Forest classifier which was used to predict the diagnosis of the validation subjects. In both cases random forests were grown with 500 trees, a number large enough for the out-of-bag error to reach the typical training plateau.
At each split $\sqrt{f} = \sqrt{4392}$ were randomly sampled, thus, for each cross validation round, different features were selected; to determine the most important features, the overall occurrence rate of each feature was measured and it was interpreted as a success rate. As a consequence, the probability to observe such occurrence with a binomial distribution was compared and an experimental p-value was computed to test the randomness hypothesis. To select a more exiguous number of features a p-value < 0.01 threshold was set, then which ones had shown a significant probability of occurrence were established . Once

the best features had been selected, they were used them to train a new ensemble model on $\mathcal{D}_{\mathrm{train}}$ and tested it on $\mathcal{D}_{\mathrm{test}}$ to assess the method robustness and evaluate the informative content carried by multiplex features. Specifically two binary classification were performed: controls versus patients and controls versus converter mild cognitive impairment subjects. An analogous learning procedure was carried out for the feature retrieved by salient skeleton multiplex networks in order to compare the informative content and the capacity of discrimination of the salient network features, the original network features and their combination.

It is worth noting that features like strength and inverse participation have a direct interpretation, being directly related to a single patch of the brain network whilst conditional means, by definition, are related to several nodes sharing a common degree k. For classification purposes this is not an issue, being based on computed features; on the contrary this is relevant in order to provide an anatomical interpretation and a diagnostic value of the features selected.

PD Assessment and validation

As to PD Parkinson a the beginning, 177 controls and 177 PD subjects, randomly selected, were used. After the removal of null mean and variance features and highly correlated features, 4048 features remained. Before using network feature matrix (subject × features) for the classification process, a feature selection was necessary to reduce high dimensionality of the model, increase its generalization and give a deep insight into underlying mechanisms. All three categories (filtering, wrapper and embedded methods) in which the existing feature selection methods can be divided, were explored to understand which was the most appropriate for this type of data. Filter methods are fast, scalable, independent from learning but they disregard possible interactions among the features; both wrapper and embedded methods take relationships among features into account but they are computationally intensive. Wrapper methods use search algorithms to evaluate all possible subsets of features and select the best set in terms of classification performances obtained from a model independent from the training one; instead in embedded methods, selection of the important features and learning are performed within the same model[121].

As filter method, ReliefF algorithm was used. It iteratively evaluates feature importance choosing a random instance i in the dataset and estimating how well each feature value of this instance distinguishes between instances close to i. Step by step, the weight assigned to each feature is updated on the basis that a change in feature value accompanied by a change in class leads to a weight increase and a change in feature value accompanied by no change in class leads to a weight decrease [122]. As to wrapper method, Sequence Forward Search (SFS) was the strategy used to explore the various features sub-sets, adding, step by step, to an initial empty set a new feature which in combination with the others gives the best performance evaluated by a first Random Forest classifier [123]. After obtaining the best feature sub-set, a second independent Random Forest classifier was used for training a classification model in order to classify blindly the validation set. For embedded method, a unique Random Forest classifier was used both for selecting the important features and for training the learning model.

Feature selection and classification step were implemented within the same cycle of 100 cross-validation 10-fold. It is worth noting that for each cross validation round, the dataset was split into a training and validation set, thus, feature selection and model learning were carried out on the training set blind to the validation set. While in AD study, the training dataset and the independent dataset were fixed, it is worth to notice that, in PD study, for each round of cross-validation training and validation set change as well as the important features. Moreover every time, multiplex network features for the validation set were computed using the aggregate degree computed over the whole training cohort.

Feature selection and classification was carried out for different scales and for the combination of three scales (125, 1000, 3000 $voxels$) chosen in order to be comparable with the dimensions of the anatomical regions mostly affected during the first stages of the disease. In this way it was possible to evaluate if information obtained from more scales could give an add value to the discrimination of disease and healthy subjects. For a given feature selection method, the whole procedure of machine learning is represented in Fig. 32.

A standard configuration was used for Random Forest classifiers because this experiment was aimed at evaluating the informative power of several feature sets and their combination, more than refining the method performances for an accurate PD diagnosis. Given f features, at each split the $\sqrt{f}$ features are randomly chosen. Forests were grown with 500 trees, a number large enough to reach the typical training plateau for the out-of-bag error.

For each cross-validation round, again, features of which importance exceeded the third quartile of the whole normalized importance distribution were recorded and the occurrence rate was computed. Then, the features of which occurrence significantly ($p < 0.01$) confuted the hypothesis of randomness were selected and the related cerebral regions were outlined, investigating

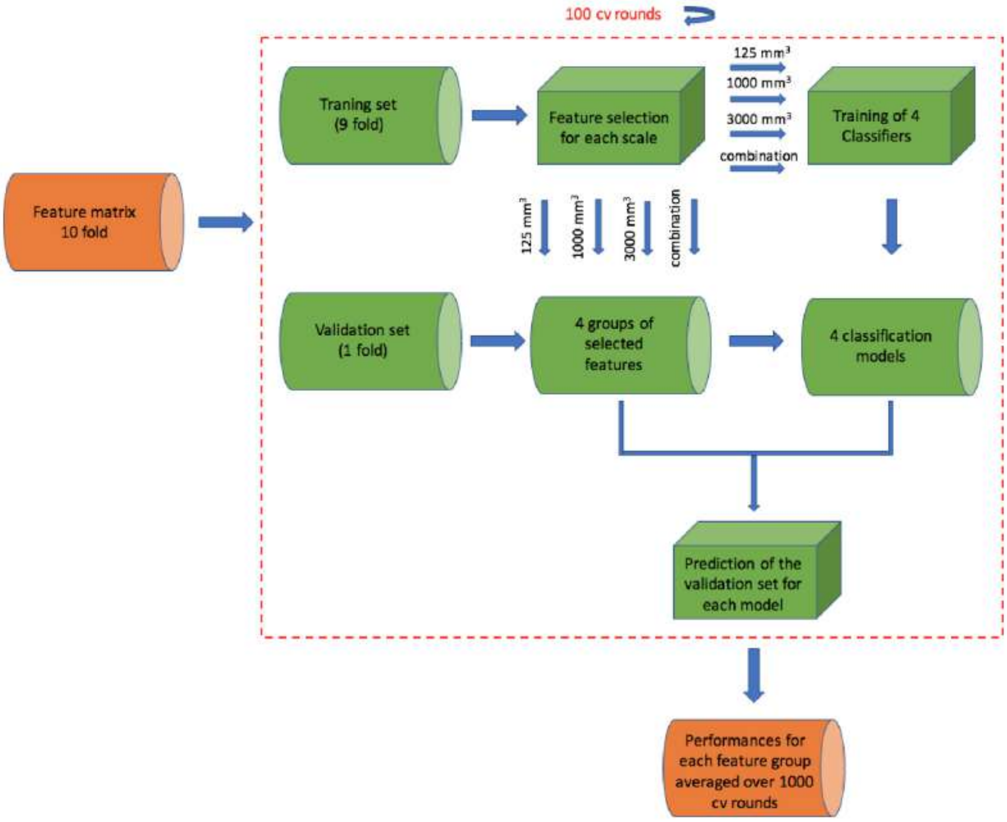

Fig. 32: The different steps within the cross-validation rounds are represented. Starting from the feature matrix, for each round, data are divided into training and validation set. Then for each scale, feature selection, classifier training and model construction are performed. Finally, models are used for classifying patients and controls.

which anatomical districts were identified for each supervoxel. In this way, important anatomical regions were found for each scale.

Then, to evaluate the informative power of the the clinical features along with the network features that are more numerous, all the PPMI available subject were used. This time, for the feature selection and the classification process, a hybrid approach was carried out in that Random Forest was used for feature selection but not for model.

Firstly, a hierarchical combination of two forests was adopted. For each 10-fold cross-validation round, the first forest, for each split, randomly picked $\sqrt{f} = \sqrt{4048}$ features and selected a number of features, which were on average 60. Features found by the first classifier were chosen always in order to exceed in importance the third quartile of importance distribution. Then, a second forest within the same cross-validation cycle, was trained with these selected features to obtain a diagnostic score. The results of this procedure were twofold: on one hand the most discriminative features were selected and on the other hand the whole information content provided by the complex network measures was summarized in a unique score. Thus, this score outlined which subjects had shown brain structural and topological changes significantly associated to the diagnosis. The presence of brain regions with modified centrality, the disrupt of connectivity, along with the previously mentioned properties concerning for example the intensity or the uniformity of connections, were all included in the classification score.

This approach was designed as the number of features deriving from the complex network description can be overwhelming if compared to the available clinical features of Table 4. It is important to remind that for each subject clinical features are provided at the baseline. Accordingly, a diagnostic model relying on clinical features typical of prodromal PD or early PD and structural MRI data is designed. Finally, for the discrimination of PD patients and normal controls, a third radial support vector machine (SVM) classifier was trained combining the diagnostic score and the clinical features and keeping fixed the training and validation set of the previous 10-fold cross-validation. For our analysis, the *R* package *e1071 version 1.6-7* with the default implementation (cost $= 1$ and gamma $= 0.003$) was used. A schematic overview of the method is shown in Fig. 33

In addition, in order to verify the robustness of the method, the classification was carried out with different learning techniques and once the best in performance terms was established, the optimal value of its parameters was found. Another interesting aspect it was the comparison of the network approach with other two standard method to demonstrate the added value provided by networks. FreeSurfer, a ROI-based method and the Voxel-Based Morphometry

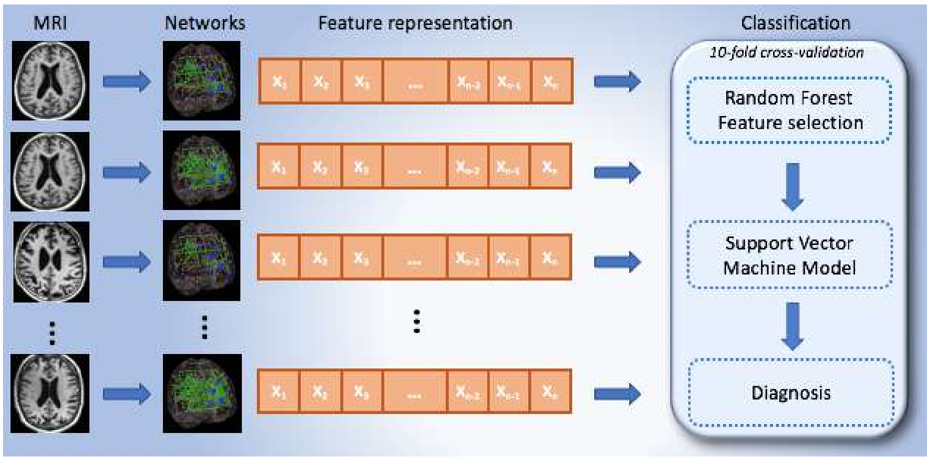

Fig. 33: Methodology flowchart. MRI scans are processed to obtain a network representation. Then, for each node several features are computed and a feature representation is obtained. These features evaluate the node importance within the networks. The most informative features are selected with Random Forest wrapper and summarized in a classiffication score, then a Support Vector Machine combines these score with clinical features to distinguish NC and PD groups. The entire classiffication process is performed in 10-fold cross-validation.

description were used. All these analyses will be described in the following experimental results chapter.

Anatomical Interpretation

It is well known from literature that several specic brain regions are interested by Alzheimer's disease and as many specific regions by Parkinson's disease. Thus, another crucial aspect of the present thesis work consisted in the evaluation of features significantly related to diseases and their visual inspection in order to establish whether these features represented anatomical regions of interest for the disease. Nodes straightforwardly related to the significant features were easily pinpointed on the reference template and the corresponding atlas (Harvard-Oxford cortical and subcortical structural atlases [124] for AD case and Talairach atlas [125] for PD). As to the significant conditional features, which intrinsically encode the information contained in different nodes with the same degree, it was recorded, for each cross-validation, whether or not a particular anatomic district had been selected subject by subject and it was tested the hypothesis that these occurrences had happened by chance. The null hypothesis is that a region should be selected according to a Bernoulli distribution with an a priori probability which is the reciprocal of the total number of patches. in other words, it was tested the occurrence of a specific region against the random occurrence rate equal to $1/patchnumber$ (1/549 for AD case and 1/12219 for PD case) which is in conclusion the number of the carried tests.

Besides both for AD and PD, this anatomic study of the areas connected to diseases was compared in terms of region number with a conventional method: Voxel Based Morphometry (VBM). for VBM with the publicly available SPM 12 suite [126], the standard prescription was followed. Firstly, a segmentation of brain tissues was performed, followed by non-linear normalization with the SPM tool DARTEL to create a study specific template. Secondly, a smoothing with an isotropic Gaussian filter with a full width at half maximum of 8 mm was performed. Lastly, a two-sample analysis was performed with a t statistics to investigate significant group-wise differences in atrophy between NC and AD on training subjects. VBM results are reported in the paragraphs 'Multiplex Networks vs Voxel Based Morphometry' of the sections 5.2.5 and 5.5.4.

4.4 EMPLOYMENT OF DISTRIBUTED INFRASTRUCTURES

The different elaborations required in this doctorate work were performed using ReCaS (Rete di Calcolo per SuperB e altre applicazioni) Bari Data Center , shown in Fig 34.

ReCaS [127] has been built by the University of Bari 'Aldo Moro' and the National Institute of Nuclear Physics (INFN) in the framework of the ReCaS project (PON Research and Competitiveness $2007-2013$). It was completed in July 2015 and inaugurated on July 9, 2015; PON aim was the upgrade of four Data Center in Southern Italy, namely Catania, Cosenza, Naples and Bari.
This Data Center is one of the most important Italian super-computer financed by public funds. It is composed of 180 servers for a total of about 12000 CPU-cpre. Indeed,each server hosts 64 core and and 256 GByte of RAM memory that is 4 GByte for core. In addition, the Recas Data Center houses a cluster dedicated to High Performance Computing (HPC) for running applications that use a large number of cores in parallel with an intense exchange of messages between cores. The HPC cluster consists of 20 servers, interconnected by means of a low latency Infiniband network, each with 40 cores, for a total of 800 cores; each server is equipped with a graphics accelerator NVIDIA K40. In Fig. 35, an image of some machine of the Data Center is reported.

All the resources are connected to each other through GARR (Gruppo per l'Armonizzazione delle Reti della Ricerca) network, Italian national computer

Fig. 34: Frontal image of the building built to host the ReCaS-Bari Data Center at the Bari Physics Department and the Bari local section of INFN.

Fig. 35: The image portrays a series of racks containing the machines that make up the Data Center.

network for universities and research. Point-point connections on optical fiber deliver information at a line rate of 10 Gbps.
In addition, the computing center also supports multiple communities that deal with life science and related projects NextMR [128], Elixir project [129], LifeWatch [130]. The centre offers:

- the possibility to require specific services on virtual machines;
- to execute parallel jobs based on MPI library;
- the possibility to manage the execution of a bunches of independent jobs also by means of WebServices;
- Cloud Storage services based on WebDav and ownCloud

As this scientific research involved the processing of large amounts of data, the use of a data center has become essential. For performing this study, ReCaS was used through the distributed computing paradigm. Indeed, it was possible to execute hundreds of independent elaborations simultaneously. Each autonomous process is sent through a server to some computational resources that the system has allocated to run the process itself. These resources are called worker nodes. A scheme of the process execution is depicted in Fig. 36

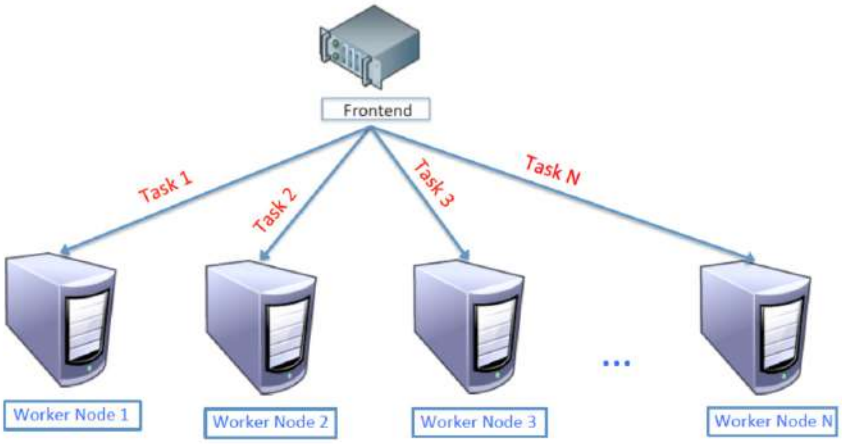

Fig. 36: Scheme of the process execution on the computing infrastructure. Processes are required by a server (Frontend), delivered and executed on the various nodes (Woker Node).

In particular for this work, each of the thousands of processes required 4/6 CPUs and 10/12 GBytes of RAM for a total of over 1000 CPUs and about 3000 GB of memory, with a machine time of about a week for each. It is clear that the use of ReCaS and the computing on a distributed infrastructure allowed the optimization of the necessary time for the analyses. To have an idea of the saving time, it is possible to consider the following estimation: 1000 processes, as those carried out in this research, on a common workstation (middle-high quality: 4-core CPUs and 12 GByts of RAM) would be performed in series with a significant increase in machine time, estimated for the entire analysis of about 1 week × 1000 processes = 19 years. Therefore, the use of distributed resources reduces considerably the running time of the entire pipeline.

5

EXPERIMENTAL RESULTS

5.1 FIVE FUNDAMENTAL QUESTIONS

In this chapter will be shown the experimental results obtained through the methodologies developed during the three years of doctorate and described in detail in chapter 4. The results regards the investigation of two neurodegenerative diseases (Parkinson and Alzheimer) through implementation and the analysis of a novel physical model of structural brain connectivity. The findings were achieved through a combined study of the complex network and several machine learning techniques and answer to the following fundamental questions:

- Is it possible to develop a physical brain model studying brain connectivity through complex networks?
- Is it possible to find quantitative biomarkers, associated to complex networks, able to reveal structural changes related to neurodegenerative disease such as Alzheimer and Parkinson?
- Do structural variations involve anatomical regions associated with the disease in question and can indicators, that describe them, be used by artificial intelligence for the early diagnosis of Alzheimer's disease?
- Is it possible to reduce a very dense and large network in order to preserve the most relevant information of the original network?
- To what extent can the combined use of multiple networks and artificial intelligence become a valid support for early diagnosis of Parkinson's disease?

In this section a detailed overview of the experimental results concerning the the assessment of the informative power of the multiplex network in the context of the early diagnosis and characterization of the AD. Firstly it will be illustrated the study to find the optimal threshold and similarity metric to build brain connectivity networks. Then, it will highlighted there is a privileged scale that is a best network node size to study AD. Moreover classification performances of the method in the discrimination respectively of controls versus patients and controls versus converter mild cognitive impairment subjects, are shown along with some experiments aimed at demonstrating the robustness and consistency of the methodology developed. Finally, there will be a precise review of the all anatomical regions that are resulted significantly connected to the disease.

5.2.1 *Optimal threshold and similarity metric*

Since this approach could in principle heavily depend on the threshold value adopted to discard negligible correlations, the threshold values ranging from 0 to 0.8 were explored with a 0.1 step. Then, for each threshold value a different multiplex was constructed. The patch dimension adopted was 3000 mm^3. The training classification performance was measured in terms of accuracy, see Fig. 37

The classification accuracy reached its maximum value with a 0.3 threshold value and it remained stable over 0.85 for a large range of correlations $[0.2, 0.5]$. Outside this range a performance drop was observed. With lower or higher threshold performances showed a significant decrease, especially above the 0.8 threshold; in which case more of the 50% of the networks resulted empty. This is because lower threshold values introduced noisy correlations within the model, thus concealing the effective network information, whilst greater threshold values were too penalizing as informative links were neglected.

It was investigated different similarity measurements to find which was the most suitable for defining edges between the supervoxels s_i and s_j of the examined multiplex, with $i, j = 1, .., N$ and voxel position within a supervoxel $k, l = 1, .., V$. In addition to the Pearson's correlation, it was studied:

- *Mutual Information* (MI) between the supervoxel pairs that is the sum of the Shannon entropies $H(S_i)$ and $H(S_j)$ of the single supervoxels random variables Si, Sj minus their joint Shannon entropy $H(Si, Sj)$. Mutual

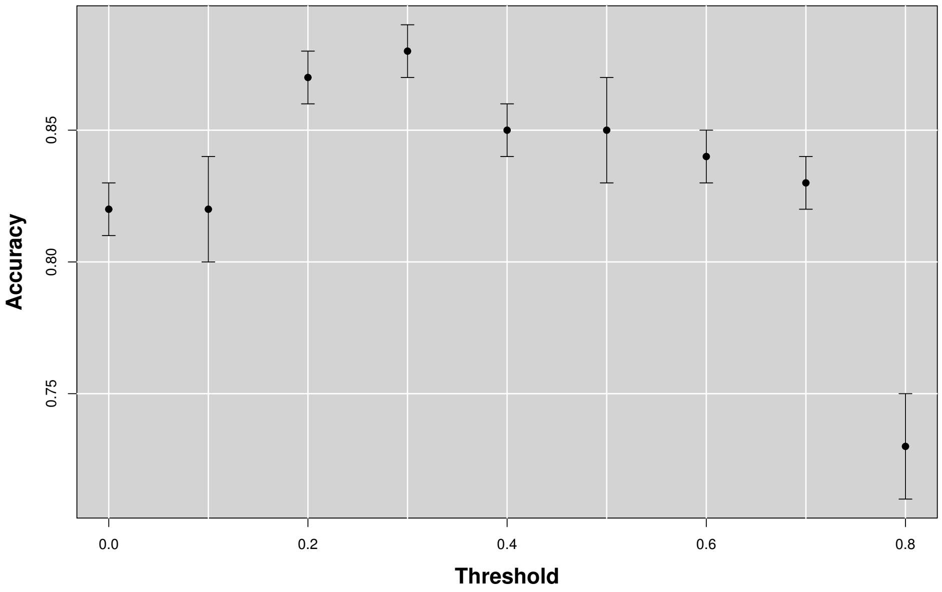

Fig. 37: The figure shows the accuracy as a function of the threshold that changes from 0 to 0.8. The best accuracy is obtained in correspondence of a threshold value of 0.3.

Information between two supervoxels can be defined alternatively as the Kullback-Leibler divergence between the pairwise supervoxel joint probability density function $p(s_i^k, s_j^l)$ and the product of their single probability density functions $p(s_i^k)$ and $p(s_j^l)$:

$$MI_{ij} = H(S_i) + H(S_j) - H(Si, Sj) = \sum_{k=1}^{V} \sum_{l=1}^{V} p(s_i^k, s_j^l) \log \frac{p(s_i^k, s_j^l)}{p(s_i^k) p(s_j^l)} \tag{70}$$

In other words this metric measures the amount of information that a random variable contains about another, minimizing joint entropy.

- *Mean Square Differences* (MSD) that is the sum of squared differences between voxel intensity pairs divided by the total number of voxels within a supervoxel:

$$MSD_{ij} = \frac{1}{V} \sum_{k=1}^{V} (s_i^k - s_j^k)^2 \tag{71}$$

- *Hellinger distance* (HD) that is the root of the sum of squared differences between the root of voxel intensity pairs divided by the total number of voxels within a supervoxel:

$$HD_{ij} = \frac{1}{V} \sqrt{\sum_{k=1}^{V} (\sqrt{s_i^k} - \sqrt{s_j^k})^2} \tag{72}$$

- *Kolmogorov Smirnov non parametric statistic test* (KST) quantifying the shape difference between gray level distributions of the supervoxel pairs.

- *Unpaired t statistic test* (UtT) evaluating the difference between means of the supervoxel gray level distribution pairs in terms of standard error.

For MI 0.3 threshold used for Pearson's Correlation was adopted. For MSD and HD, adjacency matrix was obtained computing the complementary of the normalized MSD and HD matrix and placing to zero the values < 0.3. Finally for KS and unpaired t test, adjacency matrix was given by the complementary of the normalized test statistic matrix, setting to zero the elements rejecting respectively the null hypothesis of supervoxel distribution and mean equality with a p-value< 0.05. Classification accuracies and relative standard errors, for each similarity measurement, are shown in Figure. 38.

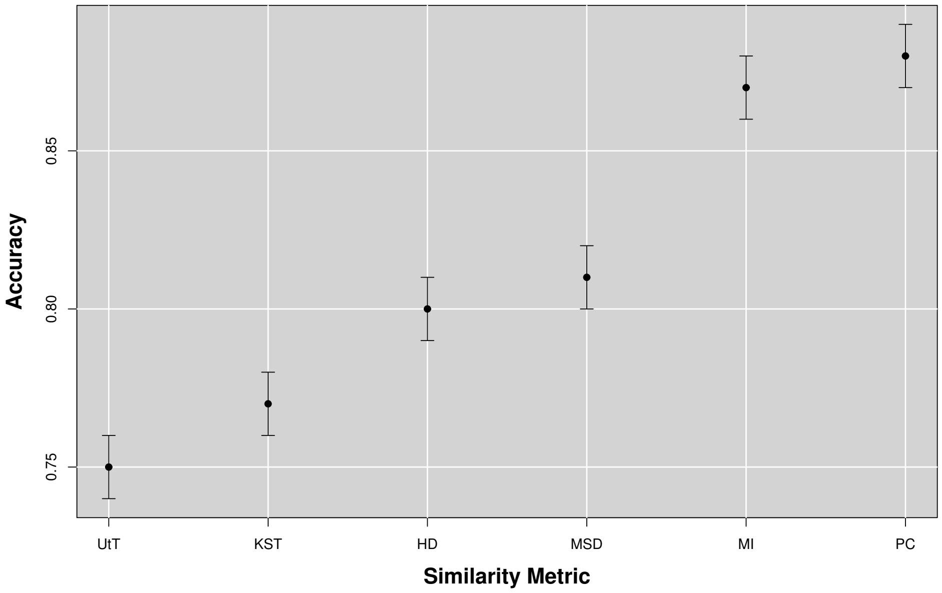

Fig. 38: The figure represents the accuracies with relative standard error for each similarity measurements: Pearson's Correlation (PC), Mutual Information (MI), Mean Square Difference (MSD), Hellinger Distance (HD), Kolmogorov Smirnov non parametric statistic Test (KST), Unpaired t statistic Test (UtT). The best accuracy was obtained using Mutual Information and Pearson's Correlation.

From the figure, it is possible to notice as PC, MI, MSD and HD, being metrics that compare corresponding voxels within a supervoxel, achieved higher performances. On the contrary, KST and UtT considering respectively only shape and mean of the supervoxel gray level distribution gave lower performances. This confirms the importance of keeping spatial information when multiplex network edges are constructed. Besides, the decrease in accuracy, for MSD and HD, underlines that unlike PC and MI they lack an intrinsic normalization and thus they are less robust and suffer a significant performance drop.

A complete summary of the metric study performances including sensitivity and specificity is reported in Table 10 from which Pearson's Correlation turned out to be the best similarity measurements.

Similarity Metric	**Accuracy**	**Sensitivity**	**Specificity**
Pearson's Correlation	**0.88 ± 0.01**	**0.90 ± 0.01**	**0.88 ± 0.02**
Mutual Information	0.87 ± 0.01	**0.90 ± 0.01**	0.87 ± 0.02
Mean Square Differences	0.81 ± 0.01	0.84 ± 0.01	0.80 ± 0.02
Hellinger Distances	0.80 ± 0.01	0.83 ± 0.01	0.77 ± 0.02
Kolmogorov Smirnov test	0.77 ± 0.01	0.82 ± 0.01	0.77 ± 0.02
Unpaired t test	0.75 ± 0.01	0.79 ± 0.01	0.73 ± 0.02

Table 5: For each similarity measurement, accuracy sensitivity and specificity with relative standard errors are shown. Best performing metrics are indicated in bold.

5.2.2 *A privileged scale for Alzheimer's disease*

Firstly, the informative content of the proposed multiplex framework was explored by investigating what is the optimal supervoxel size. The motivation behind this experiments are two: (i) it is not possible to *a priori* establish an optimal value for the supervoxel size, (ii) It is not known if there is a specific dimension related to the AD. Accordingly, the supervoxel size was varied and the classification accuracy in discriminating healthy and AD subjects from the training dataset was measured (see Fig. 39).

From this analysis it was found that the optimal size for the supervoxel size was of 3000 mm^3 and the corresponding accuracy value was on average 0.88 with a 0.01 standard error and a sensitivity and a specificity respectively of 0.90 ± 0.01 and 0.88 ± 0.02. This performance was compared with structural

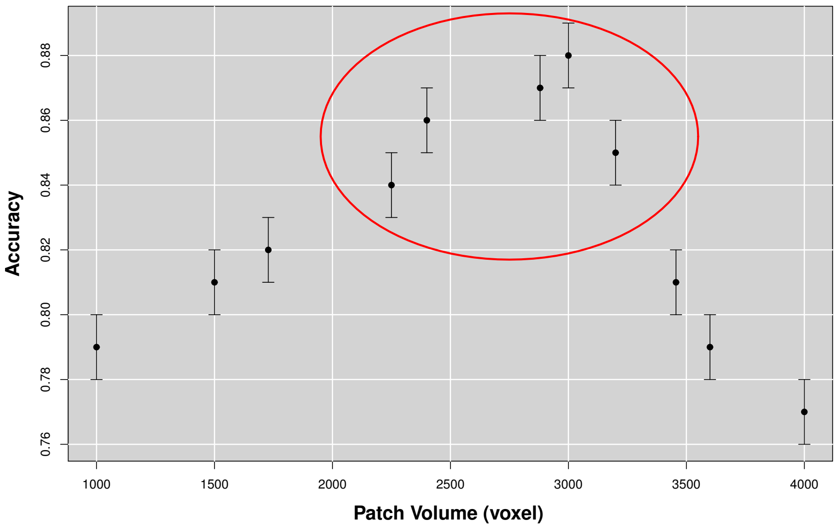

Fig. 39: The figure represents the accuracy for the NC-AD classification as a function of the supervoxel size. The existence of a robust plateau, in correspondence of $[2250, 3200]$ voxels, is circled. These results suggest the existence of a typical scale for multiplex describing the Alzheimer's disease patterns.

morphological features, obtained by FreeSurfer [9], through the same classification strategy, including a first random forest wrapper for feature selection and a second random forest classifier for prediction. In this case classification performance was on average significantly lower 0.83 ± 0.01 confirming the effectiveness of the multiplex characterization.

Moreover, accuracy pattern showed that accuracy increased with the supervoxel dimension until the range $[2250, 3200]$ mm^3 was reached. This behavior can be explained with the fact that at smaller scales, important local differences may not be appreciated because confused with misregistration errors and subject variability; at scales too large, subtle and diffused changes characteristic of the AD, may not be captured . However, it is worthwhile to note that this range was remarkably comparable with the dimensions of several brain structures related to AD, such as the hippocampal volume, therefore this result would suggest the emergence of a typical multiplex dimension for AD characterization.

5.2.3 *Robustness and consistency assessment*

To investigate if classification performance was related to the random permutation of voxels inside a patch, firstly, a varying number of voxel within each patch was shuffled affecting the Pearson's correlation pairwise measurement, while the patch decomposition was kept stable. Then, the classification accuracy was measured. The training results are presented in Fig. 40.

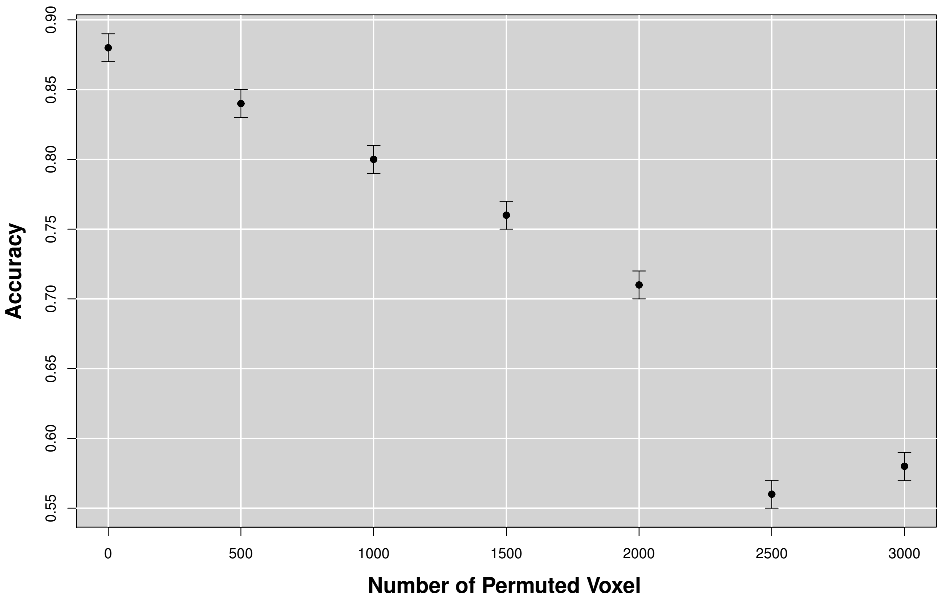

Fig. 40: Accuracy varying with the number of permuted voxel within a patch. Classification performance decreased as the number of shuffled voxels was increased. Noticeably, a drastic drop was observed when the shuffle reached values of about 2500 ~ 3000 voxels.

The test was repeated 100 times increasing the size of the shuffle by 500 voxels at the time. It could be noticed that for small variations, under 1000 voxels, performance did not suffer a significant deterioration; but with 2500 voxel permutation a drastic drop of the performance was observed, a value comparable with the dimensional scale determined in section 5.2.2.

To further assess the method robustness, also a classical non-parametric statistical permutation test was performed. This consisted in the permutation of the clinical labels of each subject belonging to $\mathcal{D}_{\text{train}}$. 1000 random permutations were performed obtaining a consistent decrease of the classification performance, as displayed in Fig. 41. This result suggest that the selected features characterize the disease.

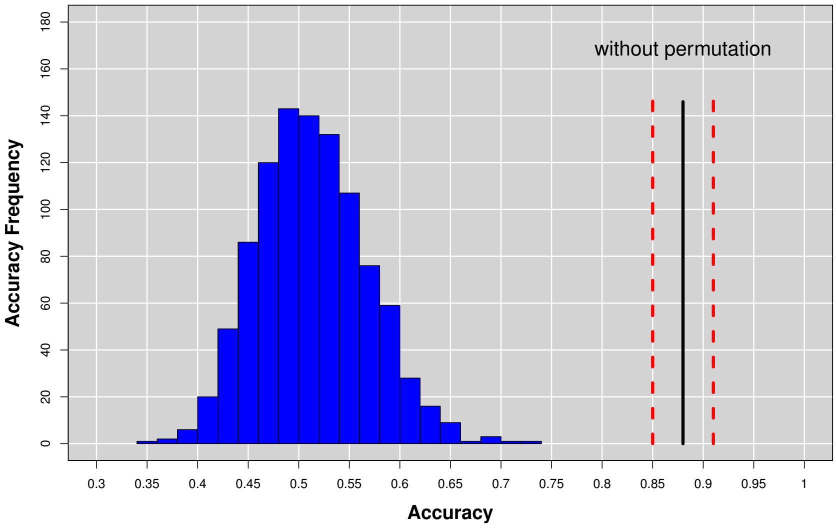

Fig. 41: The accuracy distribution for the binary problem NC-AD on the $\mathcal{D}_{\mathrm{train}}$ with a random permutation of the subject labels. The average value (continuous line) and the relative uncertainty (dotted lines) of best training performances obtained without permutation are also represented for direct comparison.

Training set randomization effectively established that the multiplex framework was able to model a significant relationship between the multiplex features and the clinical label of the $\mathcal{D}_{\mathrm{train}}$ data. Indeed, given the normality of the performance distribution obtained by permuting the labels, it was possible to assign a p-value to the performance obtained without permutations. The statistic test showed that the multiplex model was able to identify a significant ($p < .001$) class structure within the $\mathcal{D}_{\mathrm{train}}$ data. Otherwise, it would not have been possible to reject the underlying null hypothesis: mean values of the performances with and without permutations are equal and thus labels and features were independent.

5.2.4 *Validation results*

As a further assessment on the $\mathcal{D}_{\mathrm{test}}$ dat, two binary classifications: NC vs AD subjects and NC vs cMCI subjects respectively, were performed. The analysis was repeated using 100 bootstrapped $\mathcal{D}_{\mathrm{test}}$ sets to provide a measurement of the performance uncertainty. The accuracies were respectively of 0.86 ± 0.01 and 0.84 ± 0.01. The respective specificity were 0.74 ± 0.01 and 0.72 ± 0.01, while sensitivity reached higher values for both cases: 0.96 ± 0.01 and 0.94 ± 0.01. Remarkably, the NC-cMCI classification performance compared well with NC-AD classification confirming the method reliability and its informative content.

A summary of the classification performances obtained for the different groups are shown in Table 6.

Groups	**Accuracy**	**Sensitivity**	**Specificity**
$\mathcal{D}_{train}(NC - AD)$	0.88 ± 0.01	0.90 ± 0.01	0.88 ± 0.02
$\mathcal{D}_{test}(NC - AD)$	0.86 ± 0.01	0.96 ± 0.01	0.74 ± 0.01
$\mathcal{D}_{test}(NC - cMCI)$	0.84 ± 0.01	0.94 ± 0.01	0.72 ± 0.01

Table 6: Summary of the classification performances in terms of accuracy, sensitivity specificity and relative standard errors for the different groups: NC-AD used for the training, NC-AD and NC-cMCI considered for the validation.

It is worth noting that these performances were obtained using a subset of 70 features including both single-layer and multiplex features.

5.2.5 *Clinical assessment*

Once optimal dimension of multiplex network had been fixed, the most important features were selected. As previously explained in section 4.3.8, those features whose contribution to the classification was considerably far from the null hypothesis of a random behavior were chosen, see Figure. 42 for a typical example.

Starting from the initial network of 549 nodes, for each subject it was determined which regions, basing on their importance, were able to effectively distinguish controls from patients. 32 significant patches were outlined, 18 (~ 56%) in the left hemisphere and 14 in the right; these regions included 27 different cortical and sub-cortical regions listed in Fig. 43 in order of significance. As a region can be included in different patches (provided at least one of its voxels belongs to the considered patch), only most significant p-value entries are reported.

In Fig. 44 some representative brain axial planes are shown, as well as the Harvard-Oxford atlas used for this assessment. In the left hemisphere, patches corresponding to amygdala, hippocampus, para-hippocampal gyrus, pallidum and putamen showed the strongest association to AD ($p = 0.0001$). For cingulate and para-cingulate giri, pre-cuneus, cuneus, and occipital cortex $p = 0.001$.

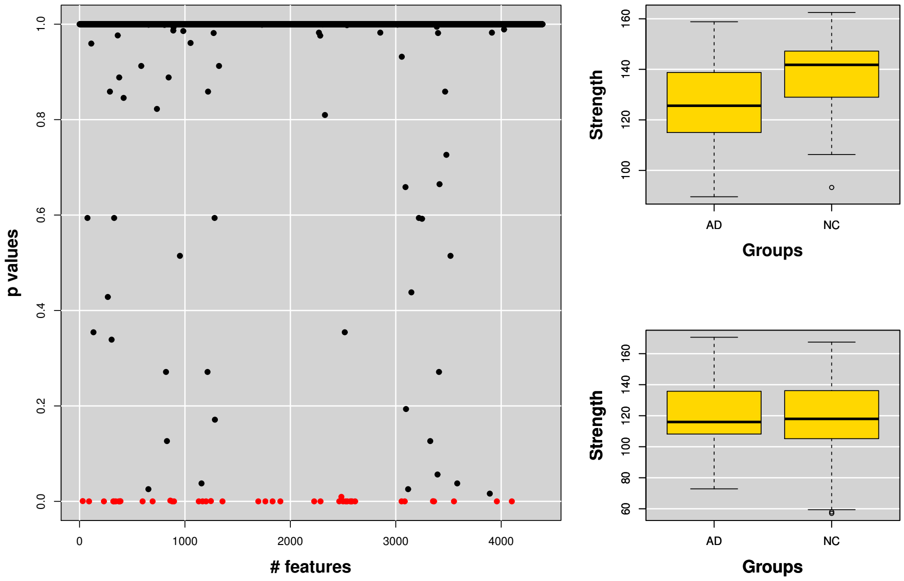

Fig. 42: The figure shows (left panel) the p-values assigned to each feature. Each feature representing a topological property, as for example the strength of a node. The same analysis was then performed for the related nodes. It is also shown (right panel) a typical example of a strength feature for a node significantly related (top) or not (bottom) to AD.

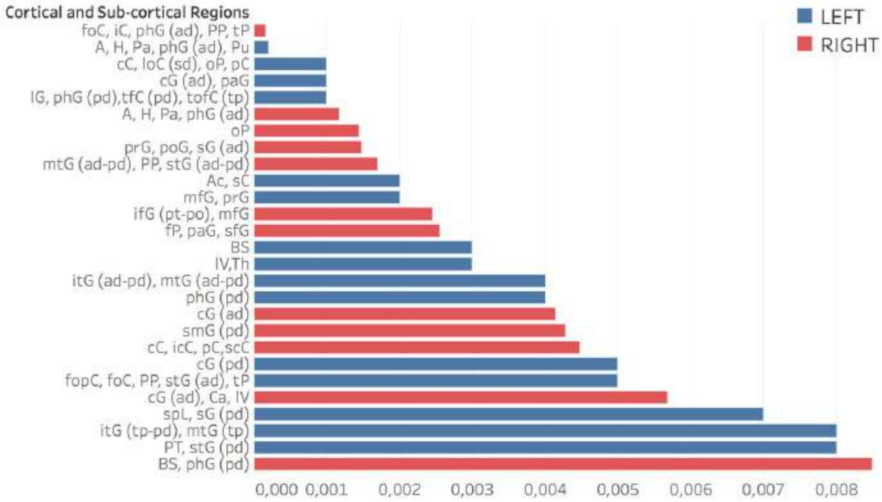

Fig. 43: Regions related to AD in order of significance. Accumbens (Ac), Amygdala (A), Brain-Stem (BS), Caudate (Ca), Cingulate Gyrus (cG) anterior division (ad), Cuneal Cortex (cC), Frontal Operculum and Orbital Cortex (fopC) and (foC), Frontal Pole (fP), Hippocampus (H), Inferior Frontal Gyrus (ifG) pars opercularis and pars triangularis (po) and (pt), Inferior Temporal Gyrus (itG) anterior division and temporoccipital part (tp), Insular Cortex (iC), Intracalcarine Cortex (icC), Lateral Occipital Cortex (loC) superior division (sd), Lateral Ventrical (lV), Lingual Gyrus (lG), Middle Frontal and Temporal Gyrus (mfG) and (mtG), Occipital Pole (oP), Pallidum (Pa), Paracingulate and Parahippocampal Gyrus (paG) and (phG), Planum Polare and Temporale (PP) and (PT). Postcentral and Precentral Gyrus (poG) and (prG), Precuneous Coretx (pC), Putamen (Pu), Subcallosal Cortex (sC), Superior Frontal Gyrus (sfG), Superior Parietal Lobule (spL), Superior Temporal Gyrus (stG), Supracalcarine Cortex (scC), Supramarginal Gyrus (sG), Temporal Fusiform and Temporal Occipital Fusiform Cortex (tfC) and (tofC), Temporal Pole (tP), Thalamus (Th). In parentheses: anterior, posterior and superior division (ad,pd,sd) and temporooccipital part (tp).

Other significant patches ($p = 0.002$) were located in middle frontal gyrus and pre-frontal gyrus, nucleus accumbens, brain stem and thalamus. On the right, $p = 0.0001$ for orbito-frontal cortex, insular cortex, prarahippocampal gyrus, planum polare and planum temporale; $p = 0.001$ for the parahippocampal-amygdalar complex, occipital pole, pre- and post-central gyri, supramarginal gyrus, middle and superior temporal gyri; $p = 0.002$ for inferior, middle and superior frontal gyri, frontal pole, and paracingulate gyrus. Interestingly, the right frontal lobe involvement was more evident.

Multiplex Networks vs Voxel Based Morphometry

In order to establish if this new approach may offer any advantages in terms of region number over existing widely used methods, the same dataset was analyzed with Voxel Based Morphometry (VBM). Significant voxels, with 5% family-wise correction, are represented in Fig. 45.

The VBM analysis showed significant reduction in grey matter density only in bilateral peri-hippocampal regions, more prominent of the left. Compared to the proposed methodology, able to detect 32 significant regions, VBM showed a largely decreased sensitivity. Since the VBM analysis confirmed that left-sided changes were more prominent, two dedicated tests were carried out to

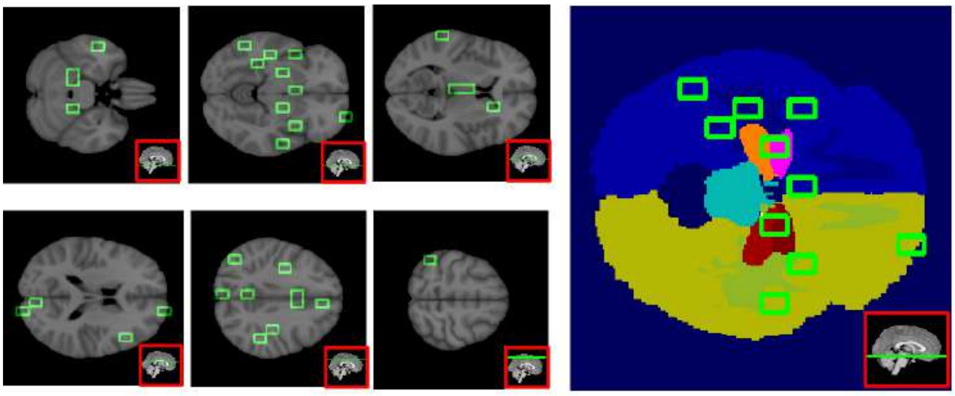

Fig. 44: This figure shows six axial planes (left panel) with the significant patches outlined in green (p-value < 0.01), and on the right, the Harvard-Oxford Atlas used for the patch anatomical localization.

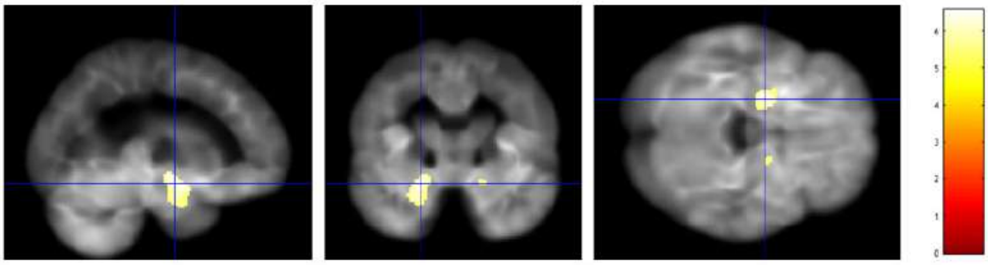

Fig. 45: A voxel based morphometry analysis shows bilateral areas of significantly reduced grey matter density in patients with AD, in medial temporal lobe structures, such as hippocampus and amygdala, more prominent on the left as expected.

further explore the lateralization. Firstly, only multiplex features inherent to the left (right) hemisphere and trained the classification models were used. It was found that left patches were able to discriminate NC from AD patients with an accuracy of 0.87 ± 0.01 while right hemisphere features were able to reach the accuracy value 0.85 ± 0.01. Left hemisphere remained responsible for a greater part of the overall information of the multiplex framework, which was 0.88 ± 0.01.

However, each patch summarizes a network of interrelationships with other patches independently from its spatial collocation. As an example, the strength of a node denotes the sum of its connections, the fact that a node of the left hemisphere is significantly related to AD does not prevent its strength to be the result of its correlation with the right hemisphere. As a consequence, a second test was performed. The multiplexes of left and right hemispheres were separately, thus disregarding one hemisphere. Classification accuracy for NC-AD when using left built (right) multiplex was 0.83 ± 0.01 (0.81 ± 0.01), thus confirming a greater involvement of the left hemisphere but also signaling a definite deterioration of the information content if compared with the whole brain multiplex.

5.3 TWO MAIN BRAIN TOPOLOGIES

There is not an a priori reason for which the scale-free topology of the super-voxel correlation networks should emerge with the same threshold value from different subjects. However, the following Fig. 46 shows how the adjusted R-squared metric averaged over all subjects, adopted to measure the goodness of fit for degree distributions with a power-law function, reaches a high and stable plateau for correlations higher than 0.6.

Observing small-worldness indicator averaged over all subjects (see Fig. 47), it can be noticed as at a threshold grater than 0.6 in addition to a scale-free topology also a small-world structure emerges.

The correlation threshold chosen for the subsequent analyses was of 0.65, the value within the adjusted R-squared plateau and at same time corresponding to the highest and minimum variance small-worldness. This makes natural finding what is called the salience skeleton, the backbone structure including all the most efficient links of the networks, after using a 0.65 threshold granting both scale-free and small world topologies.

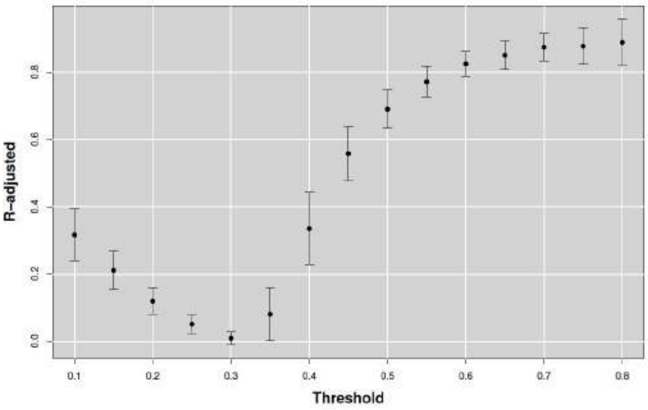

Fig. 46: The brain networks show a consistent power-law degree distribution for thresholds above 0.6. The goodness-of-fit is measured by means of adjusted R-squared coefficient. For each threshold value is represented the mean R-squared coefficient over all subjects and the relative standard deviation.

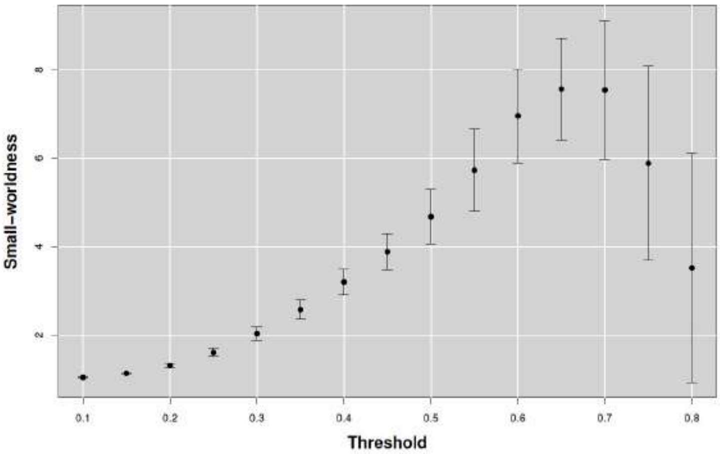

Fig. 47: The brain networks manifest an evident small-worldness behaviour for thresholds above 0.6. For each threshold value is represented the mean small-worldness coefficient over all subjects and the relative standard deviation.

5.4 INFORMATIVE POWER OF THE SALIENT SKELETON

Informative content of the high salient skeletons was compared with the complete multiplex model. In the NC-AD discrimination, the important features extracted from the skeleton, gave accurate results reported here in terms of area under the receiver-operating-characteristic curve (AUC) and its relative standard error. An AUC of 0.93 ± 0.01 and an accuracy of 0.83 ± 0.01 were found for the skeleton multiplex network, an AUC of 0.94 ± 0.01 and an accuracy of 0.88 ± 0.01 were reached for the complete multiplex. Corresponding sensitivities and specificities were respectively 0.83 ± 0.11 and 0.91 ± 0.13, for the first multiplex, and 0.85 ± 0.09 and 0.95 ± 0.10, for the second one. Moreover, the two approaches were combined and used to distinguish NC and AD. Results are presented in Fig. 48.

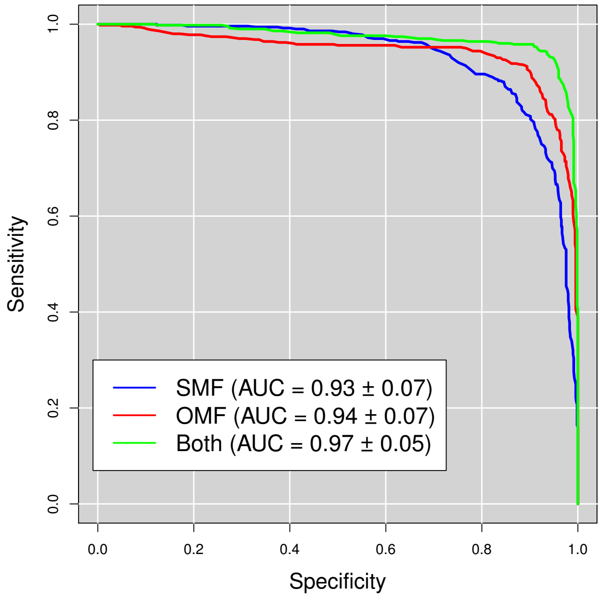

Fig. 48: In figure are reported, for the binary classification normal controls versus Alzheimer's disease patients, the receiver-operating-characteristic (ROC) curves and the corresponding areas under the curve (AUC) relative to skeleton (blue curve), original (red curve) and both (green curve) multiplex network features.

This study showed that the feature combination makes it possible to achieve an higher classification performance with an AUC of 0.97 ± 0.05, an accuracy of 0.91 ± 0.01 and a sensitivity and specificity respectively of 0.88 ± 0.06 and

0.98 ± 0.06, demonstrating that the skeleton extraction from the network is able to provide additional important information within the multiplex framework. Besides nodes involved in the feature extraction from the skeleton multiplex networks are the middle of the original network nodes providing a relevant computational saving. It is worthwhile to note as significant patches, represented in Fig. 49, covered regions such as hippocampus, amygdala, ventricles, thalamus, brain stem, cerebral sulci, inter-hemispheric portions and separation areas between gray and white matter. These findings are a further confirm that salient skeletons are an alternative way of detecting AD patterns, able to give supplemental information regarding the disease.

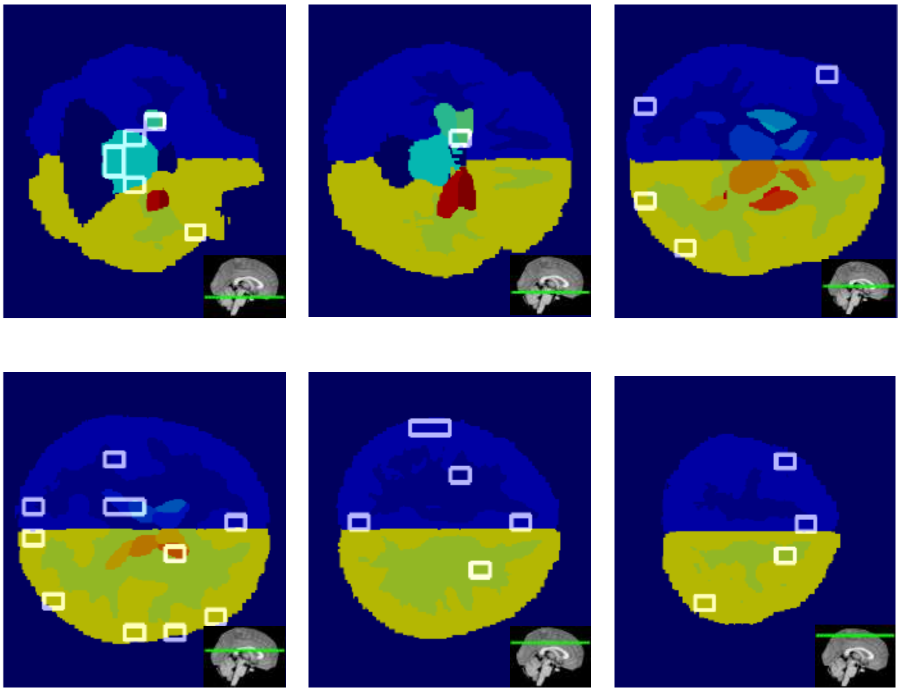

Fig. 49: The figure shows the supervoxels (white boxes), significantly connected with Alzheimer, along the different axial planes of the Harvard-Oxford atlas. The insets depicted in the bottom right corners represent position of each axial plane along sagittal plane.

In addition, to test the reliability of the skeleton multiplex network features (SMF), the classification models obtained during the training phase were then used to carried out, on the independent dataset used for validating the original networks, two binary classifications: controls vs patients and controls vs cMCI subjects . The performances shown in Table 7, along with the ones obtained from the original multiplex network features (OMF), show as the method is reliable and as the most information power of the complete multiplex network is retained by the multiple skeleton network.

Groups	**Accuracy**			**Sensitivity**			**Specificity**		
	SMF	OMF	Both	SMF	OMF	Both	SMF	OMF	Both
AD (48) - NC (52)	0.79 ± 0.01	0.86 ± 0.01	0.87 ± 0.01	0.81 ± 0.02	0.96 ± 0.01	0.93 ± 0.02	0.77 ± 0.02	0.74 ± 0.01	0.78 ± 0.02
MCIc (48) - NC (52)	0.76 ± 0.02	0.84 ± 0.01	0.82 ± 0.01	0.82 ± 0.02	0.94 ± 0.01	0.86 ± 0.02	0.71 ± 0.02	0.72 ± 0.01	0.76 ± 0.02

Table 7: Comparison of the classification performances of the salient multiplex network (SMF), the original multiplex network (OMF) and their combination (Both) in terms of accuracy, sensitivity specificity and the relative standard errors for the different groups: AD-NC and cMCI-NC.

5.5 INFORMATIVE POWER OF THE MULTIPLEX NETWORKS IN PARKINSON

Different experiments were conducted to investigate the effectiveness of the proposed digital image processing study, within the multi-layer network framework, in accurate identification of PD subjects from NC. Also if a privileged scale for studying PD disease exists, exploring the informative content of more scales leads to a significant improvement of the performances in the discrimination between disease and healthy subjects. In particular, it will be given a comprehensive overview of brain regions statistically affected by the disease at different scale. Moreover, it will be shown as the strategy of combining clinical and network features provides an added value to the classification of the subjects in the first stages of the PD disease. Finally, there will be a comparison of the developed method with other conventional approaches in terms of both classification performances and region number connected to the disease.

5.5.1 *Scale study results for several selection methods*

A first experiment was performed to test whether there was an optimal supervoxel size to reveal pathological changes in PD subjects. Thus, it was evaluated how the accuracy varies with the supervoxel volume for three types of feature selection methods (filter, embedded, wrapper). In Fig. 50, it is possible to notice that the higher performances are obtained for smaller scales, then as supervoxel dimension increases, the performance drops and remains broadly constant at a supervoxel volume of $4000\ voxels$ and an accuracy of 0.688 ± 0.004 and 0.692 ± 0.004 respectively for the embedded and the wrapper method. This performance trend happens for all three feature selection methods, also if the wrapper and the embedded work better for each supervoxel size, while the filter method gives performance significantly lower and presents a stable region starting before, around a supervoxel volume of 2000 $voxels$ and an accuracy of 0.674 ± 0.004. The best accuracy value was equal to 0.832 ± 0.004 at a volume of 125 $voxels$ using the embedded method.

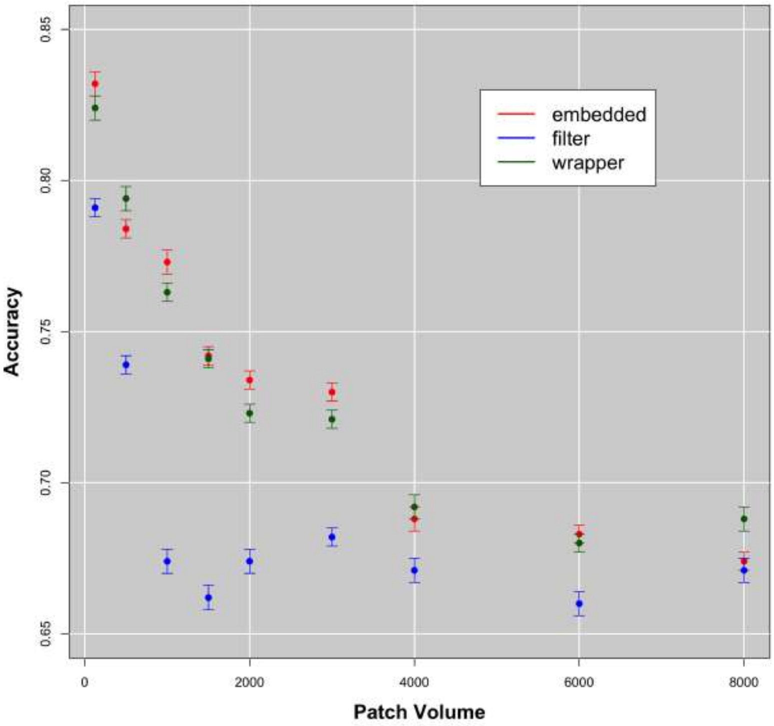

Fig. 50: Scale study for different feature selection methods: filter, wrapper and embedded method. The optimal supervoxel volume was obtained at the smallest scale of 125 $voxels$, for all selection methods also if the embedded reaches higher performance. The best performance in terms of accuracy and its relative standard error was 0.832 ± 0.004.

In Table 8, the most widely used metrics (accuracy, sensitivity, specificity, AUC) to evaluate classification goodness are reported for the best performance of each feature selection method. The major accuracies, obtained with the feature selection techniques based on learning, suggest that taking into account feature interaction is very important in the classification process.

5.5.2 *Feature combination from different scales*

Another experiment was conducted for a double aim: (i) to understand if also the interaction among the features selected from different scales could play an

method	**AUC**	**ACC**	**sens**	**spec**	**features**
Embedded	**0.902 ± 0.004**	**0.832 ± 0.004**	**0.831 ± 0.006**	**0.835 ± 0.006**	43
Wrapper	0.894 ± 0.004	0.824 ± 0.004	0.816 ± 0.006	0.826 ± 0.006	28
Filter	0.876 ± 0.004	0.791 ± 0.003	0.772 ± 0.006	0.807 ± 0.006	156

Table 8: A performance comparison between different feature selection methods (Embedded, Wrapper and Filter method) for the optimal scale, shows that proposed study is robust for the different methods so long as they can take relationships among features. Performances are described in the first three columns for several metrics: area under the receiving operating characteristics (AUC), accuracy (ACC), sensitivity (sens) and specificity (spec). In the last column, the number of the selected features, for each method, is reported. The best performances, reached be the embedded method, are in bold.

important role in the discrimination of PD subjects from NC and (ii) if features belonging to more scales could give complementary information and make more accurate the classification. Therefore, using the combination of the important features picked from three different scales, three classifiers, each one relative to a different feature selection method, were training giving very interesting classification results. A significant improvement of the classification performances was achieved with both the combination of the features selected for each scale through the embedded method and the feature combination obtained for each scale through the wrapper. While the combination of different scale features picked through the filter method gives no improvements. In Table 9, performances obtained with the combination of the features deriving from the scales of 125 voxels, 1000 voxels and 3000 voxels, are shown for all feature selection methods.

Therefore it is possible to deduce that although there is a privileged scale to study Parkinson's disease, the features related to the other scales are found to have an informative content partly different. Indeed, combining the important features belonging to the three different scales (including the best performing scale), accuracies of 0.884 ± 0.004 and 0.862 ± 0.003 was achieved, respectively for the embedded and the wrapper method, significantly greater than the accuracies obtained with features relative to the best performing single-scale equal to 0.832 ± 0.004 using the embedded method and 0.824 ± 0.004 using the wrapper. Thus the most of the changes due to the disease are detectable at small scales which, however, can neglect, some pathological variations within greater anatomical structures. In fact, these variations, using too small scales, can be mostly confused with the natural cerebral variability.

method	**AUC**	**ACC**	**sens**	**spec**
Embedded	**0.925 ± 0.004**	**0.884 ± 0.004**	**0.861 ± 0.006**	**0.908 ± 0.006**
Wrapper	0.911 ± 0.004	0.862 ± 0.003	0.846 ± 0.006	0.886 ± 0.006
Filter	0.878 ± 0.004	0.795 ± 0.003	0.765 ± 0.006	0.810 ± 0.006

Table 9: A performance comparison between different feature selection methods (Embedded, Wrapper and Filter method) for the combination of the features obtained from three different scales (125,1000 and 3000 voxels) taken before the stable region and comparable with the characteristic dimension of anatomical structures affected by PD in the prodromal phase. Performances are reported for each metric: area under the receiving operating characteristics (AUC), accuracy (ACC), sensitivity (sens) and specificity (spec). The best performance (in bold) is attained using the combination of features selected using for each scale the embedded method.

Besides, it is worth noting as features of a same scale by themselves may have little correlation with subject diagnosis, but when they are combined with features of other scales, they can be strongly correlated with the classes. Involuntary removal of these features can result in the loss of useful information that may make less efficient the identification of disease and healthy subjects. The better performances obtained with the feature selection approaches considering interactions among features demonstrate really the importance to recognize inter-connections among features of several scales for studying Parkinson's disease and achieving more accurate classification performances. In Fig. 51, some representative axial planes, containing significant supervoxels for each scale, are reported.

These supervoxels, corresponding to the most important classification features, underlined anatomical structures strictly connected with the PD, such as midbrain, putamen and substantia nigra. These are also the regions first affected by the disease.

5.5.3 *Combination of clinical and network features*

In order to evaluate the effectiveness of the proposed procedure on the complete PPMI dataset and the potentiality of the extracted features to become a diagnostic tool, the informative content provided by the complex network approach and expressed in a unique Random Forest diagnostic score, was combined with the clinical features. Besides, the informative content of network measures and clinical features were separately evaluated. All the presented re-

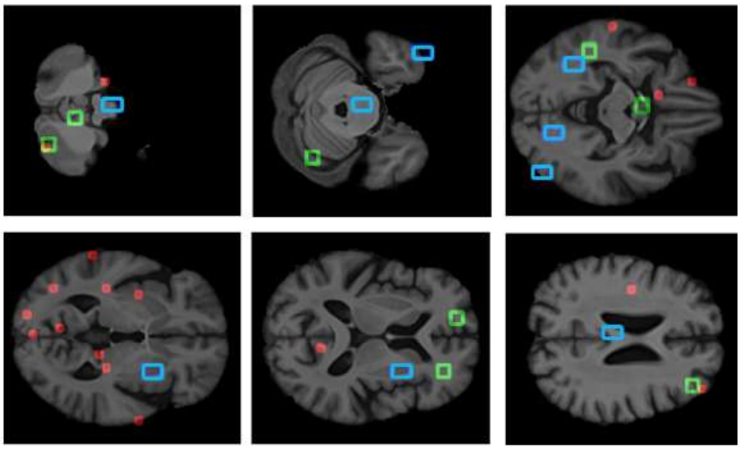

Fig. 51: This figure shows six axial planes with the significant patches (p-value < 0.01) outlined in red for the smaller scale (125 $voxels$), in green for the intermediate scale (1000 $voxels$) and in cyan for the larger scale (3000 $voxels$).

sults were acquired with a 10-fold cross-validation framework and are shown in Fig. 52.

The combined use of MRI and clinical features gives the best performance as summarized in Table 10.

features	**AUC**	**ACC**	**sens**	**spec**
Network measures	0.94 ± 0.01	0.88 ± 0.06	0.85 ± 0.09	0.88 ± 0.09
Clinical scores	0.77 ± 0.01	0.70 ± 0.08	0.65 ± 0.12	0.75 ± 0.11
Both	$\mathbf{0.97 \pm 0.02}$	$\mathbf{0.93 \pm 0.04}$	$\mathbf{0.93 \pm 0.06}$	$\mathbf{0.92 \pm 0.07}$

Table 10: NC vs PD classification performances for feature typology. Area under the receiver operating characteristics (AUC), accuracy (ACC), sensitivity (sens) and specificity (spec) are reported with the relative standard deviations. Best performance (bold) is obtained with a combined use of network and clinical features.

Area under the receiver operating characteristics $AUC = 0.97 \pm 0.02$, accuracy $ACC = 0.93 \pm 0.04$, sensitivity $sens = 0.92 \pm 0.06$ and specificity $spec = 0.93 \pm 0.07$. These results were significantly higher than those obtained by using only the complex network measures; in fact, in this case $AUC = 0.94 \pm 0.01$, $ACC = 0.88 \pm 0.06$, $sens = 0.85 \pm 0.09$ and $spec = 0.88 \pm 0.09$ were found. These results were averaged on 1000 cross-validation rounds and significance was assessed with z-tests; for all comparisons 1% significance was obtained.

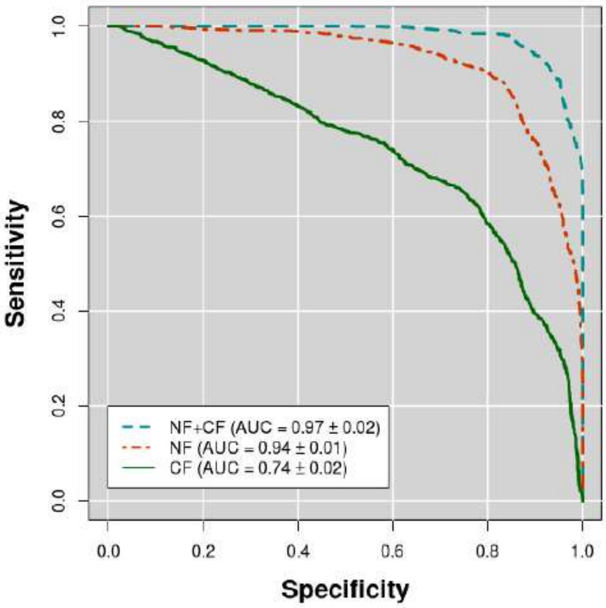

Fig. 52: Classification performance in terms of area under the receiver operating characteristics curve (AUC). The combined use of network features (NF) and clinical features (CF) reaches the highest AUC = 0.97 ± 0.02 (cyan dashed line). CF on their own reach an AUC = 0.74 ± 0.02 (dark green continuous line), which is reasonable for baseline subjects whose symptoms are mild. NF provide effective markers for PD, in fact basing on these feature it is possible to diagnose PD with an AUC = 0.94 ± 0.01 (dark red dash-dot line).

Clinical features resulted in a classification performance significantly lower than those previously reported. Thus, the information content provided by the proposed model gives a significant contribution. In particular, it was found when using only the clinical features: $AUC = 0.77 \pm 0.01$, $ACC = 0.70 \pm 0.08$, $sens = 0.66 \pm 0.12$ and $spec = 0.73 \pm 0.11$. Fig. 53 allows to appreciate this effect from a different perspective.

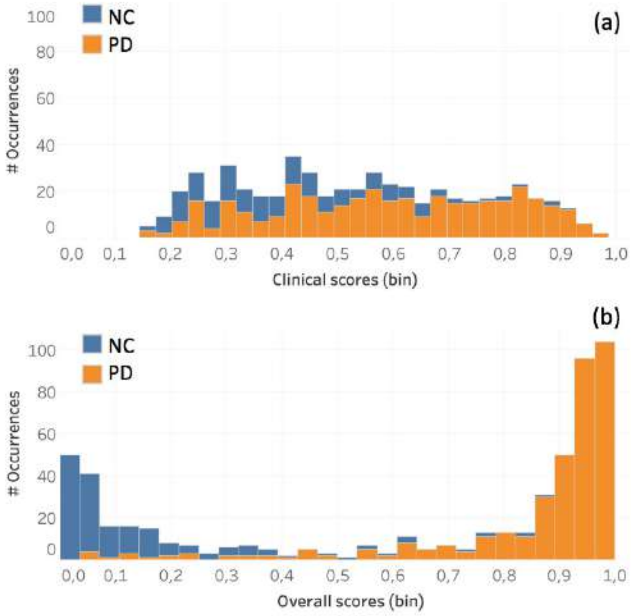

Fig. 53: The classification score distribution using (a) only clinical features and (b)combining them with network measures. Each column of the histogram contains the number of NC subjects (blue) and PD patients (orange) whose score lies in that bin. Classification scores obtained using both network and clinical features show a greatly enhanced class separation.

Classification scores based only on clinical features consistently tend to overlap and assign to PD subjects low scores. In fact, sensitivity, which is basically the discriminative power for positive subjects, is lower than specificity. On the contrary, the discrimination of the two classes is greatly enhanced when introducing complex network markers.

5.5.4 *Anatomical interpretation*

The high discriminative power shown by the features and evaluated in the previous section demonstrate the reliability of complex network measures as PD markers. Beside the diagnostic support these features can provide, it is interesting to evaluate which regions result to be affected by the disease and eventually rank them according to their statistical significance.

For each cross-validation round, features selected as the most important for classification were tested and the brain regions related to them, were outlined. Accordingly, for each cross-validation round it was counted whether or not a particular anatomic district had been selected and was tested the hypothesis that these occurrences had happened by chance. The null hypothesis is that a region should be selected according to a Bernoulli distribution with an *a priori* probability 1/12219, which is the total number of patches. It was tested the occurrence of a specific region against the random occurrence rate which was in our case equal to 1/patch number (1/12219), which is in conclusion the number of the carried tests. Fig. 54 shows some of the significant regions associated to the diagnosis.

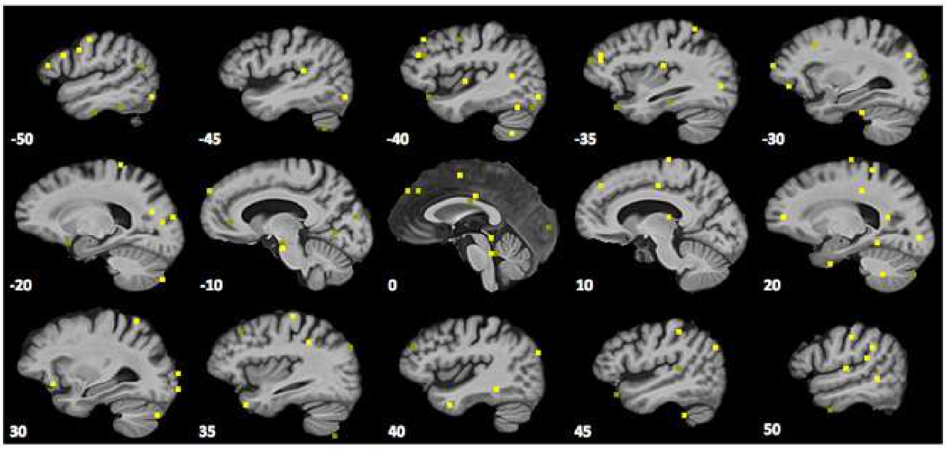

Fig. 54: A qualitative overview of significant PD related patches as they are outlined by complex network measures is represented along sagittal planes. More significant patches have a lighter shade of yellow and little by little less significant patches have a darker shade of yellow. The negative planes belong to the left hemisphere while the positive planes belong to the right one.

According to the proposed method, 186 significant patches were detected: 43% within the right hemisphere and 40% within the left one; 9% of patches were located in the cerebellum; brainstem regions appeared in 3% of cases. There is not a great difference between GM and WM regions, in fact the GM regions selected are 53%, the WM ones 46%. The vast majority of brain regions affected by PD lies in the Frontal (27%), Occipital (21%) and Temporal (15%)

lobes. Instead, in the following Table 11 a compact overview of the first 15 selected patches was reported.

Region	p-value
(L) Temporal Lobe. Middle Temporal Gyrus. GM-WM. Ba 39.	$7.2 \cdot 10^{-7}$
(R) Temporal Lobe. Superior and Inferior Temporal Gyrus. GM-WM. Ba 22.	$7.2 \cdot 10^{-7}$
(R) Occipital Lobe. Sub-Gyral (WM).	$3.6 \cdot 10^{-6}$
(L) Occipital Lobe. Superior Occipital Gyrus GM-WM. Ba 19.	$4.6 \cdot 10^{-6}$
(R) Frontal Lobe. Middle Frontal Gyrus (WM).	$5.3 \cdot 10^{-6}$
(L-c) Anterior Lobe. Culmen (GM).	$5.9 \cdot 10^{-6}$
(R) Frontal Lobe. Medial Frontal Gyrus (WM).	$5.9 \cdot 10^{-6}$
(R) Frontal Lobe. Precentral Gyrus. GM. Ba 44.	$5.9 \cdot 10^{-6}$
(L) Limbic Lobe. Cingulate Gyrus. GM. Ba 24.	$5.9 \cdot 10^{-6}$
(R) Parietal Lobe. Precuneus (WM).	$5.9 \cdot 10^{-6}$
(L) Frontal Lobe. Middle Frontal Gyrus. GM. Ba 46.	$6.2 \cdot 10^{-6}$
(L) Brainstem. Midbrain	$6.6 \cdot 10^{-6}$
(L) Temporal Lobe. Fusiform Gyrus GM-WM. Ba 37.	$7.2 \cdot 10^{-6}$
(R-c) Posterior Lobe. Declive. GM.	$7.2 \cdot 10^{-6}$
(R) Temporal Lobe.Fusiform Gyrus. WM.	$7.2 \cdot 10^{-6}$

Table 11: The regions selected according to complex network measures and the inherent level of significance with respect of diagnosis. (L) and (R) denotes the left and right hemispheres; cerebellum regions are denoted with c. Brodmann areas (Ba) are also outlined when appropriate.

These regions have been already detected in several PD studies, another indirect validation of the proposed methodology in that the selected regions consistently correspond to regions whose relationship with the disease is established.

5.5.5 *Comparison with standard methods*

In order to assess the effectiveness of the proposed approach, its classification accuracy was compared with that of two standard approaches. In particular, FreeSurfer [9] was used to extract some structural features, such as grey matter and white matter volumes of subcortical brain structures or the average cortical thickness of specific regions obtaining a 181 ROI feature representation. Then, a standard VBM pipeline [12] to detect voxels showing a significant (p-value < 0.01) association with the diagnosis was performed, these voxels provided another feature representation. Both, the ROI and the VBM descriptions were used to feed the classification framework previously described. In this way, a direct comparison was obtained evaluating the informative power of the proposed methodology and ROI/VBM approaches, see Fig. 55.

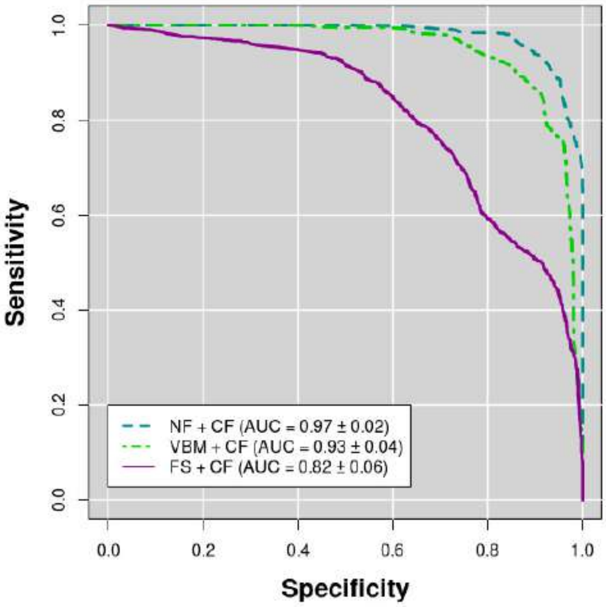

Fig. 55: Classification performance in terms of area under the receiver operating characteristics curve (AUC). The combined use of network features (NF) and clinical features (CF) reaches the highest AUC = 0.97 ± 0.02 (cyan dashed line). The combination of Voxel Based Morphometry (VBM) features and CF reaches an AUC $= 0.93 \pm 0.06$ (light green dash-dot line). Finally, FreeSurfer features (FS) combined to CF give an AUC $= 0.82 \pm 0.06$ (magenta continuous line).

The proposed method (AUC = 0.97 ± 0.02) favorably compared with VBM (AUC = 0.93 ± 0.04) and ROI (AUC = 0.82 ± 0.06) description. It is worth mentioning that even when not considering clinical features, the network description remained the most effective (AUC = 0.94 ± 0.01) respect both VBM (AUC = 0.87 ± 0.05) and ROI (AUC = 0.70 ± 0.06) approaches. A summary of this comparison is presented in Table 12.

A standard VBM pipeline to segment gray and white matter of each MRI scan was used. Each subject was normalized to the MNI152 template and extracted the t-Student maps to determine if some clusters of voxels (> 30) would exhibit an association with the diagnosis. A good agreement with the regions detected with our complex network description was found, see Fig. 56

features	AUC	ACC	sens	spec
NF + CF features	**0.97 ± 0.02**	**0.93 ± 0.04**	**0.93 ± 0.06**	**0.92 ± 0.07**
VBM + CF	0.93 ± 0.04	0.86 ± 0.06	0.88 ± 0.08	0.86 ± 0.08
FS + CF	0.82 ± 0.06	0.72 ± 0.07	0.74 ± 0.10	0.71 ± 0.12
NF	0.94 ± 0.01	0.88 ± 0.06	0.85 ± 0.09	0.88 ± 0.09
VBM	0.87 ± 0.05	0.79 ± 0.08	0.77 ± 0.12	0.77 ± 0.11
FS	0.70 ± 0.06	0.63 ± 0.07	0.60 ± 0.11	0.66 ± 0.11

Table 12: NC vs PD classification performances for network features (NF), Voxel Based Morphometry (VBM) and FreeSurfer (FS) obtained with and without the combination of clinical features (CF). Area under the receiver operating characteristics (AUC), accuracy (ACC), sensitivity (sens) and specificity (spec) are reported with the relative standard deviations. Best performance (bold) is obtained with a combined use of network and clinical features (NF+CF). The combination with the clinical features improves the performances in all three cases (NF+CF,VBM+CF,FS+CF).

for an overview, but, remarkably, the number of regions showing an association with the clinic was consistently reduced.

However, voxel-wise approaches have an intrinsic drawback in that the need for thousands of multiple comparisons dramatically lowers the statistical power of commonly available datasets. This is also the case here, as no significant association after Bonferroni corrections was found.

5.5.6 *Robustness and content evaluation*

The relation between the threshold used to remove some edges from the network and the classification accuracy was evaluated; a wide range of thresholds was explored, see Fig. 57. Threshold was varied from 0 to 0.9 with 0.1 steps and with a fixed patch volume of 125 voxels. The maximum value of the classification accuracy was obtained at a 0.3 threshold. It is worth noting that this value also corresponded to minimum variance. Accuracy remained constant at 0.93 for a wide range of correlations ($[0.3; 0.5]$), thus confirming that the method does not require a fine tuning of correlation threshold values. For higher threshold values, there was a significant performance drop, suggesting that too high threshold values cause the loss of important links.

As demonstrated, the proposed complex network approach significantly enhances the discriminative power of clinical features. The network measures de-

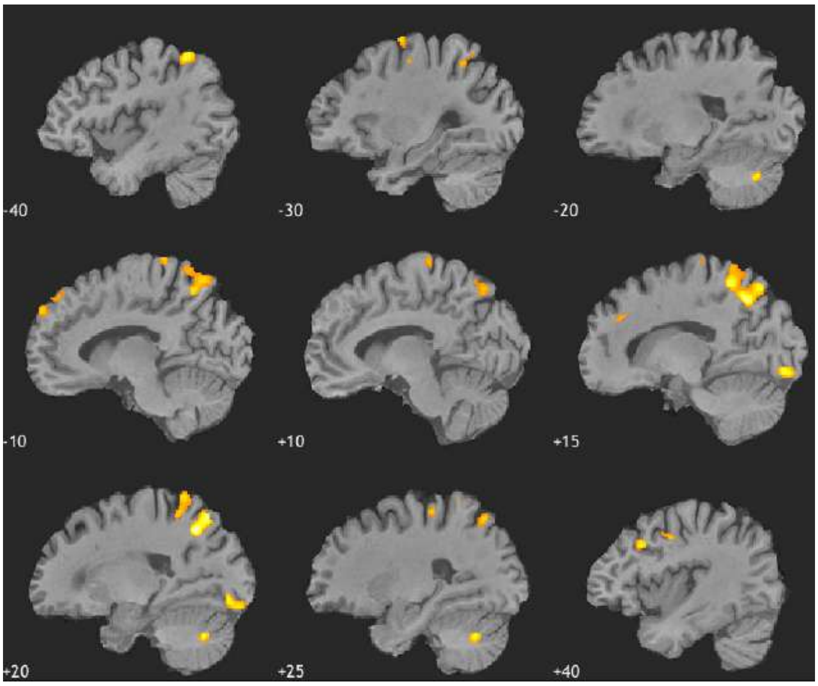

Fig. 56: Voxel based morphometry shows the presence of some clusters (the sagittal plane is reported), however these regions represent only a subset of those outlined by our approach.

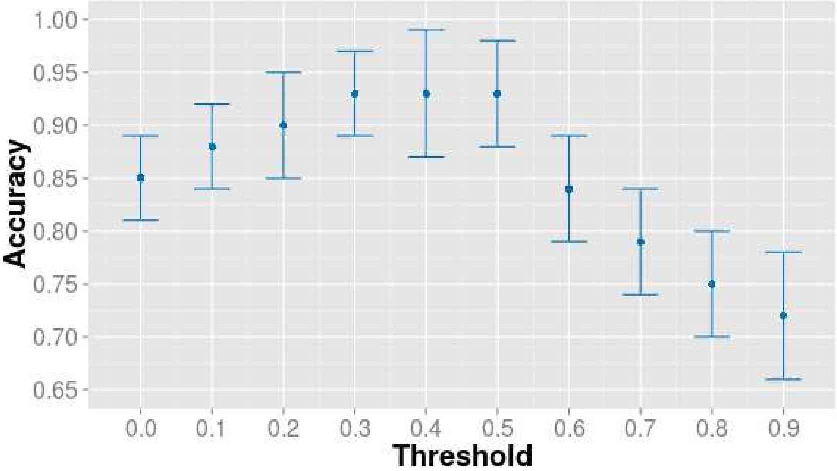

Fig. 57: The figure shows the accuracy as a function of the threshold that changes from 0 to 0.9. In correspondence of a threshold value of 0.3, the best accuracy and the minimal standard deviation was reached.

rived from MRI data effectively characterize PD patterns. A not secondary goal of this work was to demonstrate that the proposed complex network description did not require any fine tuning, accordingly, for both image processing and classification standard configurations were used , then it was evaluated, a posteriori, that choosing other configurations would have not affected the obtained results. To evaluate the robustness of the informative content provided by our method, the hyperparameter space was expired. Firstly, the cost parameter which plays a fundamental role for Support Vector Machines was evaluated,see Fig. 58.

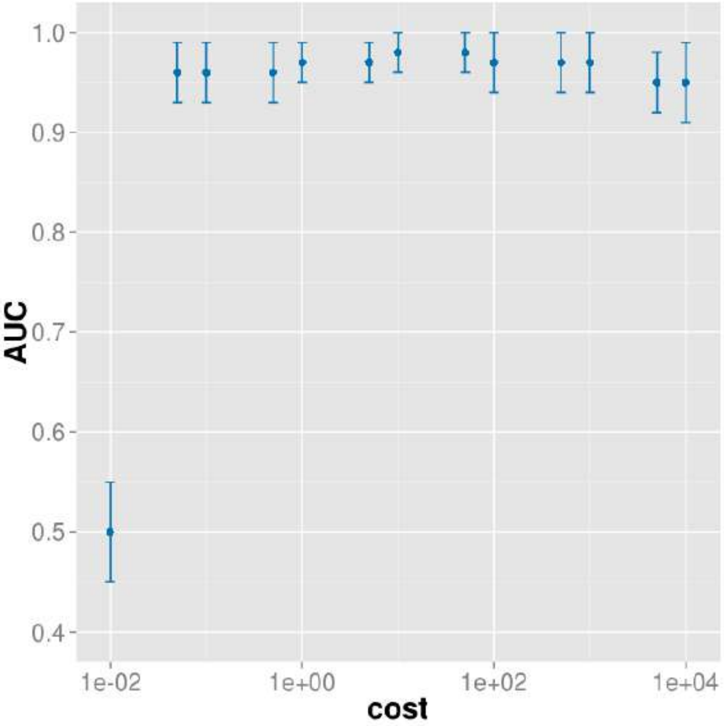

Fig. 58: The classification performance in terms of AUC remains stable by varying the cost parameter. For tiny cost values the decision hyperplane margins are too large and the performance drops.

In fact, the cost determines how much the SVM model should fit the training data by varying the margins of the decision hyperplane, larger values of cost correspond to smaller margins. The results show that for a wide range of cost values the classification performance remains stable granting robust results. When the cost reaches the 0.01 value the performance drops, this means that the margins have become so large that the model cannot just fit the data. Moreover, it was investigated the model robustness with respect of the gamma parameter which defines how far the region of influence of each training example should extend, see Fig. 59.

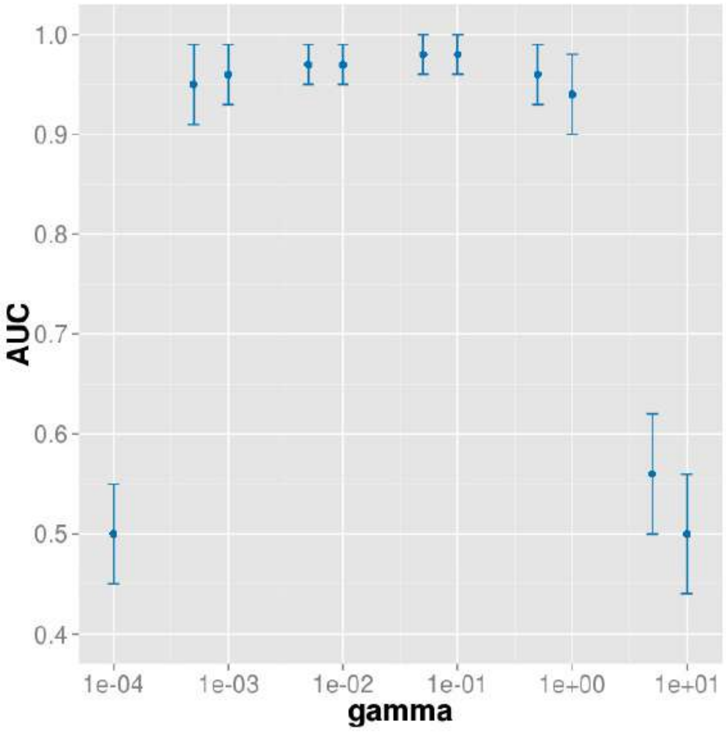

Fig. 59: Varying the gamma parameter it is possible to switch from high-bias to high-variance models, in this case however the optimal classification region extends for more than 4 orders of magnitude.

As gamma controls the variance of the model, by varying gamma one can move from a high-bias to a high-variance model. Of course, the optimal classification region stands between these two cases. The results show that for the present model a wide stability region exists, in fact the AUC consistently remains over the 0.90 value for a gamma variation of more than 4 orders of magnitude.

To evaluate the goodness of complex network measures as PD markers it is not sufficient to demonstrate that these features allow an accurate and robust classification. First of all, looking at the classification performance it is not possible to guess if credit should be given to the SVM classification model or the informative content of the features. As such, the classification performance of several state-of-the-art classifiers was compared, specifically Random Forest (RF), Naive Bayes (NB) and Neural Network (NN) classifiers were investigated. For each method within a nested cross-validation the hyperparameter space and several configurations was explored, only optimal configuration results are reported in the following Table 13.

method	**AUC**	**ACC**	**sens**	**spec**
Neural Network	0.94 ± 0.04	0.89 ± 0.05	0.90 ± 0.08	0.88 ± 0.07
Random Forest	0.97 ± 0.02	0.91 ± 0.05	0.90 ± 0.07	0.91 ± 0.07
Naive Bayes	0.97 ± 0.03	0.92 ± 0.05	0.91 ± 0.07	$\mathbf{0.93 \pm 0.07}$
Support Vector Machine	$\mathbf{0.97 \pm 0.02}$	$\mathbf{0.93 \pm 0.04}$	$\mathbf{0.93 \pm 0.06}$	0.92 ± 0.07

Table 13: A comparison between different machine learning methods (Neural Networks, Random Forests, Naive Bayes and Support Vector Machine classifiers) shows that the proposed complex network approach allows a robust diagnosis independently from the choice of the classifier, although Support Vector Machines reach slightly better results (in bold) for almost each metric: area under the receiving operating characteristics (AUC), accuracy (ACC), sensitivity (sens) and specificity (spec).

The table shows that no significant difference can be found between different models, even if SVM would seem to perform slightly better than the others. These results demonstrate that beyond the differences due to the machine learning models adopted, the proposed approach yields an outstanding base of knowledge for PD discrimination.

However, this test does not evaluate the agreement between the models. In fact, in principle two distinct models could perform equally but misclassifying different subjects. In order to assess the agreement between our chosen SVM model and the other models, the relationships existing between the classification scores were investigated. The results are presented in Fig. 60.

The classification scores are densely distributed in the top right and bottom left quadrants. The top right quadrant includes subjects whose classification scores exceed the 0.5 value, it is the case of subjects diagnosed with PD from both the SVM model, which is always reported on the x axis, and the other models, which are reported instead on the y axis. Analogously, the bottom left quadrant includes those subjects, whose classification scores are lower than 0.5, for which the models agree assigning a NC status. As expected from previous measures, the models correctly distinguish the two classes, in fact in the top right quadrant the vast majority of subjects is shown in orange, as they have a PD diagnosis, and in in the bottom left in blue, as they are mainly NC subjects.

The top left and bottom right quadrants are the regions of disagreement. In these regions, in fact, the SVM model assigns a diagnosis different from other models. For example, a subject belonging to the bottom right quadrant has

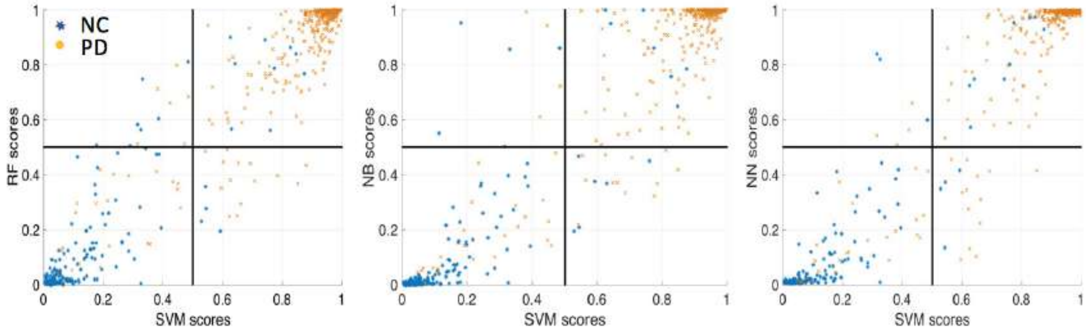

Fig. 60: From left to right the agreement between the scores obtained with Support Vector Machines (SVM) and those obtained by: Random Forest (RF), Naive Bayes (NB) and Neural Network (NN) classifiers. The scores are densely distributed in top right and bottom left quadrants, where their predictions agree. Looking at the top left and bottom right quadrants it can be noted that, when in disagreement, SVM scores tend to be slightly more accurate than other scores.

an SVM score > 0.5 and it is accordingly diagnosed as PD but a RF score (or NB/NN) < 0.5 and it is therefore labeled as NC. First of all, it is worth noting that these two quadrants are sparsely populated, especially compared to the top right and bottom left ones, which is a further demonstration of the agreement between the models; besides, SVM tends to be more accurate.

Keeping on with our example, in the bottom right quadrant for all the three cases the majority of subjects is orange, meaning that their true label is PD. This means that the SVM predictions is the right one. The same consideration holds for the top left quadrant, where the majority of subjects is represented in blue, and, again SVM correctly labels them as NC.

6

CONCLUSIONS

6.1 SCIENTIFIC MOTIVATIONS

Alzheimer's disease (AD) and Parkinson's disease (PD) are progressive, neurodegenerative diseases accounting for most cases of dementia after the age of 65. It is expected that over 115 million people will develop AD and over 8.7 million people develop PD by 2050 [131, 132]. Illness related brain changes can be detected *in vivo* with magnetic resonance imaging (MRI) that has been playing an increasingly important role for the diagnosis of neurodegenerative disorders to the extent that it has been incorporated in the diagnostic criteria for AD and PD citepMcKhann2011Thediagnosis. It is now accepted that the neurodegenerative cascade in AD and PD begins in the brain years, decades even, before the clinical and radiological manifestations of the illness. The dementia is preceded by a prodromal phase of mild cognitive impairment [25], and this, in turn, by a pre-clinical phase [22] of variable duration. Understanding the biological changes, occurring in these early phases, is of paramount importance, as it would open a window of opportunity for future disease-modifying treatments. Indeed for both the diseases there is not a cure, yet. Moreover it is not known exactly what drives the propagation of these diseases within an individual, and what there is behind the variations in the patterns of atrophy between individual. To which extent neurodegeneration propagates through anatomical contiguity, or through preferential pathways of structural or functional connectivity is yet to be clarified. The two most hallmarks used to define the diagnosis of AD upon autopsy are the amyloid plaques and the neurofibrillary tangles. Some autopsies obtained when an elderly person dies not of AD, show that at the onset of the disease symptoms over 50% of the nerve cells are already lost in specific part of the entorhinal cortex. This demonstrates that the neurodegenerative process starts probably 20 years before the onset of the symptoms [133]. The presymptomatic phase of PD from the onset of neuronal loss was estimated to be about 5 years. This phase is represented by incidental Lewy body cases: individuals who die without clinical signs of PD or dementia, but who are found to have Lewy bodies at post-mortem [134]. Nowadays, cerebrospinal fluid analyses and brain imaging using radioactive tracers, can tell us to what extent the brain is covered with plaques and tangles, in case of AD, or synucleinopathies in case of PD and these modalities are able to predict relatively accurately who is at high risk of developing the two diseases about ten years later [135]. However, these methods are very invasive, expensive and only available at highly specialized centers.

Standard imaging techniques with structural MRI have allowed the observation of gray matter reductions. For example, voxel based morphometry (VBM) has detected significant structural changes in PD subjects and longitudinal studies have also shown progressive atrophy already distinguishable in the early stages of the disease Even tough, features extracted from these methods give excellent classification results, they are often based on the supervised identification of the regions of interest associated to the disease. Therefore, these approaches can be limiting because they do not allow the investigation of novel brain regions and the relative mechanisms involved in the disease [136].

Recently, the application of the complex networks and in particular, the use of the innovative tool of the multiplex networks for studying brain magnetic resonance imaging (MRI) have been playing a fundamental role in the field of the neurodegenerative diseases. Multiplex networks, allowing a statistical analysis of the inter and intra-subject differences, are particularly appropriate for the identification of abnormal structural changes which in pathologies as PD can be part of one anatomical district or span over multiple atlas-based regions of interest (ROI) [137]. Even if novel strategies for accurate brain segmentation have been investigated, they remain a time consuming task and do not take into account the whole brain, so that an approach like the one proposed which does not require any segmentation can be very useful [138]. Besides, the use of complex networks allows the description of a set of different data through the use of quantitative indicators and permits an efficient management of the great quantity of data that a scientist inevitably has to face, in the context of neuroimaging, given the increasing technological progress.

6.2 RESULTS AND THEIR DISCUSSION

During these three years of doctorate, it was demonstrated that both the multiplex network approach and the reduction method of the salient networks have all the potential of becoming a predictive tool.

The proposed methodologies allowed the modelling of brain atrophy in AD through inter-subject multiplex networks whose nodes are represented by brain patches and edges by pairwise Pearson's correlations. Basing on multiplex features learning models were constructed and used allowing a robust classification performance over a broad range of Pearson's correlations and with the use of other similarity metrics (see sections 5.2.1).

An optimal volume size for the detection of AD effects able ti maximize the informative content of the multiplex, was identified in the range from 2250 to 3200 mm^3. It is worthwhile to note that this range was comparable with the size of several brain structures related to AD, such as the hippocampus, therefore suggesting it as an ideal multiplex dimension for AD characterization.

The high sensitivity of the method in the detection of illness related brain changes was demonstrated by the number of regions, 32, that were identified as significantly associated with AD. The detected regions comprised hippocampus and para-hippocampal-amygdalar complex, pallidum and putamen, cingulate and paracingulate giri, pre-cuneus, cuneus, and occipital cortex, middle frontal gyrus, pre-central gyrus, accumbens, sub-callosal cortex and brain stem.

It was striking how VBM on the same data set was able to detect only atrophy of the perihippocampal regions. The method here described seems more sensitive than standard VBM [139]. The method outlined the involvement of 32 significant brain regions, but only 22 concerned single-layer measures; thus, the multiplex model allowed a consistent increment (+46%) in sensitivity. The results also confirmed asymmetry in the spatial distribution of significant patches, mostly located in the left hemisphere, in keeping with several other studies [140, 141]. This asymmetry has a direct effect on the informative content (see 5.2.5).

As to the application of this methodology to disease classification studies, the method robustness on an independent set $\mathcal{D}_{\text{test}}$ was evaluated and its reliability for discriminating both AD patients and cMCI subjects from controls was confirmed. Classification performances are accurate, indeed after a training phase that gives in cross-validation an accuracy of 0.88 ± 0.01, the multiplex base of knowledge, on the independent dataset $\mathcal{D}_{\text{test}}$, is able to accurately distinguish between NC and AD subjects with an accuracy of 0.86 ± 0.01 and can be suitably employed also for NC and cMCI classification with an accuracy of 0.84 ± 0.01. It should be noticed the obtained results are comparable with recent classification-focused studies [142, 143, 144].

In this work, it was also proposed an innovative method for reducing complexity of dense brain networks, preserving and especially enhancing, at the same time, their informative content. A study was carried out to bring out the two principal topologies (small-world and scale-free) of the brain networks, able to model and describe the most of the behaviors involved in brain organization. In particular, it was studied how a neurodegenerative pathology, such as Alzheimer disease, can modify structural organization of these brain networks by extracting for each high salient skeleton in order to focus on the strategic hubs and highways of these correlation networks.

We proved, using multiplex network approach, the forcefulness of this reduction methodology, in fact feature extracted from the skeleton multiplex network give in cross-validation, for the discrimination of AD subjects from normal controls, an accuracy of 0.83 ± 0.07 which reaches 0.91 ± 0.01 combining these features with those obtained using the original multiplex network. The improvement of the classification performances and the different features extracted from the skeletons demonstrate that salient links highlight an information, at least in part, distinct from that obtained with the initial networks. Informative power of the skeletons was also confirmed by the study of the anatomical regions selected by the significant supervoxels in correspondence of areas, such as hippocampus, notably connected with Alzheimer disease progression.

In addition, validating the promising multiplex network method on the salient skeletons of an independent dataset, an accuracy of 0.79 ± 0.01, for AD versus NC and 0.76 ± 0.01, for cMCI and NC, was obtained. These results turned out, salient networks allow a link reduction of 92% and a node reduction of 82% respect to the original network providing performances that compared to those obtained with the original network, are reduced by less than 15

The multiplex network method was adapted and calibrated for another very widespread neurodegenerative disease: Parkinson's disease. It was shown how the multiple network method can be effectively used to characterize structural changes associated with Parkinson's disease. The informative power provided by multi-layer networks was investigated by training an independent classifier for each scale through which a subject scan can be analyzed. This base of knowledge was evaluated for different feature selection methods and it was studied what happens when the features associated to more scales are used together to distinguish PD subjects from NC.

It was demonstrated how this approach is able to discriminate patients from healthy subjects with classification results (accuracy of 0.832 ± 0.004) in line with the literature. It was also proved that, despite there is a privileged scale ($125\ voxels$) to study PD, combining features of different scales compatible with size of anatomical structures, mostly affected by PD, makes classification significantly ($p < 0.01$) more accurate (0.884 ± 0.004) allowing the state-of-the-art to be outperformed. Besides, comparison among several feature selection methods highlights how it is important to use selection approaches, taking into account interactions among features. Indeed, the embedded and the wrapper methods, satisfying this condition, work better and allow a relevant performance increase when more scales are combined. This study allowed also the identification of multi-scale supervoxels associated to the important features and corresponding to cerebral regions which literature reports as connected to PD.

The robustness and the accuracy of the proposed methodology were evaluated with both a direct evaluation, involving the measure of classification metrics, and an indirect check, regarding the brain regions mostly affected by the disease. This always through a comprehensive analysis of the whole head, without the aid of a preliminary segmentation. Method was validated on the complete cohort of controls and patients from the PPMI dataset; for what concerns NC/PD classification, the proposed methodology compared well (accuracy of 0.88 ± 0.06) with other state-of-the-art approaches basing on MRI data and compares well with methodologies including other imaging modalities such as SPECT.

In addition, our method results more accurate than conventional ROI- and voxel based-method and it is more sensitive than a standard VBM in terms of region number connected to PD. As reported in previous studies, see for example [145], with VBM no significant changes in cortical morphology were observed when comparing NC and PD subjects after Bonferroni correction. On the other hand the regions outlined with the proposed methodology are consistent with previous studies. Temporal and Frontal Gyri changes have shown atrophic patterns, especially in patients with dementia [146**?**]; cognitive impairment seems to acquire a relevant role for diagnosis also because of the inclusion of Brodmann areas $24, 37, 44$ and 46 as shown also in [147, 148]. As expected, Substantia Nigra and brain midstem also play a relevant role for the diagnosis. It was found that most of the significant regions are not adjacent, indeed there is no a priori reason for adjacent patches (which can often include distinct anatomical district) to share the same informative content, as neurodegenerative diseases may have a diffuse effect involving multiple voxels not necessarily belonging to the same anatomical region [146].

Our results outline the important role of combining MRI network and clinical features for an accurate early diagnosis. In fact, besides the increment of the classification accuracy (0.93 ± 0.04), it is manifest that the use of clinical features is biased towards the NC class. It was demonstrated that classification scores based only on clinical features resulted poorly sensitive, with lots of PD patients misclassified as controls; this effect is reasonable as in the early phase of the disease clinical symptoms are mild. On the other hand this result underlines the importance of complex network markers to improve both sensitivity and specificity of the classification.

The method, developed in this thesis, has great versatility and lends itself to a variety of purposes, including the identification of 'disease signature' for more anatomically heterogeneous forms of neurodegenerative disorder, such as tauophathies or synucleinopathies, where the model could be enriched with additional clinical or genetic data. The method is also very suitable to application to longitudinal studies, ideally in association with other modalities of imaging, to improve our understanding of the different patterns of neurodegeneration in different diseases. The impact of variables such as the degree of atrophy, disease duration, site or scanner type could also be investigated in further studies. It would also be interesting to deepen the study of the relationship between the features derived from different scales and investigate the discrimination of healthy and ill subjects from subjects without evidence for dopaminergic deficit (SWEDD), despite they manifests the characteristic PD symptoms.

6.3 OVERVIEW OF THE INNOVATIVE CONTRIBUTE

The strengths of this thesis work are the following:

- It was developed a novel physical brain model by studying cerebral connectivity through the complex networks.
- By means of the combined use of machine learning techniques and the study of network measures, it was obtained a method able to reveal structural changes in the early phase of Alzheimer and Parkinson and provide a privileged scale to study a certain disease.
- The implemented methodology is able to find regions connected to the disease basing on a segmentation completely unsupervised over all the brain. Indeed, a patch-based approach that segments the whole brain in rectangular boxes was used. It combines the interpretability of a voxel-wise description, without its intrinsic computational burden and noise sensitivity, and the robustness of region of interest methods, avoiding *a priori* assumptions in terms of disease effects or segmentation accuracy. In addition, as brain disease has often a diffuse effect, affecting multiple voxels, but not necessarily corresponding to entire anatomical structures, the proposed approach has the potential to better suit the description of pathological changes in the brain, reflecting biological variability.
- this approach results also to be an useful and powerful tool support to early diagnosis of neurodegenerative disease. Indeed for the discrim-

ination of healthy from diseased or prodromal subjects, performances were comparable to those obtained with usually more accurate methods like single photon emission computed tomography (SPECT) and positron emission tomography (PET) and outperform the state-of-the-art relative to MRI-based method.

LIST OF FIGURES

LIST OF TABLES

BIBLIOGRAPHY

[1] Mark F Bear, Barry W Connors, and Michael A Paradiso. *Neuroscience*, volume 2. Lippincott Williams & Wilkins, 2007.

[2] Ian P Johnson. Age-related neurodegenerative disease research needs aging models. *Frontiers in aging neuroscience*, 7, 2015.

[3] G-Science Academies' meetings. The challenge of neurodegenerative diseases in an aging population, 2017. URL `https://royalsociety.org/about-us/international/international-work/g-science-academies-meetings/`.

[4] Susie MD Henley, Gillian P Bates, and Sarah J Tabrizi. Biomarkers for neurodegenerative diseases. *Current opinion in neurology*, 18(6):698–705, 2005.

[5] Andreas Jeromin and Robert Bowser. Biomarkers in neurodegenerative diseases. In *Neurodegenerative Diseases*, pages 491–528. Springer, 2017.

[6] Alzheimer's Disease Neuroimaging Initiative. About biomarkers, . URL `http://adni.loni.usc.edu/study-design/background-rationale/`.

[7] Marion Delenclos, Daryl R Jones, Pamela J McLean, and Ryan J Uitti. Biomarkers in parkinson's disease: Advances and strategies. *Parkinsonism & related disorders*, 22:S106–S110, 2016.

[8] Honglei Chen, Edward A Burton, G Webster Ross, Xuemei Huang, Rodolfo Savica, Robert D Abbott, Alberto Ascherio, John N Caviness, Xiang Gao, Kimberly A Gray, et al. Research on the premotor symptoms of parkinsonÕs disease: clinical and etiological implications. *Environmental health perspectives*, 121(11-12):1245, 2013.

[9] Bruce Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.

[10] Bharti Rana, Akanksha Juneja, Mohit Saxena, Sunita Gudwani, S Senthil Kumaran, RK Agrawal, and Madhuri Behari. Regions-of-interest based automated diagnosis of parkinsonÕs disease using t1-weighted mri. *Expert Systems with Applications*, 42(9):4506–4516, 2015.

[11] John Ashburner, Gareth Barnes, Chun-Chuan Chen, Jean Daunizeau, Guillaume Flandin, Karl Friston, Volkmar Glauche, Rik Henson, Chloe Hutton, Stephan Kiebel, et al. Statistical parametric mapping. 1994.

[12] John Ashburner and Karl J Friston. Voxel-based morphometryÑthe methods. *Neuroimage*, 11(6):805–821, 2000.

[13] Heiko Braak and Eva Braak. Neuropathological stageing of alzheimer-related changes. *Acta neuropathologica*, 82(4):239–259, 1991.

[14] Alzheimer's Association. Seven stages of alzheimer's, 2017. URL `https://m.alz.org/stages-of-alzheimers.asp`.

[15] University of California, Irvine Institute for Memory Impairments and Neurological Disorders (UCI MIND). Mild cognitive impairment, 2017. URL `https://www.mind.uci.edu/dementia/mild-cognitive-impairment/`.

[16] John C Morris and Eugene H Rubin. Clinical diagnosis and course of alzheimer's disease. *Psychiatric Clinics of North America*, 1991.

[17] G. McKhann, D. Drachman, F. Marshall, R. Katzman, D. Price, and E. M. Stadlan. Clinical Diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 34(7):939 – 944, 1984.

[18] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC, American Psychiatric Association, 2000.

[19] M. F. Folstein, S. E. Folstein, and P. R. McHugh. Mini-Mental State: a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3):189 – 198, 1975.

[20] Gérard Emilien, Cécile Durlach, Kenneth L Minaker, Bengt Winblad, Serge Gauthier, and Jean-Marie Maloteaux. *Alzheimer disease: neuropsychology and pharmacology*. Birkhäuser, 2012.

[21] David S Knopman, Steven T DeKosky, JL Cummings, H Chui, J Corey-Bloom, N Relkin, GW Small, B Miller, and JC Stevens. Practice parameter: Diagnosis of dementia (an evidence-based review) report of the quality standards subcommittee of the american academy of neurology. *Neurology*, 56(9):1143–1153, 2001.

[22] Reisa A Sperling, Paul S Aisen, Laurel A Beckett, David A Bennett, Suzanne Craft, Anne M Fagan, Takeshi Iwatsubo, Clifford R Jack, Jeffrey Kaye, Thomas J Montine, et al. Toward defining the preclinical stages of alzheimerÕs disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia*, 7(3):280–292, 2011.

[23] Ronald C Petersen and Selamawit Negash. Mild cognitive impairment: an overview. *CNS spectrums*, 13(1):45–53, 2008.

[24] B. Dubois, H. H. Feldman, C. Jacova, S. T. DeKosky, P. Barberger-Gateau, J. Cummings, A. Delacourte, D. Galasko, S. Gauthier, G. Jicha, K. Meguro, J. O'Brien, F. Pasquier, P. Robert, M. Rossor, S. Salloway, Y. Stern, P. J. Visser, and P. Scheltens. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDSÐADRDA criteria. *Lancet Neurology*, 6(8):734 – 746, 2007.

[25] Marilyn S Albert, Steven T DeKosky, Dennis Dickson, Bruno Dubois, Howard H Feldman, Nick C Fox, Anthony Gamst, David M Holtzman, William J Jagust, Ronald C Petersen, et al. The diagnosis of mild cognitive impairment due to alzheimerÕs disease: Recommendations from the national institute on aging-alzheimerÕs association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia*, 7 (3):270–279, 2011.

[26] Guy M McKhann, David S Knopman, Howard Chertkow, Bradley T Hyman, Clifford R Jack, Claudia H Kawas, William E Klunk, Walter J Koroshetz, Jennifer J Manly, Richard Mayeux, et al. The diagnosis of dementia due to alzheimer's disease: Recommendations from the national institute on aging-alzheimerÕs association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia*, 7(3):263–269, 2011.

[27] Clifford R Jack, David S Knopman, William J Jagust, Leslie M Shaw, Paul S Aisen, Michael W Weiner, Ronald C Petersen, and John Q Trojanowski. Hypothetical model of dynamic biomarkers of the alzheimer's pathological cascade. *The Lancet Neurology*, 9(1):119–128, 2010.

[28] Ronald C Petersen. Mild cognitive impairment as a diagnostic entity. *Journal of internal medicine*, 256(3):183–194, 2004.

[29] Alzheimer's Disease Neuroimaging Initiative. Sharing Alzheimer's Research Data with the World, . URL `http://adni.loni.usc.edu`.

[30] The R Project for Statistical Computing, url = https://www.r-project.org,.

[31] WR Gibb and AJ Lees. The relevance of the lewy body to the pathogenesis of idiopathic parkinson9s disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 51(6):745–752, 1988.

[32] Joseph Jankovic. ParkinsonÕs disease: clinical features and diagnosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 79(4):368–376, 2008.

[33] Chandar Singaram, EA Gaumnitz, C Torbey, W Ashraf, EMM Quigley, A Sengupta, and Ronald Pfeiffer. Dopaminergic defect of enteric nervous system in parkinson's disease patients with chronic constipation. *The Lancet*, 346(8979):861–864, 1995.

[34] J-F Gagnon, M-A Bédard, ML Fantini, D Petit, M Panisset, S Rompre, J Carrier, and J Montplaisir. Rem sleep behavior disorder and rem sleep without atonia in parkinsonÕs disease. *Neurology*, 59(4):585–589, 2002.

[35] K Ray Chaudhuri, Daniel G Healy, and Anthony HV Schapira. Non-motor symptoms of parkinson's disease: diagnosis and management. *The Lancet Neurology*, 5(3):235–245, 2006.

[36] J Friedman and H Friedman. Fatigue in parkinson's disease. *Neurology*, 43(10):2016–2016, 1993.

[37] Steven J Huber, Edwin C Shuttleworth, and George W Paulson. Dementia in parkinson's disease. *Archives of Neurology*, 43(10):987–990, 1986.

[38] Stewart A Factor and William Weiner. Parkinson's disease: Diagnosis & clinical management. chapter 22. Demos Medical Publishing, 2007.

[39] Eduardo Tolosa, Gregor Wenning, and Werner Poewe. The diagnosis of parkinson's disease. *The Lancet Neurology*, 5(1):75–86, 2006.

[40] Douglas J Gelb, Eugene Oliver, and Sid Gilman. Diagnostic criteria for parkinson disease. *Archives of neurology*, 56(1):33–39, 1999.

[41] Rajesh Pahwa and Kelly E Lyons. Early diagnosis of parkinson's disease: recommendations from diagnostic clinical guidelines. *The American Journal of Managed Care*, 16:S94–9, 2010.

[42] Stewart A Factor and William Weiner. Parkinson's disease: Diagnosis & clinical management. chapter 14. Demos Medical Publishing, 2007.

[43] Christopher G Goetz, Barbara C Tilley, Stephanie R Shaftman, Glenn T Stebbins, Stanley Fahn, Pablo Martinez-Martin, Werner Poewe, Cristina Sampaio, Matthew B Stern, Richard Dodel, et al. Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): Scale presentation and clinimetric testing results. *Movement disorders*, 23(15):2129–2170, 2008.

[44] Ronald B Postuma, Dag Aarsland, Paolo Barone, David J Burn, Christopher H Hawkes, Wolfgang Oertel, and Tjalf Ziemssen. Identifying prodromal parkinson's disease: Pre-motor disorders in parkinson's disease. *Movement Disorders*, 27(5):617–626, 2012.

[45] L. V. Kalia and A. E. Lang. Parkinson's disease. *Lancet*, 386(9996):896–912, April 2015. doi: http://dx.doi.org/10.1016/S0140-6736(14)61393-3.

[46] Angelo Antonini, Klaus L Leenders, Peter Vontobel, R Paul Maguire, John Missimer, Maria Psylla, and Ilonka Günther. Complementary pet studies of striatal neuronal function in the differential diagnosis between multiple system atrophy and parkinson's disease. *Brain: a journal of neurology*, 120(12):2187–2195, 1997.

[47] Joseph C Masdeu. Future directions in imaging neurodegeneration. *Current neurology and neuroscience reports*, 17(1):9, 2017.

[48] Ehsan Adeli, Guorong Wu, Behrouz Saghafi, Le An, Feng Shi, and Dinggang Shen. Kernel-based joint feature selection and max-margin classification for early diagnosis of parkinsonÕs disease. *Scientific reports*, 7, 2017.

[49] Diane B Miller and James P OÕCallaghan. Biomarkers of parkinsonÕs disease: present and future. *Metabolism*, 64(3):S40–S46, 2015.

[50] Simon Duchesne, Yan Rolland, and Marc Vérin. Automated computer differential classification in parkinsonian syndromes via pattern analysis on mri. *Academic radiology*, 16(1):61–70, 2009.

[51] Marcos Hortes N Chagas, Vitor Tumas, Márcio A Pena-Pereira, João Paulo Machado-de Sousa, Antonio Carlos dos Santos, Rafael Faria Sanches, Jaime EC Hallak, and José Alexandre S Crippa. Neuroimaging of major depression in parkinson's disease: Cortical thickness, cortical and subcortical volume, and spectroscopy findings. *Journal of Psychiatric Research*, 90:40–45, 2017.

[52] Christopher Summerfield, Carme Junqué, Eduardo Tolosa, Pilar Salgado-Pineda, Beatriz Gómez-Ansón, Maria José Martí, Pau Pastor, Blanca Ramírez-Ruíz, and José Mercader. Structural brain changes in parkinson disease with dementia: a voxel-based morphometry study. *Archives of Neurology*, 62(2):281–285, 2005.

[53] Carlo Tessa, Claudio Lucetti, Marco Giannelli, Stefano Diciotti, Michele Poletti, Sabrina Danti, Filippo Baldacci, Claudio Vignali, Ubaldo Bonuccelli, Mario Mascalchi, et al. Progression of brain atrophy in the early stages of parkinson's disease: A longitudinal tensor-based morphometry study in de novo patients without cognitive impairment. *Human brain mapping*, 35(8):3932–3944, 2014.

[54] Andrea Cherubini, Maurizio Morelli, Rita Nisticó, Maria Salsone, Gennarina Arabia, Roberta Vasta, Antonio Augimeri, Maria Eugenia Caligiuri, and Aldo Quattrone. Magnetic resonance support vector machine discriminates between parkinson disease and progressive supranuclear palsy. *Movement Disorders*, 29(2):266–269, 2014.

[55] Gurpreet Singh and Lakshminarayanan Samavedham. Unsupervised learning based feature extraction for differential diagnosis of neurodegenerative diseases: a case study on early-stage diagnosis of parkinson disease. *Journal of neuroscience methods*, 256:30–40, 2015.

[56] Ehsan Adeli, Feng Shi, Le An, Chong-Yaw Wee, Guorong Wu, Tao Wang, and Dinggang Shen. Joint feature-sample selection and robust diagnosis of parkinson's disease from mri data. *NeuroImage*, 141:206–219, 2016.

[57] Angela EP Bouwmans, Annemarie MM Vlaar, Werner H Mess, Alfons Kessels, and Wim EJ Weber. Specificity and sensitivity of transcranial sonography of the substantia nigra in the diagnosis of parkinson9s disease: prospective cohort study in 196 patients. *BMJ open*, 3(4):e002613, 2013.

[58] Andrea Pilotto, Rezzak Yilmaz, and Daniela Berg. Developments in the role of transcranial sonography for the differential diagnosis of parkinsonism. *Current neurology and neuroscience reports*, 15(7):43, 2015.

[59] John C Morgan, Shyamal H Mehta, and Kapil D Sethi. Biomarkers in parkinsonÕs disease. *Current neurology and neuroscience reports*, 10(6):423–430, 2010.

[60] Margaret M Hoehn and Melvin D Yahr. Parkinsonism onset, progression, and mortality. *Neurology*, 17(5):427–427, 1967.

[61] ParkinsonÕs Progression Marker Initiative. A Landmark Study of Parkinson's Disease, . URL http://www.ppmi-info.org.

[62] F. Bloch. Nuclear induction. *Physical Review*, 70(7-8):460 – 474, 1946.

[63] E. M. Purcell, H. C. Torrey, and R. V. Pound. Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Physical Review*, 69(1-2):37 – 38, 1946.

[64] E.M. Haacke, R.W. Brown, M.R. Thompson, and R. Venkatesan. *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. Wiley, 1999.

[65] Richard R Ernst, Geoffrey Bodenhausen, Alexander Wokaun, et al. Principles of nuclear magnetic resonance in one and two dimensions. 1987.

[66] Marinus T Vlaardingerbroek and Jacques A Boer. *Magnetic resonance imaging: theory and practice*. Springer Science & Business Media, 2013.

[67] Center for Complex Network Research (CCNR). Barábasi Lab: Think Networks. URL https://www.barabasilab.com.

[68] Leonhard Euler. Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, 8:128–140, 1741.

[69] P Erdös and A Rényi. ton random graphs, upubl. *Math. Debrecen*, 6: 290–297, 1959.

[70] Jeffrey Travers and Stanley Milgram. The small world problem. *Phychology Today*, 1:61–67, 1967.

[71] Duncan J Watts and Steven H Strogatz. Collective dynamics of Ôsmall-worldÕnetworks. *nature*, 393(6684):440–442, 1998.

[72] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[73] Center for Applied Internet Data Analysis (CAIDA. URL https://www.caida.org/projects/.

[74] The Human Connectome Project. URL http://www.humanconnectomeproject.org.

[75] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424 (4):175–308, 2006.

[76] Marc Barthelemy, Bernard Gondran, and Eric Guichard. Spatial structure of the internet traffic. *Physica A: statistical mechanics and its applications*, 319:633–642, 2003.

[77] Stephen C Johnson. Hierarchical clustering schemes. *Psychometrika*, 32 (3):241–254, 1967.

[78] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99 (12):7821–7826, 2002.

[79] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.

[80] Béla Bollobás. Random graphs. In *Modern Graph Theory*, pages 215–252. Springer, 1998.

[81] Jörn Davidsen, Holger Ebel, and Stefan Bornholdt. Emergence of a small world from local interactions: Modeling acquaintance networks. *Physical Review Letters*, 88(12):128701, 2002.

[82] Albert-László Barabási. *Network science*. Cambridge university press, 2016.

[83] Jonathan PK Doye. Network topology of a potential energy landscape: A static scale-free network. *Physical review letters*, 88(23):238701, 2002.

[84] Ginestra Bianconi and A-L Barabási. Competition and multiscaling in evolving networks. *EPL (Europhysics Letters)*, 54(4):436, 2001.

[85] Cornelis J Stam and Jaap C Reijneveld. Graph theoretical analysis of complex networks in the brain. *Nonlinear biomedical physics*, 1(1):3, 2007.

[86] Peter J Mucha, Thomas Richardson, Kevin Macon, Mason A Porter, and Jukka-Pekka Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *science*, 328(5980):876–878, 2010.

[87] Miao Cao, Ni Shu, Qingjiu Cao, Yufeng Wang, and Yong He. Imaging functional and structural brain connectomics in attention-deficit/hyperactivity disorder. *Molecular neurobiology*, 50(3):1111–1123, 2014.

[88] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3), 2009.

[89] Marina Boccardi, Martina Bocchetta, Félix C Morency, D Louis Collins, Masami Nishikawa, Rossana Ganzola, Michel J Grothe, Dominik Wolf, Alberto Redolfi, Michela Pievani, et al. Training labels for hippocampal segmentation based on the eadc-adni harmonized hippocampal protocol. *Alzheimer's & Dementia*, 11(2):175–183, 2015.

[90] Sayan Mukherjee, Pablo Tamayo, Simon Rogers, Ryan Rifkin, Anna Engle, Colin Campbell, Todd R Golub, and Jill P Mesirov. Estimating dataset size requirements for classifying dna microarray data. *Journal of computational biology*, 10(2):119–142, 2003.

[91] Claudia Beleites, Ute Neugebauer, Thomas Bocklitz, Christoph Krafft, and Jürgen Popp. Sample size planning for classification models. *Analytica chimica acta*, 760:25–33, 2013.

[92] Kenneth Marek, Danna Jennings, Shirley Lasch, Andrew Siderowf, Caroline Tanner, Tanya Simuni, Chris Coffey, Karl Kieburtz, Emily Flagg, Sohini Chowdhury, et al. The parkinson progression marker initiative (ppmi). *Progress in neurobiology*, 95(4):629–635, 2011.

[93] Murray W Johns. A new method for measuring daytime sleepiness: the epworth sleepiness scale. *sleep*, 14(6):540–545, 1991.

[94] Karin Stiasny-Kolster, Geert Mayer, Sylvia Schäfer, Jens Carsten Möller, Monika Heinzel-Gutenbrunner, and Wolfgang H Oertel. The rem sleep behavior disorder screening questionnaireÑa new diagnostic instrument. *Movement disorders*, 22(16):2386–2393, 2007.

[95] Jerome A Yesavage and Javaid I Sheikh. 9/geriatric depression scale (gds) recent evidence and development of a shorter version. *Clinical gerontologist*, 5(1-2):165–173, 1986.

[96] JA Yesavage, TL Brink, TL Rose, et al. Geriatric depression scale (GDS). *Handbook of psychiatric measures. Washington DC: American Psychiatric Association*, pages 544–6, 2000.

[97] Ronald B Postuma, Daniela Berg, Matthew Stern, Werner Poewe, C Warren Olanow, Wolfgang Oertel, José Obeso, Kenneth Marek, Irene Litvan, Anthony E Lang, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Movement Disorders*, 30(12):1591–1601, 2015.

[98] Irene Litvan, Jennifer G Goldman, Alexander I Tröster, Ben A Schmand, Daniel Weintraub, Ronald C Petersen, Brit Mollenhauer, Charles H Adler, Karen Marder, Caroline H Williams-Gray, et al. Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force guidelines. *Movement Disorders*, 27(3):349–356, 2012.

[99] JC Dalrymple-Alford, MR MacAskill, CT Nakas, L Livingston, C Graham, GP Crucian, TR Melzer, J Kirwan, R Keenan, S Wells, et al. The MoCA well-suited screen for cognitive impairment in Parkinson disease. *Neurology*, 75(19):1717–1725, 2010.

[100] FMRIB Analysis Group & MGH. Lecture: Brain Extraction, Registration, Motion Correction and EPI Distortion, 2000-2017. URL `https://fsl.fmrib.ox.ac.uk/fslcourse/`.

[101] Derek LG Hill, Philipp G Batchelor, Mark Holden, and David J Hawkes. Medical image registration. *Physics in medicine and biology*, 46(3):R1, 2001.

[102] Mark Jenkinson, Christian F Beckmann, Timothy EJ Behrens, Mark W Woolrich, and Stephen M Smith. Fsl. *Neuroimage*, 62(2):782–790, 2012.

[103] Jörn Diedrichsen. A spatially unbiased atlas template of the human cerebellum. *Neuroimage*, 33(1):127–138, 2006.

[104] Marianna La Rocca, Nicola Amoroso, Roberto Bellotti, Domenico Diacono, Alfonso Monaco, Anna Monda, Andrea Tateo, and Sabina Tangaro. A multiplex network model to characterize brain atrophy in structural mri. In *Emergent Complexity from Nonlinearity, in Physics, Engineering and the Life Sciences*, pages 189–198. Springer, 2017.

[105] Giulia Menichetti, Daniel Remondini, Pietro Panzarasa, Raúl J Mondragón, and Ginestra Bianconi. Weighted multiplex networks. *PloS one*, 9(6):e97857, 2014.

[106] Betty M Tijms, Christiane Möller, Hugo Vrenken, Alle Meije Wink, Willem de Haan, Wiesje M van der Flier, Cornelis J Stam, Philip Scheltens, and Frederik Barkhof. Single-subject grey matter graphs in alzheimer's disease. *PloS one*, 8(3):e58921, 2013.

[107] Olaf Sporns. The human connectome: a complex network. *Annals of the New York Academy of Sciences*, 1224(1):109–125, 2011.

[108] Gyorgy Buzsaki. *Rhythms of the Brain*. Oxford University Press, 2006.

[109] Betty M Tijms, Alle Meije Wink, Willem de Haan, Wiesje M van der Flier, Cornelis J Stam, Philip Scheltens, and Frederik Barkhof. Alzheimer's disease: connecting findings from graph theoretical studies of brain networks. *Neurobiology of aging*, 34(8):2023–2036, 2013.

[110] Christian Thiemann, Fabian Theis, Daniel Grady, Rafael Brune, and Dirk Brockmann. The structure of borders in a small world. *PloS one*, 5(11): e15422, 2010.

[111] Olivia Woolley-Meza, Christian Thiemann, Daniel Grady, Jake Jungbin Lee, Hanno Seebens, Bernd Blasius, and Dirk Brockmann. Complexity in human transportation networks: a comparative analysis of worldwide air transportation and global cargo-ship movements. *The European Physical Journal B*, 84(4):589–600, 2011.

[112] Bin Zhang and Steve Horvath. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology*, 4(1), 2005.

[113] Danielle Smith Bassett and ED Bullmore. Small-world brain networks. *The neuroscientist*, 12(6):512–523, 2006.

[114] Daniel Grady, Christian Thiemann, and Dirk Brockmann. Robust classification of salient links in complex networks. *Nature communications*, 3: 864, 2012.

[115] Sotiris B Kotsiantis, Ioannis D Zaharakis, and Panayiotis E Pintelas. Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, 26(3):159–190, 2006.

[116] L. Breiman. Random Forests. *Machine Learning 45*, (1):5 – 32, 2001.

[117] Yunjing An, Shutao Sun, and Shujuan Wang. Naive bayes classifiers for music emotion classification based on lyrics. In *Computer and Information Science (ICIS), 2017 IEEE/ACIS 16th International Conference on*, pages 635–638. IEEE, 2017.

[118] Graziella Orrù, William Pettersson-Yeo, Andre F Marquand, Giuseppe Sartori, and Andrea Mechelli. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neuroscience & Biobehavioral Reviews*, 36(4):1140–1152, 2012.

[119] Muriel Gevrey, Ioannis Dimopoulos, and Sovan Lek. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological modelling*, 160(3):249–264, 2003.

[120] Kellie J Archer and Ryan V Kimes. Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis*, 52(4):2249–2260, 2008.

[121] Yvan Saeys, Iñaki Inza, and Pedro Larrañaga. A review of feature selection techniques in bioinformatics. *bioinformatics*, 23(19):2507–2517, 2007.

[122] Kenji Kira and Larry A Rendell. The feature selection problem: Traditional methods and a new algorithm. In *AAAI*, volume 2, pages 129–134, 1992.

[123] Sabina Tangaro, Nicola Amoroso, Massimo Brescia, Stefano Cavuoti, Andrea Chincarini, Rosangela Errico, Paolo Inglese, Giuseppe Longo, Rosalia Maglietta, Andrea Tateo, et al. Feature selection based on machine learning in MRIs for hippocampal segmentation. *Computational and mathematical methods in medicine*, 2015, 2015.

[124] Rahul S Desikan, Florent Ségonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson, Deborah Blacker, Randy L Buckner, Anders M Dale, R Paul Maguire, Bradley T Hyman, et al. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980, 2006.

[125] Jack L Lancaster, Diana Tordesillas-Gutiérrez, Michael Martinez, Felipe Salinas, Alan Evans, Karl Zilles, John C Mazziotta, and Peter T Fox. Bias between mni and talairach coordinates analyzed using the icbm-152 brain template. *Human brain mapping*, 28(11):1194–1205, 2007.

[126] Statistical Parametric Mapping (SPM. URL `http://www.fil.ion.ucl.ac.uk/spm/`.

[127] ReCaS. URL `https://www.recas-bari.it/index.php/it/`.

[128] nextMR. URL `https://web.infn.it/nextmr/index.php/en/`.

[129] Elixir. URL `https://www.elixir-europe.org`.

[130] Life Watch. URL `http://www.lifewatch.eu/)`.

[131] M Prince, E Albanese, M Guerchet, and M Prina. World Alzheimer Report 2014. AlzheimerÕs Disease International (ADI), London, 2014.

[132] Stacey L Kowal, Timothy M Dall, Ritashree Chakrabarti, Michael V Storm, and Anjali Jain. The current and projected economic burden of parkinson's disease in the united states. *Movement Disorders*, 28(3):311–318, 2013.

[133] M.G. Bondio, F. Sporing, and J.S. Gordon. *Medical Ethics, Prediction, and Prognosis: Interdisciplinary Perspectives*. Routledge Annals of Bioethics. Taylor & Francis, 2017. ISBN 9781351802581. URL `https://books.google.it/books?id=2qW8DgAAQBAJ`.

[134] Julian M Fearnley and Andrew J Lees. Ageing and parkinson's disease: substantia nigra regional selectivity. *Brain*, 114(5):2283–2301, 1991.

[135] Alastair John Noyce, Andrew John Lees, and Anette-Eleonore Schrag. The prediagnostic phase of parkinson's disease. *J Neurol Neurosurg Psychiatry*, pages jnnp–2015, 2016.

[136] Christian Salvatore, Antonio Cerasa, Isabella Castiglioni, F Gallivanone, A Augimeri, M Lopez, G Arabia, M Morelli, MC Gilardi, and A Quattrone. Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and Progressive Supranuclear Palsy. *Journal of Neuroscience Methods*, 222:230–237, 2014.

[137] Mojtaba Zarei, Naroa Ibarretxe-Bilbao, Yaroslau Compta, Morgan Hough, Carme Junque, Nuria Bargallo, Eduardo Tolosa, and Maria Jose Martí. Cortical thinning is associated with disease stages and dementia in Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 84 (8):875–882, 2013.

[138] N Amoroso, R Errico, S Bruno, A Chincarini, E Garuccio, F Sensi, S Tangaro, A Tateo, R Bellotti, Alzheimers Disease Neuroimaging Initiative, et al. Hippocampal unified multi-atlas network (HUMAN): protocol and scale validation of a novel segmentation tool. *Physics in medicine and biology*, 60(22):8851, 2015.

[139] Catriona D Good, Ingrid S Johnsrude, John Ashburner, Richard NA Henson, KJ Fristen, and Richard SJ Frackowiak. A voxel-based morphometric study of ageing in 465 normal adult human brains. In *Biomedical Imaging, 2002. 5th IEEE EMBS International Summer School on*, pages 16–pp. IEEE, 2002.

[140] Christine Fennema-Notestine, Donald J Hagler, Linda K McEvoy, Adam S Fleisher, Elaine H Wu, David S Karow, and Anders M Dale. Structural MRI biomarkers for preclinical and mild Alzheimer's disease. *Human brain mapping*, 30(10):3238–3253, 2009.

[141] Sabine Derflinger, Christian Sorg, Christian Gaser, Nicholas Myers, Milan Arsic, Alexander Kurz, Claus Zimmer, Afra Wohlschläger, and Mark Mühlau. Grey-matter atrophy in Alzheimer's disease is asymmetric but not lateralized. *Journal of Alzheimer's Disease*, 25(2):347, 2011.

[142] Elaheh Moradi, Antonietta Pepe, Christian Gaser, Heikki Huttunen, Jussi Tohka, et al. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *NeuroImage*, 104:398–412, 2015.

[143] Christian Salvatore, Antonio Cerasa, Petronilla Battista, Maria C Gilardi, Aldo Quattrone, Isabella Castiglioni, Alzheimer's Disease Neuroimaging Initiative, et al. Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach. *Frontiers in neuroscience*, 9, 2015.

[144] Esther E Bron, Marion Smits, Wiesje M van der Flier, Hugo Vrenken, Frederik Barkhof, Philip Scheltens, Janne M Papma, Rebecca ME Steketee, Carolina Méndez Orellana, Rozanna Meijboom, et al. Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: The CADDementia challenge. *NeuroImage*, 111: 562–579, 2015.

[145] Niels K Focke, Gunther Helms, Sebstian Scheewe, Pia M Pantel, Cornelius G Bachmann, Peter Dechent, Jens Ebentheuer, Alexander Mohr, Walter Paulus, and Claudia Trenkwalder. Individual voxel-based subtype prediction can differentiate progressive supranuclear palsy from idiopathic Parkinson syndrome and healthy controls. *Human brain mapping*, 32(11):1905–1915, 2011.

[146] Emma J Burton, Ian G McKeith, David J Burn, E David Williams, and John T OÕBrien. Cerebral atrophy in Parkinson's disease with and without dementia: a comparison with Alzheimer's disease, dementia with Lewy bodies and controls. *Brain*, 127(4):791–800, 2004.

[147] Andrew J Hughes, Susan E Daniel, Linda Kilford, and Andrew J Lees. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *Journal of Neurology, Neurosurgery & Psychiatry*, 55(3):181–184, 1992.

[148] A Nagano-Saito, Y Washimi, Y Arahata, T Kachi, JP Lerch, AC Evans, A Dagher, and K Ito. Cerebral atrophy and its relation to cognitive impairment in Parkinson disease. *Neurology*, 64(2):224–229, 2005.

AKNOWLEDGMENT